Improvement in Young Adults' Second-language Pronunciation after Short-term Immersion

Anders Højen
Aarhus University

Abstract
This study examined the effect of short-term immersion in English-language communities in England on young native Danish adults' English pronunciation. Pronunciation ratings by a group of native judges revealed significantly higher pronunciation ratings when compared before and after 3-10 months of English immersion. A native Danish control group received virtually identical ratings by the judges at two different time points. The pronunciation gain score for the immersion group was significantly correlated with length of residence (LOR) in England. However, a stronger correlation (r=0.81) was found between pronunciation gain score and a weighted input measure, *viz.* LOR weighted by self-reported proportion of English vs. Danish use during the immersion period. The results suggest that second-language (L2) learners' phonetic system is highly malleable and responds readily to new L2 input.

1. Introduction

Bilinguals typically speak their second language (L2) with a foreign accent. The age of L2 learning (AOL) is almost always found to be the strongest predictor of the degree of foreign accent, although a range of other factors – some of them often confounded with AOL – may also influence degree for foreign accent in the L2 (for a review, see Piske, MacKay, & Flege, 2001). The purpose of the present small-scale study was to examine the effect of experiential factors during young adults' short-term immersion

Anne Mette Nyvad, Michaela Hejná, Anders Højen, Anna Bothe Jespersen & Mette Hjortshøj Sørensen (Eds.), ***A Sound Approach to Language Matters – In Honor of Ocke-Schwen Bohn*** (pp. 543-559). Dept. of English, School of Communication & Culture, Aarhus University.
© The author(s), 2019.

in an L2 community. The specific interest was in the immediate effect of the duration of L2 immersion experience and the degree of L2 use on L2 pronunciation.

AOL effects on L2 acquisition have been attributed to the passing of a critical period, after which it is no longer possible to make use of language input to build linguistic representations (Lenneberg, 1967). On the critical period account, biological or maturational changes specific to language underlie these difficulties, and the changes are often proposed to happen around the end of childhood although a number of different cut-off ages have been proposed (for an overview, see e.g., Singleton, 2005). In addition different critical periods have been proposed for different language domains (e.g., Granena & Long, 2013).

On other accounts, there is no biologically or maturationally defined endpoint that marks a categorical difference in the way an L2 is acquired. Rather, AOL effects on acquisition are assumed to arise from the state of development of the L1 when the L2 is acquired (e.g., Ellis, 2002; Flege, 1995; Hernandez, Li, & MacWhinney, 2005; MacWhinney, 2016). Moreover, these accounts generally assume that linguistic representations for L2 perception and production are acquired and entrenched via a high frequency of use and practice, just like other complex cognitive skills, using general cognitive processing mechanisms. The role of frequency for language acquisition is summarized thus by Gries and Ellis (2015, p. 230): "The most fundamental factor that drives learning is the frequency of repetition in usage. This determines whether learners are likely to experience a construction and, if so, how strongly it is entrenched, accessible, and its processing automatized." Consistent with this notion, input frequency effects have been observed in the acquisition of L1 skills (e.g., Hart & Risley, 1995; Hurtado, Marchman, & Fernald, 2008; Weisleder & Fernald, 2013) as well as L2 skills (Højen & Flege, 2006; MacKay & Flege, 2004; Piske et al., 2001; Suter, 1976). Also consistent with the importance of input, effects of length of L2 use have been shown to influence L2 proficiency (Flege & Liu, 2001; Flege, MacKay, & Piske, 2002). In addition, the effect of duration of L2 experience has been found to be moderated by intensity of L2 input (Flege & Liu, 2001).

However, in their large-scale study of foreign accent in Italian immigrants who learned English as an L2 in Canada, Flege, Munro, and MacKay (1995), found no effect of the immigrants' length of residence (LOR), which was a measure of the duration of their L2 experience. Flege et al. (1995) suggested that this was due to ceiling effects; all the

immigrants had lived in Canada for decades. Indeed, Flege & Fletcher (1992) had already suggested that most improvements in L2 pronunciation occur during the first year of exposure to native-produced L2 input in an L2 environment. In spite of this, few studies have examined L2 pronunciation changes occurring in the early phase of L2 input in and L2 environment. Previous studies examining the effect of L2 experience on degree of foreign accent have typically compared groups who had many years of L2 experience with groups who had about 6-12 months of L2 experience (e.g., Flege, 1991; Flege et al., 1995; Piske et al., 2001; Thompson, 1991; Yamada, 1995). Therefore, those studies were not designed to assess any immediate effects of experience in the initial phase of L2 exposure.

One study that examined L2 pronunciation development in the very early phase was that of Snow and Hoefnagel-Höhle (1977). They examined pronunciation of Dutch words by native English children and adults who had moved to the Netherlands. The participants were tested three times during their first 10-11 months of learning, and a significant improvement in the pronunciation of Dutch words was found for both child and adult learners. However, it is unclear how closely their first time of test coincided with the onset of native Dutch input. The participants were tested within six weeks after they started to speak Dutch (Snow & Hoefnagel-Höhle, 1977, p. 361), but presumably six months after they moved to the Netherlands (Snow & Hoefnagel-Höhle, 1978, p. 1115).

When examining the effect of native-produced L2 input on foreign accent, participants may differ according to the extent of nonnative L2 input. In the study by Snow and Hoefnagel-Höhle (1977), the participants had not learned Dutch in school before immigrating, whereas individuals immigrating to an English speaking country have often learned English in school to various degrees. In Denmark, students begin learning English after a few years in school and often become relatively proficient English speakers. However, the typical student only receives sporadic authentic input in inter-personal communication, because most teachers are non-native speakers of English. At the same time, English is very much present on various media platforms. As noted, Flege and Fletcher (1992) suggested that pronunciation typically improves during the first year of *authentic* input. However it is not clear how authentic input, e.g., via immersion in an L2 community, affects pronunciation of L2 speakers who are already relatively skilled. This would be the scenario for Danish students, who learned English in school from non-native teachers and heard native English on various media platforms. Therefore, the present study examined

the effect of short-term immersion in England on degree of foreign accent in native Danish young adults, who had learned English in school to a relatively high degree of receptive and expressive proficiency.

A foreign accent is manifested as the realization of phones in the L2 in a different way than native speakers typically do. In addition, L2 speech may differ on prosodic dimensions. Such deviances from the native phonetic norms can be identified using acoustic analysis. However, the above-mentioned studies of foreign accent generally examined degree of foreign accent using pronunciation ratings by native listener judges, which has been shown to be a reliable metric. Piske et al. (2001) reported a strong correlation between the pronunciation scores assigned to a set of sentences by two different groups of native judges on two different rating scales. This led Piske et al. to conclude that foreign accents can be scaled reliably by native listener judges. In addition, judges give highly similar pronunciation ratings across different sentences (e.g., Flege, 1988; Flege & Fletcher, 1992; Flege et al., 1995), suggesting that native listeners can identify and reliably scale a foreign accent based on a short speech sample.

2. Methods

1.1 Participants

Thirty female talkers participated in the study (which was part of a PhD dissertation based on a series of L2 speech perception and production experiments, Højen, 2003). Only females were recruited because the original intention was to examine only au pairs (who are mostly female). However, exchange students were added to the sample because the number of au pairs that could be recruited was insufficient. To keep the sample relatively homogenous, only exchange students who were females were recruited. The participants were assigned to three different groups. The Experience Group consisted of native Danish au pairs or exchange students (N=14) who spent 3-11 months in England (the LOR). The No-experience Group (N=11) served as an age-matched native Danish control group. The Native English Group (N=5) consisted of native English speakers and served as a native English reference group. All participants reported normal hearing.

The native Danish participants had grown up in Denmark with native Danish parents. They had learned English in school for 7-10 years, but none of them ever had a native English teacher. Ten of the native Danish participants had never lived in an English speaking environment, and could

thus be said to be phonetically inexperienced with English with respect to everyday communication. However, four members of the Experience Group had previously lived in an English speaking environment for up to 12 months; this happened at a minimum age of 18 years. Participant characteristics are shown in Table 1.

	N	Age	T1 Prof	T1 Exp	LOR	EngUse
Experience Group	14	21.3 (3.4)	4.1 (0.8)	3.0 (5.0)	7.1 (3.2)	3.9 (0.8)
No-Experience Group	11	20.1 (2.4)	3.9 (0.8)	0.0 (0.0)	–	–
Native English Group	5	20.0 (2.2)	–	–	–	–

Table 1. Participant characteristics (*SD*). *Age*: years of age at Time 1. *T1 Prof*: Self-reported ability to speak English at Time 1 (1 = a little, 5 = very well). *T1 Exp*: Months of English-language experience at Time 1. *LOR*: Length of Residence during stay in England (in months) . *EngUse*: Self-reported proportion of Danish and English use during their stay in England (1 = Danish only, 5 = English only).

The participants were informed about the purpose of the study, namely to examine effect of immersion on L2 speech perception and speech production. The participants in the Experience Group were tested in Denmark before and after their stay in England. The participants in the No-experience Group were tested two times with an interval of one week to 5 months. The No-experience Group stayed in Denmark during the interval between the two times of testing. The Native English Group was tested only once.

An additional three participants in the Experience group were tested before immersion but could not be tested after their stay; one participant returned to Denmark already after one month, and two participants did not return to Denmark at the time of retests. In addition, one more participant in the Native English Group was tested but recording failed.

1.2 Speech materials

Previous research found that listener judges gave similar pronunciation ratings across different sentences (e.g., Flege, 1988; Flege & Fletcher, 1992; Flege et al., 1995). Therefore, to minimize the burden on the judges it was decided to base the listener judgments on just one sentence. The sentence *3. Are "shock" and "hot" words?* was used to obtain listener ratings of foreign accent (the number *3* was read out and included in the sentence that was rated). The sentence was pragmatically odd because

it originally served to elicit specific speech sounds for acoustic analysis along with 11 other similar sentences. The specific sentence chosen for the present purpose was chosen because it contains several speech sounds with no direct Danish counterpart, namely [θ ɹ ʃ ɒ z] as well as syllable final [d] and syllable initial [w]. For the participants in the Experience Group and the No-experience Group, a production of the sentence was recorded at Time 1 and once again at Time 2. The talkers in the Native English Group were only tested once; therefore, two physically different repetitions of the sentence were recorded in one session.

1.3 Procedure

Recordings of the target sentences by different talkers were compiled in a block of 60 sentences containing the Time 1 and Time 2 recordings from each native Danish speaker and the two single-session recordings from the native English speakers. The order of sentences was quasi-randomized. However, no two repetitions of the target sentence by the same participant were allowed to follow one another. The block was presented twice to ten 20-43-year-old listener judges, who were native speakers of British English. Because the block of sentences was presented twice, rating consistency within each judge could be assessed. Each of the 10 judges rated 120 sentences (30 participants × 2 sentence tokens × 2 blocks) for a total of 1200 judgments.

The judges heard the sentences over headphones and rated each sentence on a 100 point scale which was labeled ”Strong foreign accent” at one end (corresponding to a rating of 1), and ”No foreign accent” at the other end (a rating of 100). The judges were instructed to use the whole scale and to give the maximum score if they were sure they were listening to a native English speaker, and to give the minimum score to the most accented talkers. The participants were given 10 practice sentences to familiarize themselves with the task and the range of foreign accents. These sentences were produced by the participants of each group but differed from the test sentences. The judges indicated their rating of each sentence using a scale on a computer screen. The software UAB-soft was used to present the sentences and store the ratings.

3. Results

1.4 Rating consistency in judges

Before addressing the effect of immersion on foreign accent ratings, the judges’ rating consistency was examined. The two times the block of

sentences was rated by the judges will be called the *first round* and the *second round*. Figure 1, left panel, shows the mean ratings assigned to all sentences by each judge in the first vs. second round. As shown, some judges were stricter than others, but their general rating level was similar in the two rating rounds and did not differ significantly ($t(9)=0.903$, $p<0.39$, $d=0.57$). Mean ratings given in the first vs. second round to each group were also similar (Experience group, 39 vs. 40; No-experience Group, 29 vs. 32; Native English Group, 93 vs. 93).

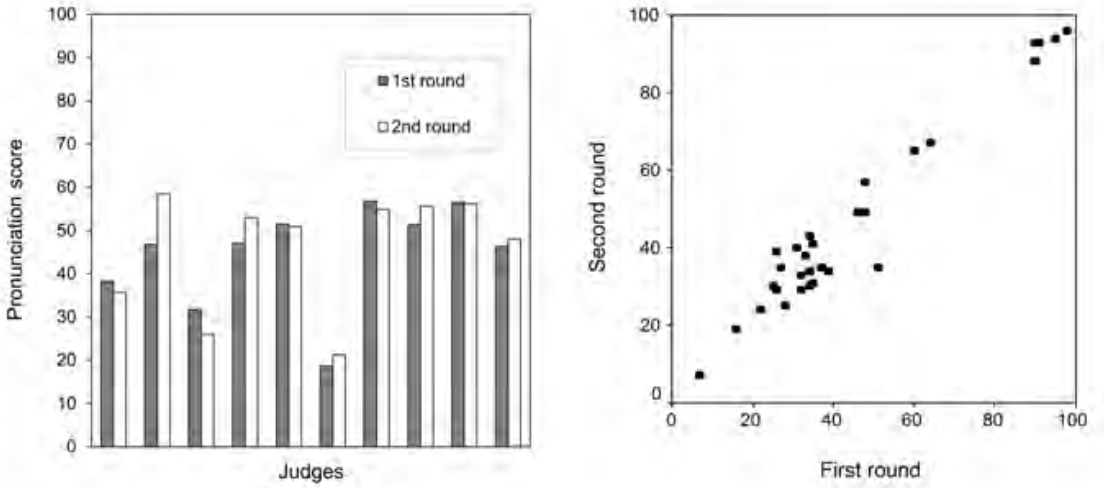

Figure 1. Left panel: The mean pronunciation score given on a 100-point scale by each judge to sentences in the first vs. the second round (not equivalent to Time 1 vs. Time 2). Right panel: The correlation between the pronunciation scores of each participant (averaged across test-time) at the first vs. the second rating round.

In order to test intra-judge consistency of rating of each participant, each judge's mean ratings given in the first vs. the second round were submitted to a Pearson correlation coefficient analysis. The first and second round judgments were strongly correlated ($r(8)=0.93$, $p<0.001$). This suggested that each judge was consistent in assigning scores. In addition, in order to test whether the judges as a group assigned the same pronunciation score to each participant in the two rating rounds, the talker-based mean scores were submitted to a Pearson correlation analysis. The correlation was highly significant ($r(28)=0.94$, $p<0.001$). A scatter plot of the talker-based round 1 and 2 correlation is shown in Figure 1, right panel. Note that, as expected, the five native English speakers, in the top right corner of the graph, consistently received very high scores, and that the variation among the native English speakers was similar in round 1 and round 2.

1.5 Pronunciation scores before and after immersion

The mean pronunciation scores given to each group at Time 1 and Time 2 are shown in Figure 2. The Native English Group received mean scores which approximated the maximum score of 100. This suggested that the judges successfully identified the native English speakers as speaking without a foreign accent, although the native speakers did not receive a perfect rating of 100. This suggests that on a few occasions, the judges were not completely sure that the speaker was a native speaker. This result aligns with previous studies of foreign accent rating (e.g., Flege et al., 1995; Yeni-Komshian, Flege, & Liu, 2000).

The Experience Group received higher scores at Time 2 than at Time 1. Before testing the significance of the difference, Figure 3 shows the distribution of scores in the three groups of participants. The figure shows that while the native English speakers (top panel) most often received near-maximum scores, the scores given to the No-experience Group (bottom panels) were skewed towards low scores at both Time 1 and Time 2.

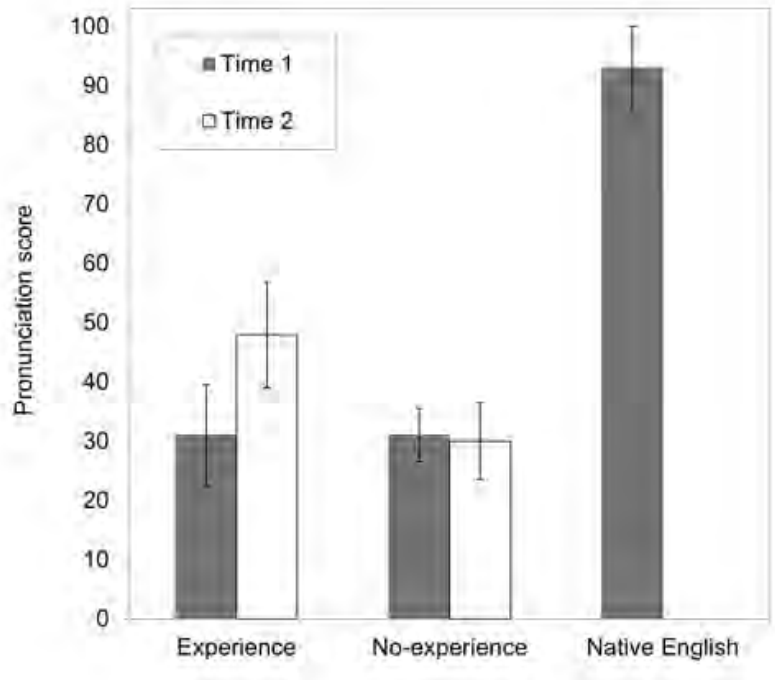

Figure 2. The mean pronunciation scores given to each group. A score of 100 indicates ”no foreign accent; a score of 1 indicates ”strong foreign accent”. Error bars denote +/– 1 SD.

However, the scores given to the Experience Group (mid panels) were less skewed towards low pronunciation scores at Time 2. Figure 3 provides important information about the pronunciation scores which are not apparent in Figure 2, namely that there is an appreciable spread of the scores. Notably, in each group of Danish speakers some sentences received high pronunciation scores, and with very few exceptions, all sentences produced by speakers in the Native English Group received very high scores.

The Native English Group was recruited mainly as a reference group, it was only tested once, and had an *N* of only 5. Therefore, only the scores of the Experience Group and the No-Experience Group were submitted to a two-way 2 (Group) × 2 (Test time) ANOVA with Test time as a repeated measure. The main effect of Group was nonsignificant ($F(1, 23)=0.25$, $p=0.621$, $\eta^2=0.10$), the main effect of Test Time was significant ($F(1, 23)=4.92$, $p=0.037$, $\eta^2=0.18$), and the Group × Test time interaction was significant ($F(1, 23)=6.30$, $p=0.020$, $\eta^2=0.22$). As expected, the source of the interaction was a significant simple effect of Test time for the Experience Group ($t(13)=2.81$, $p=0.015$, $d=1.51$), but not the No-experience Group ($t(10)= .42$, $p=0.684$, $d=0.27$), and a significant effect of Group at Time 2 ($t(23)=2.81$, $p=0.010$, $d=1.13$) but not at Time 1 ($t(23)<0.01$, $p=0.999$, $d<0.01$).

Whereas Figure 3 shows that the Experience Group (the middle panels) generally received more favorable pronunciation scores at Time 2 (right) than at Time 1, there were still many scores at the low end of the scale at Time 2. This indicates that some participants in the Experience Group did not improve their pronunciation during their stay in England. Individual scores for each of the 14 participants in the Experience Group at Time 1 and 2 are shown in Figure 4.

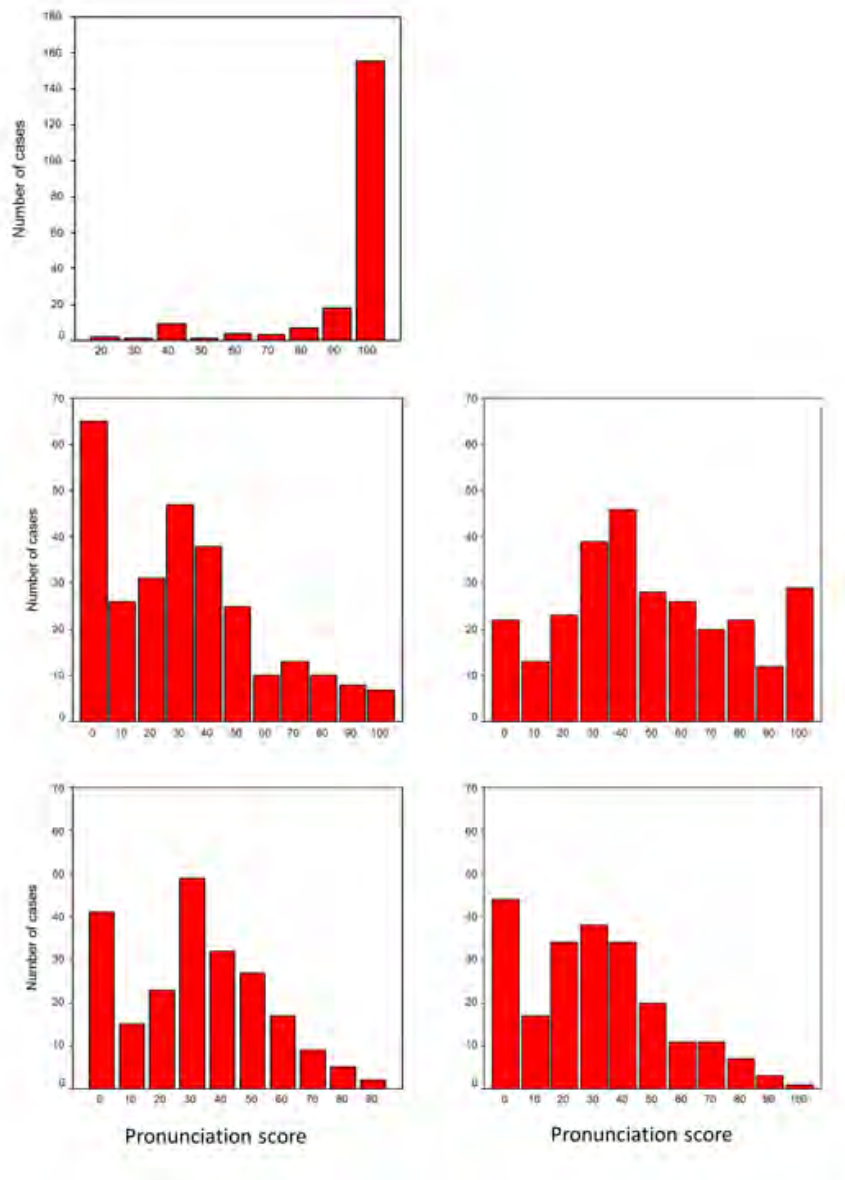

Figure 3. The frequency of pronunciation scores given to the Native English Group (top), the Experience Group (mid) and the No-experience Group (bottom). Time 1 results are shown in left panels, Time 2 results in right panels. The scores are bins of 10. Note that the frequency of scores is on different scales.

At Time 2, most (10 of 14) participants received higher scores than at Time 1; only 4 participants received lower scores. Recall that four of the participants had previous English immersion experience of between 9 and 12 months already at Time 1. These participants were number 1, 2, 5, and 14 in Figure 4. Participant 14 had a mediocre pronunciation score at Time 1, but received a high score at Time 2, which was comparable to the mean score of the Native English Group. However, the two participants who

showed the least progress – actually, they had nominally lower scores at Time 2 than Time 1 – were also two previously experienced participants. This suggests that amount of previous experience did not exert a uniform influence on pronunciation progress. However, participant 14 had an LOR in England of 11 months between Time 1 and Time 2, whereas participants 1 and 2, who also had previous immersion experience, had an LOR of only 3-4 months between Time 1 and Time 2. This suggests that LOR may also be an important factor in pronunciation score gain.

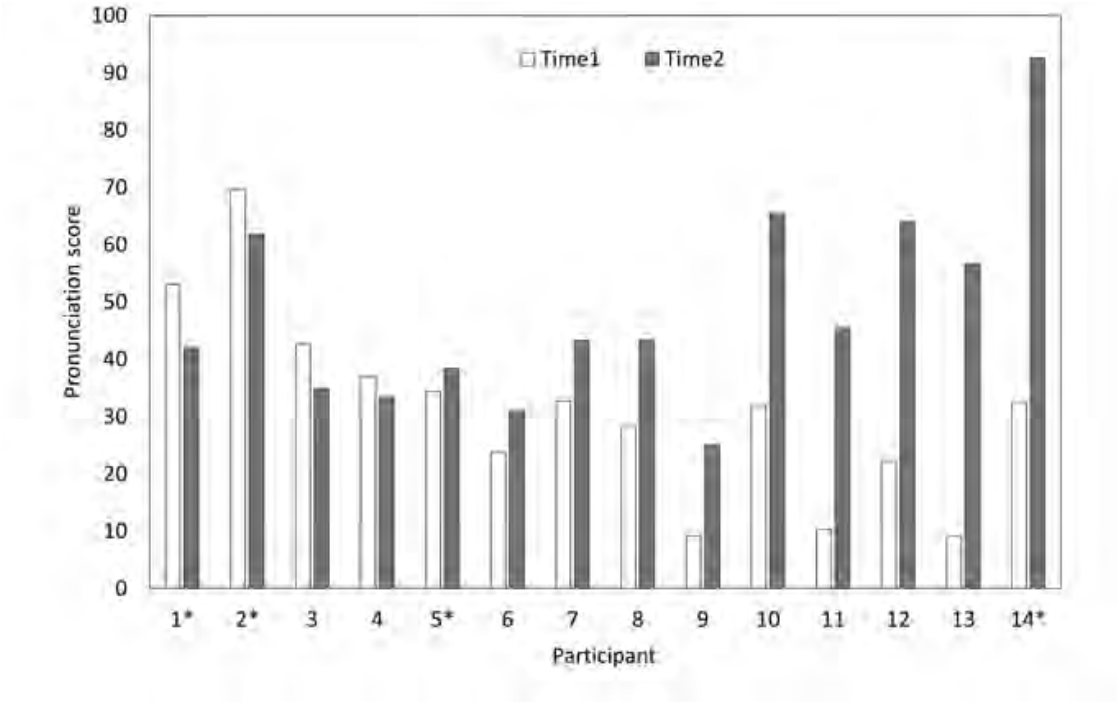

Figure 4. The mean pronunciation score assigned to each participant in the Experience Group at Time 1 and Time 2. Along the x-axis, the participants are ordered according to score increase from Time 1 to Time 2. Previously experienced participants are marked by a star.

To examine influences on participants' degree of improvement in pronunciation, a pronunciation gain score was derived by subtracting the Time 1 rating from the Time 2 rating. Figure 5, left panel, shows a scatter plot of the pronunciation gain score as a function of LOR for the 14 participants in Experience Group. In spite of some variation, the correlation between LOR and pronunciation gain score was significant ($r(12)=0.61$, $p<0.022$). Note that no participant with an LOR of less than five months improved their pronunciation. Also note that three of the six participants with an LOR of 9-11 months showed little improvement. Why did they not?

As mentioned, Flege & Liu (2001) found that the LOR effect on L2 acquisition was modulated by the intensity of L2 input which the learners were likely to have had. For the present study, the participants in the Experience Group rated their use of Danish vs. English use in active interaction on a scale from 1 to 5 (1 = Danish only, 3 = equal use of Danish and English, 5 = English only). A weighted English-language input measure was derived by multiplying the LOR in months by self-reported proportion of English use. Although it is not known whether the relative importance of LOR and language use is reflected accurately in the weighted input score, it is likely to be a better measure of participants' total amount of L2 input during their stay in England than LOR or English use alone.

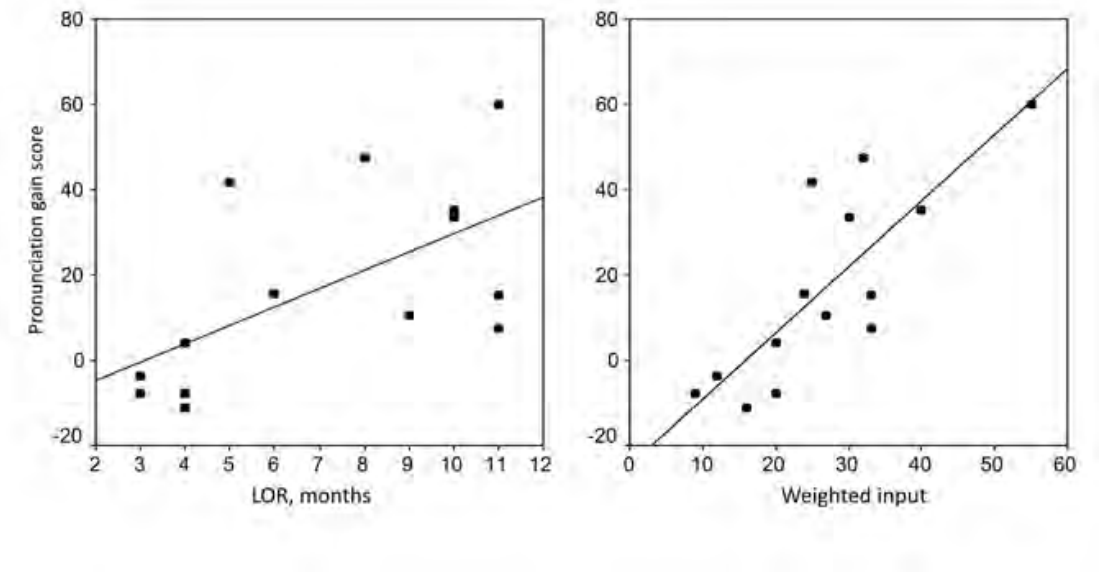

Figure 5. A scatter plot of pronunciation gain score as a function of LOR from Time 1 to Time 2 (left panel) and as a function of weighted input (LOR weighted by proportion English use, right panel).

Figure 5, right panel, shows a scatter plot of the pronunciation gain score as a function of weighted input. As shown, weighted input was quite successful at predicting pronunciation score gain, and the strong correlation between the two variables was significant ($r(12)=0.81$, $p<0.001$).

4. Discussion

The purpose of this small-scale study was to examine the effect of short-term immersion in England on young native Danish adult females' pronunciation of English. The results showed a significant effect after an

average of just 7.1 months of immersion on native judges' pronunciation rating of a single sentence produced by the participants in the Experience Group before and after immersion. On the other hand, the No-experience Group received virtually identical mean ratings for their two productions of the same sentence at two different time points.

The improvement in pronunciation after immersion varied greatly in the Experience Group, such that some participants did not improve at all, whereas others went from the low end of the rating scale to the high end. Importantly, the pronunciation gain score was significantly correlated with LOR. The data suggest that an LOR of at least five months is needed for a detectable improvement in L2 pronunciation, but note that the low *N* means that this result should be interpreted with caution.

A stronger correlation was found between pronunciation gain score and a composite measure of input derived by a simple multiplication of LOR and self-rated degree of L2 use. These results indicate that even L2 learners who have learned an L2 as a foreign language in a school setting and spoken it for about 10 years, can improve their L2 pronunciation as a rather direct function of the amount of L2 input they receive during immersion in an L2 community. The strong correlation was likely to be due to selection of a highly motivated and relatively homogenous Experience Group, i.e. all females with similar ages who were self-selected for an interest in traveling abroad. The results support the suggestion by Flege & Liu (2001) that LOR may provide only a coarse measure of L2 input and that degree of L2 use moderates the effect of LOR (see also Flege, this volume).

The results suggest that the young adult L2 learners were able to perceive at least certain phonetic differences between their own pronunciation of English and the pronunciation of the English that they encountered during their immersion period. Moreover, the improved pronunciation after short-term immersion suggests that the organization of the phonetic system in L2 learners is malleable and responds readily to new input, allowing for an approximation to the native norm of the L2.

Some of the participants in the Experience Group did not receive higher accent scores at Time 2. Two of the participants who did not improve were previously experienced (9-12 months of English immersion in England and the United States, respectively). This might indicate that L2 learners do not improve L2 pronunciation much after the first year of immersion, as suggested by Flege & Fletcher (1992). However, the participant who improved the most also had 9 months of prior English immersion

(in British Columbia, Canada). This suggests that the drop in pronunciation scores of the two non-improvers was not simply explained as a slowing down of the rate of learning following their initial period of immersion experience.

The improvement in pronunciation scores in adult learners after short-time immersion in an L2 speaking environment and the improvements' close association with duration and intensity of L2 input suggest a quite malleable phonetic system underlying speech production. This finding runs counter to the critical period hypothesis, at least in its original formulation, which suggests that adult L2 learners cannot make automatic use of input and build L2 representations based merely on L2 exposure (Lenneberg, 1967). It is true that the general pattern of this study, and that of previous research (e.g., Flege et al., 1995; Yeni-Komshian et al., 2000), is that a foreign accent is extremely difficult to avoid for adult learners, and this is in accordance with more recent and less stringent formulations of the critical period hypothesis (e.g., DeKeyser & Larson-Hall, 2005; Long, 2005). The more recent formulations of the critical period hypothesis merely claim as evidence for the critical period hypothesis that the AOL function is not strictly linear across the lifespan (i.e., a *sensitive* rather than a strictly critical period).

However, as noted by Vanhove (2013), one problem with the multiple and watered down formulations of the critical period hypothesis is that it may in essence be impossible to falsify the hypothesis. But at the very least, it seems possible to state with certainty that a biologically or maturationally defined critical period does not suffice to explain bilinguals' deviances from (monolingual) native norms. This conclusion is supported by work showing deviances in the L2 of very early bilinguals, who should not have passed their critical period (e.g., Flege et al., 1995; Yeni-Komshian et al., 2000), and even in the L1 of early bilinguals (Ivanova & Costa, 2008; Yeni-Komshian et al., 2000) as well as late bilinguals (Ammerlaan, 1996; Pavlenko, 2000; Pavlenko & Jarvis, 2002; Pelc, 2001).

As mentioned in the introduction, other accounts of bilingual deviances stress the importance of L2 use and input as well as the interaction between the L1 and L2 systems, which may vary with age or state of entrenchment of the L1 system at the onset of L2 acquisition (e.g., Flege, 1995; MacWhinney, 2016). In addition, domain-general cognitive aging has been proposed to explain AOL effects on L2 acquisition (Hakuta, Bialystok, & Wiley, 2003). Probably all L2 researchers acknowledge the existence of use and interaction effects on L2 skills and perhaps also

cognitive aging effects. What seems to remain controversial is whether unexplained variance could or should be attributed to, as yet, unidentified language-specific biological/maturational changes during childhood.

In summary, the results of the present small-scale study suggest that L2 pronunciation improves in immersed adult L2 learners as a function of a measure of L2 input (LOR weighted by degree of L2 use). Even though most or all late bilinguals continued to speak with a foreign accent, the present findings also suggest that even late bilinguals possess a readily malleable phonetic system.

References

Ammerlaan, T. (1996). *'You get a bit wobbly...': exploring bilingual lexical retrieval processes in the context of first language attrition.* (PhD), PhD dissertation, Radboud University, Nijmegen. Retrieved from http://repository.ubn.ru.nl/bitstream/handle/2066/146125/mmubn000001_215781937.pdf

DeKeyser, R. M., & Larson-Hall, J. (2005). What does the critical period really mean? In J. F. Kroll & A. M. B. de Groot (Eds.), *Handbook of Bilingualism* (pp. 88-108). Oxford: Oxford University Press.

Ellis, N. C. (2002). Frequency effects in language processing: A review with implications for theories of implicit and explicit language acquisition. *Studies in Second Language Acquisition*, *24*(2), 143-188.

Flege, J. E. (1988). Factors affecting degree of perceived foreign accent in English sentences. *Journal of the Acoustical Society of America*, *84*(1), 70-79.

Flege, J. E. (1991). Age of learning affects the authenticity of voice-onset time (VOT) in stop consonants produced in a second language. *Journal of the Acoustical Society of America*, *89*(1), 395-411.

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* (pp. 233-277). Timonium, MD: York Press.

Flege, J. E., & Fletcher, K. L. (1992). Talker and listener effects on the degree of perceived foreign accent. *Journal of the Acoustical Society of America*, *91*(1), 370-389.

Flege, J. E., & Liu, S. (2001). The effect of experience on adults acquisition of a second language. *Studies in Second Language Acquisition*, *23*(4), 527-552.

Flege, J. E., MacKay, I. R. A., & Piske, T. (2002). Assessing bilingual dominance. *Applied Psycholinguistics*, *23*(4), 567-598.

Flege, J. E., Munro, M. J., & MacKay, I. R. A. (1995). Factors affecting strength of perceived foreign accent in a second language. *Journal of the Acoustical Society of America*, *97*(5), 3125-3134.

Granena, G., & Long, M. H. (2013). Age of onset, length of residence, language

aptitude, and ultimate L2 attainment in three linguistic domains. *Second Language Research*, *29*(3), 311-343. doi:10.1177/0267658312461497

Gries, S. T., & Ellis, N. C. (2015). Statistical measures for usage-based linguistics. *Language Learning*, *65*(Suppl 1), 228-255. doi:10.1111/lang.12119

Hakuta, K., Bialystok, E., & Wiley, E. (2003). Critical evidence: A test of the critical period hypothesis for second-language acquisition. *Psychological Science*, *14*(1), 31-38.

Hart, B., & Risley, T. R. (1995). *Meaningful differences in the everyday experience of young American children*. Baltimore: Brookes

Hernandez, A., Li, P., & MacWhinney, B. (2005). The emergence of competing modules in bilingualism. *Trends in Cognitive Sciences*, *9*(5), 220-225.

Hurtado, N., Marchman, V. A., & Fernald, A. (2008). Does input influence uptake? Links between maternal talk, processing speed and vocabulary size in Spanish-learning children. *Developmental Science*, *11*(6), F31-F39. doi:10.1111/j.1467-7687.2008.00768.x

Højen, A. (2003). *Second-language speech perception and production in adult learners before and after short-term immersion.* (PhD dissertation), Aarhus University, Aarhus, Denmark.

Højen, A., & Flege, J. E. (2006). Early learners' discrimination of second-language vowels. *Journal of the Acoustical Society of America*, *119*(5), 3072-3084.

Ivanova, I., & Costa, A. (2008). Does bilingualism hamper lexical access in speech production? *Acta Psychologica*, *127*(2), 277-288. doi:10.1016/j.actpsy.2007.06.003

Lenneberg, E. (1967). *Biological foundations of language*. New York: Wiley.

Long, M. H. (2005). Problems with supposed counter-evidence to the Critical Period Hypothesis. *International Review of Applied Linguistics*.

MacKay, I. R. A., & Flege, J. E. (2004). Effects of the age of second language learning on the duration of first and second language sentences: The role of suppression. *Applied Psycholinguistics*, *25*(3), 373-396.

MacWhinney, B. (2016). Entrenchment in second language learning. In H.-J. Schmid (Ed.), *Entrenchment and the psychology of language learning* (pp. 343-366). New York: American Psychological Association.

Pavlenko, A. (2000). L2 influence on L1 in late bilingualism. *Issues in Applied Linguistics, 11*(2).

Pavlenko, A., & Jarvis, S. (2002). Bidirectional transfer. *Applied Linguistics*, *23*(2), 190-214.

Pelc, L. A. (2001). *L1 lexical, morphological and morphosyntactic attrition in Greek-English bilinguals.* PhD dissertation, City University of New York. Retrieved from https://academicworks.cuny.edu/gc_etds/1782

Piske, T., MacKay, I. R. A., & Flege, J. E. (2001). Factors affecting degree of foreign accent in an L2: a review. *Journal of Phonetics*, *29*(2), 191-215.

Singleton, D. (2005). The Critical Period Hypothesis: A coat of many colours *In-*

ternational Review of Applied Linguistics in Language Teaching (Vol. 43, pp. 269).

Snow, C. E., & Hoefnagel-Höhle, M. (1977). Age differences in the pronunciation of foreign sounds. *Language and Speech*, *20*(4), 357-365.

Snow, C. E., & Hoefnagel-Höhle, M. (1978). The critical period for language acquisition: Evidence from second language learning. *Child Development*, *49*, 1114-1128.

Suter, R. W. (1976). Predictors of pronunciation accuracy in a second language learning. *Language Learning*, *26*, 233-253.

Thompson, I. (1991). Foreign accents revisited: The English pronunciation of Russian immigrants. *Language Learning*, *41*(2), 177-204.

Vanhove, J. (2013). The critical period hypothesis in second language acquisition: A statistical critique and a reanalysis. *PloS One*, *8*(7), e69172.

Weisleder, A., & Fernald, A. (2013). Talking to children matters: Early language experience strengthens processing and builds vocabulary. *Psychological Science*, *24*(11), 2143-2152.

Yamada, R. A. (1995). Age and acquisition of second language speech sounds. Perception of American English /r/ and /l/ by native speakers of Japanese. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 305-320). Timonium, MD: York Press.

Yeni-Komshian, G. H., Flege, J. E., & Liu, S. (2000). Pronunciation proficiency in the first and second languages of Korean-English bilinguals. *Bilingualism: Language and Cognition*, *3*(2), 131-149.