PHONETIC SIMILARITY PREDICTS ULTIMATE ATTAINMENT QUITE WELL: THE CASE OF DANISH /i, y, u/ AND /d, t/ FOR NATIVE SPEAKERS OF ENGLISH AND OF SPANISH

Camila Linn Garibaldi[a] and Ocke-Schwen Bohn[b, c]
Department of English, Aarhus University, Denmark
[b] also at ConAmore, Department of Psychology, Aarhus University
[a] camila.linn.garibaldi@post.au.dk; [c] ocke.bohn@dac.au.dk

ABSTRACT

This study examined how well phonetic similarity predicted ultimate attainment in the production of Danish /i, y, u/ and /d, t/ by highly experienced native speakers of English and of Spanish. Experiment 1 compared the VOT of /d, t/ and the F2 of /i, y, u/ as produced by native speakers of the three languages. Experiment 2 examined the perceptual assimilation of Danish [i, y, u] to native /i, u/ by native English and native Spanish listeners. Experiment 3 indicated that native English speakers used their native (non-Danish) /t/ and /u/ in Danish, and did not produce Danish /y/ correctly. Native Spanish speakers produced Danish /y/ correctly, and they produced /t/ and many /d/ tokens differently in Spanish and in Danish. These results are interpreted in terms of current speech learning models which claim that phonetic similarity predicts speech learning success at ultimate attainment.

Keywords: L2 speech learning, ultimate attainment, L2 Danish, VOT

1. INTRODUCTION

Previous studies of nonnative (L2) speech learning (e.g., [5, 6, 9]) have shown that experienced learners' production and perception accuracy largely depends on the phonetic relation between segments in the L2 and the native language (L1). These studies have generally confirmed the predictions of both Flege's [8] Speech Learning Model (SLM) and Best's [1, 2] Perceptual Assimilation Model (PAM, and its extension to L2 learning, PAM-L2). Very briefly and somewhat oversimplified, these models predict that ultimate learning success is a function of the phonetic similarity of the sounds of the L1 and the L2. Similar sounds of the L2 (in SLM terms), or L2 contrasts that are perceived as equally good exemplars of an L1 category (Single-Category assimilation in PAM terms) are unlikely to be produced and perceived accurately, even after years of L2 experience. However, the more phonetically dissimilar the sounds of the L2 are, the more likely it is for experienced L2 learners to accurately perceive and produce these sounds which SLM labels as "new", or which in terms of PAM reveal Category-Goodness difference or an Uncategorized assimilation.

The present study further explores the role of the phonetic relation between L1 and L2 sounds for the production accuracy of highly experienced L2 learners who are likely to be close to their ultimate level of attainment. Specifically, we examined the production of L2 Danish (DK) /i, y, u/ and /d, t/ by speakers with English (EN) and Spanish (SP) as their L1.

From a phonological viewpoint, the learning task for SP and for EN learners appears to be the same: Both EN and SP have /i/ and /u/, and both EN and SP lack front rounded vowels like DK /y/. However, a review of the literature suggests that /u/ is implemented quite differently in EN and SP: EN /u/ can be quite fronted with a high F2 frequency, whereas SP /u/ is a true back vowel with a low F2 [7, 13]. Regarding /t, d/, the DK stop voicing contrast in initial position is implemented with short-lag vs. long-lag voice onset time (VOT). DK /t/ is affricated and has a very long VOT [11]. EN has a short-lag vs. long-lag contrast, whereas SP has a prevoiced vs. short-lag voicing contrast [10, 12].

The present study first compared the production of word-initial /d, t/ and of /i, u/ by L1 speakers of SP, EN, and DK, and of /y/ by L1 DK speakers. We focused on the most important acoustic correlate of voicing, VOT, for /d, /t/ productions, and the most important correlate of tongue position (front-back), F2, for productions of the close vowels /i, y, u/. Secondly, we examined the perceptual assimilation of DK /i, y, u/ to L1 SP and L1 EN close vowels /i, u/. The results of the first two experiments were then used to predict production accuracy for DK /d, t/, and /i, y, u/ by highly experienced learners of DK with L1 EN and L1 SP, which was examined in Experiment 3.

2. EXPERIMENT 1

2.1. Methods

Participants were 10 L1 SP speakers (5 f, 5 m; M_{age} = 43.4), 10 L1 EN speakers (4 f, 6 m; M_{age} = 48.8), and 3 L1 DK speakers (2f, 1 m; M_{age} = 41.0). DK language experience was comparable for the EN and SP participants with a mean length of residence in DK of 14.4 years (SD = 8.7) for the SP group, and 15.7 years (SD = 9.1) for the EN group.

Digital recordings were obtained from the participants reading target $C_1V_1C_2(V_2)$ words from their L1 in L1 carrier sentences[1]. C_1 and V_1 were the segments of interest (C_1: /t, d/, V_1: /i, u/, and /y/ for the L1 DK speakers), C_2 was an obstruent, and V_2 could be any vowel. Each speaker produced 3 randomly arranged repetitions of each of the 16 (EN and SP) or 15 (DK) target words.

Acoustic measurements were conducted using *praat* [3]. VOT for /t, d/ was measured from the release burst to the onset of voicing for the following vowel, or, in the case of prevoicing, from the point at which periodicity was detectable in the waveform until the release burst. F2 frequency was measured near the temporal center of the vowel, where formant frequencies appeared most stable in the spectral display.

2.2. Results

We compared VOT and F2 values across the three groups using one-way ANOVAs. Significant results were further explored using post-hoc pairwise multiple comparisons (Holm-Sidak method). Statistical comparisons were conducted on both Hertz values for F2 and on Bark-transformed Hertz values because of the unequal gender composition of the three groups. The results did not differ for the Bark-transformed or untransformed Hertz values. For ease of reference, Hertz values will be reported.

Separate ANOVAs for the VOTs of /t/ and of /d/ in the three languages revealed that the three groups produced /t/ with different VOTs in their L1s ($F(2,20)=50.43$, $p < .001$). The mean VOT for DK /t/ was significantly longer (140.0 ms) than for the EN /t/ (91.7 ms), which was significantly longer than the SP /t/ (22.6 ms). The significant ANOVA result for the VOT of /d/ ($F(2, 20)=222.13$, $p > .001$) was due to the negative VOT for the prevoiced SP /d/[2] (-105.0 ms), which differed from /d/ in both DK (28.3 ms) and EN (22.8 ms). The DK and the EN VOT for /d/ did not differ significantly.

Separate ANOVAs for the F2 of /i, y, u/ revealed that the F2 of /i/ as produced by the three language groups (DK: 2668 Hz, EN: 2458 Hz, SP: 2540 Hz) did not differ significantly, $p = .418$). However, F2 was higher for EN /u/ (1416 Hz) than for both SP /u/ (911 Hz) and DK /u/ (932 Hz), ($F(2,20) = 100.29$, $p< .001$), whose F2 did not differ significantly. This suggests a more anterior tongue position for EN /u/ than for SP and DK /u/. The F2 of DK /y/ (1943 Hz) differed significantly both from the F2 values of DK, EN, and SP /i/ and /u/.

2.3. Conclusion and predictions

The comparisons of the VOT values for /t/ and /d/ leads to the following predictions: Because of the large difference between L1 SP and DK /t/, the SP learners will be more successful at producing DK /t/ accurately than the EN learners, whose L1 /t/ is more similar to DK /t/ than the SP learners' L1 /t/. We also expect that SP learners will produce different /d/s in SP and in DK because of the large difference between prevoiced SP /d/ and short-lag DK /d/, whereas EN learners will not have to change the implementation of /d/ when speaking DK.

The comparison of the F2 of /i, y, u/ suggests that, in terms of tongue position, /i/ is an identical vowel in DK, EN, and SP, whereas DK /u/ is an identical vowel only for SP learners of DK. EN /u/ is considerably more fronted than both SP and DK /u/. This leads to the expectation that, because the two Spanish close vowels /i/ and /u/ are acoustically (in terms of F2) quite distinct from DK /y/, whereas EN /u/ is acoustically almost exactly half way between DK /y/ and /u/, SP speakers will be more successful at producing the very different ("new" in terms of SLM) DK /y/, whereas EN speakers will be less successful at producing DK /y/. We also expect that EN speakers will not produce a DK-like /u/ but rather a compromise between the similar (in SLM terms) back DK /u/ and fronted EN /u/.

3. EXPERIMENT 2

Experiment 2 examined the perceptual assimilation of DK [i, y, u] tokens to L1 /i, u/ by L1 EN and SP participants.

3.1. Methods

Participants were 8 L1 SP speakers (4 f, 4 m) and 8 L1 EN speakers (3 f, 5 m) who had also participated in Experiment 1.

Using the *praat* [3] stimulus presentation module, the participants individually first identified tokens of DK /di, dy, du/ as EN <doo> or <dee> (EN listeners) or as SP <di> or <du> (SP listeners). Immediately after labeling, the listeners rated the goodness of fit on a scale from 1 (*bad*) to 5 (*perfect*). Instructions were given orally and presented on the screen in the L1 of the participants. Three tokens each of DK [di, dy, du], recorded from a male L1

DK speaker, were presented five times each in random order.

3.2. Results

All SP and EN listeners assimilated DK [i] to their L1 /i/, and DK [u] to their L1 /u/. The SP and EN listeners' ratings indicated that the DK [i] and [u] tokens were equally good exemplars of /i/ and /u/ in SP and EN (with mean ratings of 3.7 for [i] and 3.6 for DK [u] by the SP listeners, and mean ratings of 3.3 for [i] and 3.1 for [u] by the EN listeners. Interestingly, the EN listeners classified all DK [y] tokens as EN /u/, but five of the eight SP listeners assimilated DK [y] variously to both SP /i/ and /u/. (One SP listener assimilated to only SP /i/, two SP listeners assimilated to only SP /u/). Both the EN and the SP listeners indicated that the fit for DK [y] was not as good as for DK [i, u]: The mean rating for DK [y] as EN /u/ was 2.3, and the mean ratings for DK [y] were 2.1 for SP /i/ and 2.0 for SP /u/. The fit for [y] relative to [i u] was judged to be worse by the SP than the EN listeners.

3.3. Discussion

The results from the perceptual assimilation experiment reflect the results of the acoustic comparisons of Experiment 1. The F2 for DK and SP /i/, and DK and SP /u/, did not differ significantly, which is reflected in the goodness ratings for the DK vowels which were much the same. However, for DK and EN, F2 differed for /u/, but not for /i/, which is not reflected in lower goodness ratings for DK [u] than DK [i] tokens. The goodness ratings for DK [y] suggest that EN listeners do not rate the fit as quite as bad relative to [i, u] as the SP listeners do. The goodness ratings for [y] suggest that the EN listeners are aware of the less than perfect match between DK [y] and EN /u/, indicating that this is a "similar" (SLM) vowel which is CG-assimilated (PAM). For the SP listeners, DK [y] is not a good match to two different SP categories, suggesting that [y] is so different as to be "new" (SLM) or Unassimilated (PAM), so that it is ultimately learnable.

Overall, the predictions derived from Experiment 2 are the same as those derived from Experiment 1: SP speakers will be more successful at producing /y/ and /u/ correctly in DK than EN speakers (for details, see 2.3 above).

4. EXPERIMENT 3

Experiment 3 tested the predictions derived from Experiment 1 and Experiment 2 by examining the production of DK /t, d,/ and /i, y, u/ by the same L1 SP and L1 EN speakers as in Experiment 1. Specifically, we compared the non-native speakers' VOT for DK /t, d/ and F2 for DK /i, y, u/ to the L1 DK speakers' values, and to the corresponding values for /t, d/ and /i, u/ as produced by the same speakers in their respective L1.

4.1. Methods

The methods, including the participants, were the same as in Experiment 1, except that the SP and the EN speakers now read the DK word list. Acoustic analyses were conducted using the same software and criteria as for Experiment 1.

4.2. Results

We used the same statistical procedures as for Experiment 1 (one-way ANOVAs with post-hoc pairwise multiple comparisons (Holm-Sidak method)). As for Experiment 1, Hertz values will be reported for /i, y, u/ because comparisons did not differ when conducted on the Bark-transformed or untransformed Hertz values of F2.

Separate ANOVAs for the VOTs of DK /t/ and of /d/ as produced by the three groups revealed that /t/ was produced with different VOTs ($F(2,20)=5.251$, $p = .015$) because the mean VOT of /t/ as produced by the DK group was significantly longer (140.0 ms) than the mean VOT of DK /t/ as produced by both the EN speakers (90.7 ms) and the SP speakers (86.1 ms). The VOTs for DK /t/ as produced by the EN and the SP speakers did not differ significantly. Paired t-tests revealed that the EN group did not produce L1 EN /t/ (M = 91.7 ms) with different VOTs than L2 DK /t/ M = 90.7 ms), $p = .805$). However, the SP speakers produced a significant difference between L1 SP /t/ (M = 22.6 ms) and L2 DK /t/ (M = 86.1 ms).

Regarding DK /d/, we first compare only the short-lag tokens (all DK /d/ tokens produced by the DK and EN speakers, 48.3 % of the SP speakers' DK /d/ tokens[3]). The VOT for DK /d/ as produced by the DK speakers (M = 28.3 ms) did not differ significantly from that of the EN speakers (M = 24.7 ms) and from the short-lag /d/s of the SP speakers (M = 21.5 ms, $F(2,20) = 1.438$, $p = .262$). Paired t-tests revealed that the EN group did not produce their L1 EN /d/ (M = 22.8 ms) with different VOTs than L2 DK /d/ M = 24.7 ms), $p = .185$). However, the SP speakers produced a significant difference between L1 SP /d/ (M = -105.0 ms) and the short-lag L2 DK /d/ (M = 21.51 ms), $p = < .0001$). The mean VOT of the 51.7 % of the SP speakers' DK /d/ tokens (-110.0 ms) which were prevoiced did not

differ significantly from the SP speakers' SP /d/ tokens (-105.0 ms).

These results suggest that the EN speakers implement /t/ in the same range of VOT in both DK and EN, even though DK /t/ has a much longer VOT than EN /t/. However, the SP speakers implement both /t/ and many /d/ tokens differently in DK and SP: Nearly one half of their DK /d/ tokens have short-lag VOT (as opposed to a prevoiced /d/ in SP), and they produce a long-lag /t/ in DK (as opposed to a short-lag /t/ in SP). However, this long-lag /t/ does not have the extremely long VOT values of DK /t/.

The F2 for /i/, which did not differ in the cross-language comparison (Experiment 1), did not differ when produced in DK by the DK, EN, and SP speakers ($F(2,20)=.912$, $p = 418$). However, the F2 for DK /u/ differed significantly ($F(2,20) = 100.299$, $p <.001$) because the EN speakers produced DK /u/ with a higher mean F2 (1350 Hz) than the DK and SP speakers, whose mean F2 for DK /u/ (DK speakers: 932 Hz, SP speakers: 1037 Hz) did not differ significantly. Paired t-tests revealed that the EN speakers did not produce a significant difference between the F2 for EN /u/ (M = 1416 Hz) and DK /u/ (M = 1350 Hz), $p = .296$. Surprisingly, the SP speakers produced DK /u/ with a higher F2 (M = 1037 Hz) than SP /u/ (M = 911 Hz), $p < .05$. Finally, the ANOVA revealed a significant difference between the F2 for DK /y/ ($F(2,20)=8.01$, $p = .003$) because the EN speakers produced DK /y/ with a lower F2 (M =1712 Hz) than the DK (M = 1943 Hz) and the SP (M = 1902 Hz) speakers. The difference in F2 between DK /u/ and DK /y/ as produced by the EN group was significant, $p < .001$.

4.3. Conclusions

Experiment 3 revealed that the EN speakers produced DK /t, d/ with VOTs which did not differ from the VOTs of the L1 EN /t, d/. This was expected for /d/, which is implemented in the same VOT range in DK and in EN, and it shows that the rather large difference in VOT of ca. 50 ms between EN /t/ and DK /t/ had no effect on the EN speakers production of DK /t/. The SP speakers, whose L1 /t/ is short-lag, produced a long-lag /t/ in DK which, however, fell short of the very long-lag VOT values for /t/ in DK. The SP speakers produced DK /d/ with two modal values. Roughly one half of these did not differ from the DK target, and the other half did not differ from the L1 SP values.

The results for the vowels showed that the SP speakers produced a DK-like /y/, whereas the EN speakers did not produce /y/ correctly. Both non-native groups produced DK /u/ with higher F2 values than the L1 DK speakers, which was expected for the EN speakers, but came as a surprise for the SP speakers.

5. DISCUSSION AND CONCLUSION

The present study first presented an acoustic comparison of DK /i, y, u/ and /t, d/ to EN and SP /i, u/ and /t, d/, and then examined the perceptual assimilation of DK [i, y, u] to EN and SP /i, u/. The results of Experiment 1 and Experiment 2 were interpreted in terms of SLM and PAM to yield predictions on how accurately highly experienced L1 speakers of SP and of EN would produce DK /i, y, u/ and /t, d/. These predictions were tested in Experiment 3, which examined the production of these DK sounds by the same participants as Experiment 1 and Experiment 2.

Overall, the predictions were quite correct. The SP speakers, but not the EN speakers, produced DK /y/ accurately, which was predicted based on the acoustic comparisons and the assimilation patterns. Also predicted was the failure of the EN speakers to produce DK /u/ accurately, but the fronting of DK /u/ by the SP speakers was unexpected. It could be that this result is due to different coarticulatory effects of flanking consonants across languages, as described in [4].

The predictions were also largely correct for the production of DK /t, d/. EN speakers produced these stops with L1 EN values, which was expected because of the cross-language similarity of /t/ and identity of /d/. SP produced long-lag /t/s, which, however, did not match the very long lag of DK /t/. This could be interpreted as an indication of a limit on how far even advanced learners can stretch their phonetic system to accommodate a new sound of the L2. Finally, the mixed results for the SP speakers' productions of DK /d/ (both prevoiced and short-lag) are interesting because some of the L2 learners produced a voicing contrast (prevoiced vs. long-lag) which is apparently unattested in the languages of the world, see [11]. Future studies will have to show how widespread this typological anomaly is in L2 speech.

---

[1] For EN: *I say the word _ again*, SP: *Digo la palabra __ otra vez*, DK: *Jeg siger ordet __ igen*.

[2] 8 (of 240) SP /d/ tokens produced with short-lag VOT (by 2/10 SP speakers) were excluded from the analyses.

[3] 8 (of 10) SP speakers produced both short-lag and prevoiced DK /d/s, one speaker produced only short-lag /d/s and one speaker produced only prevoiced /d/s

7. REFERENCES

[1] Best, C.T. 1995. A direct realist perspective on cross-language speech perception. In: Strange, W. (eds), *Speech Perception and Linguistic Experience: Issues in Cross-language Research*. Timonium MD: York Press, 167-200.

[2] Best, C.T., Tyler, M.D. 2007. Nonnative and second-language speech perception: Commonalities and complementarities. In: Bohn, O.-S., Munro, M.J. (eds), *Language Experience in Second language Speech Learning. In Honor of James Emil Flege*. Amsterdam: John Benjamins, 13-34.

[3] Boersma, P., Weenink, D. 2013. *Praat version 5.3.56*, retrieved 20.9.2013, http://www.fon.hum.uva.nl/praat/download_win.html.

[4] Bohn, O. S. 2004. How to organize a fairly large vowel inventory: The vowels of Fering (North Frisian). *J. Intern. Phon. Ass.* 34, 161-173.

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

[6] Bohn, O.-S., Flege, J. E. 1992. The production of new and similar vowels by adult German learners of English. *Studies in Second Lg. Acq.* 14, 131-158.

[7] Bradlow, A. R. 1995. A comparative acoustic study of English and Spanish vowels. *J. Acoust. Soc. Am.* 97, 1916-1924.

[8] Flege, J.E. 1995. Second language speech learning: Theory, findings, and problems. In: Strange, W. (eds), *Speech Perception and Linguistic Experience: Issues in Cross-language Research*. Timonium MD: York Press, 233-277.

[9] Flege, J. E., Bohn, O.-S., Jang, S. 1997. Effects of experience on non-native speakers' production and perception of English vowels. *J. Phonetics* 25, 437-470.

[10] Højen, A. 2000. Detection of Danish accent in English by native speakers of Danish. In: Leather, J., James, A. (eds), *Proceedings of the Fourth International Symposium on the acquisition of second-language speech*. Amsterdam: New Sounds, 80-86.

[11] Lisker, L., Abramson, A. S. 1964. A cross-language study of voicing in initial stops: Acoustical measurements. *Word* 20, 384-422.

[12] López, V.G. 2012. Spanish and English Word-initial voiceless stop production in code-switched vs. monolingual Structures. *Second Lang. Res.* 28, 243-263.

[13] Schmidt, A.M., Flege, J.E. 1996. Speaking rate effects on stops produced by Spanish and English monolinguals and Spanish/English bilinguals. *Phonetica* 53, 162-179

[14] Strange, W., Bohn, O. S., Trent, S. A., Nishi, K. 2004. Acoustic and perceptual similarity of North German and American English vowels. *J. Acoust. Soc. Am.*, 115, 1791-1807.