OVERCOMING NON-NATIVE OVERRELIANCE ON DURATION:
A STUDY ON ENGLISH VOWEL MANIPULATION AND NEUTRALIZATION

Gemma Moya-Galé

Supervisor: Dr. Joan Carles Mora

MA Dissertation (July 2010)

Presented at the 21st Annual European Second Language Association Conference
(Stockholm, 8-10th September, 2011)

Applied Linguistics and Second Language Acquisition in Multilingual Contexts
Department of English and German Philology
University of Barcelona

Table of contents

Index of Figures and Tables

Figures

Tables

Abstract

The accurate production of an L2 sound is determined by the adequate perception of the target speech sound (Flege, 1993, 1995, 2009). In this respect, Spanish and Catalan speaking learners of English are believed to struggle in their correct identification and subsequent production of the English phonemic vowel contrast /i:/ and /ɪ/, as they assimilate these speech sounds to their single vocalic category /i/ (Best, 1995). Furthermore, the acoustic cues that are used by native speakers when producing those vowels (spectral or quality parameters mainly, together with temporal or duration cues) appear to escape Spanish and Catalan L2 users of the language, who display an overreliance on duration when identifying and producing the targeted phonological distinction. The present study examined both the overwhelming use of temporal cues by 62 non-native listeners of L1 Spanish and Catalan by means of a word identification task, which aurally presented them with words that contained either natural or duration-manipulated /i:/ or /ɪ/ vowels, as well as the possible effects that the neutralization of natural duration values could have on those listeners. This effect was analysed by means of a word discrimination task, which exposed L2 listeners to natural and duration-neutralized tokens in order to make them focus their attention on the spectral values of those vowels. Results of the first test confirmed the tendency evinced by Spanish and Catalan speaking learners of English of overrelying on temporal parameters rather than on quality ones when perceiving the targeted phonemic contrast. The second test, however, yielded inconclusive results, as ceiling effects were obtained, which hindered the comparison of both the discrimination accuracy of duration-natural and duration-neutralized stimuli and the listeners' performance in the two tasks.

Keywords: speech perception and production, cue weighting, spectral cues, temporal cues.

1. Introduction: Literature Review

Phonetic learning ability is thought to be maintained across the human lifespan, which entails the acknowledgement that the acquisition of L2 sounds is possible even well into adulthood. Nonetheless, not all L2 sound categories are equally learnt and this dissimilarity in L2 speech learning is due to L1 influence to a considerable extent. Flege's Speech Learning Model (1995), for instance, poses the idea that, according to equivalence classification, L2 speech sounds that differ from those present in the native phonological inventory of the L2 speaker will be easier to master as compared to those that display certain similarities and which, thus, are susceptible of being confounded with an existing native category. Best's Perceptual Assimilation Model or PAM (Best, 1995; Best and Tyler, 2007) was originally based on naïve non-native listeners, who "are likely, due to their native language experience, to perceptually *assimilate* the non-native phone to the most articulatorily-similar native phoneme" (Best & Tyler, 2007: 22). PAM presented several possibilities of sound assimilation, among which category-goodness difference (two non-native sounds being assimilated to a single native category but not being perceptually equally distant from the L1 sound) was considered to result in moderate speech sound perception. This model was later on adapted to encompass second language perceptual learning as well (PAM-L2), which would entail the assimilation of L2 speech sounds to L1 categories (e.g. assimilating English vowels /i:/ and /ɪ/ to Spanish/Catalan /i/, which is the focus of the present study). Regarding an even third model of speech perception, the Native Language Magnet model (Kuhl et al., 1992, 2008; Kuhl, 1994; Iverson and Kuhl, 1996, 2000), L2 sounds that come close to the speaker's L1 sound prototype will be less discriminable than those that remain in a more peripheral position and that, therefore, are more perceptually salient in nature. Hence, Spanish and Catalan learners of English will be less able to discriminate between English vowel /i:/ and Spanish /i/, as the former comes close to their L1 prototype.

The concept of L1 transfer and its linkage to cross-language speech categorization was already claimed by the Russian linguist Trubetzkoy (1939), who posed the idea that perceived L2 speech sounds would receive an "incorrect interpretation", as they would have “to pass through the *sieve* of the native phonological system” (*cf*. Bohn 1995, p. 282). The metaphorical conception of the L1 phonological inventory as a sieve that filters incoming L2 sounds involves interpreting target or new sounds in terms of already existing L1 categories, which further leads to the conclusion that both the perceptual decoding and productive encoding of L2 sounds is bound to be influenced by native categories, and, thus, is bound to be non-native like. This line of thought is also shared by Strange, who claimed that “adults are language-specific perceivers who show a profound influence of the mother tongue” (*cf*. Bohn 1995, p. 279) and by Major, whose idea of transfer entails that precisely because similarity between native and target speech sounds may make subtle cross-language differences go unheeded, learning of new sounds becomes more difficult and transfer is, then, more likely to occur (2008). In the same vein would be Andersen’s Transfer to Somewhere Hypothesis (1983), which poses the existence of similarities among linguistic structures (i.e. native and target speech sounds) as a prerequisite for cross-language transfer. Cross-linguistic similarity is, hence, regarded as a source of difficulty for L2 learners, even more when they are linguistically inexperienced (i.e.: when they lack a good command of the target language). Moreover, if this statement is narrowed down to the purpose of the present study, it is found that the Spanish/Catalan-speaking learners’ need to perceive (and subsequently produce) two distinct English vowels (i.e.: /i:/ & /ɪ/) that occupy the portion of vowel space of a single native category (i.e.: /i/) becomes quite a challenging and arduous task.

So far the concept of similarity has been pinpointed as an explanation for misperceptions of L2 sounds. Nonetheless, within the domain of phonetic cue weighting in particular (i.e.: the relative weight that is bestowed upon acoustic parameters when perceiving and producing a given speech sound [Holt & Lotto, 2006]), it has been argued that “the most difficult speech-sound contrasts for foreign-language learners often are the ones that have multiple phonetic cues, especially if the cues are weighted differently in the foreign and native languages” (Ylinen et al, 2009:1). Thus, L2 users may encounter difficulty not only in surmounting cross-language similarities but also in bestowing the correct weight to the appropriate target cues, as tends to be the case with Spanish/Catalan speaking learners of English. In English, and even though length may be conceived as a phonological trait, it is mainly quality rather than duration that marks the difference in both vowel perception and production. Thus, the vowels /i:/ and /ɪ/ are distinguished by spectral or quality and temporal or duration cues, albeit it is the former that has a heavier weight in vowel categorization. In Spanish or Catalan, nonetheless, duration is not used as a means to distinguish separate phonemes (i.e.: phonologically), and it is quality the parameter that allows for vowel distinctions. Interestingly, several studies have come to prove that, despite not having duration as an acoustic parameter in their native language, Spanish speakers of L2 English make use of temporal, and not spectral, cues when perceiving English vowels (Aliaga-García, 2009, 2010; Cerviño, 2008; Escudero & Boersma, 2004; Boersma & Escudero, 2005; Cebrian, 2006; Mora & Fullana, 2007). These studies run counter to the aforementioned Transfer Hypothesis, which, applied to the domain of phonetic cue weighting, would assume the restriction of the use of a given acoustic cue to the fact of its being implemented phonologically in the L2 listener's L1. McAllister's Feature Hypothesis (McAllister et al., 2002; McAllister, 2007) would also accommodate the influence of the native language in the ability to accurately perceive and produce target sounds, as it puts forwards that "a feature used to signal a phonological contrast in L1 may facilitate the perception (and eventually the production) of a contrast in the L2 which uses that feature in the realization of the contrast" (2007: 155). A study that instantiates these two theories is Bohn and Flege's 1990 research on experienced and inexperienced German speaking learners of L2 English, which took heed of the use of both spectral and temporal cues by German native speakers when producing English vowels /i/, /ɪ/, /ɛ/ and /æ/. Their results brought to light that, because of the similar nature of the first three L2 sounds with the source language and German's phonemic use of duration, both experienced and inexperienced German speakers were able to use both parameters in a target-like way. In other words, their study upheld the view that "non-native listeners whose native language differentiates vowels both spectrally and temporally should use both spectral and duration cues to differentiate the non-native contrast" (Bohn, 1995: 287). Thus, how could the Spanish/Catalan native speakers' trend to use non-native duration cues be explained? As L1 transfer can be safely discarded as a viable explanation in this case, literature on phonetic cue weighting has established that certain phonetic cues may simply be preferred over others when perceiving new L2 speech sounds (Bohn, 1995); in other words, it may be thought that temporal acoustic parameters are more easily accessed than others.

Bohn's Desensitization Hypothesis (1995), which may be regarded as an L1-independent assessment of phonetic cue weighting, presents itself as another feasible alternative. Bohn's hypothesis claims that "whenever spectral differences are insufficient to differentiate vowel contrasts because previous linguistic experience did not sensitize listeners to these spectral differences, duration differences will be used to differentiate the non-native vowel contrast" (1995: 294). Applying this claim to the perception of English vowels by Spanish/Catalan listeners, it would be stated that, precisely because native speakers of Spanish and Catalan undergo a process of category-goodness difference when perceiving English vowels /i:/ and /ɪ/ (i.e.: both vowels assimilate into a single native speech sound /i/), they are believed to be "desensitized" to spectral values when being exposed to that phonological contrast and, thus, they are forced to resort to the use of duration cues as a means to compensate for the insufficient weight of the already existing L1 spectral parameters.

Be that as it may, Spanish and Catalan learners of L2 English are confronted by several factors when exposed to the English phonemic contrast of /i:/ and /ɪ/. On the one hand, these two English vowels share a portion of vowel space that is occupied by a single native vowel /i/, which tends to be matched for similarity with its target counterparts, and, hence, perceived and produced in a non-standard manner. On the other hand, learners have to cope with two principal acoustic cues in the L2, vowel quality and duration, which differs from their native phonological trend, in which spectral parameters, but not temporal ones, are exploited. Hence, and alluding to Bohn's Desensitization Hypothesis, this group of L2 learners are not experienced in the use of duration as a phonetic cue to perceive L1 vowels but, because using quality would not be enough to distinguish both target phonemes, which are, in addition, assimilated into a single native category, the temporal parameter would be resorted to when perceiving and creating new L2 categories for this pair of English vowels.

2. The Present Study

This study follows the aforementioned line of research in that it deals with the perception of the two English vowels /i:/ and /ɪ/ by non-native speakers; more specifically, by Spanish/Catalan speaking learners of the target language. This group of L2 users has been selected in the present case because of the blatant difficulty that Spanish/Catalan native speakers have shown when perceiving the quality difference in the English vowels under analysis and because of their overwhelming preference for the use of temporal cues when distinguishing between the /i:/ and /ɪ/ sounds, an acoustic parameter which is not used phonologically in their L1.

The goal of this study is twofold: on the one hand, it aims at confirming and expanding previous research, by means of introducing a higher degree of phonetic variability, regarding Spanish/Catalan listeners' overreliance on the use of duration cues when perceiving the target English phonological contrast. On the other hand, it also intends to contribute innovatively to the domain of phonetic cue weighting by observing whether vowel duration neutralization enhances L2 listeners' capacity to entirely focus on spectral cues, as the temporal ones would have been rendered perceptually ambiguous (i.e.: neutralized). Data collection time constraints hindered the possibility of developing an ideal research design, which is then left for future research. Under optimal conditions, this study would have developed two different experiments containing the exact number and type of stimuli: a word identification task, which would have aurally presented L2 listeners with three sets of stimuli (duration-natural, duration-manipulated and duration-neutralized tokens) so that they identified the targeted vowel that appeared as acoustic input, and a word discrimination task, which would have also included the three different sets of stimuli and would have aimed at testing listeners' ability to accurately discriminate between the /i:/ and /ɪ/ contrast in those three conditions. Hence, it would have been interesting to compare the two tests having the same type of tokens, as that comparison would have yielded a more complete view on the effects that vowel duration manipulation and duration neutralization have on L2 listeners when they appear in combination. Nonetheless, and albeit these two experiments were kept, it was decided to use only natural and manipulated vowels for Experiment I and natural and neutralized tokens for Experiment II if time requirements were to be met. Thus, both tasks were finally analysed as separate tests instead of being studied in conjunction.

The first experiment was based on the use of a word identification task, which required L2 listeners to choose between a pair of words that constituted an /i:/-/ɪ/ minimal pair after having been aurally exposed to one of the two words that appeared on a computer screen. Thus, the aim of this first task was to achieve an understanding of the participants' mental representation of the phonological categories encompassed in the acoustic input they heard, which was based on both duration-natural and duration-manipulated tokens in ten different minimal pairs, as produced by six different native speakers of English.

The experiment was driven by the following research question:

- Do Spanish/Catalan listeners overrely on duration when perceiving the English vowel contrast /i:/-/ɪ/?

From this question, one directional hypothesis could be drawn: it was expected that listeners' vowel perception would be heavily influenced by overreliance on duration rather than on quality cues, as listeners would resort to temporal parameters due to, and according to the desensitization hypothesis, the insufficiency of spectral information available to them (because of the difference in goodness of fit between the non-native and native vowels). Drawing on the aforementioned temporal overreliance, it was expected that duration manipulation would worsen participants' performance and would, therefore, lead them to make more errors.

Experiment II was based on a word discrimination task, in which L2 listeners had to decide whether the second word of a three-word string was the same as either the one that preceded it or followed it. In contrast with the former experiment, stimuli here were both duration-natural and duration-neutralized tokens and the task was aimed at observing how accurate the participants' discrimination of the contrast /i:/-/ɪ/ was. In other words, this second test was designed to analyse whether listeners would be able to better perceive the spectral parameters of the targeted vowels when duration could not be used as an acoustic cue after having been neutralized. The experiment was prompted by the ensuing research question:

- To what extent are L2 listeners capable of perceiving the vowel contrast /i:/-/ɪ/ when vowel duration is neutralized and, hence, when the spectral cue is the only acoustic parameter available?

It was hypothesized that temporal cues interfere in the perception of vowel quality. Therefore, if these underwent a process of duration neutralization -by which duration was kept constant for words in each minimal pair- spectral cues would be the only ones available and, hence, participants would be able to identify L2 vowels more accurately and reliably than when both acoustic parameters appeared in combination. Furthermore, this would also yield a measure of learners' ability to perceive /i:/-/ɪ/ spectral differences in F1 (tongue height) and F2 (tongue frontness/backness). Thus, it was believed that duration neutralization would allow for the reallocation of learners' attention (Ylinen et al., 2009) to these spectral differences to enhance their vowel discrimination abilities.

3. Method

The study design was divided into two main tests, which aimed at observing the effects of two different vowel modification methods (manipulation and neutralization, respectively) on Spanish and Catalan speaking L2 listeners. The premise underlying these tasks was that Spanish and Catalan non-native speakers of English would overrely on duration cues when perceiving the English phonological contrast /i:/-/ɪ/ in the first experiment and that duration neutralization would enhance their vowel discrimination accuracy in the second experiment, as they would be forced to take heed of the critical (spectral) values. Consequently, this would imply that once duration cues were controlled for and no longer interfered with quality, the spectral information in a vowel would be more easily accessed and would allow for more native-like perception and subsequent production of L2 speech sounds.

Participants

The participants in this study (N=64) were a group of Spanish/Catalan speaking undergraduates in the English Studies Degree who were taking the subject of "English Phonetics and Phonology I" at the University of Barcelona at the time this research was conducted and were given course credit for their participation. All of them reported having either Catalan, Spanish or both languages as L1s, except for two participants, whose native languages were Romanian and Italian, and who, consequently, were excluded from further analysis, as having Spanish and/or Catalan as base languages was a prerequisite. Hence, the final number of participants to be included in the study was 62 (mean age: 22.09 years, range= 18-35 years). This initial group of 62 L2 users was further reduced during the data coding process of the second test, the word discrimination task, as 4 outliers were identified on the grounds of their poor performance in the test[1], as compared to the remaining 58 participants, who all performed at ceiling. Therefore, and due to this lack of consistency among subjects (the outliers being probably a consequence of a misunderstanding of the task), it was gauged necessary to leave them out of the data analysis, as results and the study's internal validity could have been compromised otherwise. This decision, however, was only applicable for the latter experiment, which was finally left with 58 participants, as ceiling effects were obtained here that prevented any reliable comparison between the two tasks and, thus, it was deemed preferable to keep them both separate with their own number of subjects. The listeners' estimated proficiency level was established to range from an intermediate to an upper-intermediate level, as measured by a questionnaire on their linguistic background and L2 experience.

This experimental group was afterwards compared to a control group of 5 native speakers (2 males and 3 females). Two of them were native speakers of British English whilst the other three spoke American English[2]. Their mean age was 28.6 years (range= 24-41 years).

This experimental group was afterwards compared to a control group of 5 native speakers (2 males and 3 females). Two of them were native speakers of British English whilst the other three spoke American English[2]. Their mean age was 28.6 years (range= 24-41 years).

Stimuli

The stimuli (360) were based on ten English minimal pairs containing the tense/lax phonemic contrast of the first two English vowels /i:/ and /ɪ/, which is

[1] Their performance did not reach 40% of correct vowel discrimination and was sometimes even 0.
[2] American English mainly relies on spectral cues rather tan on duration as compared to British English, whose use of both parameters is more balanced (Escudero & Boersma, 2004). This dialectal variation may lead to slightly different results regarding vowel duration manipulation and neutralization within the native speaker control group.

characterised by both temporal (i.e.: duration) and spectral (i.e.: quality) differences and which is deemed to be particularly difficult to perceive and produce by Spanish/Catalan L2 speakers of English. These 10 minimal pairs were drawn from a total pool of 16 pairs, after having taken heed of the context in which the target vowel was embedded; in other words, the initial selection of minimal pairs followed a context-driven criterion, by which both prevocalic consonants' place of articulation and postvocalic consonants' voicing state were taken into account.

This latter dichotomy (voiced versus voiceless consonants in coda position) was also placed under analysis, as vowel duration is affected by the voicing state of the ensuing consonant, in that a vowel becomes noticeably shorter when followed by a devoiced final obstruent in a process known as pre-fortis clipping (e.g. "bead" → [bi:d] versus "beat" → [bi·t]). Because this specific phonetic phenomenon entails the reduction of the natural duration of a tense vowel, its temporal values may, consequently, be similar to the ones of a lax vowel, which is why pre-fortis clipping was gauged to be necessary in the analysis of vowel duration.

As the principal goal in the stimuli selection was to achieve a high degree of phonetic variability so that speech samples were as natural and diverse as possible, not only was syllabic context controlled for, but also the speakers who produced the speech samples, as they were 10 native speakers of Southern British English (5 male and 5 female), who had been previously recorded by another researcher[3].

Hence, the initial total number of tokens was 200 (10x2x10), which were subsequently converted from stereo into mono and filtered so as to reduce low frequency noise by means of sound editing software. Tokens were then segmented[4] using the Speech Analysis software, Praat (Boersma & Weenink, 2007), which rendered them ready for spectrogram analyses.

All natural samples were analysed for vowel duration and for their first and second formants (tongue height and tongue frontness/backness, respectively). From the

[3] The stimuli were borrowed from Cristina Aliaga-García's PhD Thesis and were subsequently measured and modified for the purposes of the current study.
[4] Segmentation was based on the following criteria: words beginning and ending with plosives were taken from the initial closure phase and up to the aspiration phase following the release of the obstruent sound. Words ending with a liquid were segmented when a noticeable decrease in amplitudee was observed in the speech signal.

total of 10 native speakers, and due to time constraints affecting test length, it was decided to take only 3 males and 3 females as the source of the L2 learners' acoustic input. To this purpose, a series of independent-samples t-tests were run on the three measures under study (duration, F1 and F2) to ensure that tense /i:/-lax /ɪ/ differences in all dimensions were statistically significant and that the final 6 speakers made both temporal and spectral distinctions in their vowel productions. Furthermore, vowel duration as affected by pre-fortis clipping was also scrutinized. Running this series of tests was gauged to be necessary as it would have made no sense to have included native speakers who did not produce significant spectral and temporal distinctions, as neither duration manipulation nor duration neutralization would have played a role in the tasks to be conducted subsequently. After this preliminary process of stimuli preparation, the number of natural tokens which would serve as baseline data for the experiments was eventually 120 (10x2x6).

NS	**Duration**		**F1**		**F2**	
	Tense	**Lax**	**Tense**	**Lax**	**Tense**	**Lax**
NS_1	293 (157)	125 (39)	276 (39)	419 (33)	2569 (74)	2238 (119)
NS_2	296 (96)	175 (28)	259 (18)	405 (33)	2592 (68)	2085 (146)
NS_3	267 (109)	158 (27)	322 (19)	453 (34)	2120 (34)	1980 (78)
NS_4	236 (105)	145 (35)	425 (41)	508 (41)	2721 (100)	2397 (96)
NS_5	324 (191)	136 (61)	345 (38)	512 (41)	2894 (54)	2206 (151)
NS_6	164 (65)	100 (18)	383 (36)	478 (40)	2656 (190)	2236 (75)
Means	263 (120)	140 (35)	335 (32)	463 (37)	2592 (87)	2190 (110)

Table 1. Means (SD in parenthesis) for duration (ms), F1 & F2 (Hz) values for NS (*N*=6

NS	Duration			
	Tense		Lax	
	+ voiced C	+ unvoiced C	+ voiced C	+ unvoiced C
NS_1	389 (129)	148 (13)	151 (29)	88 (6)
NS_2	358 (75)	205 (6)	176 (36)	174 (16)
NS_3	341 (74)	157 (2)	173 (24)	135 (8)
NS_4	290 (104)	154 (12)	159 (37)	123 (18)
NS_5	441 (157)	148 (7)	160 (70)	100 (5)
NS_6	205 (51)	102 (6)	112 (10)	82 (11)
Means	337 (98)	152 (7)	155 (34)	117 (11)

Table 2. Means (SD in parenthesis) for duration (ms) by consonantal context for NS (*N*=6)

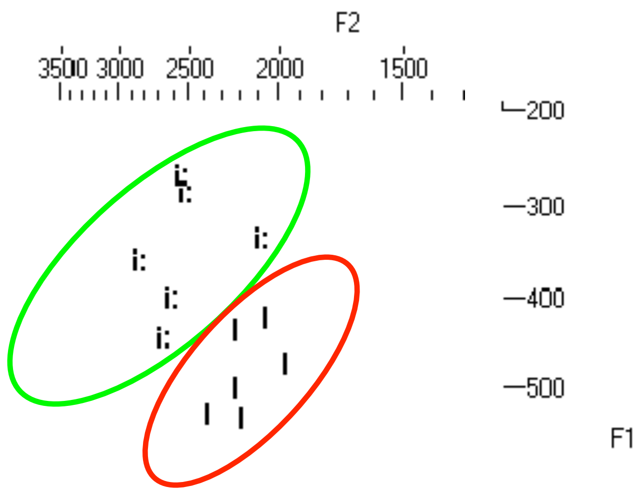


Figure1. Plot Formant for F1 & F2 means for /i:/ and /ɪ/ as produced by the 6 NS in the 10 minimal pairs

The modification of vowel duration was performed using Praat and was developed in terms of both duration manipulation and neutralization; nonetheless, the premise of maintaining quality values unchanged was kept constant throughout the whole process. The first part of the procedure followed the criterion employed by Ylinen et al.'s 2009 study, in which vowels were manipulated so that the temporal values of the phonemic tense/lax contrast of each minimal pair were interchanged; in other words, in each minimal pair, a tense vowel /i:/ was given the duration value of its lax counterpart /ɪ/ whilst the lax vowel was given the duration value of its corresponding tense vocalic sound. The manipulation of duration was performed by obtaining the relative duration of the vowels to be modified, whose value stemmed from the division between the target duration and the real duration figures (i.e.: relative_dur.= target_dur/real_dur). Paralleling the number of natural tokens, at the end of this first phase of vowel duration modification, 120 manipulated tokens were obtained, which, together with the former 120 natural ones, added up to a total of 240 items to be used in Experiment I. The second phase of vowel modification was based on the neutralization of duration, which was performed by obtaining the means of both vowels in each minimal pair and subsequently calculating their relative duration value, which eventually led to the neutralized duration for the two vowels in each minimal pair. Hence, and again, a total sum of 120 items was obtained and, taking the former natural tokens as items for later comparison, 240 tokens were available for Experiment II.

Materials and Procedure

In order to develop both experimental tasks, the DMDX display software (Forster & Forster, 2003) was used. As briefly noted earlier on in this paper, Experiment I required participants to decide between a pair of words, which were only distinguishable by a phoneme, upon hearing one of them through headphones in a quiet computer room. Participants were faced with forced-choice identifications within 5 seconds after having been aurally exposed to the input. If they failed to answer, the obtained score for that item was negative and the next pair of words was presented on the computer screen. Instructions appeared visually and in the target language, i.e.: English, and previous piloting of the task had estimated 15 minutes of test performance. The task began after some examples had been provided for practice.

As for the second experiment, the word discrimination task, three-word strings were also presented aurally through headphones following an AXB combination with 1-second interstimulus interval (ISI), which rendered 4 alternative vowel orders for each of the ten minimal pairs (/i:/-/i:/-/ ɪ/; /i:/-/ ɪ-/ɪ/; /ɪ/-/i:/-/i:/, and /ɪ/-/ɪ/-/i:/). Hence, the final total number of triads for this latter task was 480 (10x4x6x2). As with the former experiment, instructions appeared in English on the screen and some examples were provided for participants to understand the dynamics of the test. Again, listeners had 5 seconds to respond before the next triad was presented and if they failed to do so, negative scores were automatically obtained for that stimulus. Previous piloting of this latter task had estimated 40 minutes of task performance.

Both experiments took place in the same session and in the same order, the word identification task being the first test to be taken. The decision of not having a counterbalanced order was reached after having considered that participants would be easier to control if they all followed the same test sequence (especially if it is taken into account that there was only one researcher in the room) and because having the word discrimination task first could sensitize listeners to quality values and influence their performance in the word identification task, which exposed them to both duration-natural and duration-manipulated stimuli. However, placing the word identification task at the beginning was not thought to have any impact on the participants' subsequent performance in the discrimination test, precisely because in the latter duration appeared either in conjunction with spectral cues (in natural tokens) or was stabilized.

After the completion of the tests, participants were required to fill in a personal questionnaire that elicited information on both their linguistic background and their language use and L2 experience.

4. Data Analysis

Word Identification Task

Data for the first test, the word identification task, were preliminarily explored for normal distribution by means of descriptive, Kolmogorov-Smirnov (K-S) statistics. The dependent variables (the /i:/ and /ɪ/ vowels) were classified into two subgroups according to the initial distinction that determined the native speaker selection; that is, according to vowel duration (i.e.: the tense-lax dichotomy) in natural and manipulated tokens and the postvocalic consonants' voicing state. The independent variables were kept constant through this exploration and were classified regarding "L1" (Catalan, Spanish or Catalan & Spanish), "Bilingual Status" ("Balanced Bilingual", "Catalan Dominant" or "Spanish Dominant"), "Self-reported or Estimated Proficiency in Listening and Speaking in the L2" (high vs. low) and "Previous Phonetic Training in English" (yes vs. no).

A second statistical step in the data analysis procedure was to determine the existence of possible correlations between variables and to establish the direction of those correlations, for which Pearson product-moment correlation coefficients were calculated. A paired-samples t-test was then run to test for significant differences in the percentage of correct identification of natural and manipulated English vowels, which constituted the first goal of this study.

The effect of individual variables on the percentage of correct vowel identification was analysed by means of both independent-samples t-tests and one-way analysis of variance (ANOVA), depending on the number of levels each individual factor was divided into.

Word Discrimination Task

The second test, the word discrimination task, was also analysed for both normal distribution by means of K-S statistics and for Pearson correlations. Subsequent non-parametric tests were used in this case, as data were not normally distributed: a Wilcoxon Signed Ranks test was applied to analyse the means of all vowel-related variables and a Mann-Whitney U test was then run in order to analyse the interaction between individual variables and the scores on correct vowel discrimination, as well as to compare the performance of both non-native and native speakers of English.

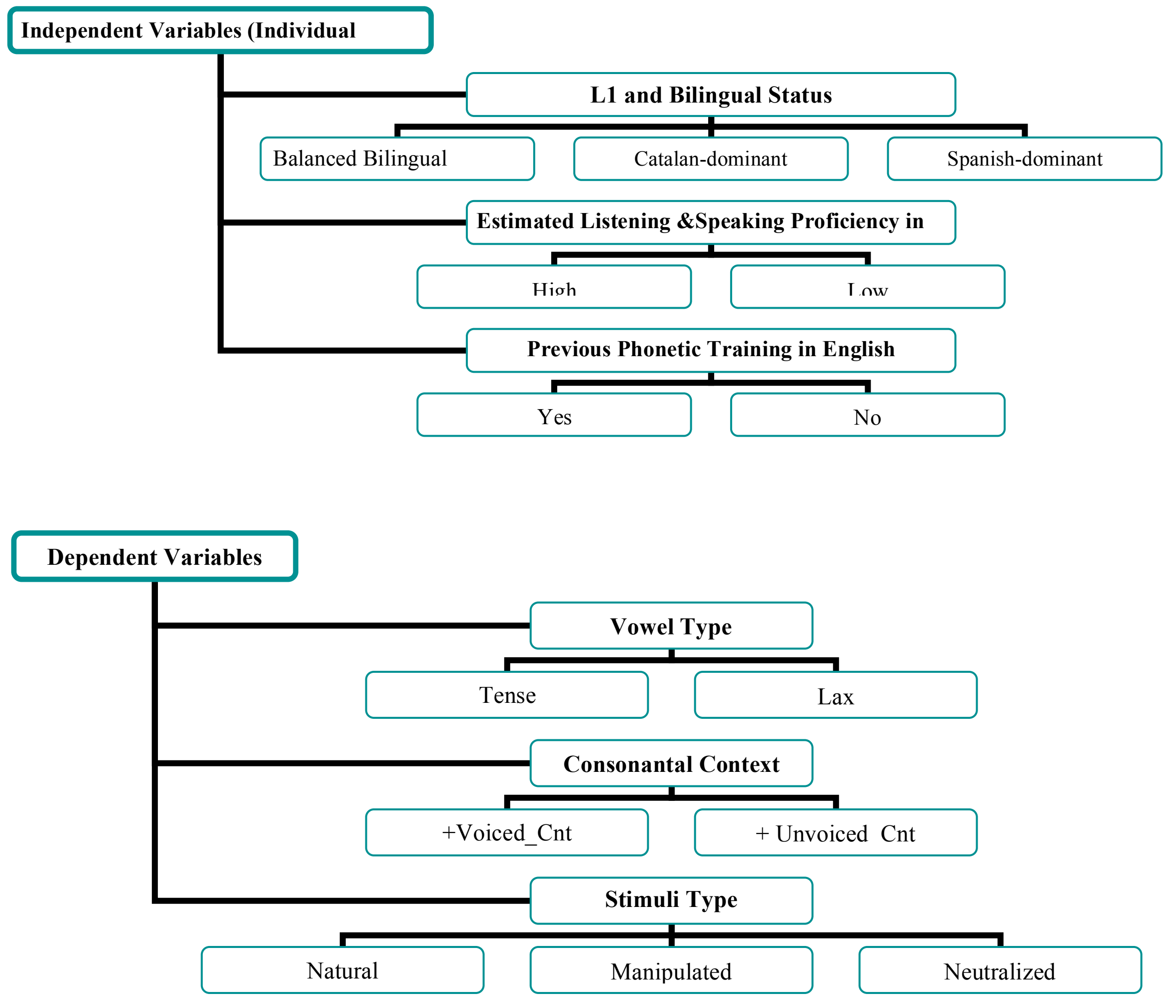


Figure2. List of Dependent and Independent Variables

5. Results

Word Identification Task

Data were deemed to be generally normally distributed. As for the presence of outliers, no extreme scores were found.

All correlations obtained for both vowel tenseness versus vowel laxness and duration-natural as opposed to duration-manipulated tokens were positive and strong and significant at the 0.01 level ($r > .5$). Hence, and as expected, the better L2 listeners were able to perceive the /i:/-/ɪ/ phonemic contrast in natural tokens, the better their categorization of the target vowels when these had been manipulated for duration. Instances of those correlations were as follows:

	tense_man	lax_nat	lax_man
tense_nat	,851**	,538**	
tense_man			,681**
lax_nat			,632**

Table 3. Pearson Correlations for NNS (*N*=62). **Correlation is significant at the 0.01 level (2-tailed)

Pearson correlations were also run combining vowel variables and information on learners' linguistic background that had been elicited by means of a personal questionnaire. Thus, the aforementioned vowel-related variables were explored for possible correlations with "Bilingual Status", "Estimated Listening & Speaking Proficiency in the L2" and "Phonetic Training in English". Nonetheless, no significant correlations were found between these individual variables and the percentage of correct vowel identification.

Regarding the results yielded by the paired-samples t-test, and as expected, Spanish and Catalan learners of English overrelied on duration when perceiving the target contrast and, actually, differences in the percentage of correct vowel identification between natural and manipulated tokens were statistically significant. Hence, L2 learners made more errors when identifying manipulated vowels (overall correct scores for manipulated vowel identification: *M*=47,19, *SD*=16,29, *t*(61)=19,68), whose original duration values had been interchanged for those of the other vowel (tense for lax and lax for tense, respectively), than when the locus of attention was drawn to natural tokens (overall correct scores for natural vowel identification: *M*=73,43, *SD*=12,39, *t*(61)=19,68), those whose temporal values had been left unchanged. These results, therefore, are in accordance with those obtained in Ylinen et al.'s 2009 study with Finnish L2 learners of English.

			Means		Sig. (2-tailed)	
			NNS	NS	NNS	NS
	vowels	nat	73,43 (12,39)	96,50 (2,24)		
		man	47,19 (16,39)	92,67 (4,22)	,000	,056
	tense	nat	70,47 (15,22)	94,00 (3,03)		
		man	44,48 (20,94)	89,00 (6,93)	,000	,194
	lax	nat	76,39 (13)	99,00 (1,49)		
		man	49,89 (14,72)	96,33 (6,39)	,000	,317
tense	nat	voiced	84,60 (10,71)	93,89 (4,56)		
		voiceless	49,26 (28,49)	94,17 (8,64)	,000	,960
tense	man	voiced	47,18 (17,77)	83,89 (10,83)		
		voiceless	40,46 (28,67)	96,67 (3,49)	,005	,055
lax	nat	voiced	74,37 (12,6)	98,89 (1,52)		
		voiceless	79,44 (16,25)	99,17 (1,86)	,001	,704
lax	man	voiced	34,32 (17,53)	95,56 (6,97)		
		voiceless	73,25 (18,19)	97,50 (5,59)	,000	,080

Table 4. Means (SD in parenthesis) and Paired-Samples T-Test results for NNS (*N*=62) and NS (*N*=5)

Furthermore, significant differences were also observed for vowel duration as affected by postvocalic consonants' voicing state. In the case of manipulated vowels, it is worth mentioning that only the lax /ɪ/ sound followed by a voiceless consonant appeared to have a high percentage of correct identification on average (*M*=73,25 , *SD*= 16,25, *t*(61)=-15) whilst all other modified tokens did not even reach chance level, which means that listeners were misled by the process of vowel duration manipulation. Incorrect identification was also observable for natural, tense vowels when followed by an unvoiced consonant (*M*=49,26 , *SD*=28,49, *t*(61)=10,71), as the latter shortened the tense /i:/ sound by means of pre-fortis clipping.

The results for this group of L2 learners were compared to those obtained by the control group (*N*=5) of native speakers and later illustrated by means of a series of boxplots, which classified the two groups using the "Native" criterion and which are available in Appendix C. As expected, and because temporal cues do not have the same weight as spectral cues in vowel perception for native speakers of English, the manipulation of vowel duration did not affect the native listeners' percentage of correct vowel identification to a significant extent. Furthermore, the differences between the phonemic (the tense versus lax distinction) and the phonetic (voiced versus voiceless postvocalic consonants) contrasts were minimal and, therefore, did not reach statistical significance.

It is worth mentioning that only the difference between the overall correct scores for duration-natural vowel identification when focused on the tense versus lax distinction reached statistical significance in the native group (M=94 vs. 99, SD= 3,03 vs. 1,49, $t(4)$=-6,7, p=,003). Such difference was not significant, however, when manipulated tokens or the consonantal context in which the vowel was embedded were taken into consideration. It is also interesting to note, however, that -albeit not statistically significant- native speakers' overall scores for correct vowel identification were lower for both manipulated tense vowels (M=89, SD=6,93) and manipulated tense vowels when followed by a voiced consonant (M=83,89, SD=10,83), the latter percentage of correct identification being even lower due to pre-fortis clipping. As for the role played by dialectal variation among the native speakers, it was reflected in the results in that British speakers made slightly more errors when perceiving manipulated vowels than American speakers did, although that difference in performance was minor and both subgroups reached high percentages of correct vowel identification.

Independent-samples t-tests were run to find out whether there were significant differences among vowel variables and individual, two-level factors: "Estimated Proficiency Level" (high vs. low), "Phonetic Training in English" (yes vs. no) and "Bilingual Status" (Spanish-dominant vs. Balanced bilingual/Catalan-dominant). This latter variable was included as it was believed that, because the Spanish vocalic system is smaller than the Catalan one, a possible difference in scores among people differing in their daily use of Spanish and Catalan could be expected. However, none of the three individual variables applied proved to be statistically significant for the current analysis. In a similar vein, an ANOVA test was also run to find out whether there were significant differences among scores in correct vowel identification and individual variables with more than two levels of classification ("L1" and "Bilingual Status"), but no significant results were here obtained either.

Word Discrimination Task

Kolmogorov-Smirnov statistics proved data not to be normally distributed (as highly significant results (<.05) were obtained). Rather, figures were highly skewed to the right, indicating very high scores on vowel discrimination, which was due to subjects performing at ceiling. Therefore, non-parametric tests were applied in further analyses. Regarding Pearson correlations, and as had occurred with the word identification task, these were positive, strong and highly significant (r >.5). The variables that were used were based on vowel type (tense or lax) and stimuli type (natural or neutralized); as table 5 shows, and taking into account the high percentages of correct vowel discrimination obtained by all 58 participants, listeners could discriminate both duration-natural and duration-neutralized tokens equally well, which means that no enhancing effect was apparently at work with the neutralized tokens. Nonetheless, in the current case these results were not deemed to reflect reality, as such a high degree of significance was undoubtedly influenced by the aforementioned ceiling effects. The percentage of correct vowel discrimination was then explored in conjunction with three individual factors ("Bilingual Status", "Estimated Listening & Speaking Proficiency in the L2" and "Phonetic Training in English"), but no significant correlations were found.

	tense_neu	**lax_nat**	**lax_neu**
tense_nat	,949**	,827**	
tense_neu			,859**
lax_nat			,965**

Table 5. Pearson Correlations for NNS (N=58). **Correlation is significant at the 0.01 level (2-tailed)

A Wilcoxon Signed Ranks Test was run to find out possible significant differences among the three different subsets of vowel-related variables (vowel type, consonantal context and stimuli type). As can be observed in Table 6, differences between the overall scores for correct duration-natural and duration-neutralized vowel discrimination are statistically non-significant. That is, vowels that had been reproduced as natural tokens and vowels that had undergone a process of duration neutralization were both perceived equally well (overall correct scores for natural vowel discrimination was:

M=97,01, *SD*=5,64 and overall correct scores for neutralized vowel discrimination was: *M*=96,88, *SD*=5,36). This high percentage of correct vowel discrimination was also maintained when vowel type was detailed: tense_neu (*M*=96,81, *SD*=4,72); tense_nat (*M*=96,65, *SD*=6,05); lax_neu (*M*=96,95, *SD*=6,40) and lax_nat (*M*=97,37, *SD*=5,75). The difference in vowel discrimination, however, was significant ($p = <.05$) when the consonantal context was specified, both for tense and lax vowels when followed by either a voiced or voiceless consonantal sound (*M* range=96,09-97.84; *SD* range: 4,51-7,25). Comparing these results to the ones obtained by the control group, and upholding the view that English native speakers make use of duration cues to a much lesser extent, the difference in the scores obtained for both natural and neutralized vowels was not significant.

A Mann-Whitney U Test followed, which was used to explore the relationship between individual variables ("Estimated Listening & Speaking Proficiency in the L2", "Phonetic Training in English" and "Bilingual Status") and scores on correct vowel discrimination and to compare the test performance by non-native and native speakers. All these interactions, however, were not statistically significant. Regarding the native versus non-native interaction, the Mann-Whitney U Test revealed no significant differences between the two groups for percentage of correct vowel discrimination, due to the fact that the non-native group had performed at ceiling; hence, any comparison on the grounds of the "Native-Non-Native" distinction was futile (overall correct scores for natural vowel discrimination for non-native speakers: *M*=97,01, *SD*=5,64 and overall correct scores for neutralized vowel discrimination for non-native speakers: *M*=96,88, *SD*=5,36; overall correct scores for natural vowel discrimination for native speakers: *M*=96, *SD*=6,64 and overall correct scores for neutralized vowel discrimination for native speakers: *M*=97,16, *SD*=3,85) .

In order to analyse data in more depth, reaction times for the word discrimination task were also explored. The premise was that, albeit percentage of correct vowel discrimination did not convey reliable information due to the ceiling effects, reaction times could provide an explanation about the relationship among the different types of vowel-related variables and between these and listeners' individual factors. In other words, it was believed that reactions times could also be informative about the degree of difficulty involved in the discrimination of different English minimal pairs and about the interaction between vowel perception and several individual variables. Thus, and once more, K-S statistics were applied to check for normality values, which were in this case both normal and skewed. However, and because some normal distribution was indeed found, parametric tests were run for these analyses.

A paired-samples t-test displayed non-significant differences in scores for either general vowel values (neutralized vowels: *M*=469,65, *SD*=147,86 versus natural vowels: *M*=461,77, *SD*=150,89713, *t*(56)=129, *p*= >.05) or consonantal context, both voiceless and voiced (*p*= >.05). Nevertheless, this lack of statistical significance is in tune with the results obtained for correct scores in the word discrimination task, which means that reaction times are also uninformative due to the participants' performance at ceiling. Independent-samples t-tests and one-way ANOVAs were finally run to observe possible differences between individual variables (two- and three-level factors, respectively) and correct scores on vowel discrimination. No statistically significant values were obtained, however.

No comparison was developed between the word identification task and the word discrimination task because the latter's ceiling effects rendered that comparison misleading and unreliable; in other words, and on the whole, it was deemed appropriate to maintain both tests separate and analyse them as individual experiments because the participants' performance at ceiling obscured the understanding of both the processes of vowel discrimination and the interaction between them and the different individual factors that served as independent variables.

6. General Discussion

In accordance with previous research (Cebrian, 2006; Escudero and Boersma, 2004; Ylinen et al., 2009, among others) and as proved by Experiment I, the word identification task, Spanish and Catalan speaking learners of L2 English were shown to overrely on temporal cues when identifying the English phonemic contrast /i:/-/ ɪ/, despite not using that acoustic parameter phonologically in their L1. Hence, and corroborating the study's initial prediction, vowel duration manipulation did affect L2 listeners' perception of the targeted contrast, leading them to make more errors than when they were being exposed to natural tokens of the same vowels. This was so despite the phonetic variability in the stimuli, which had been included in the study to provide for more diverse and realistic data in terms of duration and vowel formants (F1 and F2). Going back to Bohn's Desensitization Hypothesis, Spanish and Catalan listeners would be desensitized to the spectral values that serve to distinguish the target phonemic contrast because their L1 would not have provided them with any previous experience using quality cues to perceive those two English vowels that occupy a portion of the vocalic space shared by their L1 speech sound /i:/. Therefore, they would be forced to resort to duration acoustic parameters when being aurally exposed to English /i:/ and /ɪ/, which explains their being misled by vowel duration manipulation in the current experiment. This, and not surprisingly, was not the case for the control group of native speakers, as English speakers mainly use spectral (or quality) cues when perceiving and producing the speech sounds under focus. Dialectal variation among the control group (British versus American English) did not influence results in that all native speakers reached high levels of correct vowel identification.

As for vowel discrimination (Experiment II), which was based on the acoustic perception of both duration-natural and duration-neutralized tokens, inconclusive results were obtained, as almost all participants performed at ceiling, which not only prevented an in-depth exploration of the data but a later comparison of the two tasks. In order to achieve a more detailed insight of the results, reaction times were also measured but resulted not to be informative either. The word identification task had been designed so as to predict discrimination scores; that is, it was believed that those subjects who were less affected by vowel duration manipulation (i.e.: were more native-like in the perception of the contrast /i:/-/ɪ/ because they had developed more accurate representations for the target L2 vowel sounds) would do better in the word discrimination task, as they would not rely on duration parameters to the extent that those who performed poorly in the first experiment did. In a similar vein, this latter group of L2 listeners could not be explored for positive neutralization effects because they had also performed at ceiling in the discrimination task. As a final consequence, the study's second hypothesis -the fact that participants' performance in Experiment II would be enhanced by keeping duration constant- could not be proved and was left unanswered.

7. Conclusions: Limitations and Further Research

Time constraints were the main limitation of this experimental research, which was reflected in the study design in several ways. To begin with, an ideal design would have included the three sets of stimuli type (natural, manipulated and neutralized) in both perception tasks, in order to analyse how they were identified and discriminated by the same listeners depending on the condition they were set in (either identification or discrimination). Regarding the word discrimination task, it proved to have methodological flaws, which could not be detected in the piloting sessions, as these had been developed to control for timing and stimuli presentation mainly. As a result, the task came to be too easy and this led to the yielding of ceiling effects. Moreover, its ISI may have been too short, as it lasted one second due to the limited time allowed to conduct the experiments (1 hour). Actually, it is gauged convenient to leave 2 seconds or beyond between each stimulus presentation so as to avoid sensory memory effects, as "the longer the interstimulus delay, the greater would be the uncertainty as to the exact identity of the first vowel, and the more difficult the comparison would be" (Cowan & Morse, 1986:506). That is, making the sensory trace (Hojen & Flege, 2006) fade away by preventing listeners from resorting to their phonological short term memory would force them to retrieve their mental categorization of the target phoneme, which would yield a more precise and realistic insight of the learners' own L2 category representations. Therefore, in optimal conditions (limited time constraints and unlimited resources) these two tasks would have been run so that they emulated the final experimental session to control for possible flaws in the study design that could have a detrimental impact on the results and to ensure the correct understanding of the instructions. In this way, not only would the experimental tasks be more fine-grained in their final implementation but the study's internal validity would not be threatened.

A main improvement for future research in this area would definitely be the design of the second test, the word discrimination task, which had been created using 10 real, one-syllable-long English minimal pairs. One of the reasons why ceiling effects may have been obtained is probably due to the "lexical bias hypothesis" (Mora, 2005; Pisoni et al., 1994) in that participants' lexical knowledge may have facilitated the discrimination of the contrast. Thus, and in order to avoid this, low frequency words (i.e.: words that are not normally used on a daily basis) or non-words following English phonotactics (e.g.: *drid*) could be used as acoustic input in the future. Consequently, L2 listeners would not be able to automatically activate and retrieve the word form from their mental lexicon without taking heed of the vocalic sound and would need to focus on the vowel under study in order to be able to either identify the targeted speech sound or discriminate between L2 vowels. Making the task more demanding would also be achieved by combining different words in a single triad as uttered by different speakers instead of having each triad containing the same minimal pair and being uttered by the same native speaker, which would additionally contribute to even higher stimuli variability. Moreover, and regarding the implementation of the tests, it would be valuable for subsequent research to develop the tasks in individual sessions, as participants would be controlled all throughout the process and clearly instructed to follow the design as required. Nonetheless, and under a realistic light, this latter methodological issue would entail not only taking a heavy toll on time and human resources but also possible scheduling conflicts if each participant was to take the test individually, especially when large population samples are studied (like 62 learners in the present case).

To conclude, there are also pedagogical implications to be considered. If vowel duration neutralization is proved to help L2 listeners focus their attention on spectral acoustic parameters, which are the ones mainly exploited by native speakers of English and therefore critical for native-like perception and production, then not only would their perceptual abilities improve but also their L2 speech production. Hence, controlling duration by keeping it constant (i.e: neutralized) and emphasizing the spectral dimension of the target sounds would contribute to the development of phonetic training techniques and ultimately to the enhancement of L2 speech learning.

Acknowledgements

My heartfelt thanks to my supervisor, Dr. Joan Carles Mora, for his helpful guidance and priceless advice, which made this research possible. I am also very grateful to Cristina Aliaga, as she lent me her own stimuli from her PhD Thesis, to Eva Cerviño, for her support throughout the process and for sharing with me her experience with her MA Thesis writing and to Eva Utrilla, who helped me during the data collection process, which I could not have done on my own. Last but not least, I am very grateful to all participants for their contribution as L2 listeners, to the native speakers that formed the control group for sharing their time with me and, on the whole, to everyone who has helped me in the process.

References

Aliaga-García, C. 2009. Effects of audiovisual auditory vs. articulatory training on L2 vowel category learning. *Proceedings of the PTLC 2009 Phonetics Teaching and Learning Conference*. London: Dept. of Phonetics and Linguistics, University College London.

Aliaga-García, C. 2010. measuring perceptual cue weighting after training: a comparison of auditory vs. articulatory training methods. A K. Dziubalska-Kolaczyk, M. Wrembel and M. Kul. (eds.) *New Sounds 2010 Proceedings of the 6th International Symposium on the Acquisition of Second Language Speech*, pp. 77-82. ISBN: 978-83-928167-9-9.

Best, C. T. and Tyler, M. D. 2007. "Nonnative and second-language speech perception". In O.-S. Bohn and M.J. Munro. (eds.) Language Experience in Second Language Speech Learning: In honor of James Emil Flege. 13-35. Amsterdam/Philadelphia: John Benjamins Publishing Company.

Best, C. T. 1995. "A Direct Realist View of Cross-Language Speech Perception". In W. Strange. (ed.) *Speech perception and linguistic experience: Theoretical and methodological issues in cross-language speech research*. 171-206. Timonium, MD:York Press.

Boersma, P. and Escudero, P. 2005. Measuring Relative Cue Weighting: a Reply to Morrison. *Studies in Second Language Acquisition*. 27: 607-617.

Boersma, P. and Weenink, D. 2007. *Praat: doing phonetics by computer* (version 5.1.27). http://www.praat.org/

Bohn, O.-S. 1995. "Cross-Language Speech Perception in Adults: First Language Transfer Doesn't Tell it All". In W. Strange. (ed.) *Speech perception and linguistic experience: Theoretical and methodological issues in cross-language speech research*. 279-304. Baltimore: York Press.

Bohn, O.-S. and Flege, J. E. 1990. "Perception and Production of a new vowel category by adult second language learners". In J. Leather and A. James. (eds.) *New Sounds 90: Proceedings of the 1990 Amsterdam Symposium on the Acquisition of Second Language Speech*. 37-56. Amsterdam: University of Amsterdam Press.

Bohn, O.-S. and Flege, J. E. 1990. Interlingual identification and the role of foreign language experience in L2 vowel perception. *Applied Psycholinguistics*. 11: 303-328.

Cebrian, J. 2006. Experience and the use of non-native duration in L2 vowel categorization. *Journal of Phonetics*. 34: 372-387.

Cerviño, E. 2008. "Investigating the effect of duration manipulation on Spanish/Catalan learners' perception of English /◌ˊ/ /-/⋂/ and word-final /˩/-/□/". Unpublished MA Thesis, University of Barcelona.

Cowan, N. and Morse, P. A. 1986. The use of auditory and phonetic memory in vowel discrimination. *Journal of the Acoustical Society of America*. 79/2: 500-507.

Escudero, P. and Boersma, P. 2004. Bridging the gap between L2 speech perception research and phonological theory. *Studies in Second Language Acquisition*. 26: 551-585.

Flege, J. E. 1993. Production and perception of a novel, second-language phonetic contrast. *Journal of the Acoustical Society of America*. 93: 1589-1608.

Flege, J. E. 1995. "Second-language Speech Learning: Theory, Findings, and Problems". In W. Strange. (ed.) *Speech Perception and Linguistic Experience: Issues in Cross-language research*. 229-273. Timonium, MD: York Press.

Flege, J. E. 2009. "Give Input a Chance". In T. Piske and M. Young-Scholten. (eds.) *Input Matters in SLA*.175-191. Bristol: Multilingual Matters.

Forster, K. I. and Forster, J. C. 2003. DMDX: A Windows display program with millisecond accuracy. *Behavior Research Methods, Instruments, & Computers*. 35/1: 116-124.

Hojen, A. and Flege, J. E. 2006. Early learners' discrimination of second-language vowels. *Journal of the Acoustical Society of America*. 119/5: 3072-3084.

Holt, L. L. and Lotto, A. J. 2006. Cue weighting in auditory categorization: Implications for first and second language acquisition. *Journal of the Acoustical Society of America*. 119/5: 3059-3071.

Iverson, P. and Kuhl, P. K. 1996. Influences of phonetic identification and category goodness on American listeners' perception of /r/ and /l/. *Journal of the Acoustical Society of America*. 99: 1130-1140.

Iverson, P. and Kuhl, P. K. 2000. Perceptual magnet and phoneme boundary effects in speech perception: Do they arise from a common mechanism? *Perception and Psychophysics*. 62: 874-886.

Kuhl, P. K.; Williams, K. A.; Lacerda, F.; Stevens, K. N. and Lindblom, B. 1992. Linguistic experience alters phonetic perception in infants by 6 months of age. *Science*. 255: 606-608.

Kuhl, P. K. 1994. Learning and representation in speech and language. *Current Opinion in Neurobiology*. 4: 812-822.

Kuhl, P. K.; Conboy, B. T.; Coffey-Corina, S.; Padden, D.; Rivera-Gaxiola, M. and Nelson, T. 2008. Phonetic learning as a patway to language: new data and native language magnet theory expanded (NLM-e). *Philosophical Transactions of the Royal Society B*. 363: 979-1000.

Major, R. C. 2008. “Transfer in second language phonology (a review)”. In J.G. Hansen Edwards and M. Zampini. (eds.) *Phonology and Second Language Acquisition*. 63-95. Philadelphia: John Benjamins Publishing Company.

McAllister, R.; Flege, J. E. and Piske, T. 2002. The influence of L1 on the acquisition of Swedish quantity by native speakers of Spanish, English and Estonian. *Journal of Phonetics*. 30: 229-258.

McAllister, R. 2007. Strategies for realization of L2-Categories: English /s/-/z/. In O. Bohn and M. J.Munro. (eds.) *Language Experience in Second Language Speech Learning: In honor of James Emil Flege*. 153-167. Amsterdam/Philadelphia: John Benjamins Publishing Company.

Mora, J. C. 2005. Lexical Knowledge Effects on the Discrimination of Non-Native Phonemic Contrasts in Words and Non-Words by Spanish/Catalan Bilingual Learners of English. *Proceedings of the ISCA Workshop on Plasticity in Speech Perception*, pp. 43-48. London, United Kingdom.

Mora, J. C. and Fullana, N. 2007. Production and perception of English /ʘ/ /-/⋂/ and /∵/-/⌈/ in a formal setting: Investigating the effects of experience and starting age. *Proceedings of the 16th International Congress of the Phonetics Sciences*, pp. 1613-1616. Saarbrücken, Germany.

Pisoni, D. B., Lively, E.S. and Logan, J.S. 1994. "Perceptual learning of nonnative speech contrasts: implications for theories of speech perception". In J.C. Goodman and H.C. Nusbaum. (eds.) *The Development of Speech Perception: the Transition from Speech Sounds to Spoken Words*. 93-120. The MIT Press, Cambridge: Mass.

Ylinen, S., Uther, M., Latvala, A., Vepasalainen, S., Iverson, P., Akahane-Yamada, R. and Naatanen, R. 2010. Training the Brain to Weight Speech Cues Differently: A Study of Finnish Second-Language Users of English. *Journal of Cognitive Neuroscience*. 22/6: 1319-1332.

APPENDIX A

LINGUISTIC BACKGROUND QUESTIONNAIRE

Family Name: **Name:** **Age:**

e-mail:

Language Background:

1. Mother tongue(s) (First Language or L1):

Catalan ☐ Spanish ☐ Catalan and Spanish ☐ Other (specify):

2. As a bilingual speaker, I regard myself as:

Balanced bilingual ☐ Dominant in Catalan ☐ Dominant in Spanish ☐ Other (specify):

3. Home language:

Catalan ☐ Spanish ☐ Catalan and Spanish ☐ Other (specify):

4. Parents' First Language (mother tongue or L1):

Catalan ☐ Spanish ☐ Catalan and Spanish ☐ Other (specify):

5. Second Language (L2) learnt after age 5:

Catalan ☐ Spanish ☐ Catalan and Spanish ☐ Other (specify):

6. Third Language (L3) learnt after age 5:

Catalan ☐ Spanish ☐ Catalan and Spanish ☐ Other (specify):

7. I am fluent in the following languages:

Catalan ☐ Spanish ☐ English ☐ Other(specify):

Language Use and L2 experience:

8. Estimate the % of daily use of the following languages:

Catalan: ___% + Spanish: ___% = 100%

9. Estimate the % of daily use of the following languages:

Catalan: ____% + Spanish: ____% + English: ____% + Other (............): ___% = 100%

10. Specify if you have ever been abroad in an English-speaking country (where, when, for how long and the purpose of that stay):

11. Rate your command of English according to the following scale:

1 (very poor) - 9 (near-native)

Reading: 1☐ 2☐ 3☐ 4☐ 5☐ 6☐ 7☐ 8☐ 9☐

Listening: 1☐ 2☐ 3☐ 4☐ 5☐ 6☐ 7☐ 8☐ 9☐

Speaking: 1☐ 2☐ 3☐ 4☐ 5☐ 6☐ 7☐ 8☐ 9☐

Writing: 1☐ 2☐ 3☐ 4☐ 5☐ 6☐ 7☐ 8☐ 9☐

12. I normally use English outside the university setting: Yes ☐ No☐

With my family ☐ Hours per week _____

With native friends ☐ Hours per week _____

Reading/ Watching English books/films ☐ Hours per week _____

13. Specify if you have previously received any type of Phonetic Training in English:

Informed Consent

Any information of data which is obtained from you during this research/study which can be identified with you will be treated confidentially. I understand what is involved in this research/study and I agree to participate.

____

Signature

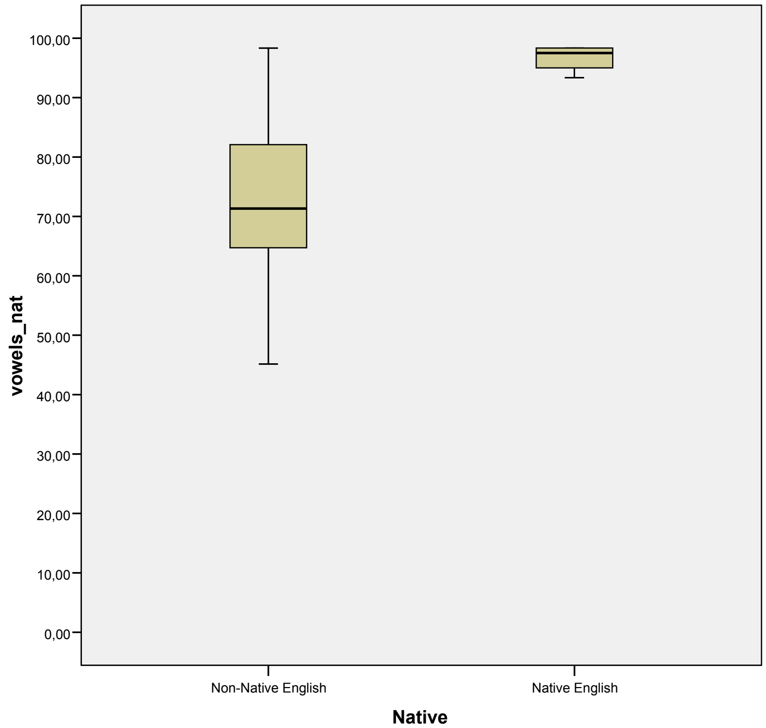
APPENDIX B

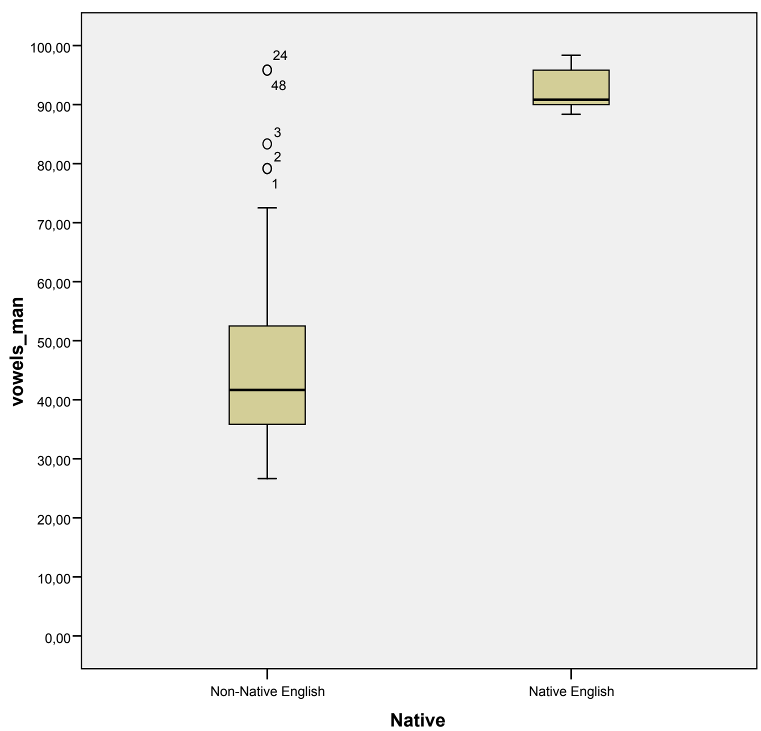
Native Speaker	FINAL MINIMAL PAIRS								
NS_1	**Bead** 498.6 ms F1:221.59 F2:2651.59	**Bid** 168.63 ms F1:383.76 F2:2236.7	**Beat** 167.23 ms F1:279.62 F2:2579.1	**Bit** 97 ms F1:423.89 F2:2284.6	**Deed** 496.62 ms F1:220.68 F2:2475.3	**Did** 177.58 ms F1:383.75 F2:2316.3	**Feel** 214.03 ms F1:304.33 F2:2588.9	**Fill** 139.09ms F1:447.9 F2:2002.1	**He_d** 452.42 ms F1:231 F2:2490.6
NS_2	**Bead** 429ms F1:279.47 F2:2616.4	**Bid** 190 ms F1:370.17 F2:2036.8	**Beat** 205 ms F1:240.83 F2:2457.3	**Bit** 192 ms F1:405.71 F2:2052.9	**Deed** 437 ms F1:278.39 F2:2639.2	**Did** 232 ms F1:441.73 F2:1954.4	**Feel** 240.26 ms F1:253.74 F2:2533.5	**Fill** 138.06 ms F1:445.06 F2:1846.9	**He_d** 380.46 ms F1:226.06 F2:2655.9
NS_3	**Bead** 430.61 ms F1:295.56 F2:2136.1	**Bid** 182.85 ms F1:403.25 F2:1964.5	**Beat** 158.81 ms F1:351.34 F2: 2141.6	**Bit** 148.37 ms F1:471.37 F2:2384.2	**Deed** 393.66 ms F1: 308.71 F2:2114.4	**Did** 216.56 ms F1:415.68 F2:2078.4	**Feel** 255.36 ms F1:348.53 F2:2094.9	**Fill** 158.18 ms F1:495.19 F2:1839.6	**He_d** 375.7 ms F1:315.39 F2:2142.5
NS_4	**Bead** 619.08 ms F1:439.68 F2:2798.1	**Bid** 183.49 ms F1:506.67 F2:2386	**Beat** 163.2 ms F1:442.55 F2:2708.3	**Bit** 143.37 ms F1:497.61 F2:2384.2	**Deed** 384.42 ms F1: 415.6 F2:2824.2	**Did** 187.62 ms F1:472.31 F2:2421.2	**Feel** 181.25 ms F1:435.91 F2:2574.4	**Fill** 116.86 ms F1:572.81 F2:2194.6	**He_d** 369.54 ms F1:481.08 F2:2911.7
NS_5	**Bead** 619.08 ms F1:324.51 F2:2830	**Bid** 275.38 ms F1:482.65 F2:2257	**Beat** 156.5 ms F1:300.42 F2:2970	**Bit** 83.98 ms F1:508.09 F2:2295.6	**Deed** 549.25 ms F1:316.32 F2:2909.9	**Did** 187.55 ms F1:457.45 F2:2213.1	**Feel** 222.67 ms F1:386.05 F2:2797.4	**Fill** 76.73 ms F1:566.1 F2:1920.4	**He_d** 369.54 ms F1:359.35 F2:2865.6
NS_6	**Bead** 259.13 ms F1:339.3 F2:2753.2	**Bid** 125.52 ms F1:451.9 F2:2282	**Beat** 110.45 ms F1:397.06 F2:2664.6	**Bit** 86.23 ms F1:532.2 F2:2227.5	**Deed** 250.23 ms F1:372.37 F2:2760.3	**Did** 109.11 ms F1:409.23 F2:2236	**Feel** 136.89 ms F1:453.21 F2:2268.6	**Fill** 115.5 ms F1:496.72 F2:2107.1	**He_d** 214.02 ms F1:327.58 F2:2707.6
MEANS	**Bead** 438.79 ms F1:316.68 F2:2630.9	**Bid** 187.64 ms F1:433.06 F2:2193.9	**Beat** 160.19 ms F1:335.3 F2:2586.8	**Bit** 125.15 ms F1:473.14 F2:2206.5	**Deed** 418.53 ms F1:318.67 F2:2620.5	**Did** 185.07 ms F1:430.02 F2:2203.2	**Feel** 208.41 ms F1:363.62 F2:2476.3	**Fill** 124.07 ms F1:503.96 F2:1985.1	**He_d** 385.81 ms F1:263.51 F2:2629

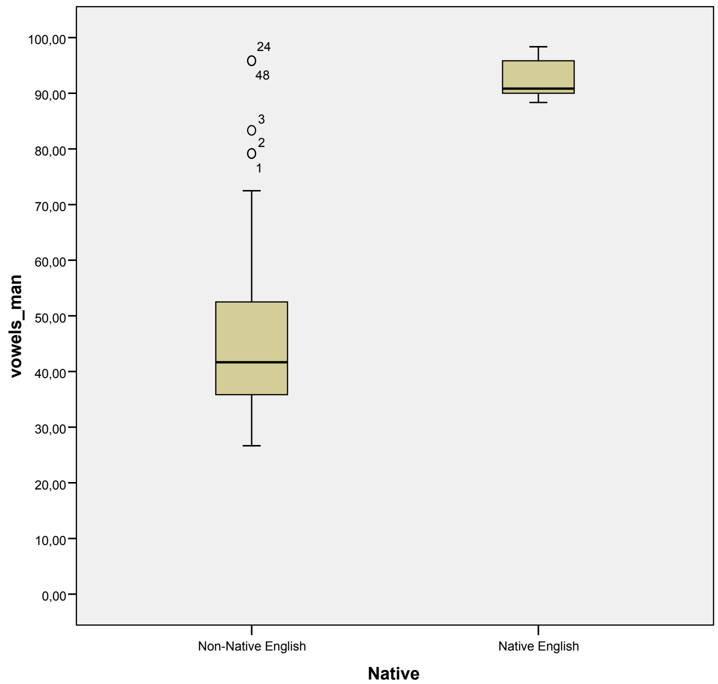
Hid 157.3 ms F1:395.7 F2:2369.7	**Heel** 238.55 ms F1:311.74 F2:2482.8	**Hill** 97.28 ms F1:412.78 F2:2109.7	**Peak** 136.03 ms F1:308.4 F2:2605.8	**Pick** 86 ms F1:466.74 F2:2400.8	**Peat** 139.96 ms F1:309.79 F2:2547	**Pit** 90 ms F1:421.9 F2:2189.4	**Seed** 437.51 ms F1:267.7 F2:2704.1	**Sid** 166.34 ms F1:383.04 F2:2266	**Teak** 149 ms F1:312.12 F2:2570.4	**Tick** 80.49 ms F1:471.95 F2:2205.2
Hid 184.03ms F1:374.82 F2:2302.9	**Heel** 306.86 ms F1:251.57 F2:2645.1	**Hill** 132.72 ms F1:368.75 F2:2177.6	**Peak** 212 ms F1:266.54 F2:2515.3	**Pick** 172 ms F1:419.71 F2:2114.9	**Peat** 196 ms F1:269.53 F2:2577.3	**Pit** 180 ms F1:456.29 F2:2010.8	**Seed** 355 ms F1:248.83 F2:2624.6	**Sid** 183 ms F1:378.69 F2:2040.7	**Teak** 208.1 ms F1:279.8 F2:2655.9	**Tick** 153.65 ms F1:399.4 F2:2313.1
Hid 174.88 ms F1:432.8 F2:2056.6	**Heel** 249.49 ms F1:346.26 F2:2109.5	**Hill** 147.09 ms F1:489.15 F2:1853.4	**Peak** 161.11 ms F1:308.57 F2:2122.4	**Pick** 130.35 ms F1:460.19 F2:2001.4	**Peat** 156.19 ms F1:304.67 F2:2157.6	**Pit** 129.64 ms F1:486.27 F2:2029.4	**Seed** 342.09 ms F1:320.1 F2:2150	**Sid** 160.79 ms F1:414.78 F2:2029.4	**Teak** 155.42 ms F1:327.41 F2:2039.6	**Tick** 134.36 ms F1:468.47 F2:1980.1
Hid 157.37 ms F1:453.73 F2:2529	**Heel** 173.85 ms F1:438.69 F2:2701	**Hill** 112.8 ms F1:558.64 F2:2278.9	**Peak** 167.35 ms F1:419.7 F2:2694.5	**Pick** 99.99 ms F1:549.69 F2:2432.9	**Peat** 149.12 ms F1:423.82 F2:2621.1	**Pit** 119.11 ms F1506.61: F2:2445.4	**Seed** 239.11 ms F1:439.06 F2:2730.1	**Sid** 199.25 F1:456.35 F2:2482	**Teak** 139.44 ms F1:318.36 F2:2655.2	**Tick** 130.27 ms F1:511.81 F2:2421.5
Hid 170.3 ms F1:476.74 F2:2423.2	**Heel** 280.85 ms F1:364.25 F2:2873.4	**Hill** 99.24 ms F1:593.53 F2:1964.2	**Peak** 149.25 ms F1:364.19 F2:2893.5	**Pick** 102.37 ms F1:514.9 F2:2241.7	**Peat** 137.73 ms F1:278.8 F2:2963.5	**Pit** 105.33 ms F1:530.18 F2:2252.8	**Seed** 452.01 ms F1:371.1 F2:2912.1	**Sid** 155.64 ms F1:486.09 F2:2226	**Teak** 100.09 ms F1:384.86 F2:2748.6	**Tick** 111.92 ms F1:505.09 F2:2272.1
Hid 94.36 ms F1:441.31 F2:2390.3	**Heel** 148.82 ms F1:421.81 F2:2338.5	**Hill** 112.75 ms F1:467.47 F2:2160.9	**Peak** 95.99 ms F1:382.2 F2:2758.2	**Pick** 65.34 ms F1:508.6 F2:2260.3	**Peat** 104.78 ms F1:375.19 F2:2755.3	**Pit** 88.26 ms F1:524.04 F2:2273.9	**Seed** 222.62 ms F1:384 F2:2810.9	**Sid** 118.87 ms F1:450.44 F2:2208.3	**Teak** 100.09 ms F1:384.86 F2:2748.6	**Tick** 91.22 ms F1:505.82 F2:2222.5
Hid 156.37 ms F1:429.19 F2:2345.3	**Heel** 233.07 ms F1:355.72 F2:2525.1	**Hill** 112.75 ms F1:481.72 F2:2090.7	**Peak** 153.62 ms F1:341.6 F2:2598.3	**Pick** 109.34 ms F1:486.63 F2:2242	**Peat** 147.29 ms F1:326.85 F2:2603.6	**Pit** 118.72 ms F1:487.55 F2:2200.3	**Seed** 341.39 ms F1:338.47 F2:2655.3	**Sid** 163.98 ms F1:428.23 F2:2205.7	**Teak** 150.43 ms F1:335.61 F2:2599	**Tick** 116.88 ms F1:477.02 F2:2235.8

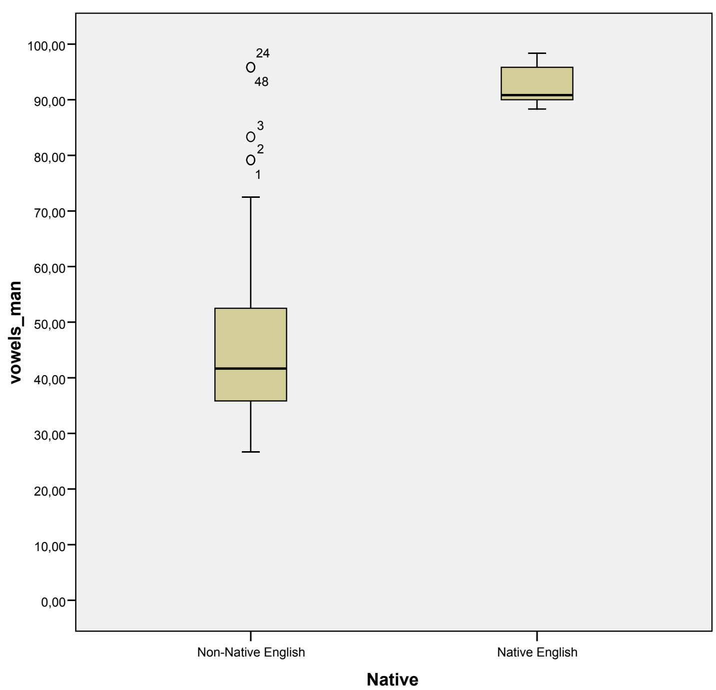
Table 5. Duration, F1 & F2 Values for 10 Minimal Pairs

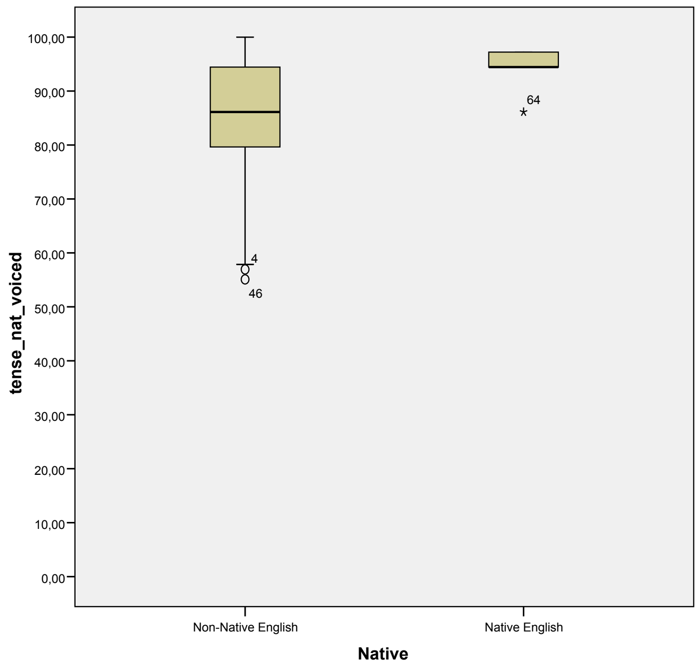
APPENDIX C

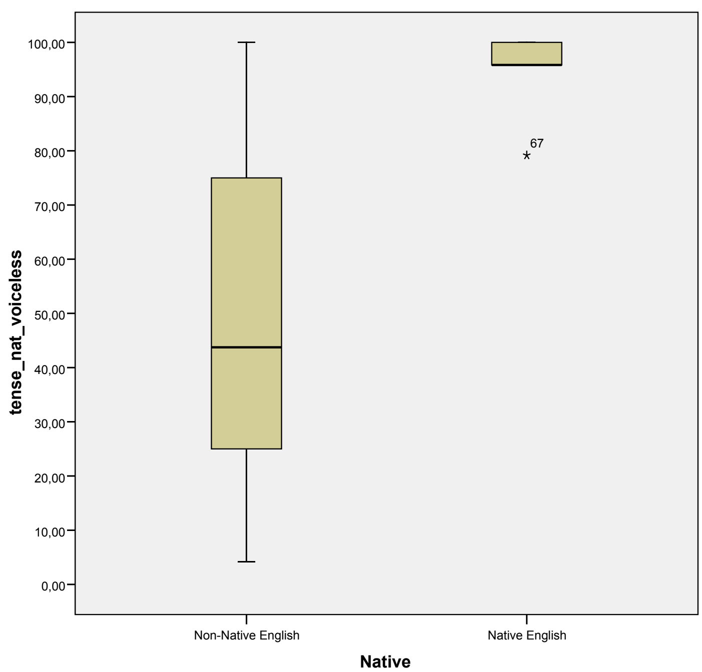

100,00
90,00
80,00
70,00
60,00
50,00
40,00
30,00
20,00
10,00
0,00
vowels_nat
Non-Native English
Native English
Native

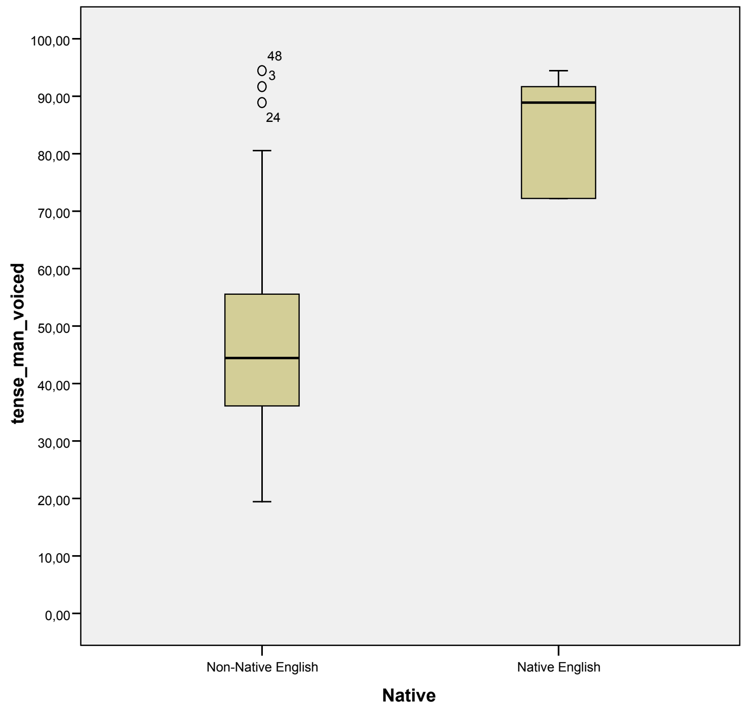

100,00
90,00
80,00
70,00
60,00
50,00
40,00
30,00
20,00
10,00
0,00
vowels_man
24
48
3
2
1
Non-Native English
Native English
Native


vowels_man
100,00
90,00
80,00
70,00
60,00
50,00
40,00
30,00
20,00
10,00
0,00
24
48
3
2
1
Non-Native English
Native English
Native

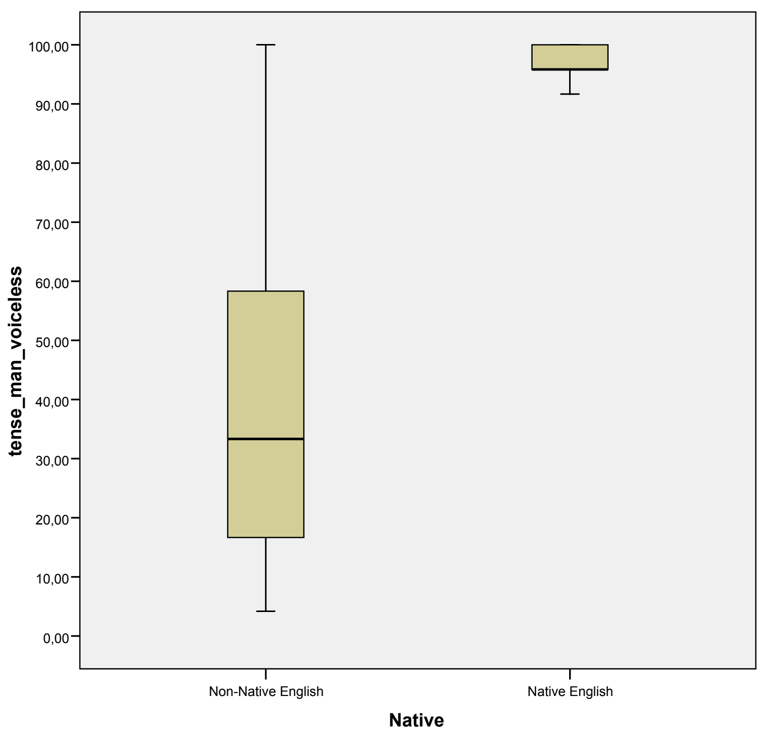

vowels_man
100,00
90,00
80,00
70,00
60,00
50,00
40,00
30,00
20,00
10,00
0,00
24
48
3
2
1
Non-Native English
Native English
Native


100,00
90,00
80,00
70,00
60,00
50,00
40,00
30,00
20,00
10,00
0,00
tense_nat_voiced
64
4
46
Non-Native English
Native English
Native

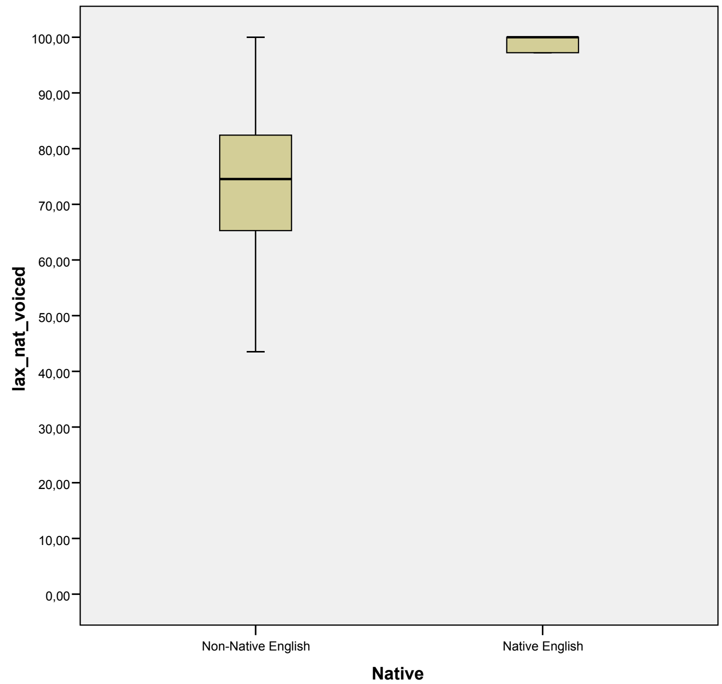

100,00
90,00
80,00
70,00
60,00
50,00
40,00
30,00
20,00
10,00
0,00
tense_nat_voiceless
67
Non-Native English
Native English
Native



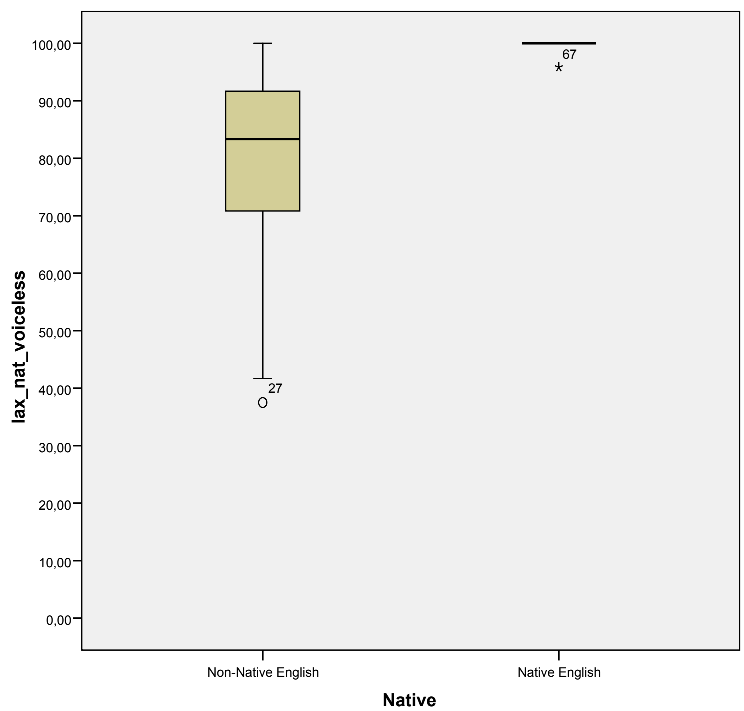
100,00
90,00
80,00
70,00
60,00
50,00
40,00
30,00
20,00
10,00
0,00
48
3
24
tense_man_voiced
Non-Native English
Native English
Native



100,00
90,00
80,00
70,00
60,00
50,00
40,00
30,00
20,00
10,00
0,00
tense_man_voiceless
Non-Native English
Native English
Native



100,00
90,00
80,00
70,00
60,00
50,00
40,00
30,00
20,00
10,00
0,00
lax_nat_voiced
Non-Native English
Native English
Native



100,00
90,00
80,00
70,00
60,00
50,00
40,00
30,00
20,00
10,00
0,00
lax_nat_voiceless
67
27
Non-Native English
Native English
Native

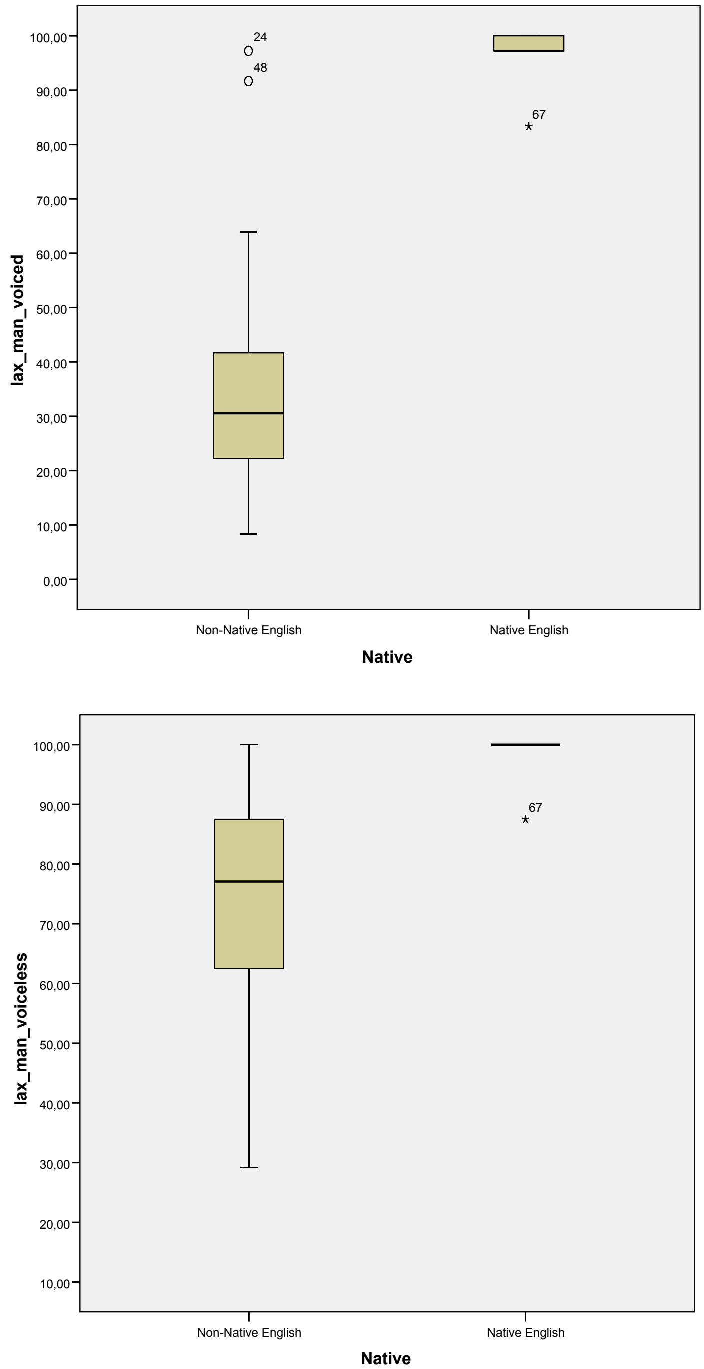


Figure 3. Boxplots Natives vs. Non-natives