

VOWEL INVENTORY SIZE MATTERS: ASSESSING CUE-WEIGHTING IN L2 VOWEL PERCEPTION

Hanna Kivistö-de Souza*

Universidade Federal de Santa Catarina (UFSC)
Florianópolis, Brazil

Angélica Carlet**

Universitat Internacional de Catalunya (UIC)
Barcelona, Spain

Izabela Anna Julkowska***

Universitat Autònoma de Barcelona
Barcelona, Spain

Anabela Rato****

University of Toronto
Toronto, Canada

Abstract

To examine whether L1 vowel inventory size could be a contributing factor to the use of temporal cues in L2 vowel perception, this study assessed the perception of English /i-ɪ/ by 66 learners of four different L1s: Danish, Portuguese, Catalan and Russian. The L2 learners performed a forced-choice identification task containing natural and duration-manipulated stimuli. Findings suggest that the participants' over-reliance on duration cues seems to be partially related to their L1 vowel inventory size. The participants with the largest L1 vowel inventory (Danish) demonstrated the most native-like vowel perception and the participants with the smallest L1 vowel inventory (Russian) over-relied on duration cues more than the other learners. Interestingly, the participants with somewhat comparable L1 vowel inventories (Portuguese and Catalan) performed similarly.

Keywords: L2 vowel perception; Vowel inventory size; Cue-weighting

* PhD in Applied Linguistics from the University of Barcelona and is an Assistant Professor at the Federal University of Santa Catarina (UFSC). Her research centers around L2 speech acquisition by looking into the individual cognitive variables that underlie L2 speech perception and production. Her e-mail address is hanna.kivistodesouza@gmail.com.

** Professor of English Phonetics and phonology at the International University of Catalonia (UIC). PhD. Research Areas: L2 Speech perception and production, phonetic training, The effect of formal instruction on L2 learners' oral skills. Her e-mail address is angelicacarlet@hotmail.com.

*** MA in Advanced English Studies from the Autonomous University of Barcelona and is a Freelance Linguist. Her main research interests include perception and production of second and third language speech, phonetic training, sociophonetics and natural language processing. Her e-mail address is julkowska.izabela@gmail.com.

**** Assistant Professor at the Department of Spanish and Portuguese of the University of Toronto. Her main research interests are L2 speech perception and production, phonetic training, and applied phonetics. Her e-mail address is asrato@gmail.com.



Introduction¹

Acquiring a target-like pronunciation in a foreign language is frequently seen as a more challenging task than acquiring second language (L2) grammar or vocabulary. One of the reasons L2 phonological acquisition is seen to be an arduous task is that it relies on “hard-wired biological processes that cannot easily be influenced by conscious learning efforts” (Jilka, 2009, p.5). Whereas native-like pronunciation is rarely the desired or realistic aim, accurate and fluent L2 speech is highly beneficial for foreign language learners in daily communicative situations.

Vowel sounds usually present great difficulties for foreign language learners and inaccurate vowel production and perception affect intelligibility. According to current influential L2 speech learning models, such as Speech Learning Model (Flege, 1995) and PAM-L2 (Best & Tyler, 2007), the perceptual similarity of the first language (L1) and L2 vowels is one of the important predictors of accurate L2 vowel acquisition. More specifically, more target-like acquisition is expected to occur when the L2 sound is perceived as different to the existing L1 speech sounds.

Some researchers have suggested that L2 vowel acquisition, apart from perceptual similarity, is affected by the learners’ L1 vowel inventory size, so that speakers of languages with large vowel inventories have a benefit over speakers of languages with smaller vowel inventories (Fox, Flege, & Munro, 1995; Frieda & Nozawa, 2007; Hacquard, Walter, & Marantz, 2007; Iverson & Evans, 2007). This is because speakers of languages with large vowel inventories would have gained experience in tuning to small-scale spectral differences from their L1, making the perception of L2 vowels an easier task. When the learners are unable to perceive the spectral differences between the L2 vowels, they may resort to duration and mistakenly employ it as a cue to discern the problematic vowels (Bohn, 1995; Escudero, 2000). For example, English instructors frequently teach that some English vowels contrast in duration (long/short) such as those in the pairs ‘feet-fit’ or ‘pool-pull’, rather than in quality or tenseness of the speech organs.

The aim of this article was to determine whether the size of L1 vowel inventory affects the use of perceptual cues used to discern between L2 vowels. In total, 62 English as a foreign language (EFL) learners were tested from four L1 backgrounds: Danish, European Portuguese, Catalan and Russian. These languages differ in the size of their vowel inventory, and also in the cues used to discern between the L1 vowels (quality only vs. quality and duration). We assumed that these differences could result in different cue-weighting strategies in the perception of English high front vowels. Whereas the role of vowel inventory on L2 vowel perception has been studied previously (e.g., Bohn, 1995; Fox et al., 1995; Hacquard et al., 2007; Iverson & Evans, 2007), to the best of our knowledge, no study has examined this particular language combination before.

1. Review of literature

1.1. Cue-weighting in English high front vowels by native and non-native speakers

The acquisition of English /i-ɪ/ has been examined amongst a wide variety of language learners. Mastering this distinction is important for EFL learners due to the high functional load of this English vowel pair. That is, many high frequency minimal pairs are distinguished by /i-ɪ/. For example, Higgins (2013) lists 466 minimal pairs. The acquisition of this vowel distinction is of particular interest for the foreign language learners of the present study as none of the L1s (Danish, Portuguese, Catalan and Russian) has a short lax vowel spectrally similar to the English /ɪ/, whereas all of them have a high front vowel similar to the English /i/. We could thus expect that accurate distinction of the vowel pair might pose some difficulties for the EFL learners of the present study.

Native speakers of American and British English have been shown to discern the /i-ɪ/ vowel pair mainly based on spectral differences, making little use of temporal cues (e.g. Bohn & Flege, 1990; Escudero & Boersma, 2004). In other words, native English speakers tend to distinguish this vowel pair based on vowel quality and not based on vowel duration. Conversely, EFL

learners from several L1 backgrounds have been shown to weight temporal cues over spectral ones (e.g. Bohn, 1995; Flege, Bohn & Jang, 1997; Ylinen et al., 2009).

Several reasons have been proposed as to why non-native English speakers seem to favor the secondary temporal cues. Some researchers have attributed the non-target like weighting of duration cues to negative L1 transfer so that when duration is employed contrastively in the L1, it can be transferred incorrectly into the L2. However, negative transfer from the L1 cannot be accounted for as the only reason for over-reliance on duration cues, since speakers of both quantity languages (Finnish: Ylinen et al., 2009; Japanese: Morrison, 2002; Korean: Flege et al., 1997) and of languages that do not use duration contrastively (Mandarin: Bohn, 1995; Flege et al., 1997; Spanish: Escudero, 2006; Russian: Kondaurova & Francis, 2008) have been shown to rely on temporal cues.

Others have suggested that insufficient instruction contributes to EFL learners' over-reliance on duration cues. Anecdotal evidence and researchers' reports (e.g. Flege et al., 1997) from several countries suggest that many EFL learners (e.g. Spanish and Brazilian Portuguese) learn that the difference between the two vowels is that of length: long/short. This might be due to the teacher's inexperience with English phonetics and phonology.

Escudero (2000) and Morrison (2008) suggest that reliance on temporal cues might be a developmental stage in the acquisition of the L2 phonology. In studying L1 Spanish EFL learners, both authors discovered that less proficient EFL learners were more inclined to rely on temporal cues than more proficient EFL learners. The authors suggest that as language proficiency and language experience increase, the learners' cue-weighting strategies tend to become more target-like.

Yet through studying L1 Spanish EFL learners, Escudero and Boersma (2004) suggested that it is more effective to add a duration distinction to the five existing L1 vowels than to acquire the new English vowels based on spectral differences. They argue that it is easier to create new categories in a dimension which is not exploited contrastively in the L1 (duration) than to split the existing L1 phonemes into several L2 phonemes.

However, even though a language would not use duration contrastively, as is the case with Spanish, language users still have some experience with duration. Kondaurova and Francis (2008) note that speakers of languages that do not use duration contrastively (L1 Russian and L1 Spanish speakers in their study) have experience with duration in their L1s through vowel lengthening before voiced consonants and in stressed syllables. Kondaurova and Francis (2008), as well as Morrison (2008), propose that this experience with allophonic vowel duration variation in the L1 could be an explanation as to why EFL learners tend to overuse temporal cues.

Another contributing reason could be perceptual salience. Temporal cues have been suggested to be more salient perceptually than spectral ones so that even if the speakers lack experience with contrastive temporal distinctions in their L1, they can still exploit this dimension in the L2 (Bohn, 1995; Bohn & Flege, 1990). Bohn's *Desensitization Hypothesis* states 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" (Bohn, 1995, p. 294-295).

The organization of the L1 vowel space *desensitizes* the language learner to spectral differences so that when the L1 does not employ small-scale spectral differences as a whole (e.g. three- or five-vowel systems) or in a given portion of the vowel space (e.g. the high front area, as in the present study), the learner has difficulties in perceiving small-scale spectral differences, such as the English /i/ and /ɪ/, and turns into duration differences instead. Consequently, following Bohn's hypothesis, speakers of languages with small vowel inventories are expected to show more difficulties in discerning differences in vowel quality than speakers of languages with large L1 vowel inventories (Bohn, 1995).

Bohn's hypothesis has received some empirical support from perception studies. Fox et al. (1995) examined L1 English and Spanish speakers' perception of English and Spanish vowels and concluded that "the structure of a listener's vowel space is significantly affected by the vowel inventory of the listener's L1"

(Fox et al., 1995, p. 2548). Similar conclusions were obtained from a study conducted by Iverson and Evans (2007) who examined L1 Spanish, French, German, and Norwegian EFL learners' perception of the English vowel system. The speakers from the languages with more complex L1 vowel systems (Norwegian and German) recognized the L2 vowels more accurately than speakers of languages with simpler vowel systems. Recent neurolinguistics research investigating the expansion of the perceptual vowel space as a function of vowel inventory size and organization has arrived at the same conclusion. The results from Hacquard et al. (2007) suggest that speakers of languages with large vowel inventories perceive the same sounds as more distinct than speakers of languages of smaller vowel inventories. The organization of the vowels within the vowel space, nevertheless, did not appear to play a role.

1.2. Vowel systems of the L1s of the present study

The participants in the study came from five L1 backgrounds: English, Danish, European Portuguese, Catalan and Russian. Our aim was to investigate languages differing in L1 vowel inventory size. Consequently we chose Danish with a large L1 vowel inventory (20+) and a crowded high front vowel space, European Portuguese and Catalan with medium-sized vowel inventories (9 & 8, respectively) with comparable high front vowel space (1 high front vowel) and Russian with a small L1 vowel inventory (5). Additionally, Danish was the only language which employs duration contrastively, and Portuguese was the only language which has nasal vowels. Interestingly, nasal vowels have been shown to be longer than oral vowels (Sampson, 1999) and it could thus be argued that speakers of languages with nasal vowels might perform differently when employing temporal cues to discern between vowels than speakers with languages which only have oral vowels. This section presents the vowel inventories of their L1s beginning with the target language.

Standard Southern British English (SSBE), the target language of the present study, has 12 vowels which are distinguished by means of spectral differences.

SSBE has two high front vowels which differ in spectral quality and tenseness: /i/ and /ɪ/. The tense high front vowel, /i/, is somewhat more fronted and more close than the lax high front vowel /ɪ/. Traditionally, the two are also described to differ in duration so that everything else being equal, /i/ is somewhat longer than /ɪ/. However, as mentioned in the previous section, native English speakers do not tend to rely on duration when differentiating between the two vowels, spectral quality being a more important cue of the vowel's identity.

Danish has at least 20 stressed monophthongs (Steinlen, 2005) distinguished primarily by means of temporal differences and by small spectral differences. Danish vowels occur in long-short pairs, namely /i-i:/, /e-e:/, /ɛ-ɛ:/, /a-a:/, /u-u:/, /o-o:/, /ɔ-ɔ:/, /y-y:/, /ø-ø:/, /œ-œ:/ . Almost all the vowels are located in the upper half of the acoustic vowel space. Danish has four high front vowels: the unrounded /i/ and /i:/ and the rounded /y/ and /y:/ . The members of each vowel pair are distinguished mainly by means of duration. According to acoustic measurements performed by Steinlen (2002), the English /ɪ/ and the Danish /e/ do not significantly differ in their quality. Moreover, her results did not find large differences between the Danish /i/ and the English /i/, and according to Bohn and Caudery (2012), the use of the Danish /i/ instead of the English /i/ does not contribute notably to a foreign accent. In examining Danish listeners' English vowel perception, Bohn and Steinlen (2003) reported that English /i/ is easily assimilated into the Danish counterpart due to its similarity, whereas the English /ɪ/ may pose more problems to Danish speakers, as it was identified with either Danish /e/ or /i/.

The European Portuguese (EP) vowel system includes nine oral vowels (/i/, /e/, /ɛ/, /ɨ/, /ɐ/, /a/, /ɔ/, /o/, /u/), and five nasal vowels (/ĩ/, /ẽ/, /ã/, /õ/, /ũ/) which differ both spectrally and temporally from their oral counterparts. All vowels, except /ɨ/, occur in stressed position. In pretonic position, all oral vowels occur, whereas in posttonic position they are reduced to four (/i/, /ɨ/, /ɐ/, /u/), and in word-final position to three (/ɨ/, /ɐ/, /u/). The five nasal vowels occur in pretonic position, but only two (/ẽ/, /ũ/) in posttonic position (Barroso, 1999; Mateus, Falé, & Freitas, 2005).

Interestingly, the nasal vowels are acoustically and perceptually longer than their oral counterparts due to the presence of the nasal murmur which lengthens them. Nasality also affects the vowels' bandwidth and modifies the frequencies of the first (F1) and second (F2) formants, resulting in larger bandwidths and different formant frequencies (Santos, 2003). The EP front vowel space has three front vowels (/i/, /e/, /ɛ/) that contrast in spectral quality and have intrinsic vowel duration differences (Escudero, Boersma, Rauber & Bion, 2009). The English vowel contrasts /i/-/ɪ/ and /e/-/æ/ present both production and perception difficulties for adult native EP speakers. English /i/ and /æ/ tend to be assimilated into the Portuguese vowel sounds /i/ and /e/ respectively, and no distinction between the two vowels of each pair is made due to their acoustic and articulatory proximity (Rato, 2014; Rato et al., 2014).

The L1 Catalan participants of the study spoke the Eastern Catalan variety of the language. Eastern Catalan has eight simple vowels (/i e ɛ u o ɔ a ə/), seven occurring in stressed positions and one schwa-like vowel occurring in unstressed positions. Catalan does not employ duration contrastively. The high front vowel region of Catalan is only occupied by vowel /i/. Acoustic comparisons indicate that English /i/ and the Catalan /i/ are highly similar, whereas the English /ɪ/ is acoustically closest to the Catalan /e/ (Cebrian, 2006). In vowel perception, Catalan speakers tend to assimilate the English /i/ into the native /i/, whereas the English /ɪ/ is perceived as more dissimilar and is assimilated into either the Catalan /e/, /i/ or /ɛ/. We could thus conclude that whereas the English /i/ can be easily assimilated into Catalan /i/, the English /ɪ/ is less straightforward and it might pose more problems to the L2 learners. This is because it could be perceived as a new vowel but its assimilation pattern could also correspond to that of two-category assimilation, single-category assimilation or category-goodness difference (Best & Tyler, 2007).

The L1 Russian participants were native speakers of Standard Russian (Moscow dialect). Standard Russian vowel inventory consists of five distinct vowel sounds /i e u o a/ occurring in stressed syllables (Barnes, 2007; Padgett, 2004). In non-stressed syllables, the vowel contrast is reduced to two to three vowels depending on

the palatalization context. A long-standing discussion has been held among linguists concerning the high central unrounded vowel [ɨ] and its status as a distinct phonological segment in the Russian vowel inventory. Following Padgett & Tabain (2005), we assume that the vowel sound [ɨ] is an allophone of /i/ occurring in complementary distribution in palatalized context. Palatalization results in the fronting of the vowels so that after palatalized consonants, the /i/ is realized as [i], whereas after non-palatalized consonants it is realized as [ɨ]. In Russian, vowel contrast is primarily distinguished by spectral differences. However, vowel duration is used to distinguish different levels of stress in Russian (Bondarko, 1998). In stressed syllables, the vowels are realized as long, before stressed syllables the vowels are produced shorter, and in unstressed syllables they are the shortest (Kondaurova & Francis, 2008). In the perception of English high front vowels, Russian listeners were found to use duration cue exclusively, whereas spectral differences were not a predictor of vowel contrast (Kondaurova & Francis, 2008). Consequently, we could expect L1 Russian speakers to show difficulties in discerning English /i-ɪ/ accurately due to their over-reliance on temporal cues.

1.3. Research question and objectives

The present research project is an extension of Kivistö-de Souza and Carlet (2014) cue-weighting study in which the cue-weighting strategies of L1 Catalan and L1 Danish learners of English were examined. In the present study, two more groups of EFL learners were tested: L1 Portuguese and L1 Russian. The study aimed at answering the following research question:

RQ: Is the size of the L1 vowel inventory associated with the use of temporal cues in the perception of English /i-ɪ/ contrast?

Following previous research, we hypothesized that speakers of languages with larger L1 vowel inventories (such as Danish and Portuguese) would expand their perceptual space in order to accommodate the large number of L1 vowels, enabling them to perceive larger

spectral distances between L2 vowels than speakers with smaller L1 vowel inventories (Bohn, 1995; Hacquard et al., 2007). Consequently, we expected that the L1 Danish speakers would identify the vowels to the most accurate extent, followed by the L1 Portuguese speakers, then the L1 Catalan speakers and finally the L1 Russian speakers, whose L1 vowel inventory is the smallest.

2. Method

2.1. Participants

Participants from five language backgrounds were tested: L1 English, L1 Danish, L1 Portuguese, L1 Catalan and L1 Russian. The demographic characteristics of the participants are shown in Table 1 below.

	English (<i>n</i> =7)	Danish (<i>n</i> =20)	Portuguese (<i>n</i> =18)	Catalan (<i>n</i> =20)	Russian (<i>n</i> =14)
Age	27.7 (4.6)	23.0 (3.5)	20.6 (5.6)	22.8 (4.9)	19.2 (1.3)
Sex	100% f	75% f, 25% m	78% f, 22% m	60% f, 40% m	22% f, 78% m
L1 daily use (%)	67.85 (13.8)	74.9 (14.0)	NA	81.1 (13.7)	84.7 (13.3)
L2 daily use (%)	31.66 (12.9)	24.7 (13.9)	NA	17.7 (14.2)	14.9 (12.9)
L2 fluency	-	100% y	63% y, 31% n	80% y, 20% n	50% y, 50% n

Table 1. Demographic characteristics of the participants grouped by L1

The L1 Danish, L1 Catalan and L1 Portuguese participants were first year English majors at the universities of Aarhus, Barcelona and Minho respectively. The L1 Russian participants were first and second year students at the Moscow State University of Mechanical Engineering with a high command of English. The participants reported using little English on a daily basis overall (“L2 daily use”): The L1 Danish participants reported employing English on average 24% of the time, being the L1 participants’ with the highest L2 use, the Catalan and the Russian participants

reported somewhat lower L2 use.² All L1 Danish participants considered themselves fluent in English, followed by 80% of the Catalans and 63% of the L1 Portuguese participants. The L1 Russian participants had the lowest self-perceived fluency with only 50% considering themselves fluent in English.

In order to compare the language learners’ vowel perception to native standards, a control group of native English speakers (*n*=7) living in Barcelona during the time of data collection was recruited. Three of the native English speakers were from Southern England and four from the United States. In a previous study (Kivistö-de Souza, 2011), the British and American participants’ vowel production data as well as perception scores were submitted to a set of Mann-Whitney U tests. None of the tests yielded significant differences, indicating that the British and American participants did not perform significantly differently, enabling their grouping together.

2.2. Vowel Identification Task

The task used to measure participants’ cue-weighting was a vowel identification task previously used in Moya-Galé (2010), Mora and Cerviño-Povedano (2010), and Gilabert, Mora and Muñoz (2010). The stimuli consisted of six /i-ɪ/ minimal pairs: *bead-bid*; *beat-bit*; *deed-did*; *peak-pick*; *Pete-pit*; *seed-Sid* produced by six SSBE speakers (three male and three female) in a sound-attenuated booth. Half of the target vowels were followed by a voiceless consonant and half were followed by a voiced consonant as voicing of the following consonant affects the vowel duration. In 50% of the trials, the duration was manipulated (“duration-manipulated trials”) and in 50% the duration was left natural (“natural trials”). In the duration-manipulated trials, the duration of the two vowels was exchanged so that the tense vowel received the mean duration of the lax vowel (155ms) and the lax vowel received the mean duration of the tense vowel (189ms). The vowel quality was left intact. The aim of the duration manipulation was to determine whether the participants were relying on duration cues when perceiving the vowel pair.

The 144 target words were presented in a random order aurally and orthographically through DmDx software (Forster & Forster, 2012). The participant heard a stimulus and had to select from two orthographically presented answering options the one corresponding to the auditory stimulus.

2.3. Procedure

The L1 English, L1 Russian, L1 Danish and L1 Catalan participants were tested individually in a

words was clear, the participants performed the vowel identification task. The vowel identification task took around 10 minutes to complete.

2.4. Results

Participants' responses to the vowel identification task were analyzed and the percentage of accurate responses for each condition (vowel/stimulus type) was computed. Descriptive statistics are presented in Table 2.

Stimulus type	Participant L1				
	English (<i>n</i> =7)	Danish (<i>n</i> =20)	Portuguese (<i>n</i> =18)	Catalan (<i>n</i> =20)	Russian (<i>n</i> =14)
ID natural	99.2 (1.5)	97.5 (4.2)	72.45 (22.2)	77.5 (12.1)	66.5 (11.6)
ID manipulated	98.0 (2.0)	94.1 (7.4)	59.1 (28.4)	51.3 (20.5)	29.5 (8.6)
ID lax natural	99.6 (1.0)	97.5 (4.4)	67.1 (22.3)	74.5 (15.7)	61.3 (13.3)
ID lax manipulated	99.6 (1.0)	95.8 (7.0)	54.3 (28.2)	49.3 (19.6)	25.1 (10.1)
ID tense natural	98.8 (2.1)	97.5 (5.0)	77.7 (25.6)	80.5 (12.7)	71.8 (11.4)
ID tense manipulated	96.4 (4.1)	92.5 (8.6)	64.0 (30.9)	53.3 (23.9)	33.9 (10.9)

quiet room in their respective universities. The L1 Portuguese participants were tested in a computer room in a joint session. Despite testing being conducted in different testing locations, due to the nature of the study, the exact same instructions were followed by the researchers, in the attempt to minimize any differences in the testing setting. At the beginning of the data collection session, the participants read and signed a consent form and filled in a language background questionnaire. Before the vowel identification task, the participants performed a familiarization task which presented the target words aurally together with pictures. The aim of the familiarization task was to make sure that participants knew the meaning of the target words. Replaying of the stimuli was allowed as many times as desired by the participant. After finishing the familiarization task, the participants could ask the researcher for clarifications, and once the meaning of all the test

Table 2. Response accuracy (%). Standard deviations in brackets.

In the identification of the natural stimuli, the L1 Danish participants performed most accurately of the EFL learners groups ($M=97.5\%$), whereas the Russian participants performed worst ($M=66.5\%$). The L1 Portuguese group showed large individual variation as revealed by the large standard deviations. In the duration-manipulated stimuli, the L1 Danish participants again outperformed the other EFL learners ($M=94.1\%$), whereas the L1 Russian participants had difficulty in identifying duration-manipulated stimuli, being able to do so accurately on average in only 29.5% of the cases. Overall, there was a tendency for the tense vowel to be better identified than the lax vowel in the natural and duration-manipulated contexts for the EFL learners.

The response data were not normally distributed (Kolmogorov-Smirnov $p < .05$) which is why non-

parametric tests were performed in order to determine whether the different L1 groups showed differences in their vowel identification in natural and duration-manipulated stimuli. The identification accuracy data for natural and duration-manipulated stimuli can be visually inspected in the following bar graph (Figure 1).

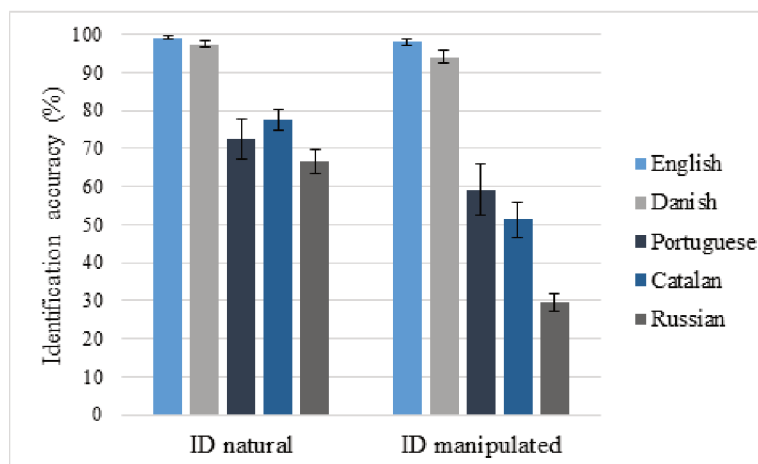


Figure 1. Identification accuracy in natural and duration-manipulated stimuli (error bars represent ± 1 standard error).

The identification accuracy in the natural stimuli was taken as an indication of the participants' base-line level in the perception of the target vowels. A Kruskal-Wallis test for the five L1 groups showed significant differences in the perception of the natural stimuli ($\chi^2(4) = 49.997$, $p < .001$). Post-hoc Mann-Whitney U-tests indicated that only the L1 Danish participants did not significantly differ from the native English speakers ($Z = -1.228$, $p = .220$). All the other EFL learner groups showed significantly lower identification accuracy to the L1 English and L1 Danish participants. The identification accuracy of natural stimuli did not significantly differ between the Portuguese and the Catalan participants ($Z = -.380$, $p = .704$) nor between the Portuguese and the Russian participants ($Z = -1.197$, $p = .231$). However, the Catalan participants identified the natural stimuli significantly better than the Russian participants ($Z = -2.189$, $p = .029$).

After seeing that there were some differences in the perception of the natural stimuli between the EFL learners, we turn to the main data, the duration-manipulated stimuli, in order to answer the research

question. A Kruskal-Wallis test for the five L1 groups showed significant differences for the duration-manipulated stimuli ($\chi^2(4) = 52.832$, $p < .001$). Post-hoc Mann-Whitney U-tests yet again showed that the L1 Danish participants did not perform differently from the native English speakers ($Z = -1.122$, $p = .262$), whereas all the other EFL learner groups had a significantly lower vowel identification accuracy in the duration-manipulated trials than the native English speakers. The L1 Portuguese and the L1 Catalan participants did not differ in the identification of the duration-manipulated stimuli ($Z = -1.111$, $p = .276$), paralleling the results for the natural stimuli. Contrary to the natural stimuli, the L1 Russian participants had a significantly lower identification accuracy in the duration-manipulated trials than the L1 Catalan ($Z = -3.574$, $p < .001$) and the L1 Portuguese participants ($Z = -3.211$, $p = .001$). Consequently, the L1 Russian participants showed the lowest performance accuracy in the duration-manipulated stimuli. These results partly support the initial hypothesis, namely, that L1 vowel inventory size has an effect on L2 vowel perception. As expected, the L1 Danish speakers performed most accurately and the L1 Russian speakers least accurately. However, no differences were observed between the L1 Portuguese and the L1 Catalan speakers, even though Portuguese has a larger vowel inventory than Catalan.

Finally, participants' identification accuracy of the two vowels was analyzed separately. Wilcoxon Signed Rank tests were performed individually for the five L1 groups in the two conditions (natural/duration-manipulated) comparing the performance in the tense and the lax target vowel. For the L1 English and L1 Catalan participants, there were no differences in the identification of the two vowels in either condition (L1 English natural: $Z = -1.34$, $p = .180$; L1 English manipulated: $Z = -1.63$, $p = .102$; L1 Catalan natural: $Z = -1.56$, $p = .118$; L1 Catalan manipulated: $Z = -1.04$, $p = .295$). In other words, the L1 English and L1 Catalan participants identified the two vowels equally accurately. For the L1 Danish participants, the target vowels were identified equally accurately in the natural condition ($Z = -.718$, $p = .473$), but not in the duration-manipulated condition ($Z = -2.78$, $p = .005$), in which the lax vowel was

identified more accurately than the tense vowel (95.8% vs. 92.5%). For the L1 Russian and L1 Portuguese participants, there were significant differences between the identification of the vowels in both conditions. The L1 Portuguese participants identified the tense vowel better than the lax vowel in the natural stimuli ($Z=-2.32, p=.020$) and in the duration-manipulated stimuli ($Z=-2.35, p=.019$). The L1 Russian participants also identified /i/ better than /ɪ/ in the natural ($Z=-2.97, p=.003$) and the duration-manipulated stimuli ($Z=-2.10, p=.035$). In other words, although the EFL learners showed a tendency towards identifying the tense vowel more accurately than the lax vowel, with the exception of the L1 Danish participants, the difference was only significant for the L1 Portuguese and the L1 Russian participants. This tendency might be due to the fact that all of the L1s in the present study had a high front vowel similar to the English /i/ whereas the lax high front vowel is more dissimilar when mapped into the L1 inventories. Thus, the acquisition of /ɪ/ would require it to be perceived phonetically as dissimilar from the L1 sound. According to SLM (Flege, 1995) and PAM-L2 (Best & Tyler, 2007), when this does not happen, the two L2 sounds are assimilated to a single L1 category.

3. Discussion and conclusions

The present study investigated the role of L1 vowel inventory size on the use of cue-weighting strategies

forced choice vowel identification task which presented English /i-ɪ/ minimal pairs in natural and duration-manipulated contexts. Participants' over-reliance on the duration cue was expected to be related to the L1 vowel inventory size so that participants with the smallest L1 vowel inventory (Russians) would rely on duration cues more than participants with bigger L1 vowel inventories (Danes, Portuguese and Catalans).

This hypothesis was partly confirmed as the results indicated that the L1 Russian participants had the least accurate vowel identification of the duration-manipulated stimuli, whereas the L1 Danish participants had the most target-like vowel identification. Nevertheless, no differences in cue-weighting strategies were observed between the L1 Portuguese and the L1 Catalan participants, even though Portuguese has a larger vowel inventory than Catalan (9 oral+5 nasal vs. 8 oral). It thus appears that the presence of the nasal vowels and one additional oral vowel did not bring a benefit for the L1 Portuguese speakers over the L1 Catalan speakers even though the L1 Portuguese speakers were hypothesized to be more exposed to both spectral and temporal differences due to their L1 vowel inventory.

When examining the participants' reliance on duration cues, we observed the behavior illustrated in Figure 2 below.

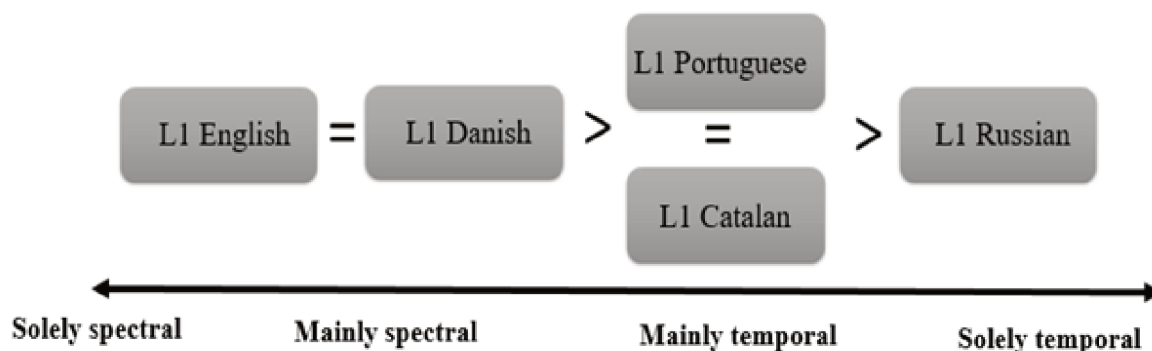


Figure 2. Participants' cue-weighting strategies in the perception of English high front vowels

(quality/quantity) in L2 vowel perception. Participants were EFL learners from four L1 backgrounds: Danish, European Portuguese, Catalan, and Russian. Participants' L2 vowel perception was measured in a

As expected from previous research with native English speakers, the L1 English participants appeared

to discern the vowel pair based on spectral differences, as no differences were observed between the identification of the natural and the duration-manipulated stimuli. The L1 Danish speakers did not differ significantly from the native English speakers in their vowel identification accuracy. This is in line with our hypothesis that the large vowel space of Danish sensitizes Danish speakers for perceiving small-scale spectral differences, a strategy which is successfully transferred into L2 vowel perception. In spite of their near native perception, the L1 Danish participants still showed a small effect of reliance on duration cues (identification natural: 97.5% vs. identification duration-manipulated: 94.16%). This might be due to the fact that duration is employed contrastively in Danish. It might be that the L1 Danish speakers also make use of their L1 contrastive duration when identifying the target L2 vowels. However, this seems to be a secondary strategy and we conclude that the L1 Danish participants of the study discerned the target vowel pair mainly based on spectral cues.

As already mentioned, the L1 Portuguese (59.1%) and L1 Catalan participants (51.3%) did not differ significantly from each other in the identification of the duration-manipulated stimuli. Both EFL learner groups manifested identification accuracy barely above the chance level in the trials in which the duration of the two vowels was exchanged. The poorer performance in the duration-manipulated trials indicates that L1 Portuguese and L1 Catalan participants were relying heavily on duration when discerning the target vowel pair. We thus conclude that the L1 Portuguese and L1 Catalan participants relied mainly on temporal cues in identifying the English high front vowels. Following the *Desensitization Hypothesis* (Bohn, 1995), we expected the L1 Portuguese participants to perform more accurately than the L1 Catalan participants, due to the former's larger L1 vowel inventory. Even though this occurred, the differences were not statistically significant. This does not clearly support the *Desensitization Hypothesis*, but it may indicate that the presence of one oral vowel will not affect the use of cue-weighting strategies. Another reason for this result could be that European Portuguese and Catalan are comparable in terms of the high front vowel space, both languages presenting only

one oral high front vowel, in comparison to Danish, for example, whose high front vowel area is more crowded. More research is required on the matter.

The L1 Russian participants had the lowest performance accuracy in the duration-manipulated stimuli (29.5%). In fact, their identification accuracy, which was well below the chance-level, might suggest that the L1 Russian participants were relying on duration solely. In other words, they might have based their responses only on the duration of the target item without attending to spectral cues at all or to a very small extent. In practice, this would mean that when they identified the duration of the vowel as short, they mapped it with /i/ and when they considered the duration of the vowel to be long, they mapped it to /i/. This type of strategy would clearly be ineffective and lead to poor identification accuracy, as the duration-manipulated tense and lax vowels had exchanged their durations.

One explanation of the poor results of the L1 Russian participants could be found in their vowel perception baseline data. The L1 Russian participants were the least accurate in identifying the target vowels in the natural condition (66.5%), suggesting that their category formation for the L2 vowels was weaker than for the other EFL groups. Another explanation could be found in how duration is employed in their L1. Russians might have more experience with the duration than the L1 Catalan and L1 Portuguese speakers, as lexical stress is phonologically distinctive in Russian so that two words may differ only in terms of stress placement (Kondaurova & Francis, 2008).³ Nevertheless, Danish is the only one of the languages studied which employs duration contrastively and the L1 Danish speakers performed with the highest accuracy. Consequently, being experienced with duration in the L1 might not be related to cue-weighting strategies in the L2.

Another issue which could have affected the L1 Russian participants' identification accuracy is the role of palatalization on the L1 high front vowel. None of the target vowels appeared after a palatal consonant, thus in L1 Russian, the pronunciation of the high front vowel would be [i] instead of [i]. Consequently, the Russians might have identified the target sounds as

poor examples of the L1 [i], which would be reflected in the poor identification accuracy of the L2 /i/ and /ɪ/. However, the palatalization context did not vary in the test words, i.e., all of the target vowels occurred in non-palatalized context, meaning that the possible effect should be similar to the duration-manipulated and natural trials as well as for the two target vowel trials. Clearly, more studies are needed to determine the effect of palatalization on L1 Russian speakers' L2 vowel perception.

Finally, we should consider an alternative explanation for the results taking into account the role of L2 proficiency. Our hypothesis stated that L1 vowel inventory size would have an effect on L2 vowel perception. However, it is possible that L2 proficiency plays a role as well, as proposed by Escudero (2000). The poor identification results of the L1 Russian speakers might be contributed to their possibly lower English proficiency level. Due to methodological limitations, we could not measure the L1 Catalan and L1 Danish participants' L2 vocabulary size. The L1 Portuguese and L1 Russian participants' L2 vocabulary size was measured with X_lex and Y_lex vocabulary size tests (Meara, 2005; Meara & Miralpeix, 2006) whose scores have been shown to relate to L2 proficiency levels (Miralpeix, 2012). We made an effort to match the Portuguese participants to the Russian participants in terms of vocabulary size, but there was still a difference between the two groups (Portuguese $M=6097$, Russian $M=5192$, max. 10,000). Nevertheless, the Portuguese and Russian participants did not significantly differ from each other in the identification of the natural stimuli, indicating that their baseline vowel perception was similar. On the other hand, if we examine the self-reported fluency data, we see some differences between the EFL groups' self-perceived English fluency that might partly explain the obtained results. Recall that all of the L1 Danish speakers considered themselves fluent in English, followed by 80% of the L1 Catalan speakers and 63% of the L1 Portuguese speakers, whereas only half of the L1 Russian learners considered themselves fluent in English. When mapping the self-reported L2 fluency data to the perception results, we see some similarities, as the L1 Danish speakers performed most

accurately and the L1 Russians least accurately. The L1 Catalan and L1 Portuguese speakers' vowel perception results were comparable, but the former's self-perceived fluency was higher than the latter's. As noted by an anonymous reviewer to this manuscript, this might be because the L1 Portuguese speakers' larger vowel inventory might have compensated for lower (self-perceived) L2-fluency.⁴ Consequently, future studies on cue-weighting should measure the participants' L2 proficiency in order to statistically control for its effect on L2 vowel perception. Furthermore, the L2 proficiency and/or fluency measures should be based on objective tests rather than on subjective measurements as was the case of self-reported fluency in the present study.

To summarize, the results of the present research offer partial support to Bohn's (1995) *Desensitization Hypothesis* and suggest that L1 vowel inventory size is a contributing factor in L2 vowel perception. The L1 Danish participants, with the largest L1 vowel inventory, showed less reliance on temporal cues than the other EFL learners, indicating that the large L1 vowel inventory had successfully sensitized the L1 Danish speakers to small-scale spectral differences. The participants from the two medium-sized vowel inventories, Portuguese and Catalan, identified the L2 vowels less accurately than the L1 Danish participants, but more accurately than the L1 Russian participants. The Russian participants, with the smallest L1 vowel inventory, relied heavily on temporal cues in discerning the target vowels. It would thus appear that the smaller vowel inventory of Russian *desensitized* the Russian participants to the small spectral differences present in the English /i-ɪ/, forcing them to rely on temporal cues instead. More support for the role of L1 vowel inventory size on L2 cue-weighting could be obtained by examining different language combinations. We especially recommend comparing languages which differ in the size of the vowel inventory (large/small) but not in the use of temporal cues. One possible pairing would then be to compare L1 Arabic (a three-vowel system) or L1 Finnish (a five-vowel system) with L1 Danish EFL learners.

Previous studies (e.g. Ylinen et al., 2009) suggest that it is possible to shift language learners' attention

towards more relevant (e.g. spectral) cues. Our results imply that L1 Russian learners of English, and to a smaller extent, L1 Portuguese and L1 Catalan EFL learners, might benefit from perceptual cue-training employing duration-manipulated English vowels. Such training might lead to differences in cue-weighting strategies, resulting in a progressive shift from over-reliance on temporal cues to more target-like employment of spectral cues. This in turn might result in more accurate perception, and possibly production, of the L2.

Acknowledgements

We would like to show our gratitude to professors Ocke Schwen-Bohn and Camilla Søballe Horslund for providing insight into the complex Danish vowel system. We would also like to thank Olga Ivanova and Varvara Molodtsova for their assistance with the data collection in Moscow. Finally, we are immensely grateful for the anonymous reviewers for their comments on an earlier version of the manuscript.

Notes

1. A previous version of this paper was presented at the 8th International Conference on Second Language Speech in Aarhus 10-12th June, 2016.
2. Due to a technical failure in the data collection, there are no data on the L1 Portuguese speakers' language use.
3. An anonymous reviewer brought to our attention the fact that in Portuguese, words can also differ in terms of lexical stress (e.g. 'sá.bi.a - sa'bi.a). This occurs in Spanish and Catalan as well ('hab.lo - hab'ló; 'par.la - par'là) but in these languages, syllable duration is not a reliable cue for word-stress (Llisterri et al. 2003). We are not aware of any empirical studies measuring the correlates of lexical stress in (European) Portuguese, thus we cannot confirm whether Portuguese would behave similarly to Russian or to Spanish and Catalan.
4. We thank the anonymous reviewer for bringing this into our attention

References

- Barnes, J. (2007). *Phonetics and phonology in Russian unstressed vowel reduction: A study in hyperarticulation*. (Unpublished Master's thesis). Boston University, Boston, MA.
- Barroso, H. (1999). *Forma e substância da expressão da língua portuguesa*. Coimbra: Livraria Almedina.
- Best, C.T., & Tyler, M.D. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In M.J. Munro & O. -S. Bohn (Eds.), *Second language speech learning: The role of language experience in speech perception and production* (pp.13-34). Amsterdam: John Benjamins.
- 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: Issues in Cross-language research* (pp.279-304). Timonium, MD: York Press.
- Bohn, O.-S., & Caudery, T. (2012). *The Sounds of English: An Activity-Based Course in English Phonetics and Phonology*. Aarhus, DK: Aarhus University.
- Bohn, O., & Flege, J. E. (1990). Interlingual identification and the role of foreign language experience in L2 vowel perception. *Applied Psycholinguistics*, 11(03), 303. doi:10.1017/s0142716400008912
- Bohn, O. -S., & Steinlen, A. K. (2003). Consonantal context affects cross-language perception of vowels. In *Proceedings of the 15th International Congress of Phonetic Sciences* (pp. 2289-2292). Barcelona: Universitat Autònoma de Barcelona.
- Bondarko L. V. (1998). "Fonetika Sovremennogo Russkogo Jazyka" (*Phonetics of Modern Russian Language*). University of St. Petersburg, St. Petersburg, Russia.
- Cebrian, J. (2006). Experience and the use of non-native duration in L2 vowel categorization. *Journal of Phonetics*, 34(3), 372-387. doi:10.1016/j.wocn.2005.08.003
- Escudero, P. (2000). *Developmental patterns in the adult L2 acquisition of new contrasts: The acoustic cue weighting in the perception of Scottish tense/lax vowels by Spanish speakers*. (Unpublished Master's thesis). University of Edinburgh, Scotland, UK.
- . (2006). The phonological and phonetic development of new vowel contrasts in Spanish learners of English. In B. Baptista & M. Watkins (Eds.), *English with a Latin Beat. Studies in Portuguese/Spanish-English Interphonology* (pp. 51-55). Amsterdam: John Benjamins.
- Escudero, P., & Boersma, P. (2004). Bridging the Gap Between L2 Speech Perception Research and Phonological Theory. *Studies in Second Language Acquisition*, 26(04), 551-585. doi:10.1017/s0272263104040021
- Escudero, P., Boersma, P., Rauber, A. S., & Bion, R. A. (2009). A cross-dialect acoustic description of vowels: Brazilian and European Portuguese. *The Journal of*

- the Acoustical Society of America*, 126(3), 1379-1393. doi:10.1121/1.3180321
- Flege, J. E. (1995). Second language speech learning: Theory, findings and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Theoretical and methodological issues in cross-language speech research* (pp. 233-277). Timonium, MD: York Press.
- Flege, J. E., Bohn, O., & Jang, S. (1997). Effects of experience on non-native speakers' production and perception of English vowels. *Journal of Phonetics*, 25(4), 437-470. doi:10.1006/jpho.1997.0052
- Forster, K. I., Forster, J.C. (2012). DMDX (Version 4.0.6.0) [Computer software]. Retrieved from <http://www.u.arizona.edu/~jforster/dmdx.htm>
- Fox, R. A., Flege, J. E., & Munro, M. J. (1995). The perception of English and Spanish vowels by native English and Spanish listeners: A multidimensional scaling analysis. *The Journal of the Acoustical Society of America*, 97(4), 2540-2551. doi:10.1121/1.411974
- Frieda, E. & Nozawa, T. (2007). You are what you eat phonetically. The effect of linguistic experience on the perception of foreign vowels. In O.-S. Bohn & M. Munro (Eds.), *Language Experience in Second Language Speech Learning. In honor of James Emil Flege* (pp.79-96). Amsterdam: John Benjamins.
- Gilbert, R., Mora, J. C., & Muñoz, C. (2010). *Phonological short-term memory and working memory capacity in foreign language acquisition*. Paper presented at SLRF 2010 (Second Language Research Forum. Invited colloquium on Aptitude and Individual Differences in SLA), University of Maryland, Washington, USA.
- Hacquard, V., Walter, M. A., & Marantz, A. (2007). The effects of inventory on vowel perception in French and Spanish: An MEG study. *Brain and Language*, 100(3), 295-300. doi:10.1016/j.bandl.2006.04.009
- Higgins, J. (2013). Minimal pairs for English RP. Retrieved from <http://myweb.tiscali.co.uk/wordscape/wordlist/minpairs.html>
- Iverson, P., & Evans, B. G. (2007). Learning English vowels with different first-language vowel systems: Perception of formant targets, formant movement, and duration. *The Journal of the Acoustical Society of America*, 122(5), 2842-2854. doi:10.1121/1.2783198
- Jilka, M. (2009). Talent and proficiency in language. In G. Dogil & S. Reiterer (Eds.), *Language talent and brain activity* (pp.1-16). Berlin: Mouton De Gruyter.
- Kivistö-de Souza, H. (2011). *Effects of speech style on the use of temporal and spectral cues in the production and perception of a non-native vowel contrast* (Unpublished Master's thesis). University of Barcelona, Barcelona, Spain.
- Kivistö-de Souza, H., & Carlet, A. (2014). Vowel inventory size and the use of temporal cues in non-native vowel perception by Catalan and Danish EFL learners. *Concordia Working Papers in Applied Linguistics*, 5, 322-336.
- Kondaurova, M. V., & Francis, A. L. (2008). The relationship between native allophonic experience with vowel duration and perception of the English tense/lax vowel contrast by Spanish and Russian listeners. *The Journal of the Acoustical Society of America*, 124(6), 3959-3971. doi:10.1121/1.2999341
- Llisterri, J., Machuca, M., de la Mota, C., Riera, M., and Rios, A. (2003). The perception of lexical stress in Spanish, in *Proceedings of the 15th International Congress of Phonetic Sciences*, edited by M. J. Sole, D. Recasens, and J. Romero, Barcelona, 3-9 August, casual productions [CDROM].
- Mateus, M. H. M., Falé, I., & Freitas, M. J. (2005). *Fonética e fonologia do português*. Lisboa: Universidade Aberta.
- Meara, P. M. (2005). *X_Lex: The Swansea vocabulary levels test* (Version 2.05.) [Computer software]. Swansea: Lognostics.
- Meara, P. M., & Miralpeix, I. (2006). *Y_Lex: The Swansea advanced vocabulary levels test* (Version 2.05.) [Computer software]. Swansea: Lognostics.
- Miralpeix, I. (2012). *X_Lex and Y_Lex: A validation study*. Paper presented at the 22nd VARG Conference (Vocabulary Acquisition Research Group), Newtown, UK.
- Mora, J.C., & Cerviño-Povedano, E. (2010). *Phonological short-term memory and L2 vowel perception*. Paper presented at the 20th Annual Conference of the European Second Language Association (EUROSLA-20), Università di Modena e Reggio Emilia, Reggio Emilia, Italy.
- Morrison, G.S. (2002). Perception of English /i/ and /ɨ/ by Japanese and Spanish listeners: Longitudinal results. In G.S. Morrison & L. Zsoldes (Eds.), *Proceedings of the North West Linguistics Conference* (pp.29-48). Burnaby, BC, Canada: Simon Fraser University.
- _____. (2008). L1-Spanish Speakers' Acquisition of the English /i /-/-/I/ Contrast: Duration-based Perception is Not the Initial Developmental Stage. *Language and Speech*, 51(4), 285-315. doi:10.1177/0023830908099067
- Moya-Galé, G. (2010). *Overcoming non-native over-reliance: a study on English vowel duration manipulation and neutralization*. (Unpublished Master's thesis). Universitat de Barcelona, Barcelona, Spain.
- Padgett, J. (2004). Russian vowel reduction and Dispersion Theory. *Phonological Studies*, 7, 81-96.

- Padgett, J., & Tabain, M. (2005). Adaptive Dispersion Theory and Phonological Vowel Reduction in Russian. *Phonetica*, 62(1), 14-54. doi:10.1159/000087223
- Rato, A. (2014). Effects of Perceptual Training on the Identification of English Vowels by Native Speakers of European Portuguese. *Proceedings of the International Symposium on the Acquisition of Second Language Speech - Concordia Working Papers in Applied Linguistics*, 5, 529-546.
- Rato, A., Rauber, A., Soares, L., & Lucas, L. (2014). *Challenges in the perception and production of English front vowels by native speakers of European Portuguese. Diacrítica - série ciências da linguagem*, 28(1), 137-155.
- Sampson, R. (1999). *Nasal vowel evolution in Romance*. Oxford: Oxford University Press.
- Santos, G. B. (2003). *Análise fonético-acústica das vogais orais e nasais do português: Brasil e Portugal*. (Unpublished Doctoral Dissertation). Universidade Federal de Goiás, Brasil. Retrieved from: <https://repositorio.bc.ufg.br>
- Steinlen, A.K. (2002). *A cross-linguistic comparison of the effects of consonantal contexts on vowels produced by native and non-native speakers* (Unpublished Doctoral dissertation). Aarhus University, Aarhus, DK.
- _____. (2005). *The influence of consonants on native and non-native vowel production: a cross-linguistic study* (Vol. 30). Tübingen: Gunter Narr.
- Ylinen, S., Uther, M., Latvala, A., Vepsäläinen, S., Iverson, P., Akahane-Yamada, R., & Näätänen, R. (2009). 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. doi:10.1162/jocn.2009.21272

Recebido em: 21/03/2017

Aceito em: 26/07/2017