IMITATING SPEECH IN AN UNFAMILIAR LANGUAGE AND AN UNFAMILIAR NON-NATIVE ACCENT IN THE NATIVE LANGUAGE

Mónica A. Wagner[1], Mirjam Broersma[2], James M. McQueen[1,3], & Kristin Lemhöfer[1]

[1]Donders Institute for Brain, Cognition and Behaviour, Radboud University, [2]Centre for Language Studies, Radboud University, [3]Max Planck Institute for Psycholinguistics
m.wagner@donders.ru.nl

ABSTRACT

This study concerns individual differences in speech imitation ability and the role that lexical representations play in imitation. We examined 1) whether imitation of sounds in an unfamiliar language (L0) is related to imitation of sounds in an unfamiliar non-native accent in the speaker's native language (L1) and 2) whether it is easier or harder to imitate speech when you know the words to be imitated. Fifty-nine native Dutch speakers imitated words with target vowels in Basque (/a/ and /e/) and Greek-accented Dutch (/i/ and /u/). Spectral and durational analyses of the target vowels revealed no relationship between the success of L0 and L1 imitation and no difference in performance between tasks (i.e., L1 imitation was neither aided nor blocked by lexical knowledge about the correct pronunciation). The results suggest instead that the relationship of the vowels to native phonological categories plays a bigger role in imitation.

Keywords: phonetic talent, individual differences, speech imitation, pronunciation, lexical representations

1. INTRODUCTION

Why is it so hard to shake the traces of your mother tongue in a new language? And why, despite this, do some people still manage to acquire a native-like accent in a non-native language? These questions have led some researchers to propose such a thing as "phonetic talent" and to investigate the cognitive and linguistic processes that underlie this ability (e.g., [6]).

If such a general phonetic talent truly exists, one would expect that talented people would demonstrate an advantage with any new sound system they encounter. Support for the claim of phonetic talent comes from studies showing that speech imitation ability is related to L2 pronunciation. A relation with L2 pronunciation has been demonstrated with the imitation of unfamiliar speech sounds (e.g., [15, 20]), unfamiliar languages (L0; e.g., [17, 18]), and non-native accents and dialects in the mother tongue (L1; e.g., [7, 18, 20]). However, stronger evidence for phonetic talent would be provided by comparing performance on two completely unfamiliar sound systems. Here we aim to test the strong claim of a general phonetic ability in an L0 and an unfamiliar accent in the L1. We assess phonetic ability in both cases (i.e., L0 and unfamiliar accent in L1) with an imitation task, contrary to previous studies which have compared imitation to L2 pronunciation. If a general phonetic ability exists, imitation performance on the two tasks should be correlated; that is, participants who are good at one task should tend to be good at the other one.

However, imitating words in an unfamiliar language and accented words in a known language are different processes. One such difference is the prior existence of lexical representations in the L1 but not in the L0 which might thus affect L1 but not L0 imitation. These representations likely differ from the way the words are said with an accent. Knowing the word could either make the task less demanding (e.g., in terms of the amount of working memory resources required) or, to the contrary, interfere with imitation, blocking accurate pronunciation. Thus, in addition to examining the relationship between the two imitation tasks, we will compare performance in them to evaluate the potential role of lexical representations in imitation.

2. METHOD

2.1. Participants

Fifty-nine female native Dutch speakers who grew up monolingually in the same region in The Netherlands were recruited. Participants who reported any current hearing or speech impairments, or who indicated familiarity with the L0 or L1 accent, were excluded, leaving 57 participants between the ages of 18 and 28 (M = 22.5, SD = 2.3).

2.2. Materials

2.2.1. Critical items

Greek-accented Dutch and Basque were chosen for the L1 and L0 imitation tasks, respectively. They were thought to be equally unfamiliar to most potential participants. The L0 target sounds [a] and [e] were selected because they do not have a direct equivalent in Dutch. Perhaps the closest equivalents are the Dutch [aː] or [ɑ] and [eː] or [ɛ], respectively, but they all vary from the Basque sounds spectrally and/or in terms of length. For the L1 imitation task, the target sound deviations [ɪ] → [i] and [ʏ] → [u] were chosen. While the target non-native pronunciations are similar to sounds from the native Dutch inventory, they also do not perfectly coincide with Dutch sounds. The differences between the unfamiliar vowels and the native Dutch ones are shown in Figures 1-3.

For both tasks, two-syllable words were chosen in which the critical sounds appeared in the first and stressed syllable: the Basque words *datu* ("data point") and *etzi* ("the day after tomorrow") for L0 imitation, and the Dutch words *disco* (id.) and *dubbel* ("double") for L1 imitation. These words do not contain other sounds that were expected to be difficult for native Dutch speakers.

The words were recorded by a 29-year-old female native Basque speaker from Donostia and a 26-year-old female native Greek speaker from Crete who had been speaking Dutch for three years.

2.2.2. Baseline items

Before the imitation tasks, participants produced baseline utterances of the two words for the L1 imitation task, as well as Dutch words that contained the closest equivalents to the target sounds described above (i.e., *datum* ["date"] and *dat* ["that"] for *datu*, *etsen* ["etchings"] and *keet* ["shed"] for *etzi*, and *dief* ["thief"] and *doedel* ["doodle"] for *dubbel*). These words were matched to the critical items in terms of phonetic context and were obtained in order to compare the critical items and imitations thereof with the native Dutch vowel inventory.

In addition, both the participants and model speakers produced words with vowels in a CVC, CVCC, or CCVC (Dutch speakers) or CVCV (Basque model speaker) context so that we could normalize the values of their vowel utterances to their vowel spaces.

2.3. Procedures

At the beginning of the experiment, participants read lists of words out loud, including two instances each of the baseline items.

Next, participants performed two imitation tasks: in the first task (L0 imitation), they imitated real words in Basque and in the second (L1 imitation), they imitated Dutch words as spoken by the native Greek speaker. The participants were informed that they would hear words, in a foreign language and in Dutch respectively, and that they should repeat the words, trying to imitate how they were pronounced to the best of their ability. They were told that they would have five consecutive attempts per item to improve their imitations, during which they would first re-hear the model utterances before each attempt. In order to ensure that participants understood the words in the L1 imitation task, before each item, they saw the word they would be imitating printed on the screen. This was not done for the L0 to prevent participants applying Dutch spelling-to-sound rules. Each task began with two practice items that did not contain either of the target sounds to familiarize participants with the task and the speaker's voice.

After the imitation tasks, participants completed a demographic and language background questionnaire that included questions about their familiarity and experience with the L0 and with the L1 accent.

As one of the research questions concerns individual differences, the order of the tasks and items was kept constant across participants.

Participants performed the tasks individually, seated in a sound-attenuated recording booth in front of a computer screen where the written instructions were presented. Their speech was recorded digitally with a microphone and they listened to the models' utterances through headphones. The sound was set at the same comfortable volume for all participants.

2.4. Analyses

Analyses of critical items were conducted on the fifth attempt of each item. Vowels and words were manually annotated in Praat [5]. Vowel boundaries were set at a zero-crossing at the beginning and end of periodicity based on auditory and visual inspection of the waveform and spectrogram. Vowel duration and formant values at vowel temporal midpoints were automatically extracted with a script employing the default settings in Praat's built-in Burg algorithm (window: 25 ms, time step: 10 ms). Automatic formant extraction was checked by outlier inspection.

F1 and F2 were normalized for speaker's vowel space by applying the Lobanov transformation [1, 21] in the phonR package ([14]; see Figures 1 and 2). Euclidean distances to the model speakers' normalized critical vowels were then calculated.

Vowel durations were first normalized for speech rate by transforming them into proportion of total word length (raw values are shown in Figure 3). Then a difference score was obtained by subtracting the model speakers' values from each participant's values.

3. RESULTS

3.1. Relationship between L0 and L1 imitation

The relationship between the Euclidean distances of the imitated tokens to the model speakers' tokens for L1 imitation and the equivalent distances for L0 imitation was estimated using Spearman's rank correlation coefficient for non-normal distributions. There was no significant correlation between performance on the two tasks, $r_s = .15$, $p = .256$, 95% CI [-.11, .40] [19].

Figure 1: L0 imitations (dark gray), native Dutch baselines (light gray), and model speakers' tokens (black circles). Ellipses represent 1 SD (68% CI).

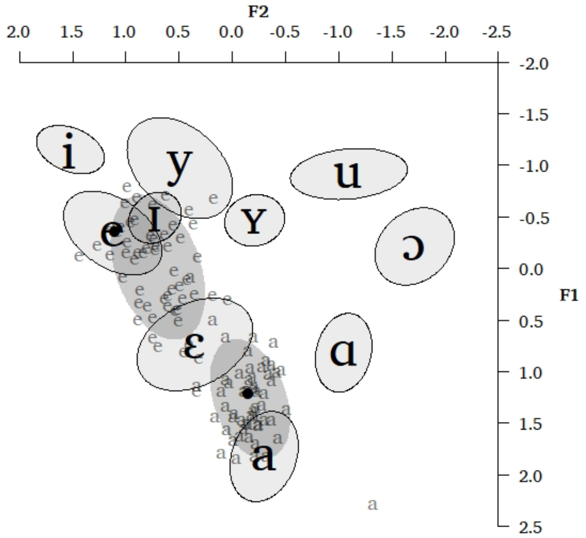
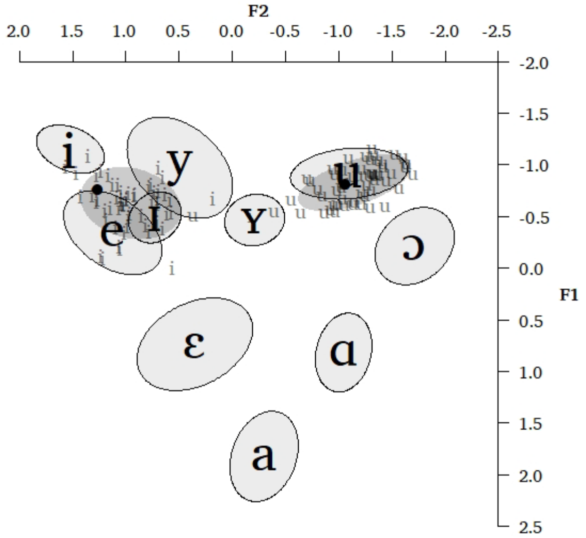
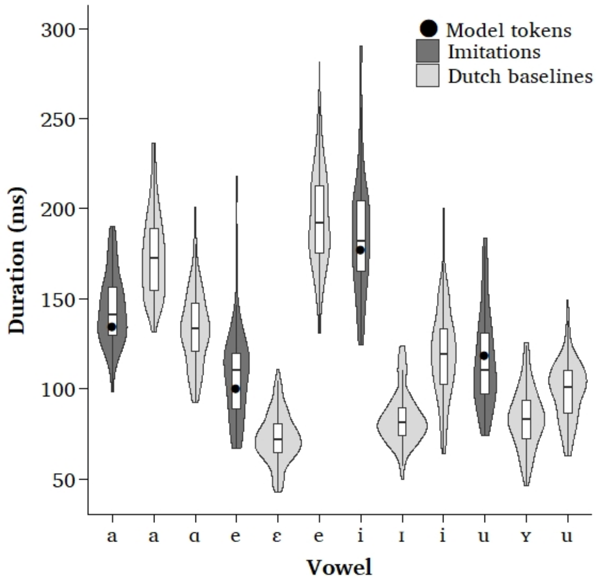


Figure 2: L1 imitations (dark gray), native Dutch baselines (light gray), and model speakers' tokens (black circles). Ellipses represent 1 SD (68% CI).



The relationship between duration differences between the imitated tokens and the model speakers' tokens for L1 imitation and the equivalent distances for L0 imitation was estimated using Pearson's product-moment correlation. Again, there was no significant correlation between performance on the two tasks, $r = .17$, $p = .211$, 95% CI [-.10, .41] [19].

Figure 3: Raw duration of imitations, native Dutch baselines, and model speakers' tokens.



3.2. Comparison between L0 and L1 imitation

Linear mixed-effects (LME) models and generalized linear mixed-effects models (GLMMs) were run using the lme4 package [2] in R [16] and p values for the GLMMs were calculated using the lmerTest package [11]. Random effect structures were determined by means of model comparison, opting for the most parsimonious model.

First, a GLMM was run on the non-normally distributed Euclidean distance data in order to assess the role of task (L0 or L1 imitation) on imitation performance. Task was a fixed factor and random effects for participant and vowel were included. There was no significant effect of task, $ß = -.23$, $SE = .27$, $t = -.85$, $p = .397$. Another model was run to test for an effect of vowel (4 levels) on the Euclidean distance data with participant as a random effect. The results (Table 1) reveal that imitations of /e/ were further from the target than all other vowels.

Table 1: Results of analysis of effect of vowel on Euclidean distances (Tukey-corrected [12]).

Contrast	Estimate	SE	z ratio	p
a - e	1.20	.25	4.72	<.001
a - i	.49	.29	1.73	.310
a - u	-.27	.33	-.83	.839
e - i	-.70	.22	-3.24	.007
e - u	-1.47	.27	-5.37	<.001
i - u	-.77	.30	-2.53	.056

A similar LME run on the durational data also found no significant effect of task, $ß = -.002$, $SE = .01$, $t = -.16$, $p = .879$. Here again there was an effect of vowel. The results (Table 2) reveal that participants tended to produce /u/ (-2.7%) relatively shorter than the model than they did /a/ (0.4%) and /i/ (2.4%), and that they extended the model's duration of /i/ compared to /e/ (-1.4%).

Table 2: Results of analysis of effect of vowel on duration differences (Tukey-corrected [12]).

Contrast	Estimate	SE	t ratio	p
a - e	.02	.01	2.08	.165
a - i	-.02	.01	-2.26	.113
a - u	.03	.01	3.60	.002
e - i	-.04	.01	-4.33	<.001
e - u	.01	.01	1.52	.426
i - u	.05	.01	5.85	<.001

4. DISCUSSION

The present study examined individual differences in speech imitation ability and the role of lexical representations in imitation. In particular, we aimed to answer two questions: 1) whether imitation of sounds in an L0 is correlated with imitation of sounds in an unfamiliar non-native accent in the L1, and 2) whether lexical knowledge makes L1 imitation easier or harder than L0 imitation.

As regards the first question, no significant correlation was observed between L0 and L1 imitation neither in terms of vowel spectral nor durational properties. This finding does not provide support for the strong claim of a general phonetic ability. It is also not consistent with the previous literature on imitative ability and L2 pronunciation. However, this may be because the analyses in those studies were done on a much more general level: while [17] and [18] used native ratings of 7-11 syllable-long utterances, [7] and [20] relied on self-ratings of the more general "mimicry ability." Even [15], which looked at accuracy in imitating specific features (e.g., aspiration following a voiceless plosive), measured performance on a rough rating scale in terms of whether the imitation was successful or not. It may be that, at least at the start of sound learning, phonetically talented people imitate cues beyond spectral or durational properties that make them sound more like the model, but which are not detected by our measures. An ongoing project in which we are obtaining ratings will be able to address this question.

With regard to the second question, L1 imitation was not found to be facilitated or hindered in comparison to L0 imitation. In fact, the language of the task was not relevant. Rather, performance on the tasks was better explained by the vowels that had to be imitated. Speakers were worst at approximating the spectral properties of the Basque /e/, while in terms of duration, they tended to undershoot the Greek-accented /u/ the most and overshoot the Greek-accented /i/ the most.

Inspection of the vowel plots reveals that the non-native targets varied in differing degrees from the corresponding native Dutch baselines. In some cases, such as the Greek-accented /u/, the spectral overlap with a native category may have aided performance, while in others, such as the Basque /e/, this overlap did not seem to help, perhaps due to the mismatching duration. These findings suggest that the relationship of the vowels to be imitated to the native phonological categories may play a bigger role during imitation than the language the words belong to [3, 4]. This is in line with studies that have found that the proximity of L2 sounds to the closest native category, as well as the compactness/dispersion of the L1 category, needs to be taken into account when assessing L2 pronunciation [9, 10]. Furthermore our results show that the effect of the relationship to native sounds is complex, sometimes perhaps facilitating accurate pronunciation, sometimes leading speakers astray.

Given the influence of native phonological representations, the large pre-existing differences between the target sounds and native Dutch vowels may have made it harder to assess the role of lexical representations in imitation. In addition, our tasks may not be fully comparable as they require different processes (e.g., because of difference in memory load). Future studies could better address the question of the role of lexical representations by employing the same sounds in known and unknown words.

Imitation has been gaining attention in the literature on individual differences in L2 pronunciation (e.g., [8, 13, 16]). However our findings paint a complex picture where several different factors such as the feature in question and the relationship of the target sounds to pre-existing phonological categories, and perhaps the existence of prior lexical representations, influence imitation. Further light needs to be shed on these intricacies before any conclusions can be made about the relationship between imitation and phonetic talent or L2 pronunciation. The current findings nevertheless show that a simple account where individual phonetic talent alone should predict imitation ability across all languages is not tenable.

5. REFERENCES

[1] Adank, P., van Hout, R., Smits, R. 2004. An acoustic description of the vowels of Northern and Southern Standard Dutch. *J. Acoust. Soc. Am.* 116, 1729-1738.

[2] Bates, D., Mächler, M., Bolker, B., Walker, S. 2015. Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software* 67, 1-48.

[3] Best, C. T. 1995. A direct realist perspective on cross-language speech perception. In: Strange, W. (ed), *Speech Perception and Linguistic Experience: Issues in cross-language research* (pp. 171-204). Baltimore: York Press.

[4] 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, 1334*, 1-47.

[5] Boersma, P., Weenink, D. 2013. Praat: doing phonetics by computer [Computer Program]. Retrieved from http://www.praat.org.

[6] Dogil, G., Reiterer, S. M. (Eds.). 2009. *Language Talent and Brain Activity*. Berlin: Walter de Gruyter.

[7] Flege, J. E. 1999. Age constraints on second-language acquisition. *J. Mem. Lang.* 41, 78-104.

[8] Hao, Y. C., de Jong, K. 2016. Imitation of second language sound in relation to L2 perception and production. *J. Phonetics* 54, 151-168.

[9] Kartushina, N., Frauenfelder, U. H. 2014. On the effects of L2 perception and of individual differences in L1 production on L2 pronunciation. *Front. Psychol.* 5:1246.

[10] Kartushina, N. Hervais-Adelman, A., Frauenfelder, U. H., Golestani, N. 2016. Mutual influences between native and non-native vowels in production: Evidence from short-term visual articulatory feedback training. *J. Phonetics* 57, 21-39.

[11] Kuznetsova A., Brockhoff P. B., Christensen R. H. B. (2017). “lmerTest Package: Tests in Linear Mixed Effects Models.” *Journal of Statistical Software* 82(13), 1-26.

[12] Lenth, R. V. 2016. Least-Squares Means: The R Package lsmeans. *Journal of Statistical Software*, 69(1), 1-33.

[13] Llompart, M., Reinisch, E. 2018. Imitation in a second language relies on phonological categories but does not reflect the productive usage of difficult sound contrasts. *Language and Speech*.

[14] McCloy, D. R. 2013. phonR: R tools for phoneticians and phonologists. R package version 1.0-7.

[15] Purcell, E. T., Suter, R. W. 1980. Predictors of pronunciation accuracy: A reexamination. *Lang. Learn.* 30, 271-287.

[16] R Core Team. 2018. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

[17] Reiterer, S. M., Hu, X., Erb, M., Rota, G., Nardo, D., Grodd, W., … Ackermann, H. 2011. Individual differences in audio-vocal speech imitation aptitude in late bilinguals: functional neuro-imaging and brain morphology. *Front. Psychol.* 2:271.

[18] Reiterer, S. M., Hu, X., Sumathi, T. A., Singh, N. C. 2013. Are you a good mimic? Neuro-acoustic signatures for speech imitation ability. *Front. Psychol.* 4:782.

[19] Signorell, A., Aho, K., Anderegg, N., Aragon, T., Arppe, A., Baddeley, A., ... Zeileis, A. 2019. DescTools: Tools for descriptive statistics. R package version 0.99.27.

[20] Thompson, I. 1991. Foreign accents revisited: The English pronunciation of Russian immigrants. *Lang. Learn.* 41, 177-204.

[21] Van der Harst, S. 2011. *The vowel space paradox: A sociophonetic study on Dutch*. Doctoral dissertation. Nijmegen, The Netherlands: Radboud University Nijmegen.