LEARNING OF A NON-NATIVE VOWEL THROUGH INSTRUCTED PRODUCTION TRAINING

Antti Saloranta[1,2], Henna Tamminen[1,2], Paavo Alku[3], Maija S. Peltola[1,2]

1) Department of Phonetics, University of Turku, Finland 2) Learning, Age and Bilingualism laboratory (LAB-lab), University of Turku, Finland 3) Department of Signal Processing and Acoustics, Aalto University, Finland

antti.saloranta@utu.fi

ABSTRACT

Models of second language acquisition predict that adults are unlikely to learn to produce a non-native vowel very quickly due to their reduced sensitivity to acoustic features not phonological in their native language. Earlier studies show that attention focusing training and articulatory instructions can facilitate learning of non-native speech sounds. We used an instructed listen-and-repeat paradigm with young Finnish adults to help them learn a non-native vowel in two days of training. The results indicated that adults learned to produce the vowel after just one session of training, likely as a result of the combination of explicit instructions and training that made them maximally responsive to the new acoustic features.

Keywords: production training, articulatory training, second language acquisition

1. INTRODUCTION

Prevailing models of speech acquisition state that adults have difficulties perceiving and producing speech sounds and contrasts that are not phonological in their native language. Two prominent models, the Speech Learning Model (SLM) by Flege [9] and the Perceptual Assimilation Model (PAM) by Best and Strange [3], both assert that the most difficult learning situations arise when the novel sound or contrast overlaps with existing native phoneme categories. This difficulty arises from reduced discrimination sensitivity, caused by the acquisition and further maturation of the native language sound system during childhood [13]. This causes the brain to essentially ignore contrasts that are not phonologically relevant in the native language.

Several training methods have been used to try and overcome this reduced sensitivity in order to facilitate second language acquisition. Training studies typically concentrate on attempting to improve discrimination. It has been shown, for example, that the so called "High Variability Phonetic Training" method which uses repetitions of stimuli uttered by native speakers of the target language is effective in training Japanese learners to discriminate between the difficult English /r/-/l/-contrast [5, 14]. Furthermore, despite the fact that the method focuses only on perception training, it was also shown to improve the subjects' production of the novel contrast [6]. Perception training can also be enhanced through orienting the subjects' attention. Pederson and Guion-Anderson showed in their two studies [12, 15] that phonetic discrimination training of Hindi sounds on English learners could be made more effective by instructing the subjects to focus on the specific sound being trained, as opposed to the meaning of the training stimuli.

Language context of the training situation can also affect the way in which the subjects perceive the stimuli. In an EEG experiment, Peltola and Aaltonen [16] showed that elicitation of an MMN (mismatch negativity) response in advanced Finnish students of English to English and Finnish vowel contrasts was affected by the language the subjects thought they were hearing. Whether they were instructed in English or Finnish also had an effect. Subjects who were instructed in Finnish and did not know whether they were listening to Finnish or English, displayed lower overall MMN amplitudes to the English contrast than subjects who were instructed in English and were told they were only hearing English vowels.

Earlier studies on production training for adults have mainly focused either on computer-assisted teaching tools for second language learners, or on rehabilitative training for those suffering from speech or hearing disabilities [1, 8]. An early study by Catford and Pisoni [7] showed, however, that production training using articulatory instructions was more successful in helping young English adults learn “exotic” consonants and vowels than auditory training. This training used native phonemes as reference points for the new articulation patterns; the subjects were told to first form a certain native sound and then move their articulators towards the new place of articulation. The learning effect was visible in both production and discrimination tests, although it was more pronounced in the former.

A neural network model of speech production, DIVA (Directions Into Velocities of Articulators) [11], suggests that speech production is controlled

through an internal model that is based on feedforward commands and the various neural feedback mechanisms that exist in the central nervous system. The internal model is formed in early childhood in two stages: first, as the infant hears speech, the phoneme categories of the native language form and become defined in the brain during the first months of life. During the babbling phase, the child compares sounds he himself produces with the existing phoneme categories, and the motor patterns and articulatory configurations are stored in the internal model for each sound as production targets which are used as feedforward commands in speech production [10]. The patterns are recorded by the proprioceptive and tactile feedback systems, and during adulthood the model is maintained by the auditory feedback system, through hearing one's own speech and comparing it to the existing model. This suggests that auditory feedback can also be used to rearrange existing neural mappings in order to change established production patterns; the effect is best seen in patients with hearing loss who receive cochlear implants [17].

In this project, we examined a training paradigm aimed at improving the production of a non-native vowel contrast with young adult learners. The purpose of the paradigm was to provide the subjects with both perceptual and production training in order to maximize the effectiveness of the relatively short training period. This was achieved through the use of natural-sounding, semi-synthetic stimuli as the production targets, and articulatory instructions that not only became more detailed as the study progressed, but also direct the attention of the subject to the relevant acoustic feature and give the training a language context. It was hypothesized that the training blocks themselves would enable the subjects to self-correct their productions through auditory feedback, in accordance with the DIVA model, and that the explicit articulatory instructions and the language context would prime the subjects to be as receptive as possible to the features of the non-native target.

Similar training with the same stimuli, though without instructions, resulted in the successful production of the new vowel after three training sessions with 7–10–year–old children [18]; it is hypothesized here that the instructions may help the less plastic adults reach a similar learning speed. Articulation changes elicited by the training are most likely to occur in the F2, as tongue backedness, the main articulatory difference between the stimuli, mainly affects F2. F1 is much less likely to change, though change is possible if the subject's jaw moves as he moves his tongue back.

2. METHODS

2.1 Subjects and stimuli

The subjects consisted of nine (six females) young Finnish adults, aged 21–30 (mean 24.3). All were normally hearing native speakers of Finnish. None of them had studied languages at university level, or lived in any of the other Nordic countries. Subjects self-evaluated their language skills with a questionnaire, and they matched the expected level for Finnish adults: all of them reported knowledge of English and Swedish, and most of them knew a third language. Usage levels were low for languages other than English. The study was approved by the Ethics Committee of the University of Turku.

The training stimuli were the close front rounded vowel /y/ as the non-target, and the close central rounded vowel /ʉ/ as the target. For Finnish speakers, this represents a difficult learning situation, as Finnish only contrasts the aforementioned /y/ with the close back rounded vowel /u/, as per SLM [9] and PAM [3]. The target vowel /ʉ/ is therefore likely to be interpreted as a poor exemplar of either /u/ or /y/. The distinction is mainly based on changes in the second formant. The vowels were embedded in semisynthetic Finnish pseudowords /ty:ti/ and /tʉ:ti/. The semisynthetic method is based on an extracted glottal pulse of a real speaker, and it produces natural sounding stimuli while simultaneously allowing for accurate control over their phonetic features. Further information on the method can be found in [2]; further information regarding the creation of the current stimuli in [18]. Both of the stimuli were 624 ms long and had a fundamental frequency of 126 Hz. Measured from the midpoint of the vowel, 190 ms from the stimulus onset, F1, F2 and F3 in /tʉ:ti/ were 338, 1258 and 2177 Hz, respectively; the same values in /ty:ti/ were 269, 1866 and 2518 Hz.

2.2 Procedure and analysis

The training protocol was divided over two days, with both days containing two training sessions and two production recordings. The first day began with a baseline recording, followed by the first training block, the second recording and the second training block. On the second day, this order was reversed so that the day began with a training block and ended with the final recording. In both the training and the recording sessions subjects were asked to repeat the words they heard to the best of their ability. In the training block subjects heard each word 30 times, and in the recording block 10 times, so that the subjects heard and repeated the stimuli for a total of 320 times.

The stimuli were presented with an interstimulus interval of three seconds.

Before each training block, the subjects were given instructions on how to improve their production of the non-native vowel. The instructions were designed so that on the first day the subjects were made explicitly aware of the existence of the non-native vowel, and on the second day they were given articulatory instructions to further improve their productions. The instructions also provided a Finnish–Swedish language context on the first day. Subjects received no feedback on their productions at any point in the study. The instructions given were as follows (translated from Finnish):

1. You will hear two words alternately. The other one has a Finnish vowel, and the other has a Swedish one.
2. The Swedish vowel can be described as a mixture of the Finnish "y" and "u".
3. Try keeping your mouth otherwise in the same position as you do for /y/, but move your tongue slightly back in your mouth.
4. There are minor differences in the roundedness of the lips. The lips are more tightly rounded in /y/ than they are in /ʉ/.

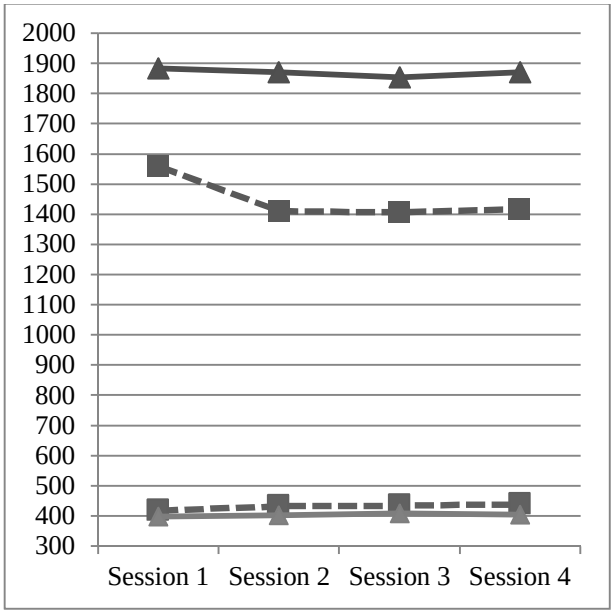
All tests were performed at a sound attenuated laboratory of the Department of Phonetics at the University of Turku. The stimulus trains were presented and the productions recorded using the SLH-07 headset provided in the Lab 100 language lab system by Sanako Corporation. The researcher and subject were separated by a divider, and communicated through an intercom system. The subjects were themselves able to set the volume of the stimuli to a comfortable level.

The results were first analyzed from the spectrograms of the individual productions using the Linear Predictive Coding (Burg) algorithm of the Praat software, version 5.3.37 [4]. Fundamental frequency and the formant values F1 and F2 were analyzed at the vowel midpoint of each word. All data was normally distributed. Mean values for these measurements were then calculated, resulting in 32 values for each subject. These means were statistically analyzed with IBM SPSS Statistics (version 22) using Repeated Measures Analysis of Variance (ANOVA).

3. RESULTS

Statistical analysis of the formant values began with a Word (2) X Session (4) X Formant (2) ANOVA in order to find out differences between the productions of the words during the experiment. This resulted in a significant main effect of Word ($F(1,8) = 135.110$; $p < 0.001$) and a significant Word x Formant interaction ($F(1,8) = 175.714$; $p < 0.001$), signifying that the words were produced consistently differently throughout the experiment and that the formants were different between the two words. No effects for session were significant. As the data seemed to indicate a change in the formant values between Session 1 and Session 2, these were next analyzed with a Word (2) X Session (2) X Formant (2) ANOVA. This resulted in a significant main effect of Word ($F(1,8) = 36.012$; $p < 0.001$) and, more importantly, Session ($F(1,8) = 6.343$; $p = 0.036$), indicating that the two words were produced differently and that change had occurred during training. This change can clearly be seen in Fig. 1. A significant Session X Formant interaction ($F(1,8) = 7.900$; $p = 0.023$) was also found, signaling more change in one of the two formants than the other. Session effects did not reach significance between sessions 2 and 3 or 3 and 4.

Figure 1: Average formant values of the subjects' productions in Hz (dashed lines = /ʉ/, solid lines = /y/. Upper lines = F2, lower lines = F1.



In order to find out whether F2 was behind the change, as suggested by the data, a Word (2) X Session (2) ANOVA was performed with just the F2 values. This resulted in a significant main effect of Word ($F(1,8) = 42.283$; $p < 0.001$) and Session ($F(1,8) = 42.283$; $p = 0.026$), indicating that the previously observed change during training was indeed focused on the F2. No session effects reached significance between sessions 2 and 3 or 3 and 4.

4. DISCUSSION AND CONCLUSION

In the present study, it was hypothesized that the paradigm used in the training would allow the subjects to learn to reliably produce the novel vowel contrast during the relatively short training period. Overall, the subjects learned to produce the target vowel, and did so very rapidly, having only gone through one of the four training blocks and one line of instructions. The performed analyses indicated that the formant structures of the two vowels were not at any point similar, suggesting that the subjects did not initially confuse the target vowel with /y/. The significant decrease in the F2 value of the target vowel /ʉ/, however, implied that the subjects had initially produced it with more /y/-like, *i.e.* higher, F2 values; after being given instructions the subjects were able to self-correct their productions and finally settle on the values seen in Session 4.

The fact that the vast majority of change in the formant values took place already between sessions 1 and 2, with no statistically significant changes occurring after this, was somewhat unexpected. It may indicate that the adult subjects learned to produce a non-native vowel even faster than 7–11–year-old children [18] that used the same stimuli and a non-instructed listen-and-repeat training paradigm. At this point in the study the subjects had undergone only one block of training, and received the first part of the four-part training instructions, in which they were told that there were two words than contained two different vowels, a Finnish and a Swedish one. This result implies that simply making the subjects explicitly aware of the two vowels and the language context made a major difference to the subjects' production abilities. This is somewhat in line with the studies on attention orienting [12, 15] and the effect of language context [16], but does not fully explain the fast adaptability this early on.

It is likely that as hypothesized, the explanation lies in the combination of explicit instructions and the actual training. In the baseline recording the subjects were simply asked to repeat what they heard and were not given any instructions or information about the stimuli. In addition to likely not being able to hear the difficult contrast very well, their attention would not have been focused on it specifically, as they did not know what was expected of them. After the first explicit instructions, however, subjects were able to fully focus their attention on the relevant part of the stimuli, which enabled them to engage their auditory feedback system for self-evaluating and correcting their productions according to the model provided by the stimulus, as suggested by the DIVA model [11]. The actual production of the vowel may not have been too difficult after this, as the articulatory changes required were not very complex and the subjects could spend the entire first training block focused solely on the vowel, unlike in [18], for example, where the child subjects had to decipher the relevant features themselves.

In conclusion, we showed that instructed production training can be used highly effectively in training non-native vowels in young adults. The result shows that with attention direction and instruction, reduced sensitivity can be overcome, and adults can acquire new production patterns at least rapidly as children with a very simple listen-and-repeat training paradigm.

5. ACKNOWLEDGEMENTS

This research was partially supported by a grant from the Utuling doctoral program. The research equipment was provided by Sanako Corp.

6. REFERENCES

[1] Akahane-Yamada, R., McDermott, E., Adachi, T., Kawahara, H., Pruitt, J.S. 1998. Computer-based second language production training by using spectrographic representation and HMM-based speech recognition scores. *Proc. of the 5th International Conference on Spoken Language Processing*, Sydney, 1747-1750.

[2] Alku, P., Tiitinen, H., Näätänen, R. 1999. A method for generating natural sounding speech stimuli for cognitive brain research. *Clinical Neurophysiology* 110, 1329–1333.

[3] Best, C.T. 1994. The emergence of native-language phonological influences in infants: a perceptual assimilation model. In: Nusbaum, H.C. (ed.), *The Development of Speech Perception: The Transition from Speech Sounds to Spoken Words*. MIT: Cambridge, MA, 167–224.

[4] Boersma, P., Weenink, D. (2011). Praat: doing phonetics by computer, version 5.2.1, retrieved 1st August 2014 from http://www.praat.org/

[5] Bradlow, A., Akahane-Yamada, R., Pisoni, D.B., Tohkura, Y. 1999. Training Japanese listeners to identify English /r/ and /l/: Long-term retention in perception and production. *Perception & Psychophysics* 61(5), 977–985.

[6] Bradlow, A.R., Pisoni, D.B., Akahane-Yamada, R., Tohkura, Y. 1997. Training Japanese listeners to identify English /r/ and /l/: IV. Some effects of perceptual learning on speech production. *J. Acoust. Soc. Am.* 101, 2299–2310.

[7] Catford, J.C., Pisoni, D.B. 1970. Auditory vs. articulatory training in exotic sounds. *The Modern Language Journal* 54(7), 477–481.

[8] Dalby, J., Kewley-Port, D. 1999. Explicit pronunciation training using automatic speech

recognition technology. *CALICO Journal* 16(3), 425–445.

[9] Flege, J.E. 1987. The production of “new” and “similar” phones in a foreign language: evidence for the effect of equivalence classification. *Journal of Phonetics* 15, 47–65.

[10] Guenther, F. 2003. Neural control of speech movements. In Schiller, N.O. and Meyer, A.S. (eds.), 2003. *Phonetics and phonology in language comprehension and production.* New York: Mouton de Gruyter, 209–239.

[11] Guenther, F., Perkell, J. 2004. A neural model of speech production and its application to studies of the role of auditory feedback in speech. In Maassen, B., Kent, R., Peters, H., van Lieshout, P., Hulstijn W. (eds.), *Speech motor control in normal and disordered speech.* New York: Oxford University Press, 29–49.

[12] Guion, S., Pederson, E. 2007. Investigating the Role of Attention in Phonetic Learning. In Bohn, O-S, Munro, M. (eds.), *Language Experience in Second Language Speech Learning: In honor of James Emil Flege.* Amsterdam: John Benjamin, 55–77.

[13] Iverson, P., Kuhl, P.K., Akahane-Yamada, R., Diesch, E., Tohkura, Y., Kettermann, A., Siebert, C. 2003. A perceptual interference account of acquisition difficulties for non-native phonemes. *Cognition* 87, 47–57.

[14] Lively, S.E., Pisoni, D.B., Yamada, R.A., Tohkura, Y., Yamada, T. 1994. Training Japanese listeners to identify English /r/ and /l/. III. Long-term retention of new phonetic categories. *J. Acoust. Soc. Am.* 96, 2076–2087.

[15] Pederson, E., Guion-Anderson, S. 2010. Orienting attention during phonetic training facilitates learning. *J. Acoust. Soc. Am.* 127, 54–59.

[16] Peltola, Maija S., Aaltonen, Olli. 2005. Long-Term Memory Trace Activation for Vowels Depends on the Mother Tongue and the Linguistic Context. *Journal of Psychophysiology* 19, 159–164.

[17] Perkell, J. 2012. Movement goals and feedback and feedforward control mechanisms in speech production. *Journal of Neurolinguistics* 25, 382–407.

[18] Taimi, L., Jähi, K., Alku, P., Peltola, M. 2014. Children Learning a Non-native Vowel – The Effect of a Two-day Production Training. *Journal of Language Teaching and Research* 5: 1229–1235.