LEARNING TO PRODUCE DIFFICULT L2 VOWELS: THE EFFECTS OF AWARENESS-RAISING, EXPOSURE AND FEEDBACK

Aurora Troncoso-Ruiz, Mirjam Ernestus and Mirjam Broersma

Radboud University Nijmegen, the Netherlands
m.troncosoruiz@let.ru.nl, m.ernestus@let.ru.nl, m.broersma@let.ru.nl

ABSTRACT

This paper investigates how alignment in conversation fosters L2 sound learning. We studied the role of (1) awareness of the difficulty of the sound, (2) exposure to a native pronunciation and (3) implicit negative feedback on the learner's production.

Forty-nine Dutch speakers interacted with an English confederate playing a game designed to raise their awareness of the difficulty of the English /æ-ɛ/ contrast. During the interaction, we exhaustively controlled for the presence versus absence of exposure and implicit negative feedback.

We analysed participants' English /æ/ and /ɛ/ as produced in sentence-completion tasks before and after the interaction. Results show that Awareness-raising led participants to produce more native-like /æ/'s, both in height and frontness. Exposure led to improvement in the frontness of the two vowels. Feedback did not have an effect on the production of these vowels; the combination of Feedback and Exposure even hindered learning, undoing the effect of Exposure.

Keywords: Second language acquisition, sound learning, alignment, feedback, vowels

1. INTRODUCTION

Achieving a native-like pronunciation is a crucial and challenging aspect of second language (L2) acquisition. Incongruences between the sound system of a learner and that of the L2 may raise difficulties in the process of learning L2 sounds. These difficulties present themselves in the form of mispronunciations. Since the sound system of a learner remains adaptive over the course of their life [5], learners might eventually overcome these mispronunciations. This paper investigates three factors which might help learners overcome these incongruences and achieve a more native-like pronunciation.

First, one of the factors helping learners overcome mispronunciations could be *awareness*. If learners become conscious about their difficulty with producing an L2 sound distinctively, they may try to actively improve their pronunciation.

Second, *exposure* is commonly taken to be of key importance for sound learning. Speakers who are exposed to unfamiliar sounds in conversations may show adaptation to those sounds in their pronunciation. This phenomenon, known as *alignment*, takes place in L1 context, as well as in L2-L1 interactions [9]. We propose that L2 learners may obtain a more native-like production when exposed to native speakers' pronunciation.

Third, *feedback* might affect L2 sound learning. Previous studies have shown inconsistent outcomes with some evidence supporting the effectiveness of feedback on L2 learning [3, 7], while there is also evidence suggesting that feedback does not substantially contribute to learning [11].

In this paper, we investigate how learners' mispronunciations due to mismatches with the native sound inventory can be improved during a single conversation. We studied the roles of (1) *awareness* of the difficulty of the L2 sounds, (2) *exposure* to a native pronunciation, and (3) *implicit negative feedback* on the learner's pronunciations. By including these factors in the same study, we can assess their influence on learning independently, and also draw fair comparisons between them.

We used the recently developed Ventriloquist paradigm [4]. In this methodology, participants believe they are having a genuine conversation, whereas in fact they are interacting with a confederate who plays pre-recorded speech samples and never actually speaks to them. This paradigm provides the possibility to elicit spontaneous speech while controlling the three factors whose effects we wished to investigate.

Our participants were Dutch learners of English. The critical contrast for the experiment was the English /æ-ɛ/ distinction. These two vowels contrast in height (/æ/ is lower than /ɛ/) and in frontness (/ɛ/ is more fronted than /æ/). This contrast has been reported to be problematic for this population [2].

2. THIS STUDY

This experiment consisted of a production pre-test, an interaction and a production post-test. In the interaction stage, Dutch speakers played a puzzle-solving game in English with a confederate for 15-20 minutes. During this game, participants were visually exposed to minimal pairs. In critical trials, the minimal pairs differed in the contrast /æ-ɛ/. Their task in the game was to instruct the confederate to select one specific word. In critical trials, this word was the /æ/ member of the /æ-ɛ/ minimal pair. Participants, thus, had to pronounce the target vowel /æ/ clearly enough for the confederate to distinguish it from /ɛ/. This task was expected to raise their awareness of the difficulty to distinguish the /æ-ɛ/ contrast in their pronunciation.

Participants performed two Sentence Completion tasks containing the target sounds /æ/ and /ɛ/ of approximately 10-15 minutes, immediately before and after the interactive game. By comparing the pronunciations of the target contrast in the pre-test and the post-test tasks, we can evaluate the effect of *awareness* on L2 learning.

Exposure to the target sounds as pronounced by a native speaker (present/absent) and implicit negative *feedback* on their pronunciation (present/absent) were fully crossed during the interaction stage, resulting in four conditions.

3. METHODOLOGY

3.1. Participants

Forty-nine female Dutch-native speakers with an average age of 21.4 years (SD= 2.27) participated in this study. They were all students from Radboud University and their average LexTale [6] score was 75.78 (SD = 10.82) which corresponds to an upper intermediate level or B2 (CEF).

3.2. Materials

3.2.1. Critical and filler minimal pairs

Critical items were 24 minimal word pairs differing in the /æ-ɛ/ contrast (e.g. *flash/flesh*, *axe/ex*). All items were high frequency words (M = 4.81, SD = .661) according to SUBTLEX-UK [12]. As fillers, we selected four sets of 24 minimal word pairs containing /ɪ, i:/, /p, b/, /m, n/, or /r, l/, which Dutch learners find easier to distinguish.

3.2.2. Ventriloquist Recordings

A 25-year-old female speaker of British English recorded a large number of utterances to be played by the confederate during the interaction, unbeknownst to participants. These items were designed to enable giving participants the impression they were part of a real, live conversation (see [4]) and they were carefully scripted not to contain words with /æ-ɛ/.

3.2.3. Sentence Completion Task

The sentence beginnings that the participants had to complete in the Sentence Completion Task contained all 48 words from the /æ-ɛ/ minimal pairs. Words from the filler minimal pairs were also included in the sentence beginnings. In this way, we ensured participants pronounced a high number of words including the critical sounds, without being aware that these were the sounds of interest.

Two sentence beginnings were generated for each word from the 24 /æ-ɛ/ minimal pairs set, resulting in a total of 96 utterances. The two sentences including the same critical word were identical in structure and in the phonetic context of the critical word, e.g (a) *I saw my* ***dad*** *was checking his watch as I arrived really(...)* and (b) *I heard her* ***dad*** *was complaining about her going to(...)*. One sentence of each pair was included in the pre-test and the other one in the post-test.

3.3. Procedure

During the interaction stage, participants played the Code Breaker game with the confederate. In this game, the two players saw different information on their screens and had to co-operate in order to solve a series of puzzles. There were two types of trials: the Code Breaker (CB) trials, designed to manipulate *awareness* and *feedback*, and the Semantic Relationship (SR) trials, designed to manipulate *exposure*.

In the CB trials, the confederate saw an incomplete sequence of shapes and a set of four words, consisting of two minimal pairs, on her screen. The participant saw four shapes, each linked to one of those four words. Every trial of the game started with the confederate using pre-recorded audio to describe the sequence. In order to solve the puzzle, the participant had to indicate which of the four words was linked to the shape that could complete the sequence described, and the confederate had to click on that word on her screen.

The four words that appeared on the participant's screen consisted of two minimal pairs. Out of the total 64 CB trials, only 12 contained the /æ/ word from a critical minimal pair as the word linked to the right shape, meaning that only 12 times participants had to distinguish this contrast in production terms during the game. For the rest of the trials, the filler contrasts were used.

For the groups with implicit negative feedback, participants could see the selection made by the confederate. In the critical trials, when participants had to say a word with /æ/, the confederate systematically chose the wrong word, containing the /ɛ/ vowel, which made participants believe their pronunciation of the contrast was not clear enough. In conditions without feedback, participants never saw the selection made by the confederate.

In the SR trials, participants described a picture they saw on their screen to the confederate. The confederate saw two minimal pairs on their screen and determined which of the four words was semantically related to the picture described. In every trial, the confederate played an audio reading all the possible options out loud to the participant. To implement exposure, we introduced /æ-ɛ/ minimal pairs among the options read out loud in the audios played by the confederate in this type of trials. In conditions without exposure, all the words read in SR trials belonged to the filler minimal pairs; therefore, participants were never exposed to the native pronunciation of these vowels.

In order to draw fair comparisons between the effects of exposure and feedback, we restricted the frequency with which they appeared during the interaction: in all 12 SR participants heard the critical contrast and only 12 of the 64 CB trials induced negative feedback on participants' production of the critical contrast.

During the pre- and post-test, participants read the sentences beginnings, one by one, from a computer screen (Calibri, 28 pt). They were recorded producing the complete sentences with a Senheisser K-6 microphone, at a sampling rate of 44.1 kHz, 16-bit quantisation.

4. RESULTS

The sentences that participants produced during the pre-test and the post-test were automatically aligned to their corresponding orthographic transcriptions using Montreal Forced Aligner (MFA) [8]. We used the default MFA pre-trained models for English and modified the CMU dictionary to include Dutch-accented English pronunciations, improving the alignment.

Participants produced an average of 87.5 /æ/ tokens and 83.7 /ɛ/ tokens during the pre-test and 86.7 /æ/ tokens and 80.3 /ɛ/ tokens in the post-test. For every token, the first (F1) and the second (F2) formants were extracted at the midpoint of the vowel. In order to exclude mistakes produced by the automatic aligner, we removed all F1 and F2 values that were 2.5 standard deviations away from the means per vowel for each participant in the pre-test and post-test separately. The F1 and F2 values were transformed using the Lobanov transformation.

An exploratory analysis comparing the vowels of the voice recorded for the Ventriloquist and the vowels of the Dutch learners of English produced in the pre-test showed differences in the articulation of the two critical sounds. While the /æ-ɛ/ vowels differed in terms of height and frontness for the Ventriloquist, they did not differ in frontness for the Dutch speakers of English. The Dutch produced both less fronted than British /ɛ/ and less back than /æ/.

Due to participants' asymmetrical knowledge of the height and frontness distinction in the pre-test, we expected the factors under study to affect the F1 and F2 distinction differently. Therefore, we fit linear mixed effects models, using R [10] and the lme4 package [1], on the data for F1 and F2 separately. We included the theoretically relevant predictors: Vowel (/æ-ɛ/), Exposure (Yes/No), Feedback (Yes/No) and Test (pre-test/post-test), and a random intercept by Speaker and by Word. The model structures were improved by removing the effects that were considered not significant, i.e. those whose t-value were smaller than absolute 1.96. The final models were always refit without model outliers (2.5 standard deviations).

4.1. Vowel Height (F1)

For the data set of the F1 values of the vowels, there were simple effects of Vowel (β=-0.329, t =-7.303) and Test (β=0.027, t=1.964) as well as an interaction between the two (β=-0.051, t=-2.551). Together they suggest that participants distinguished the two vowels already in the pre-test in terms of F1 and that they further raised the F1 of the /æ/ (intercept) in the post-test. There seems to be no difference in F1 for /ɛ/ between pre- and post-test (i.e. no learning at all), as supported by releving of the data to place /ɛ/ on the intercept: we found no simple effect of Test. There were no significant effects of Exposure nor Feedback.

The F1 of /æ/ thus rises in all the conditions in the post-test, after participants played the Awareness-raising game, regardless of the presence or absence of Exposure and Feedback. The F1 of /ɛ/ stayed the same across all conditions.

4.2. Vowel Frontness (F2)

The model fit for the F2 values revealed a 4-way interaction among all the predictors, namely Vowel, Exposure, Feedback, and Test. In order to interpret this interaction, we split the data by Vowel.

For the F2 of /æ/ tokens, we found a main effect of Test (β=-0.035, t=-2.232) with the absence of Exposure and the absence of Feedback on the intercept, which means that the F2 of this vowel decreased in the post-test, after participants played the Awareness-raising game with the confederate, even if they received no Exposure or Feedback. We also found an interaction between Test and Exposure, (β=-0.042, t=-1.976), which indicated that the decrease in the F2 was larger with than without Exposure. As for Feedback, the interaction between Test and Feedback was not significant, meaning that the presence of only Feedback without Exposure did not have an additional effect on the F2 values. We found a 3-way interaction between Exposure, Feedback and Test (β=0.067, t= 2.290) going in the opposite direction. This interaction indicates that the presence of Feedback modulated the effect of Exposure. We subset the data to include only the participants who received both Exposure and Feedback and we found no effect of Test. The absence of this effect suggests that the combination with Feedback hindered the learning triggered by Exposure.

In summary, participants in the condition where Exposure and Feedback were not present learned to produce /æ/ further back. The presence of Exposure induced a larger learning effect. Feedback, on the other hand, had no effect in isolation and hindered learning when combined with Exposure.

As regards the F2 values of /ɛ/ tokens, we did not find a main effect of Test, which indicates that for the condition in the intercept, the condition without Exposure and Feedback, the F2 values did not change between the pre-test and the post-test. There was an interaction between Exposure and Test (β=0.057, t=2.843), indicating that the presence of Exposure made the F2 values increased in the post-test compared to the pre-test. The interaction between Feedback and Test was not significant, meaning that the presence of Feedback in isolation did not affect F2 values. We also found a 3-way interaction among Feedback, Exposure and Test, (β=-0.099, t= -3.555). This interaction suggests that the effect of Exposure was, again, modulated by the presence of Feedback. We fit a model on the subset of the data only including the condition where both Exposure and Feedback were present and it revealed no effect of Test, which indicates that Feedback and Exposure did not lead to any learning in the conditions were Feedback was also present.

Briefly, Awareness and Feedback had no effect on the F2 values. Only Exposure without feedback raised the F2 of /ɛ/. However, the combination of the Feedback and Exposure led to no change between the pre-test and the post-test.

5. DISCUSSION

The current study investigated whether, in conversation, Dutch learners of English improve their pronunciation of the problematic English contrast, /æ-ɛ/, when they (1) become aware of the difficulty of the contrast, (2) were exposed to native pronunciation of these sounds and (3) obtain feedback on their production.

Awareness-raising led participants to produce more native-like /æ/'s in the post-test, both in terms of height and frontness. Participants did not show any improvement in /ɛ/ driven by awareness-raising. Exposure led to an improvement in the frontness of the vowels: /æ/ became more back and /ɛ/ more fronted. Feedback did not have an effect on the production of these vowels when implemented without exposure. The combination of Feedback and Exposure hindered learning as it undid the effect of Exposure.

One possible explanation to why feedback was not beneficial could be related to the specific type of feedback implemented in this study. For the sake of consistency, the feedback participants received was always corrective, regardless of the actual production of the vowels. With this type of feedback, participants' pronunciation was corrected in every one of the 12 critical trials, even if they were making an effort to pronounce /æ/ distinctively different from /ɛ/. Participants may have found this type of feedback confusing.

Similarly, the combination of exposure and feedback was not helpful either. The presence of the two factors together may have been too overwhelming or confusing for participants to benefit from their presence.

In general, /æ/ seems to be more affected by learning than /ɛ/. Exposure affects both vowels, but only /æ/ also improved as a result of awareness. This could be due the fact that /æ/ had more room for improvement, since Dutch-accented English /æ/ was articulated further away from their native English counterpart than Dutch English /ɛ/ is from its equivalent. If Dutch speakers are aware that Dutch English /ɛ/ is closer to the native sound, they may be less willing to change their pronunciation of that sound. Another possible explanation could be related to the fact /æ/ was the focus of the experiment since, in the critical Code Breaker trials, participants always had to produce the word containing /æ/ and never /ɛ/.

These results suggest that exposure leads to changes in the vowels' frontness but not in the height distinction. One possible explanation could be linked to the participants' previous knowledge. The pre-test analysis of the vowels shows that participants already applied the height distinction between these two sounds. Perhaps the combination of awareness and exposure during the experiment made participants realise that frontness is also a cue distinguishing the two sounds. As a result, participants started applying this cue too.

In conclusion, this study has shown that awareness of the difficulty of an L2 sound, and exposure to native pronunciation trigger short-term learning of L2 sounds. Corrective feedback did not help participants achieve a more native-like pronunciation. Moreover, this experiment has shown that conversation-intrinsic factors can be studied in a naturalistic setting that is nevertheless highly controlled.

6. ACKNOWLEDGEMENTS

This research was supported by an NWO Vidi grant awarded to the third author.

7. REFERENCES

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

[2] Broersma, M. 2005. *Phonetic and Lexical Processing in a Second Language*. Nijmegen, The Netherlands: PhD dissertation, Radboud University Nijmegen

[3] Ellis, R., Loewen, S., Erlam, R. 2006. Implicit and explicit corrective feedback and the acquisition of L2 grammar. *Studies in second language acquisition*, 28(2), 339.

[4] Felker, E., Troncoso-Ruiz, A., Ernestus, M., Broersma, M. 2018. The ventriloquist paradigm: Studying speech processing in conversation with experimental control over phonetic input. *The Journal of the Acoustical Society of America,* 144(4).

[5] 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,* 233-277.

[6] Lemhöfer, K., Broersma, M. 2012. Introducing LexTALE: A quick and valid Lexical Test for Advanced Learners of English. *Behavior Research Methods*, 44, 325-343.

[7] Lyster, R., Mori, H. 2006. Interactional feedback and instructional counterbalance. *Studies in second language acquisition,* 28(02), 269-300.

[8] McAuliffe, M., Socolof, M., Mihuc, S., Wagner, M., Sonderegger, M. 2017. Montreal Forced Aligner: trainable text-speech alignment using Kaldi. In *Proceedings of the 18th Conference of the International Speech Communication Association*.

[9] Pickering, M.J, Garrod, S. 2004. Toward a mechanistic psychology of dialogue. *Behavioral and brain science,* 27(2), 169-190.

[10] R Development Core Team. 2008. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

[11] Truscott, J. 1996. The case against grammar correction in L2 writing classes. *Language learning,* 46(2), 327-369.

[12] Van Heuven, W.J.B., Mandera, P., Keuleers, E., Brysbært, M. 2014. Subtlex-UK: A new and improved word frequency database for British English. *Quarterly Journal of Experimental Psychology*, 67, 1176-1190.