ULTRASOUND TONGUE IMAGING FOR VOWEL REMEDIATION IN CZECH ENGLISH

Tanja Kocjančič Antolík, Jan Volín

Institute of Phonetics, Charles University, Prague, Czech Republic
tkocjancic@gmail.com

ABSTRACT

Increasing comprehensibility is a common desire of many speakers of a foreign language. However, most learners have troubles improving the articulation of already acquired foreign sounds despite continuing language learning. Czech speakers of English typically struggle with the contrast between English vowels /e/ and /æ/. The present study employed ultrasound tongue imaging as a visual feedback for vowel remediation and two methods of articulatory practice. Eight adult speakers of Czech English received three 40-minute ultrasound training sessions practicing articulation of the two vowels in isolation, syllables and minimal pairs. Half were practicing only articulation of the two vowels while the other half was first practicing lingual articulatory setting for English, followed by vowel practice. Perceptual evaluation comparing pre- to post-training production shows an improvement in minimal pair contrast for most speakers but no clear difference between the training methods.

Keywords: ultrasound tongue imaging, vowel remediation, foreign language learning

1. INTRODUCTION

Czech speakers of English typically merge English vowels /e/ and /æ/ in their production into one front mid vowel since this opposition is not part of the Czech vowel system [17]. Consequently, they do not produce the contrast between these two English vowels reducing speech comprehensibility due to the existence of a number of minimal pairs.

A method yielding promising results for speech sound remediation is ultrasound tongue imaging (UTI) used as a real-time visual feedback. It allows observation of the tongue shape, position and movements during speech, while being safe, non-invasive and relatively easy to use. The method has been successfully used in a number of clinical studies where participants (children and adults) with speech sound disorders of different origins acquired new lingual articulations, transferred them to non-trained items and retained them long term (e.g. [3,5,13]).

More recently, the application of this method has been explored in foreign language learning, both for consonants and vowels (for review see [4]). Participating speakers showed improvement after even only one training session [6] and retention at two months post-training [8]. The latter study also directly compared performance of learners practicing vowel production with UTI and those following a more conventional pronunciation training, showing greater change in the UTI training group.

The present study attempts to partially replicate the above one by evaluating vowels in words and not only in isolation, while also comparing two different training methods: (1) direct training of individual vowel's lingual articulation (as done in the UTI training studies reported so far) and (2) acquisition of lingual articulatory setting (AS) for English first, followed by individual vowel training.

Several studies have reported language specific AS [7, 9, 18] and some authors claim that adopting an L2 AS is the necessary prerequisite for adequate production of L2 speech sounds [11].

According to [12] AS for English comprises correct position of the tongue, lips, jaws, pharynx and larynx. Because the tongue is the main articulator involved in the production of English /e/-/æ/ contrast, only lingual AS is the focus of the study presented here. For English this comprises active lateral bracing of the tongue against the upper (pre)molars, the tongue tip being positioned very close to the alveolar ridge but without direct contact, and the centre of the tongue, from behind the tip backwards, lying concave to the roof, creating a "butterfly" shape in coronal view [12, 11]. In contrast, for the Czech language the tongue tip is in contact with the lower incisors and/or gums which has been reported as the main reason for difficulties in the articulation of English vowels for Czech speakers [15].

In terms of the two vowels articulation, it was expected that the main articulatory difference will be noted in the very front part of the tongue, tip and blade, which will be positioned higher for /e/ than for /æ/.

The goal of the present study was thus two-fold: (1) evaluate if speakers of Czech English improve the /e/-/æ/ contrast after pronunciation training employing UTI as a real-time visual feedback, and (2) explore the effect of practicing the articulation of vowels only and of practicing the articulation of same vowels with lingual AS.

2. METHOD

2.1. Speakers

Eight participants took part in the experiment. They were all first year students of the Czech Language & Literature Programme, aged between 19 and 21 years. All took a state exam in English at the end of high school suggesting at least a B1 level of proficiency. None of them ever spent any extensive time in an English speaking country and only SST1 lived in a foreign country (18 months in the Netherlands). All participants knew ahead that the training will be focused on English vowels /e/ and /æ/, and they were very motivated to improve their pronunciation.

Participants were randomly assigned into two groups differing in the method used in the speech training sessions: four practiced only the two target segments (SST group), and four first practiced lingual AS for English and later practiced the segments with this setting (AST group).

2.2. Speech material

Because the final goal of the speech training was improvement in the production of English /e/ - /æ/ contrast, the test and training material was based on minimal pairs.

Test data collected pre- and post-training consisted of 12 minimal pairs: end-and, head-had, pet-pat, bet-bat, pen-pan, met-mat, men-man, dead-dad, ten-tan, Ken-can, said-sad, set-sat. The set was chosen because it allowed using real words while minimizing coarticulatory effect of final consonant (post-dental/alveolar place of articulation) on the vowel.

2.3. Speech training

Each participant received three 40-minute speech training session (at most a week apart) using UTI as a real-time visual feedback.

Following a brief familiarization with UTI and English vowel system, the participant produced the two vowels in isolation while observing ultrasound images and explained any similarities or differences in tongue shape and position between them. Next, the trainer (first author) produced the two vowels, with UTI, and the participant described these productions. Once the participant understood the target articulations, the speech training started and the trainer did not produce any additional modelling of the target vowels.

Importantly, at this point the participants in the AST group were explained lingual AS for English. They practiced positioning their tongue in the required configuration while observing a coronal view of their tongue. Once they felt confident holding tongue in the target setting, they used it during the entire vowel practice.

During the first training session all participants practiced the two vowels in isolation, followed by non-word syllables with CV, VC and C(C)(C)VC structure. 39 minimal pairs (78 words), including nine out of 12 pairs used in the test material, were added in the second training session. The same words were used in sentences in the third training session.

Initially, participants were asked to monitor production via tongue images. However, after several correct productions, they were asked to rely solely on the acoustic output and on the proprioception of the position of their tongue in the mouth.

Participants produced at least ten repetitions of each used item per session, the order of items varied between sessions and speakers. The trainer provided immediate feedback at the beginning of each session, however the control of production was progressively transferred to the participant.

Furthermore, the training material included filler items without the focus on the target vowels.

2.4. Data collection

Articulatory and acoustic data were recorded at the beginning of the first session and at the end of the third session. The data were recorded using Micro ultrasound system and Articulate Assistant Advanced software [2]. The system allows synchronisation of the ultrasound and audio signals. Probe stabilization headset [1] was used during the recording. Participants made two repetitions of the word list and the ultrasound data were captured in midsagittal view. Additionally, a coronal view of lingual AS was recorded while participants were speaking English and Czech.

2.5. Data analysis

Only the first production of each test item was used in the analysis.

2.5.1. Articulatory analysis

In order to extract a sagittal image of tongue contour representing a vowel, the mid-point of vowel duration was selected and tongue surface was traced in the associated ultrasound frame. Mean tongue contour was calculated from 12 tongue contours of the test word list.

A coronal image of tongue shape presenting lingual AS was extracted from the intervals between consecutive English words.

2.5.2. Perceptual analysis

Three experienced phoneticians, non-native but highly proficient in English rated the productions in a perceptual discrimination test. Target words were presented in minimal pairs (12 pairs x 2 conditions (pre, post) x 8 speakers) and the listeners had to decide whether the words in a pair are similar or different. Because some of the participants marked the contrast with vowel duration, the listeners were instructed to base their decision on the quality of the vowel and not on its duration. Fleiss kappa was calculated to evaluate interrater reliability.

3. RESULTS

3.1. Articulatory data

Figure 1 shows mean tongue contour pre- and post-training for one speaker of each method group. Firstly, just as the two speakers presented in Figure 1, all the other participants (except AST2) made almost no difference in the lingual shape and position for the two English vowels pre-training. Post-training data revealed more differences across participants. Three speakers, AST1, AST4 and SST3 (the data is less conclusive for this speaker) showed the expected lower front of the tongue for /æ/ than for /e/, the reverse was observed for SST4 and AST2, while speakers SST1, SST2 and AST3 showed no difference.

Figure 1: Mean midsagittal tongue contours for speakers SST1 (top) and AST1 (bottom) pre- (left) and post-training (right). /e/ = solid line, /æ/ = dashed line. Tongue front is on the right side.

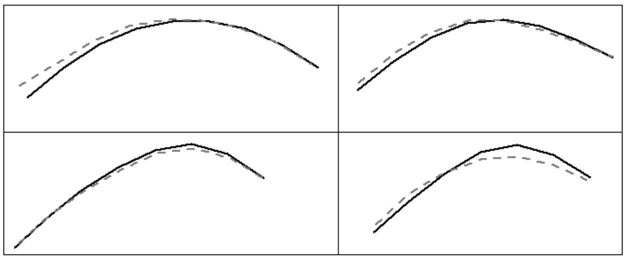
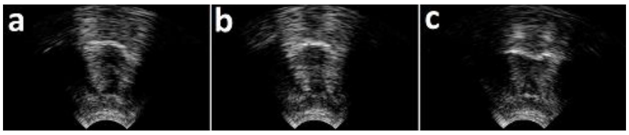


Figure 2: Lingual articulatory setting of AST4 when speaking Czech (2a), English pre-training (2b) and English post-training (2c)



Coronal ultrasound images in Figure 2 represent AST4’s lingual AS when speaking Czech (2a), English pre-training (2b) and English post-training (2c). The first image corresponds to the expected lingual shape for Czech, where the sides of the tongue are turned downwards and the imaged tongue has an upside-down U shape. The same shape is present at pre-training, while post-training the speaker uses English lingual AS, with the tongue being actively braced and approximating a “butterfly” shape. All AST speakers were using English lingual AS in post-training data collection.

3.1. Perceptual data

The Fleiss kappa test (kappa = 0.508, p-value = 0) revealed only a moderate interrater reliability in the assessment of minimal pairs sounding same or different. The same can be observed (Table 1) in the number of minimal pairs (out of total 12) that were rated by all three listeners as same, different or a mix of the two ratings. The results imply a post-training decrease in the number of pairs being perceived as the same by all the listeners for all participants, except SST2 and AST2. The number of pairs being rated as different by all three listeners (at least minimally) increased for four participants, stayed the same for two and decreased for two. Consequently, the number of pairs that received mixed ratings increased for six participants, stayed the same for one and decreased for one (AST4).

Table 1: The number of minimal pairs (out of total 12) that were rated by all three listeners as same, as different or as a mix of the two ratings.

rated as	*same*		*different*		*mix*	
speaker	*pre*	*post*	*pre*	*post*	*pre*	*post*
SST1	5	3	4	1	3	8
SST2	7	7	0	0	5	5
SST3	9	2	0	3	3	7
SST4	10	6	1	2	1	4
AST1	4	3	1	1	7	8
AST2	5	5	5	4	2	3
AST3	10	4	0	1	2	7
AST4	3	1	3	10	6	1

4. DISCUSSION

Because of the mismatch between the Czech and English vowel set, it was expected that speakers of Czech English will show little difference in the production of English vowels /e/ and /ae. This was conformed both by the articulatory analysis of midsagittal tongue contours and by perceptual evaluation. In pre-training, none of the speakers showed the expected difference in the position of the front part of the tongue, with almost all speakers having no difference in tongue shape and position. Similarly, perceptual discrimination test revealed that for each speaker more minimal pairs were rated as same by all three listeners than as different (except

SST2 and AST2 with equal distribution), and more were rated as same than having a mixed response (except AST1 and AST4). The latter suggests that listeners agreed in their evaluation of most pairs.

One of the aims of the study was to evaluate whether speakers of Czech English improve the /e/-/æ/ contrast after UTI pronunciation training. It is important to note here that all participants understood the difference between their pre-training pronunciations and the target ones once they saw their own and the trainer's productions. Moreover, they were all able to produce the two English vowels correctly in the first few attempts after the demonstration. During the entire training, all participants, except SST2, reliably produced articulatory and perceptually adequate targets. The trained vowel contrast was noted in isolated vowel productions, syllables, words and sentences, and it was present across repetitions. SST2 was the only speaker who had difficulties producing the target contrast.

Post-training data, however, did not fully capture this change. Articulatory data confirmed the expected lower front of the tongue for /æ/ than for /e/ only for three participants. There was no notable difference for the remaining five, although successful productions were observed throughout the training sessions for most of the participants. A very likely reason for this lies in one of the major limitations of UTI – raised tongue tip cannot be imaged because of the air pocket below it. The main articulatory difference between the two vowels was expected to be in the vertical position of the front of the tongue and it is very likely that that part was not imaged adequately during the recording.

Post-training perceptual evaluation is more supportive of the noted changes in vowel contrast productions during the training. Most speakers (except SST2 and AST2) had less minimal pairs evaluated as same by all three listeners, and most had a greater number of pairs receiving a mixed response by the listeners. The latter can be viewed as a direct result of acquiring new articulations. The participants had a relatively short time for practicing new lingual movements and it was not expected that the movement would become automatic by the end of the training. However, the increase in mixed responses by different listeners suggests that the speakers were trying to use new articulation but did not yet execute them correctly. Increased variability in the production of target vowels at the end of UTI pronunciation training has been reported previously [8]. Finally, the increase in the number of minimal pairs being perceived as different by all three listeners was only minimal for most speakers. The greatest increase in this value was achieved for speakers SST3 and AST4 who were also the most consistent in their correct articulations during the training.

Two more possible sources affecting the test data were noted. First, some of the participants marked the contrast between the two English vowels by vowel length which was perceptively longer for /æ/ than for /e/. Vowel length is a primary distinctive feature in Czech but only secondary in English and the usage of vowel length differs between Czech and English [10, 14, 16]. Second, it is possible that participants were not sure which vowel to produce based on the written prompts. This could be more problematic in the post-training data collection, because they were aware of the different vowels but had a very limited time to make a choice. The recording started just after the prompt appeared on the computer screen and they were asked to utter the word as soon as the recording started.

The second objective of the study was to evaluate any differences between the speakers using two different training methods. All AST participants spoke with English AS at post-training as visible in the 'butterfly'-like shape of the tongue [11, 12]. However, the data presented here does not provide clear answer whether one method is better than the other. Interestingly, all speakers in the AST group remarked that they sounded more English than when speaking without the English lingual AS. More research is needed to investigate the effect of L2 AS on the L2 production.

Finally, all participants expressed positive feelings about using UTI in the pronunciation training. They reported that it helped them to understand the difference in the articulation of the two vowels and to produce them correctly.

5. CONCLUSIONS

The study presented here aimed at investigating whether UTI helps speakers of Czech English to realize the /e/-/æ/ vowel contrast in minimal pairs and whether the gains are different for speakers practicing the new articulations with Czech or English articulatory setting. The results suggest some improvement in producing the contrast for most speakers post-training. However, the change is not uniform across the speakers. Furthermore, no clear distinction between the two methods was noted. Possible reason for a lack of observable articulatory differences is a methodological limitation of UTI.

Acknowledgements
This work was supported by the OP VVV project no. CZ.02.2.69/0.0/0.0/17_050/0008466.

6. REFERENCES

[1]Articulate Instruments Ltd. 2008. Ultrasound Stabilisation Headset Users Manual: Revision 1.4. Edinburgh, UK: Articulate Instruments Ltd.

[2] Articulate Instruments Ltd. 2012. Articulate Assistant Advanced User Guide: Version 2.14. Edinburgh, UK: Articulate Instruments Ltd.

[3] Bacsfalvi P. 2010. Attaining the lingual components of /r/ with ultrasound for three adolescents with cochlear implants. J. Speech Lang. Pathol. Audiol. 34, 206–217.

[4] Bliss, H., Abel, J., Gick, B. 2018. Computer-assisted visual articulation feedback in L2 pronunciation instruction: a review. Journal of Second Language Pronunciation 4, 129-153.

[5] Cleland, J., Scobbie, J.M., Wrench, A.A. 2015. Using ultrasound visual biofeedback to treat persistent primary speech sound disorders. Clinical Linguistics and Phonetics, 29, 575-597.

[6] Gick, B., Bernhardt, B., Bacsfalvi, P., Wilson, I. 2008. Ultrasound imaging applications in second language acquisition. In: Hansen Edwards, J.G., Zampini, M.L. (eds.), Phonology and second language acquisition. Amsterdam: John Benjamins, (pp. 309–322).

[7] Gick, B., Wilson, I., Koch, K., Cook, C. 2004. Language-specific articulatory settings: Evidence from inter-utterance rest position. Phonetica, 61, 220–233.

[8] Kocjančič Antolík, T., Pillot-Loiseau, C., Kamiyama, T. 2019. The effectiveness of real-time ultrasound visual feedback on tongue movements in L2 pronunciation training. Journal of second language pronunciation, 5, 72-79.

[9] Lowie, W., Bultena, S. 2007. Articulatory settings and the dynamics of second language production. Proceedings of the Phonetics Teaching and Learning Conference (PTLC), London, 1–4.

[10] Menhard, Z. 1982. A Workbook in English Phonetics. Příbram: Státní pedagogické nakladatelství Praha.

[11] Messum P., Young, R. 2017. Bringing the English articulatory setting into the classroom: (1) The tongue. Speak out!: newsletter of the IATEFL Pronunciation Special Interest Group, 57, 29-39.

[12] Mompeán González, J.A. 2003. Pedagogical tools for teaching articulatory setting. Proceedings of the 15th International Congress of Phonetic Sciences. Barcelona, 1603–06.

[13] Preston, J. L., Leaman, M. 2014. Ultrasound visual feedback for acquired apraxia of speech: a case report. Aphasiology 28, 278–295

[14] Roach, P. 2009. English Phonetics and Phonology. 4th edition. Cambridge: CUP.

[15] Skaličková, A. 1979. Srovnávací fonetika češtiny a angličtiny. Praha: Státní pedagogické nakladatelství.

[16] Skarnitzl, R., Šturm P., Volín J. 2016. Zvuková báze řečové komunikace: Fonetický a fonologický popis řeči. Praha: Karolinum.

[17] Šímackova, Š., Podlipský, V.J., Chládková, K. 2012. Czech spoken in Bohemia and Moravia. Journal of the International Phonetic Association 42, 225-232.

[18] Wilson, I., Gick, B. (2014). Bilinguals use language-specific articulatory settings. Journal of Speech Language and Hearing Research, 57, 361-373.