Using ultrasound tongue imaging to improve L2 English pronunciation in Dutch students

Lisanne de Jong[1] & Martijn Wieling[1,2]
[1]University of Groningen, [2]Haskins Laboratories

Introduction

Recently, interest in the application of speech-production-based technologies in pronunciation training for second language (L2) learners has increased. Similar to the application in clinical settings where studies show that the visualization of articulators (such as the tongue) can facilitate speakers in producing target sounds (e.g. Preston, Brick & Landi, 2013), this type of bio-visual feedback might also aid L2 learners in producing non-native sounds. Ultrasound tongue imaging (UTI) is a non-invasive technique that can be used to visualize tongue movements in a way that is relatively easy to interpret. Up until now, several studies have shown beneficial effects of a pronunciation training using UTI-based visual feedback on the production of non-native sounds (e.g. Ouni, 2014; Cleland et al., 2015). Building onto this research, our study investigated whether a short training using UTI could improve Dutch high school students' pronunciation of two English target contrasts, /æ/-/ɛ/ (as in *bat* - *bet*) and /k/-/g/ (as in *pick* - *pig*). These contrasts were chosen as Dutch learners of English generally have difficulties with /æ/ and /g/ and tend to substitute these with /ɛ/ and /k/ respectively. Another reason to include these particular two contrasts was to find out whether visualizing articulators can be helpful regardless of the target sound to be learned, or whether it might be more helpful for sound contrasts where the difference lies mostly in the tongue shape (i.e. more for /æ/-/ɛ/ than for /k/-/g/).

Methodology

Participants

The data was collected at a high school located in Ter Apel in The Netherlands. The 38 participants (24F, 14M) consisted of first- and second-year students with an average age of 12.7. All were native speakers of Dutch who did not have any other native languages, nor did they report any severe language, speech or hearing disorders. Ahead of the experiment, the parent(s)/guardian(s) of the participants gave their written consent and provided background information about their child through a survey. Students received 10 euros for participating. Ethical approval for the study was obtained through the University of Groningen, Faculty of Arts' Central Ethical Testing Organization.

Procedure

Participants first did a perception experiment in which they heard words from minimal pairs containing either the /æ/-/ɛ/ or the /k/-/g/ contrast and had to indicate which word they heard (i.e. *bat* or *bet*). Next, participants were asked to read a list of words (presented in randomized order), containing either the /æ/ or /ɛ/ sound or the /k/ or /g/ sound. This functioned as the pre-test. Pronunciations were recorded with a microphone (Shure WH20) attached to the ultrasound headset (Articulate Instruments Ltd). Following this, in a session of roughly 20 minutes, the researcher trained the participants on the articulatory differences between /æ/-/ɛ/ and /k/-/g/. They were able to practice the target words and listen to previously recorded pronunciations by a native speaker of (American) English. The content of this training was the same for all participants, except that only half of them had access to UTI feedback (audiovisual condition). Participants in this condition saw their own (live) UTI image as well as the UTI video of the pronunciation of the native speaker. The other group did not

receive any visual feedback (audio-only condition). After the training session, participants read the list of the (randomized) words again in the post-test.

Rating data

In order to judge the pronunciations by the Dutch speakers, native speaker judgments were collected using an online survey. A total of 248 native speakers of English (60F, 179M, 9 who indicated 'other' or preferred not to indicate their gender) with a mean age of 49.8 took part in this part of the study. They were given a set of recordings (in randomized order) and for each recording were asked to indicate which word out of two (i.e. bad or bed) they heard. None of the raters reported severe hearing issues.

Results

To assess the effect of the training session on the pronunciation, we performed a mixed-effects logistic regression analysis with the dependent variable being whether or not the target word was recognized correctly by the rater. In the model with the optimal random-effects structure, a significant effect of test phase was found (β = .21, p < .05), meaning that target words recorded in the post-test were significantly more likely (0.2 logits, corresponding to an increase of about 5% in recognition probability) to be recognized correctly by the raters than words recorded in the pre-test. However, the experimental condition (audiovisual versus audio-only) did not have a significant effect on recognition, either by itself or in interaction with test phase. Regarding the target contrasts, target words in the /k/-/g/ category were more likely to be recognized correctly than target words in the /æ/-/ɛ/ category (β = .66 , p < .01; about 16% recognition probability increase). An interaction with test phase did not improve the model and was therefore not included. However, we did find a significant effect of participants' score in the perception task and target words in the /k/-/g/ category (β = .27, p < .001). No other significant influences of personal characteristics on recognition were found.

Discussion

This study investigated whether a short training session using UTI-based visual feedback would improve Dutch learners' pronunciation of the English sound contrasts /æ/-/ɛ/ and /k/-/g/. Looking at the native speaker ratings, we found that words pronounced after the training were more likely to be recognized correctly, indicating that even in a short session explicit instruction on the articulatory differences between the target sounds could help speakers in improving their production of the sounds. Although we did not find any significant differences between the participants in the audiovisual and the audio-only condition, a reason for this might be the short length of the training session. Students might need more time to familiarize themselves with the interpretation of the UTI signal, especially for sound contrasts like /æ/-/ɛ/ where the tongue shape differences are subtle.

References

Cleland, J., J. M. Scobbie, S. Nakai & A. Wrench (2015). Helping children learn non-native articulations: The implications for ultrasound-based clinical intervention. Paper presented at the 2015 International Conference of Phonetic Sciences, Glasgow, Scotland.

Ouni, S. (2014). Tongue control and its implication in pronunciation training. *Computer Assisted Language Learning*, *27*(5), 439-453.

Preston, J., Brick, N. & Landi, N. (2013). Ultrasound Biofeedback Treatment for Persisting Childhood Apraxia of Speech. *American Journal of Speech Language Pathology*, *22*, 27-643.