What does successful L2 vowel acquisition depend on? A conceptual replication

Jurriaan Witteman[1], Ekaterina Karaseva, Niels O. Schiller[1] & James M. McQueen[2]

[1]Leiden University, [2]Radboud University
j.witteman@hum.leidenuniv.nl

ABSTRACT

It has been suggested that individual variation in vowel compactness of the native language (L1) and the distance between L1 vowels and vowels in the second language (L2) predict successful L2 vowel acquisition. Moreover, general articulatory skills have been proposed to account for variation in vowel compactness. In the present work, we conceptually replicate a previous study to test these hypotheses with a large sample size, a new language pair and a new vowel pair. We find evidence that individual variation in L1 vowel compactness has opposing effects for two different vowels. We do not find evidence that individual variation in L1 compactness is explained by general articulatory skills. We conclude that the results found previously might be specific to sub-groups of L2 learners and/or specific sub-sets of vowel pairs.

Keywords: phonetics, speech production, vowels, bilingualism, compactness

1. INTRODUCTION

Acquiring native-like production of a second language (L2) is difficult when the native language (L1) has already been acquired. A basic assumption of all speech learning models is that L2 acquisition is difficult because the neural system has been optimized to process L1 sounds. This can interfere with the processing of L2 sounds.

Processing of L2 sounds is particularly difficult for L2 sounds that do not exist in the L1 but are very similar to an L1 sound. In this case, the L2 learner may simply substitute the L1 neighbouring sound for the two L2 sounds. Therefore, the L2 sounds may deviate from the target (native speaker) sounds. Consider the English vowels /u:/ (as in GOOSE) and /ʊ/ (FOOT). In Dutch, these vowels do not exist, but the similar vowel /u/ (as in MOE 'tired') does. Both English vowels are close to the Dutch vowel, but /u:/ is closer to /u/ than /ʊ/ (Figure 1). Indeed, it is known that Dutch speakers have difficulty contrasting /ʊ/ and /u:/ because they substitute Dutch /u/ for both English vowels [1].

According to the revised Speech Learning Model (SLM-r) [2], when an L2 sound is substituted by an L1 sound, the resulting L2 sound will sound more native-like when the L1 substitute is a better exemplar of the native target sound. Since Dutch /u/ is a better exemplar for English /u:/ than /ʊ/, the corresponding L2 /u:/ is expected to sound more native-like than the L2 /ʊ/.

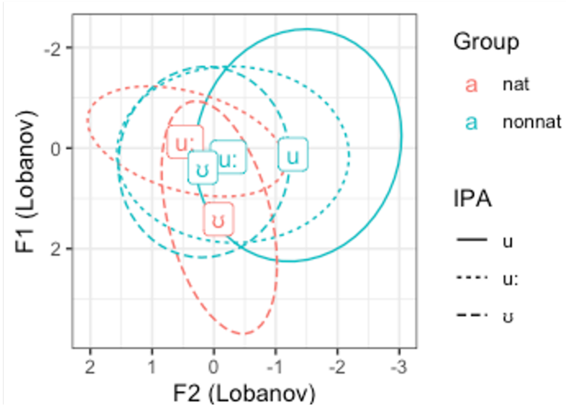

Figure 1. Vowel plot for English FOOT (/ʊ/) and GOOSE (/u:/) and Dutch MOE (/u/) for 6 native (nat) and 123 Dutch (nonnat) speakers of English. Ellipses are based on 0.67 SD's.

L2 vowel acquisition success is highly variable, with some individuals sounding more native-like than others. What might be causes of this individual variation? One hypothesis that can be derived from the SLM-r, is that if the Dutch vowel /u/ for a given individual is closer to the target English vowel in the individual's inter-phonological vowel space, the English token will be produced in a more native-like way.

A previous study [3] suggests an additional factor that could explain variation in L2 vowel production acquisition – variation in how precise individuals produce vowels in the *L1*. Some individuals show more token-to-token variation for the same L1 vowel than others. For instance, in Figure 2 below, vowel plots are shown for two individuals – one individual (left) has a relatively large token-to-token variation (the vowel is relatively 'non-compact') for Dutch /u/ while the individual on the right shows relatively little variation (the vowel is relatively 'compact').

Evidence was found [3] for the hypothesis that individuals with relatively compact L1 vowels produced L2 vowels more like natives than individuals with less compact L1 vowels. According to [3], better L2 production performance with more compact L1 vowels can be explained by the idea that a more compact L1 vowel space leaves more 'blank space' available for the accommodation of L2 vowels.

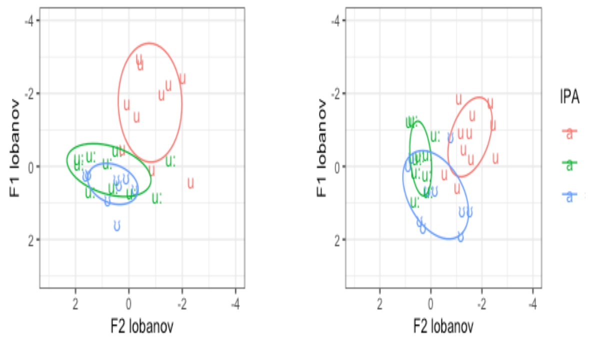

Figure 2. Vowel plot for English FOOT (/ʊ/) and GOOSE (/u:/) and Dutch MOE (/u/) for a relatively imprecise speaker of Dutch /u/ (left) and a relatively precise speaker (right). Ellipses are based on 0.67 SD's.

In line with the SLM-r above, the authors of [3] additionally found that for individuals whose L1 vowel was closer to the target L2 sound, L2 vowel production was more native-like. This result can be explained by the hypothesis that when the L1 competitor vowel in a given individual is a better exemplar of the L2 target vowel, the substitution will result in a more native-like target.

Last, an interaction was found, i.e., as the L1 competitor vowel came closer to the target native vowel for the L2, the influence of L1 vowel compactness on L2 target nativeness became larger. This result was explained as follows: when the distance between an L1 and L2 vowel is smaller, having a more compact L1 vowel is more beneficial in creating 'blank space' for the new L2 vowel then when the distance is large.

Regarding the *cause* of individual variation in L1 compactness, various factors have been proposed. One such factor suggested by [3] is that individual variation in L1 vowel compactness is explained by 'general articulatory skill' in how precisely individuals produce sounds. If true, then we would expect individuals who produce more compact *L1* vowels also to produce more compact *L2* vowels.

In the present study, we conceptually replicate [3] with a much larger sample size (almost 10 times larger), a new language pair (Dutch L1, English L2) and a new vowel pair (English /u:/-/ʊ/). From the literature discussed above we derive the following predictions:

1. Individuals with a more compact L1 /u/ vowel produce the L2 English vowels more accurately (smaller Mahalanobis distance to the native English vowels) than individuals with a less compact L1 /u/ vowel.
2. Individuals whose L1 /u/ is closer to the native speaker target L2 vowel /u:/ or /ʊ/ (smaller Mahalanobis distance) produce that L2 vowel more accurately (smaller Mahalanobis distance to the native speaker vowels).
3. When the distance between the L1 vowel and the English native speaker vowel is smaller, L1 vowel compactness will have a larger negative association with the distance between the L2 vowel and the native speakers than when the distance between the L1 vowel and the native English vowel is large.
4. There is a positive association between the compactness of L1 /u/ and the mean compactness of the L2 English vowels.

2. METHODS

2.1 Participants

A total of n = 123 (91 female) advanced Dutch (L1) learners of English (L2) were included from the first recording session of a longitudinal project investigating individual variation in L2 acquisition. All these participants were university students, had Dutch as their L1, had on average 6 years of high school education in English but had not received extensive explicit instruction in English phonetics and had not experienced immersion in an English-speaking environment. Additionally, six (3 female) native speakers of British (n = 3) and American (n = 3) English were included. All native speakers were monolingual speakers of English, grew up in an English-speaking country and used the contrast FOOT – GOOSE in their native variant of English. None of the participants reported any speech production or speech perception disorders.

2.2. Procedure

The study was approved by the Faculty's ethics committee at Leiden University and all participants gave informed consent. Participants were reimbursed 5-20 euro's for participation in the session (that ranged between 30-180 minutes). All participants read out the *North Wind and the Sun* (NWS) passage three times from a computer screen in a dimly lit sound attenuated booth. Dutch speakers additionally read out loud the Dutch version of NWS three times. Recordings were made using a RØDE NTG2 directional condenser microphone with appropriate amplification and at 16 bit resolution and a sampling frequency of 44.1 kHz.

2.3 Data analysis

All vowels were manually segmented by students who had received training in phonetics and extensive training in segmenting vowels using a segmentation manual. Vowels were segmented using visual inspection of the spectrogram and oscillogram and by listening to the vowel. F1 and F2 were extracted from vowels that did not contain creaky voice or presence of noise or interruption of the formant tracks. Mean formant frequencies were extracted using a PRAAT [4] script that used the Burg algorithm to calculate the mean F1 and F2 over the middle 50% of the vowel with a time step of 0.01 seconds, a window length of 0.025 seconds and a pre-emphasis of 50 Hz and a formant ceiling of 5,500 Hz.

Formant values that were more than 2 SD's away from the participant's vowel mean F1 or F2 were excluded. Vowel compactness was then calculated by speaker for every vowel as reported in [3]:

(1) $CS_v = SD_{F_1} \times SD_{F_2} \times \pi$

Compactness scores that were more than 2 SD's away from the mean compactness score were excluded. Mahalanobis distances (MD) between a token and a reference distribution were be calculated using the *mahalanobis* function from the *stats* package in R [5]. MDs that were more than 2 SD's away from the mean MD were excluded.

Categorical predictors were mean centered and continuous predictors were Z-normalized to promote convergence and to interpret interactions. For all models, the *lmer* function from the *lme4* package was used. The maximum random-effects structure was obtained by starting with random intercepts for Participant and Word and adding random slopes that increase model fit based on the likelihood ratio test using the *anova* function in R until the model did not converge. Statistical significance of main effects and interactions was tested using the *lmerTest* package.

For analysis involving more than one predictor and interactions, if an interaction is significant, the nature of the interaction was inspected using the *emmeans* function in R.

3. RESULTS

Descriptive statistics concerning MDs are given in Table 1 below. Note that Dutch /u/ indeed is closer to English /u:/ than English /ʊ/.

Table 1

Vowel pair	MD mean (SD)
/u/ - /ʊ/	9.33 (3.06)
/u/ - /u:/	8.37 (3.95)
L2 /ʊ/ - native /ʊ/	3.83 (4.18)
L2 /u:/ - native /u:/	4.65 (4.43)

To test hypotheses 1-3, a multilevel regression analysis was performed with the MD between the native distribution of tokens for a given English vowel and an L2 vowel token as dependent and (1) Z-normalized CSv for /u/, (2) the mean Z-normalized MD between /u/ and the native vowel, and (3) Vowel (/ʊ/ and /u:/) as predictors and Participant and Word as random intercepts and a by-participant random slope for the Vowel factor.

Contrary to expectations, there was (1) no main effect of vowel (2), no main effect of the MD between the L1 vowel /u/ and the native English vowels on production accuracy, (3) no main effect of L1 CSv on production accuracy, and (4) no interaction between L1 vowel compactness and the MD between L1 vowel /u/ and native vowels on production accuracy. However, there was a significant two-way interaction between Vowel and Compactness. In Figure 3 below, it can be observed that while for /ʊ/ there was indeed a positive relationship between L1 compactness and L2 production accuracy, for /u:/ the relationship was *negative*. However, when the interaction was followed up, the negative association for /u:/ did not reach statistical significance (β = -.22, 95 CI = [-0.72, 0.28], T(121) = -.87, p = 0.38) while the positive association for /ʊ/ approached statistical significance (β = 0.41, 95 CI = [-0.01, 0.839], T(113)= 1.93 , p = 0.06)

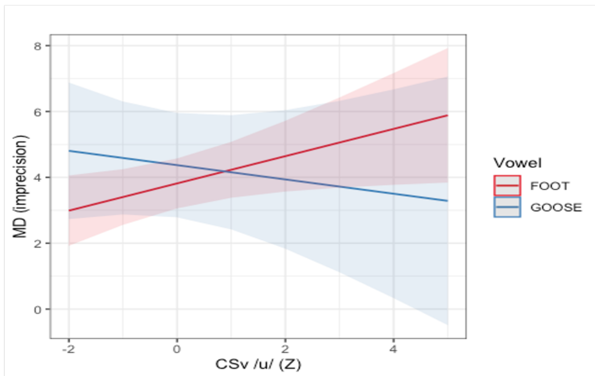

Figure 3. Interaction between Vowel and L1 compactness for L2 production accuracy.

Finally, the Pearson correlation between L1 vowel compactness and L2 vowel compactness was small (r = -0.03) suggesting no linear association between precision of L1 and L2 vowel production (see Figure 4).

4. DISCUSSION

In the present study, a conceptual replication was performed on the speech production part of [3] with an almost 10 times larger sample size (original study: n = 14, present study: n = 123). In contrast to [3], we did not find main effects of compactness or L1-L2 distance on L2 production accuracy, nor an interaction between L1-L2 distance and L1 vowel compactness on the accuracy of L2 production. We did find an interaction between Vowel and Compactness, with the expected positive association between L1 compactness and L2 production for /ʊ/ but a negative association for /u:/. However, neither simple main effect was statistically significant. Finally, as in [3], we did not find a positive association between L1 and L2 vowel compactness.

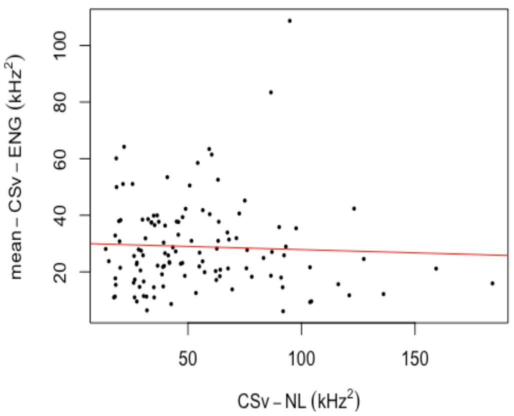

Figure 4. Vowel compactness of the L1 /u/ vowel and the mean compactness of the two L2 English vowels

Several explanations are possible for the lack of correspondence of the results with [3]. First, effects of L1 vowel compactness on L2 vowel production acquisition might be particularly prominent in subgroups of learners who produce L1 vowels with high variability – the L1 vowel compactness scores in the current study (mean: 54.1 kHz^2) were nearly five times smaller than in [3] (mean: 259 kHz^2). Perhaps due a floor effect, therefore, no effect of compactness could be observed. Since the compactness of L1 production was high in the present study, perhaps the effect of L1-L2 distance on L2 production was also smaller than in [3].

A second (non-mutually exclusive) explanation is that effects of L1 vowel compactness and distance of L1 to native target vowels in the to-be-acquired language are particularly prominent among non-advanced learners and less so in more advanced learners such as those in the present study. Perhaps vowel compactness and L1-L2 distance play a role only at the start of L2 vowel acquisition (when the L1 is relatively dominant).

Third, these effects might be limited to specific vowel pairs – in the present work, MD's between native and L2 vowel productions were larger than in [3], suggesting that the vowel contrast in the present study was more difficult than in [3] and that for relatively difficult vowel contrasts the role of L1 compactness might be smaller than for easier vowel contrasts. On the other hand, the difference in direction of the compactness effect for /ʊ/ than /u:/ suggests that the positive effect of L1 compactness found in [3] might particularly apply to L2 vowels for which the L1 competitor vowel is, at the same time, close in the inter-phonological vowel space but also a bad exemplar for the target vowel (i.e., /u/ for /ʊ/).

Fourth, given the small sample size of [3], the effects found might have been inflated [6] and the current larger study might provide more precise estimates.

In the present work, similarly to [3], we did not find a positive linear association between L1 compactness and L2 compactness. This suggests that it is unlikely that general articulatory skills explain variation in the accuracy of vowel production.

Future high-powered (preferably longitudinal) studies using various vowel-contrasts, language combinations and L2 learners in different stages of L2 acquisition are needed, however, to reach more definite conclusions.

5. REFERENCES

[1] Collins, B. & Mees, I. M. (2003). *The phonetics of English and Dutch.* Leiden – Boston.

[2] Flege, J., & Bohn, O. (2021). *The Revised Speech Learning Model (SLM-r).* In R. Wayland (Ed.), Second Language Speech Learning: Theoretical and Empirical Progress (pp. 3-83). Cambridge University Press.

[3] Kartushina, N. & Frauenfelder, U. H. (2014). On the effects of L2 perception and of individual differences in L1 production on L2 pronunciation. *Frontiers in Psychology,* 5, 1246.

[4] Boersma, P. & Weenink, D. (2022). Praat: doing phonetics by computer [Computer program]. Version 6.3.02, retrieved 29 November 2022 from http://www.praat.org/

[5] R Core Team (2022). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

[6] Button, K. S. *et al.* (2013). Power failure: Why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience,* 14, 365-376.