VOWEL PERCEPTION AND TRANSCRIPTION TRAINER FOR LEARNERS OF ENGLISH AS A FOREIGN LANGUAGE

ŠÁRKA ŠIMÁČKOVÁ
Palacký University Olomouc
sarka.simackova@upol.cz

VÁCLAV JONÁŠ PODLIPSKÝ
Palacký University Olomouc
vaclav.j.podlipsky@upol.cz

Abstract
We use the freely available program Praat to create a vowel-training application for learners of English familiar with IPA transcription. The application is easy to operate, allowing users to change the training difficulty, providing the listeners with immediate feedback, and adapting to their performance during a training session. To evaluate the effectiveness of the Trainer, performance of 59 Czech learners during a single training session and across multiple sessions was tracked. Results showed improvement both between sessions and within sessions. In the final training session, vowel identification accuracy showed considerable resistance to gradual addition of increasing levels of noise. Testing the trainer with additional 52 learners showed significantly higher error-rates for low-frequency words and supported the importance of top-down lexical effect in vowel identification.

Keywords: English as a foreign language, IPA, training application, vowel identification

1. Introduction

Our intention in this paper is to contribute towards one of Tracy Derwing's not-so-utopian goals for second language pronunciation teaching (Derwing 2010), namely to the goal of developing easy-to-use and useful software. We present a computer application "English Vowel Trainer" created by the second author with the use of the freely available program Praat (Boersma and Weenink 2019). The application, originally intended for EFL learners in tertiary education who study English as an academic subject, is designed to help upper-intermediate-to-advanced EFL learners to enhance their speech perception skills in English and improve accuracy of their phonological representations of English vowel sounds. Central to this effort is the assumption that accurate perception of L2 sounds underlies their accurate representations in memory and precedes their accurate production (Escudero 2007). Though far from being uniformly accepted in the literature (for other views see e.g. Sheldon and Strange 1982), the perception-production link has found strong empirical support (e.g. Baker and Trofimovich 2006, Flege, MacKay and Meador 1999, Jia, et al. 2006), including results of perceptual training studies (Bradlow, Akahane-Yamada, Pisoni and Tohkura 1999, Motohashi-Siago and Hardison 2009) and classroom instruction studies (Kissling 2014).

In our effort to improve the learners' perceptual abilities, we further assume that speech sound learning can stretch past puberty, into and across adulthood. Adult learners' sound categories are not necessarily fossilized and can change in response to new input even at later phases of L2 learning. Studies of immersion acquisition show that the amount of experience with an L2 crucially correlates with accuracy of segmental production (Flege, Frieda, Walley & Randazza 1998, Flege and Liu 2001). However, in foreign-language learning settings, where exposure to interactional native speech is scarce and L1 use prevails over the use of L2, a reattunement of L2 sound categories may be difficult. Limitations on input quantity and quality naturally imposed by such learning environment need to be compensated for. We see the means of such compensation in effective, i.e. sound-focused, age- and proficiency-adjusted, instruction that includes intensive (if not extensive) exposure to structured, and possibly also modified, speech input in audio-materials. A recent study indicates that explicit (computer-delivered) phonetics instruction built into a foreign language classroom can indeed lead to improved perception of L2 sounds for a range of language proficiencies but may especially benefit more advanced learners (Kissling 2015). The effect of form-focused classroom instruction and perception training on learners' L2 phonological representations is further enhanced by providing explicit corrective feedback (Lee and Lyster 2015).

A number of laboratory studies confirm that focused phonetic training facilitates L2 sound learning even in post-pubertal learners (most recently, e.g. Shinohara and Iverson 2018, Grenon, Kubota and Sheppard 2019). The very successful high variability phonetic training (HVPT) paradigm has been used to demonstrate that in a relatively short time (e.g. the total of 15-22.5 hours over 3-4 weeks in Bradlow et al., 1999; 6 hours over 1-2 weeks in Iverson, Pinet & Evans 2012; 13.5 hours over a month in Nishi and Kewley-Port 2007), adult learners whose attention is directed to specific phonetic dimensions of L2 speech sounds placed in multiple phonetic environments and recorded by multiple voices, substantially improve their ability to make L2 phonetic contrasts, they generalize the training to new instances, and show significant gains on delayed post-tests. Some HVPT studies document improvements in perception leading to improvements in speech production (e.g. Bradlow et al. 1999, Shinohara and Iverson 2018) while other studies indicate that effectiveness of HVPT is modulated by individual differences in perceptual abilities (e.g. Perrachione, Lee, Ha and Wong 2011).

Feedback, which plays an important role in the HVPT paradigm, is typically not available to listeners participating in distributional training studies that have recently started to appear (Escudero and Williams 2014). These studies are designed to test whether, like children, even adult learners can learn sound categories from passive experiencing of frequency distributions of speech sounds that vary along a specific phonetic dimension: can they form sound category representations solely based on exposure to these distributions? In the course of a distributional training session, L2 learners are predicted to learn new contrasts implicitly, just from hearing an abundance of exemplars falling near the opposite ends of an acoustic continuum and a limited number of tokens from the middle of the continuum (bimodal distribution). The bimodally-trained learners' ability to discriminate between the contrasting sounds before and after the training is compared to that of learners exposed to a unimodal distribution, in which the frequency of tokens is distributed normally along the acoustic continuum with a single peak in the middle. Some studies indicate that an exposure of a few minutes might induce a lasting change in L2 sound representations (Escudero and Williams 2014). If corroborated, such type of training could have a great pedagogical potential. However, currently, the ability of adult learners to benefit from distributional learning is far from established (Wanrooij, Boersma and van Zuijen 2014, Wanrooij, Boersma and Benders 2015, Wanrooij, De Vos and Boersma 2015). And even the HPVT paradigm, whose positive effect on L2 learners' phonetic abilities has been documented by ample research, has not yet been widely tested in pedagogical contexts (for a discussion see Barriuso and Hayes-Harb 2018).

The above mentioned laboratory experiments use rigorous research methodology, often involving manipulated and/or artificially enhanced speech input with the goal to address carefully formulated theoretical issues. In this paper, conversely, we follow a purely pedagogical aim of describing a training tool originally created for the purpose of expanding EFL learners' limited opportunities for listening practice, learning to recognize English vowel phonemes and transcribe them in the International Phonetic Alphabet (IPA). Our Vowel Trainer is based on the assumption that explicit knowledge of phonology and conscious focused attention on natural sounds used in real words can affect L2 learning. In fact, the idea of learning unconsciously from distributional statistics is in direct opposition to what we are doing when in our classes we engage our students in building metalinguistic knowledge by teaching them the English inventory of vocalic categories and the relationships of those categories in the vowel space, and by teaching them IPA. In listening and transcription exercises we give our students immediate overt feedback trying to increase their awareness of correct vowel identifications. What can such explicit knowledge do for a learner's representation of sounds, and their perception and production skills? While learning IPA symbols for English vowels may have a consciousness-raising function informing learners about all the existing sound categories, it alone does not have an effect on the learner's perceptual and production targets, i.e. the implicit knowledge of categories underlying their performance in English. However, being trained to label vowel sounds with the

IPA symbols during focused intensive listening while receiving immediate feedback may have such an effect.

Although our Vowel Trainer application does not assume any profound knowledge of English phonetics, it does improve the learner's familiarity with the IPA transcription symbols and with the notion of vowel space. While the phonetic alphabet is routinely taught to university students of English in phonetics courses, which are often a required component of English programmes, other EFL learners may first need to learn the IPA transcription symbols for vowels, although some forms of simplified phonetic transcription are in fact even part of most lower-level English language textbooks. We consider IPA a useful tool in perceptual vowel training. It is useful for a learner to think of individual vowel sounds as perceptual objects that they need to recognize and distinguish from other such objects. Conscious focus on distinct transcription symbols explicitly forces the awareness of there being distinct vowel categories. That can benefit especially an EFL learner whose L1 inventory of distinctive vowels is smaller compared to that of English and thus the L1 biases them towards perceptual assimilation of L2 vowels to L1 vocalic categories and consequently a reduced interlanguage system of vowel contrasts. We believe that IPA symbols are useful to all learners, not only those with high level of proficiency in English or with a developed metalinguistic knowledge of English. Outside academic English language studies, the effectiveness of IPA for teaching pronunciation has been acknowledged, for example, by recommendations in literature on teaching diction in choral singing (Dekaney, 2003).

2. EFL learners who used the English Vowel trainer in the current study

The learners who tested the English Vowel Trainer for the purposes of this paper were young Czech adults in their early twenties, all first-year college students majoring in English (Palacký University Olomouc). Despite individual variation within the group, they all had achieved a relatively high level of proficiency in English. The entry proficiency level into the English language programme is B2 and all students are required to pass an exam at C1 level in their first academic year (for B2 and C1 proficiency levels refer to Verhelst et al. 2009).

The degree of foreign-accentedness varies across these learners. In an earlier accent-rating study conducted with a different but equivalent learner population, 18 undergraduate English majors at Palacký University were judged by English native listeners, American mid-west college students, on a nine-point Likert scale where 1 was a strong foreign accent and 9 was native-like pronunciation (Šimáčková and Podlipský 2016). The mean accentedness scores in that study ranged between 2.3 and 6.2. Several specific pronunciation features can be identified as typical of Czech accent in English (Šimáčková and Podlipský 2012). The accent has its typical suprasegmental characteristics (Volín and Skarnitzl 2010), segmental features observable in connected speech (Šimáčková, Kolářová and Podlipský 2014), consonantal features (Skarnitzl and Šturm 2016) as well as vocalic features. Our learners' productions of English vowels, even at the relatively advanced levels of proficiency, continue to betray cross-linguistic influence of the small and relatively symmetrical Czech vocalic system, with vowel length distinctions in five long – short phoneme pairs /iː - ɪ, ɛː - ɛ, aː - a, oː - o, uː - u/.[1] Specific vocalic difficulties of Czech learners of English documented in research studies include e.g. spectral non-differentiation of GOOSE and FOOT[2], or of DRESS and TRAP vowels (Šimáčková and Podlipský 2018, Šimáčková 2003), non-reduction of unstressed vowels (Volín, Weingartová and Skarnitzl 2013), or no durational adjustments in the context of a following fortis obstruent (Skarnitzl and Šturm 2016). Although unquestionably all the various aspects of Czech EFL learners' interlanguage phonology deserve attention during pronunciation practice, our training programme targets vowels, specifically identification of vowel phonemes.

3. The English Vowel Trainer

3.1. The Praat script

The English Vowel Trainer was created with the use the Praat Demo Window. Praat is excellent for creating learning software like our Trainer because of its versatility and programmability. The potential users, both instructors and learners, will appreciate Praat's easy availability and flexibility, allowing them to customize the Trainer to their needs. In this section we describe the key features of the English Vowel Trainer and indicate all the settings that can be adjusted in the script as it is now implemented (summarized in Table 4 in the Appendix). As the first step, the user opens and runs the script in Praat and the programme's blue interface, shown as white in Figure 1 (top left), appears on the screen. The user is instructed to press the spacebar to hear the first stimulus, i.e. an English word containing the target vowel (the stressed vowel in polysyllabic words). They may be allowed to replay the stimulus. The possibility of replay and the number of replays can be adjusted in the script (see the Appendix). In response to the perceived stimulus, the user clicks on an IPA symbol on the screen to identify the target vowel. Correctness of their response and their reaction time are recorded.

[1] The high front vowels are represented by different IPA symbols [iː]-[ɪ]. The qualitative differentiation was documented in Skarnitzl & Volín (2012), and its importance for perception shown in Podlipský, Skarnitzl & Volín (2009).

[2] We use J.C. Wells' keywords to refer to English vowel phonemes (Wells, 1982).

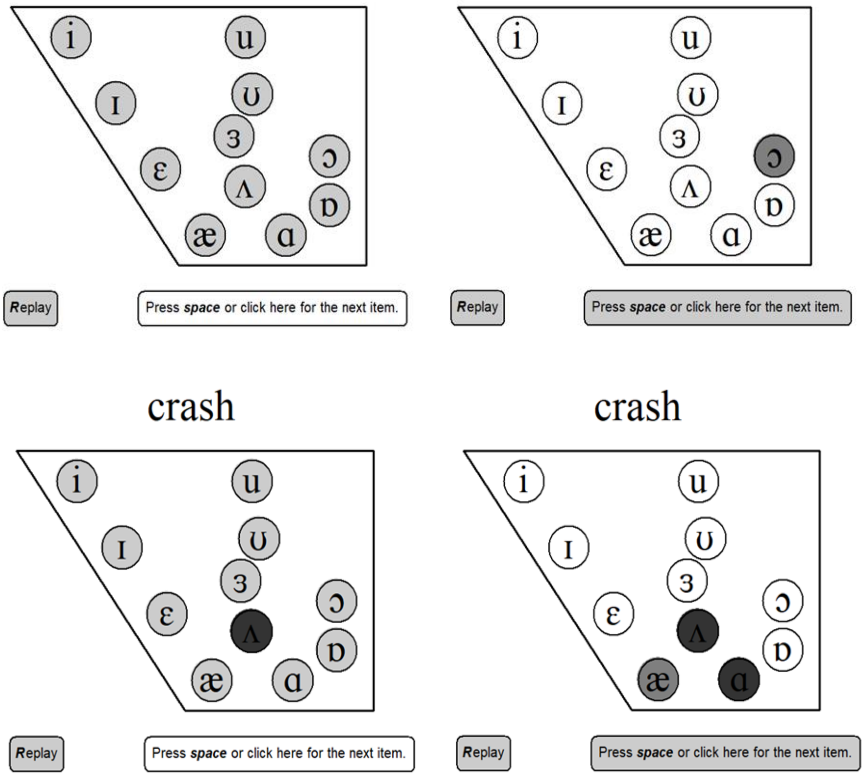
i
u
ɪ
ʊ
ɜ
ɔ
ɛ
ʌ
ɒ
æ
ɑ
Replay
Press space or click here for the next item.
i
u
ɪ
ʊ
ɜ
ɔ
ɛ
ʌ
ɒ
æ
ɑ
Replay
Press space or click here for the next item.
crash
i
u
ɪ
ʊ
ɜ
ɔ
ɛ
ʌ
ɒ
æ
ɑ
Replay
Press space or click here for the next item.
crash
i
u
ɪ
ʊ
ɜ
ɔ
ɛ
ʌ
ɒ
æ
ɑ
Replay
Press space or click here for the next item.

Figure 1. Screen shots of the English Vowel Trainer interface. Clockwise: beginning of a trial – active buttons in blue (light grey in print); correct response feedback in green (medium grey in print); incorrect response feedback (with the orthography option) in red (dark grey in print); two incorrect responses feedback with the correct answer and orthography revealed

The user receives immediate visual feedback on their response. Following a correct answer, the circle around the selected symbol turns green (Figure 1, top right). In case of a wrong answer, the circle turns red and the audio stimulus is automatically replayed. It is possible to allow the user to also have a simultaneous orthographic feedback shown on the screen outside the vowel space (Figure 1, bottom left). This is done by a simple adjustment of the script (see the Appendix). If the learner's response is wrong again, both incorrectly selected IPA symbols are shown in red. The stimulus is replayed and the correct symbol is shown in green. (Figure 1, bottom right). On completing the whole training session, the user is given a summary feedback in the form of the percentage of correct responses and the mean response time. Data about the learner's individual trials within a training session, including the vowel identity, response correctness, response time, and number of replays, are saved in a text log file in the folder Results.

Importantly, the training is adaptive; the selection of the stimuli to be presented varies as a function of how well the learner has been performing on identifying individual vowels. At the beginning of a training session, all of the 11 British English stressed monophthongs /i, ɪ, ɛ, æ, ɜ, ʌ, ɑ, ɒ, ɔ, ʊ, u/ (Ladefoged and Johnson 2011) are repeated in a randomized order for a given number of times. In the current version of the script, the length of this “diagnostic” phase is set to three occurrences of each vowel but this is easily adjustable (see the Appendix). Following the diagnostic part of the training, the stimuli are presented in rounds of maximally 17 trials. At the beginning of each round, a list of stimuli is created drawing on the 11 stressed-monophthong vowel set. As mentioned above, the stimulus list adapts dynamically to the learner’s responses so that there is more emphasis on problematic items. The vowels that were incorrectly identified in a given number (the current setting is four) of their most recent occurrences have a greater chance to be included in the next round: with our setting, if 2/3 of the recent responses to a vowel or more were wrong, then the vowel must occur at least twice in the new round; if less than 2/3 but more then 1/3 of responses were wrong, the vowel must be included at least once, if misidentified in less than 1/3 of the recent cases, then the vowel may (but does not necessarily) drop out from a round. Correctly identified vowels may be dropped completely from the rest of the training session. With our setting the number of the trial after which well-answered vowels should start dropping out is set to 60 but that again can be adjusted in the script (see the Appendix). The length of the training session is given in the total number of trials. The Trainer, used with our learners as reported below, had the number of the last trial set to 200. Also, in the course of the training, the learner is offered a pause after every x (in our case 25) trials.

3.2. The training stimuli

The current demo version of the English Vowel Trainer used citation forms of words spoken in the Standard Southern British English accent by both male and female voices. The source of the stimuli, representing the 11 stressed vowels of SSBE, was the online OneLook Dictionary Search (https://www.onelook.com/). Both open-class and closed-class words were included. The stimuli varied in length. In polysyllabic words the learners were instructed to focus on the vowel in the stressed syllable.

The user, a learner or an instructor, has the option in the Trainer to increase the difficulty of the vowel-identification task by adding noise to the recording and thus degrading the sound quality. This is done by a simple setting in the script (see the Appendix). The noise level can be constant or it can be set to change dynamically, with the signal-to-noise ratio (SNR) decreasing gradually. The onset of adding noise can also be set, e.g. the user selects the number of the trial after which noise starts to be added. The SNR then changes linearly from its initial value on trial t+1 (40 dB in the setting we used in the last session) to a separately defined value (20 dB in our setting) on the last trial.

The difficulty of the vowel identification task can also be influenced by the choice of the stimulus words if we assume that top-down lexical processing plays a role (Samuel, 2001). Training learners to recognize English vowel sounds as they are used in existing meaningful words rather than in isolation or in non-words makes for a more realistic language task but we have to consider to what extent the learner's knowledge of a word affects their perception and identification of the target vowel. When a learner hears a word they know, the mental representation of that word becomes active and they are likely to identify the target vowel phoneme also on the basis of that information rather than purely on the basis of the acoustic signal. Factors that influence availability of lexical information and thus affect the rate and accuracy of vowel phoneme identification include the overall frequency of the word, the length of the word and also its lexical complexity, which increases when the word's sound image evokes multiple semantic representations. For instance, this can happen to a member of minimal pair when a merger of two vowel categories results in perceptual confusability with its competitor (Broersma, 2012). A learner's inaccurate processing of the acoustic signal leads to an activation of an incorrect lexical representation, which in turn leads to misidentification the vowel category. Naturally, providing the user of the Trainer with the orthographic representation of a stimulus word also affects word recognition. Therefore, the Trainer never presents the spelled form of the stimulus word to the learner before they make their first response, i.e. even if the Show Orthography parameter is on, the word is shown in writing only after the user's initial response (along with the positive or negative feedback on that answer).

It can be expected that a stimulus set made up of rarely-occurring unfamiliar words and words with minimal-pair competitors will present the user with a greater challenge than a set of lexically simple and/or high-frequency monosyllabic words. In this scenario, when the lexical information is not readily available, the learner is forced to rely on the acoustic information about the target vowel. However, short frequent words should also be included in the training. Immediate negative feedback may force the learner to pay a closer attention to the quality of the vowels in some of those early-learned words and re-evaluate their possibly inaccurate representations. When Czech learners of English are consistently penalized for selecting /ɛ/ in response to words such as *map*, *fat*, or *back* but receive positive feedback if they choose the same vowel category in words such as *mess*, *pet*, or *neck*, they might begin to tune in to the qualitative differences between TRAP and DRESS vowels. Since the stimulus set can be made sufficiently large for individual lexical items not to be repeated within a single training session, any improvement (e.g. on TRAP vs. DRESS or STRUT tokens) should mean learning at the level of the vowel category rather than learning pronunciation of specific words.

The noise-masking option and the lexical difficulty option were tested with two separate groups of EFL learners during piloting of the Vowel Trainer.

4. Piloting the Vowel Trainer

4.1. Pilot 1 – with a final session using noise-masking

In total, 59 bachelor students attending the English phonetics course (i.e. three complete classes) participated in this phase of piloting the Trainer that consisted in 4 weekly training sessions. Nineteen students participated in a single training session only. Next, 26 students completed two training sessions without noise, and 14 students completed three training sessions without noise. Out of these 40 students, 36 continued to the last session with gradually increasing noise-masking of the stimuli after trial 20 (with the SNR decreasing from 40 dB on trial 21 to 20 dB on the last, 200th trial). The four sessions were one week apart for those students who participated in all of them and one or two weeks apart for those who missed a session. The training was conducted in a computer lab with students wearing circumaural headphones and proceeding through each training session at their own pace. The training was limited to the maximum of 12 minutes. Altogether 326 stimuli of varying length were used in these training sessions, pronounced by multiple female and male voices and they included 167 monosyllabic words, 112 disyllabic words, 40 words with three and 7 words with four syllables. One replay of a stimulus was allowed. In case of an incorrect answer, the same stimulus was presented once more, and if the user’s answer was wrong again, the correct answer was revealed (and the stimulus replayed automatically).

Since we do not have data to make a pre-test vs. post-test performance comparison, we demonstrate the usefulness of the English Vowel Trainer by looking at the learners’ progress, i.e. changes in correctness of their responses, over the course of a training session and also by looking at their progress across the three consecutive noise-free training sessions. A logistic regression model was fit to the data set to estimate the impact of Trial (1 through 200), Session (1 through 3), Vowel and Response latency (in s) on the likelihood that the learners would identify a stimulus correctly. Robustness of learning in the three noise-free conditions was subsequently evaluated by looking for any drop in correct identification in the final session with the gradually increasing levels of noise-masking. In a regression model, learners’ response accuracy to noisy stimuli was regressed against the number of the trial. Table 1 summarizes learners’ vowel identification performance in percentages of correct and incorrect responses in the four sessions of Pilot 1. Table 2 presents the results of logistic regression models on the data from the sessions without and with noise.

Table 1. Mean percentages of correct and incorrect responses in Pilot 1, Sessions 1-3 without noise, Session 4 with noise

	Sessions			
	without noise			**with noise**
	1	**2**	**3**	**4**
% Correct	65.69	73.35	73.94	76.38
% Incorrect	34.31	26.65	26.06	23.57

Table 2. Vowel identification accuracy regression models for Czech EFL learners' (HLM LOGIT); *p < .1; **p < .01; ***p < .001

	Sessions without noise		**Session with noise**	
	Estimate	**SE**	**Estimate**	**SE**
Session 2	0.324**	0.16		
Session 3	0.341*	0.19		
Trial	0.002***	0	-0.002***	0
Response latency	-0.148***	0.01		
KIT as reference category:				
TRAP	-2.661***	0.17		
LOT	-2.157***	0.17		
BIRD	-2.098***	0.17		
STRUT	-2.062***	0.17		
BATH	-1.951***	0.17		
THOUGHT	-1.900***	0.17		
FOOT	-1.782***	0.17		
GOOSE	-1.715***	0.17		
DRESS	-1.689***	0.17		
FLEECE	-0.819***	0.19		
Constant	2.937***	0.2	1.378***	0.11
Log Likelihood	-11615.56		-2884.33	
N	21323		5419	
N groups	113		36	

It follows from the adaptive nature of our training tool that increases in correct identification over the course of a training session as well as between sessions are unlikely to be dramatic. The emphasis on problematic vowels means that incorrectly identified vowels get preference for selection in the upcoming trials.

Therefore, improvement over time may appear smaller than it is in reality. This may be why the mean percentage of correct responses in session 3 seen in Table 1 remained virtually the same as in session 2 (along with the fact that several participants skipped session 2 and session 3 was in fact their second session). Still, the logistic regression model confirmed that the increases in accuracy between Session 1 and 2, and between Sessions 1 and 3, as represented by the means shown in Table 1, were significant, and also that the increasing number of the trial significantly predicted a small increase of the probability of correct responses within each of the noise-free sessions overall (Table 2). The significance of the response latency predictor indicates that correct decisions tended to require less time to make.

Considering the learners' ability to identify the vowel categories in the noise-masked stimuli used in Session 4, we find some support for a lasting effect of the preceding training. As shown in Table 1, the overall performance in the last training session with noise was even somewhat better compared to the last session without noise, and although the likelihood of correct vowel identification decreased over the course of the session, i.e. with the increasing trial number (see Table 2), this overall decrease was only mild considering that the SNR dropped from the initial 40 dB to the final 20 dB.

The logistic regression model also allows us to assess how accurately the learners identified each of the eleven vowel categories. Identification accuracy of the individual vowel monophthongs was measured against the KIT vowel as the reference category, since it was earlier shown that /ɪ/ is the most accurately produced English vowel in speech of Czech EFL learners (Šimáčková and Podlipský 2018), and also since it had the highest identification accuracy in the present results: as seen in Table 2, all other vowel categories had significantly lower likelihood of correct answers. In Table 2 the vowel categories are ordered from those with the greatest negative difference from KIT (the vowel TRAP) to the lowest (the vowel FLEECE). Table 3 then shows a confusion matrix for the vowel identification, pooling data across the three noise-free sessions. The order of the vowels in the matrix is motivated phonetically; their reordering according to the percentages of correct responses in grey cells corresponds exactly to the order given in Table 2 (with the exception of the BATH and THOUGHT that are reversed but have very similar values of the estimates in Table 2 as well as of the percentages in Table 3).

The poor performance on the vowel /æ/ was expected. The relatively high proportion of correct responses for its predicted competitor, /ɛ/, may reflect the phonetic similarity of the English lax /ɛ/ to the Czech short /ɛ/ and its identification accuracy parallels more accurate pronunciation of /ɛ/ compared to /æ/ in productions of EFL learners from the same learner population (Šimáčková and Podlipský 2018). However, the same production data showed that /æ/ and /ɛ/ are easily confused in these EFL learners' speech, and that compared to the other English monophthongs, both /æ/ and /ɛ/ display increased within-speaker variation, as well as substitutions between each other in pronunciation of individual lexical items. In Šimáčková and Podlipský (2018), learners' productions of /æ/ showed little lowering or retraction as compared with /ɛ/. In this respect, the production data do not match the current vowel identification data elicited in Pilot 1. The confusion matrix of identification outcomes in percentages for each English monophthong (Table 3), shows that the low front /æ/ was misidentified more often as a back or central vowel (/ɑ/, or /ʌ/) than as the front /ɛ/.

Regarding FLEECE and KIT, the current vowel identification scores (Table 3) do correspond to the production results in Šimáčková and Podlipský (2018). Qualitatively, SSBE high front /ɪ/ and /i/ closely acoustically resemble the Czech short – long pair /ɪ/-/iː/. The KIT vowel followed by the FLEECE vowel are the most accurately perceived and most authentically produced vowels by Czech EFL learners. For the last contrasting pair tested in Šimáčková and Podlipský (2018), FOOT and GOOSE, it was found that learners' productions of the two vowels were less native-like, differentiated only in duration and not significantly in spectral quality. The less reliable differentiation of /u/-/ʊ/ in the production data is paralleled by increased mutual confusability of GOOSE-FOOT compared to FLEECE-KIT in the present identification results (Table 3). Similarly, Table 3 suggests that the vowel identification errors elicited for the BATH – STRUT and LOT – THOUGHT pairs are related to poorer differentiation between the pair members in spectral quality.

It has to be noted that besides perceptual confusability of the English vowel sounds to the Czech learners, some proportion of incorrect responses was surely due to confusability of the IPA symbols themselves.

Table 3. Vowel categories selected by the learners (Response vowels) in response to a stimulus (Stimulus vowel); bold numbers in grey cells are percentages of correct responses

		Stimulus vowel										
		i	**ɪ**	**ɛ**	**æ**	**ɜ**	**ʌ**	**ɑ**	**ɒ**	**ɔ**	**ʊ**	**u**
Response vowels	**i**	**88.5**	5.0		0.3	1.3	0.3	0.4	0.6	0.5	0.2	0.9
	ɪ	7.7	**95.0**	2.0	1.7	4.9	0.7	0.9	1.5	2.0	0.6	1.0
	ɛ	2.5		**75.5**	10.0	7.6	1.0	1.1	0.4	0.3	1.0	0.7
	æ			12.3	**56.4**	4.5	2.2	4.6	0.5	0.7	0.6	1.1
	ɜ	1.3		5.7	4.5	**64.8**	2.8	2.3	1.6	1.6	1.6	1.4
	ʌ			1.7	12.0	4.7	**67.9**	15.5	5.7	1.0	1.2	1.0
	ɑ			2.0	12.2	2.3	13.0	**70.2**	4.1	3.1	0.7	0.4
	ɒ			0.2	0.2	4.0	7.1	3.1	**64.4**	17.5	4.1	0.7
	ɔ				0.6	2.8	1.9	1.7	20.3	**69.6**	3.4	1.8
	ʊ			0.2	2.2	2.0	2.4		0.6	1.8	**72.5**	17.5
	u			0.4	0.1	1.2	0.8		0.3	1.8	14.2	**73.5**

4.2. Pilot 2 – Effects of lexical knowledge

In Pilot 2, the English Vowel Trainer was tested with three different stimulus sets. The intention was to vary the difficulty of vowel identification by manipulating the availability of top down lexical cues. Three complete classes of first-year bachelor students attending the English phonetics course were engaged in Pilot 2. Only performance of students who completed all 3 consecutive weekly sessions is reported here (52 students in total). For each session a different set of audio stimuli was created. In Session 1, the stimulus set included 195 monosyllabic words varying in frequency and the availability of minimal-pair competition. All vowel categories were represented in the stimulus set, with the number of tokens in a category ranging between 17 and 21. This stimulus set was modified in Session 2 by replacing 48 words with monosyllabic lower-frequency words each having a minimal pair competitor. The replacement was biased towards the difficult vowel /æ/ and its contrast with /ɛ/ and /ʌ/, so that 10 new TRAP-words were introduced, as well as 5 new DRESS- and 5 new STRUT-words forming actual TRAP-DRESS and TRAP-STRUT minimal pairs. Five new GOOSE words and five new FOOT words were used, each potentially forming a minimal pair with FLEECE and KIT words respectively. This was done in order to test possible misperceptions due to the fronted pronunciation of /u/ and /ʊ/ that has become common in SSBE (Cruttenden 2013). Three new words for the vowels /i, ɪ, ɜ, ɔ, ɒ, ɑ/ were included. In Session 3, 48 stimulus words from Session 2 were again replaced, this time with disyllabic trochees of low frequency and if possible also forming potential or actual minimal pairs with other low-frequency words. The replacement followed the same pattern as in Session 2, i.e. there were 10 new TRAP words, 5 new words for DRESS, STRUT, GOOSE, and FOOT vowels, and 3 new words for each of the remaining six monophthongs. Thus, for example, the TRAP vowel was presented to the learners in words *act, ash, bag, bat, cash, cat, clash, crash, fact, fat, flat, chat, man, mat, rat, smash, tact, that* in Session 1, in words *gnat, ham, hatch, mag, pap, rack, rag, sac, scat, tat* as well as *ash, bag, cash, clash, crash, chat, mat, smash* in Session 2, and in *aster, bangle, barrow, clatter, lattice, mantle, pallet, raffle, shatter, tanner, gnat, mag, rack, sac, ash, bag, cash, chat* in the final Session 3. After an initial incorrect answer two attempts for a correction were allowed.

In Pilot 2 we were interested in the cross-session comparison. A Repeated Measures ANOVA on the mean proportion of correct responses with Session as the within subject variable (3 levels) confirmed the significant main effect of Session (*F* [2, 102] = 60.004, *p* < .0001). Pairwise Tukey HSD tests showed a significant decrease in the proportion of correct responses between Sessions 1 and 2 and between Sessions 2 and 3, *p* < .01 (see Figure 2).

The results indicate that lexical knowledge indeed facilitated vowel recognition. The learners were most accurate in Session 1 when they heard a mixture of short, i.e. monosyllabic, words varying in overall frequency. When

25% of the original stimuli were replaced with a set of monosyllabic words of low frequency and increased lexical complexity in Session 2, vowel identification became less accurate. Further replacement of monosyllabic words with disyllabic low-frequency words in Session 3 reduced identification accuracy even further.

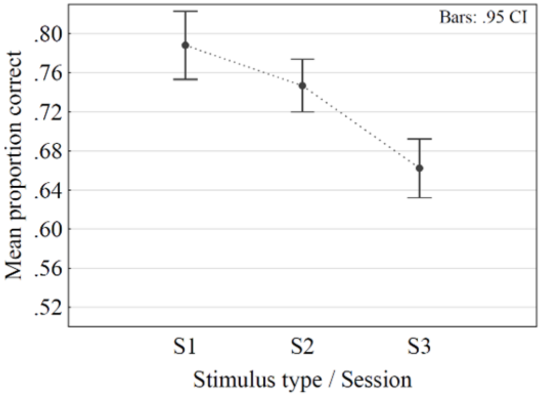

Figure 2. Mean proportion of correct responses in Pilot 2 across Sessions 1, 2 and 3

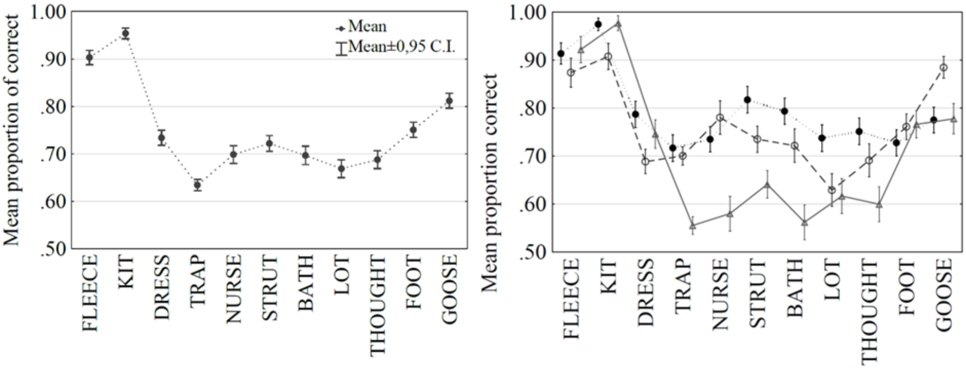

Figure 3. Mean proportion of correct responses to each English stressed monophthong in Pilot 2. Left: all sessions combined. Right: sessions separated: full circle – session 1, empty circle – session 2, empty triangle – session 3.

The group's success in identifying individual vowels represented in Figure 3 mirrors the performance of the learners in Pilot 1 as shown by the differences between individual vowel results in Table 3. The contribution of lexical knowledge to the identification rate is seen in Figure 3 Right, which shows how the learners performed on each vowel across the three sessions. Compared to Session 1, the correctness of responses changed in the expected direction, i.e. dropped, for vowels /i, ɪ, ɛ, ʌ, ɑ, ɒ, ɔ/ in Sessions 2, and for vowels /æ, ɜ, ʌ, ɑ, ɔ/ in Session 3 compared to Session 1. The vowels previously identified as less difficult for the learners (in Šimáčková and Podlipský 2018, and in Pilot 1) did not show any deterioration across sessions, which however is also due to the fact that for these vowels only few tokens were replaced between sessions.

5. Summary

The goal of this paper was to introduce a computer tool for practicing perception and identification of English monophthongs. The Praat-based English Vowel Trainer is intended for English learners at the intermediate-to-advanced level of proficiency to be used for individual practice at home or in a computer classroom, possibly with assistance from an English language instructor. It can also be used to lend variability to an English pronunciation or English phonetics class. The requirements for using the Trainer include downloading the free software Praat into one's computer, acquiring a pair of head phones, but also having basic knowledge of the IPA symbols for English vowels.

The demo version of the English Vowel trainer has been tested with the help of advanced EFL learners attending an introductory phonetics class at the English Department of Palacký University Olomouc. Tracking their performance in the course of a training session and across three consecutive training sessions has shown that using the Trainer can improve learners' vowel perception. In the future we intend use pre-training vs. post-training tests in order to corroborate the efficacy of the Trainer as a tool for learning English vowel identification.

The Praat script for the Demo version of the English Vowel trainer is available from link: https://anglistika.upol.cz/fileadmin/userdata/FF/katedry/kaa/docs/vowel_ipa_trainer.zip.

References

Baker, Wendy and Pavel Trofimovich. 2006. Perceptual paths to accurate production of L2 vowels: The role of individual differences. *IRAL–International Review of Applied Linguistics in Language Teaching* 44 (3), 231-250. https://doi.org/10.1515/IRAL.2006.010

Barriuso, Taylor Anne and Rachel Hayes-Harb. 2018. High Variability Phonetic Training as a Bridge from Research to Practice. *CATESOL Journal* 30 (1), 177-194.

Bradlow, Ann R., Akahane-Yamada, Reiko, Pisoni, David. B. and Yoh'ichi Tohkura. 1999. Training Japanese listeners to identify English /r/ and /l/: Long-term retention of learning in perception and production. *Perception & psychophysics* 61 (5), 977-985. http://dx.doi.org/10.3758/BF03206911

Boersma, Paul and David Weenink, 2019. *Praat: doing phonetics by computer* [Computer program]. Version 6.0.49, retrieved 10 March 2019 from http://www.praat.org/

Broersma, Mirjam. 2012. Increased lexical activation and reduced competition in second-language listening. *Language and cognitive processes* 27 (7-8), 1205-1224. https://doi.org/10.1080/01690965.2012.660170

Dekaney, Elisa M. 2003. The effect of computerized versus classroom instruction on the phonetic pronunciation of English. *Journal of Research in Music Education* 51 (3), 206-217. https://doi.org/10.2307/3345374

Derwing, Tracy. M. 2010. Utopian goals for pronunciation teaching. In: *Proceedings of the 1st Pronunciation in Second Language Learning and Teaching Conference*, ed. by John Levis and Kimberly LeVelle. Ames, IA: Iowa State University, 24-37.

Escudero, Paola. 2007. Second-language phonology: The role of perception. In: *Phonology in Context*, ed. by M. C. Pennington. Palgrave Macmillan, London, 109-134. https://doi.org/10.1057/9780230625396_5

Escudero, Paola and Daniel Williams. 2014. Distributional learning has immediate and long-lasting effects. *Cognition* 133 (2), 408-413. https://doi.org/10.1016/j.cognition.2014.07.002

Flege, James Emil and Serena Liu. 2001. The effect of experience on adults' acquisition of a second language. *Studies in Second Language Acquisition* 23 (4), 527-552. https://doi.org/10.1017/S0272263101004041

Flege, James Emil, Ian R.A. MacKay and Diane Meador. 1999. Native Italian speakers' perception and production of English vowels. *The Journal of the Acoustical Society of America* 106 (5), 2973-2987. https://doi.org/10.1121/1.428116

Flege, J. E., Frieda, Elaina M., Walley, Amanda C. and Lauren A. Randazza. 1998. Lexical factors and segmental accuracy in second language speech production. *Studies in Second Language Acquisition* 20 (2), 155-187. https://doi.org/10.1017/S0272263198002034

Grenon, Izabelle, Mikio Kubota and Chris Sheppard. 2019. The creation of a new vowel category by adult learners after adaptive phonetic training. *Journal of Phonetics* 72, 17-34. DOI: 10.1016/j.wocn.2018.10.005

Jia, Gisela, Strange, Winifred, Wu, Yanhong, Julissa Collado and Qi Guan. 2006. Perception and production of English vowels by Mandarin speakers: Age-related differences vary with amount of L2 exposure. *The Journal of the Acoustical Society of America* 119 (2), 1118-1130. https://doi.org/10.1121/1.2151806

Iverson, Paul, Melanie Pinet and Bronwen G. Evans. 2012. Auditory training for experienced and inexperienced second-language learners: Native French speakers learning English vowels. *Applied Psycholinguistics* 33 (1), 145-160. https://doi.org/10.1017/S0142716411000300

Kissling, Elizabeth M. 2014. What predicts the effectiveness of foreign-language pronunciation instruction? Investigating the role of perception and other individual differences. *Canadian Modern Language Review* 70 (4), 532-558. http://dx.doi.org/10.3138/cmlr.2161

Kissling, Elizabeth M. 2015. Phonetics instruction improves learners' perception of L2 sounds. *Language Teaching Research* 19 (3), 254-275. https://doi.org/10.1177/1362168814541735

Lee, Andrew H. and Roy Lyster. 2016. The effects of corrective feedback on instructed L2 speech perception. *Studies in Second Language Acquisition* 38 (1), 35-64. https://doi.org/10.1017/S0272263115000194

Ladefoged, Peter and Keith Johnson. 2011 (6th ed.). *A Course in Phonetics*. USA: Wadsworth.

Motohashi-Siago, Miki and Debra M. Hardison. 2009. Acquisition of L2 Japanese geminates: Training with waveform displays. *Language Learning & Technology* 13 (2), 29-47. http://hdl.handle.net/10125/44179.

Nishi, Kanae and Diane Kewley-Port. 2007. Training Japanese listeners to perceive American English vowels: Influence of training sets. *Journal of Speech, Language, and Hearing Research* 50 (6), 1496-1509. https://doi.org/10.1044/1092-4388(2007/103)

Perrachione, Tyler K., Lee, Jiyeon, Ha, Louisa Y. and Patrick C. M. Wong. 2011. Learning a novel phonological contrast depends on interactions between individual differences and training paradigm design. *The Journal of the Acoustical Society of America* 130 (1), 461-472. https://doi.org/10.1121/1.3593366

Podlipský, Václav J., Radek Skarnitzl and Jan Volín. 2009. High front vowels in Czech: A contrast in quantity or quality? *INTERSPEECH-2009*. [Online] Available at http://www.isca-speech.org/archive/interspeech_2009. [Accessed on 10 February 2017].

OneLook Dictionary Search. Retrieved from https://www.onelook.com/

Samuel, Arthur G. 2001. Knowing a word affects the fundamental perception of the sounds within it. *Psychological Science* 12.(4), 348-351. https://doi.org/10.1111/1467-9280.00364

Sheldon, Amy and Winifred Strange. 1982. The acquisition of /r/ and /l/ by Japanese learners of English: Evidence that speech production can precede speech perception. *Applied Psycholinguistics* 3.(3), 243-261. https://doi.org/10.1017/S0142716400001417

Shinohara, Yasuaki and Paul Iverson. 2018. High variability identification and discrimination training for Japanese speakers learning English /r/–/l/. *Journal of Phonetics* 66, 242-251. https://doi.org/10.1016/j.wocn.2017.11.002

Šimáčková, Šárka. 2003. "Engela's Eshes": Cross-linguistic perception and production of English [æ] and [ɛ] by Czech EFL learners trained in phonetics. In Solé, Maria-Josep, Daniel Recasens and Joaquín Romero (eds.), *Proceedings of the 15th International Congress of Phonetic Sciences (ICPhS 2003)*, Barcelona, 3-9 August, 2003, 2293-2296.

Šimáčková, Šárka and Václav J. Podlipský. 2012. Pronunciation skills of an interpreter. In: Jitka Zehnalová, Ondřej Molnár and Michal Kubánek (eds.), *Teaching translation and interpreting skills in the 21st century*. Olomouc: Palacký University Olomouc, 139-149.

Šimáčková, Šárka and Václav J. Podlipský. 2016. Global foreign accent rating of code-switched and L2-only sentences. *Research in Language* 14 (2), 149-164. https://doi.org/10.2478/rela-2018-0009

Šimáčková, Šárka and Václav J. Podlipský. 2018. Production accuracy of L2 vowels: Phonological parsimony and phonetic flexibility. *Research in Language* 16 (2), 169-191. https://doi.org/10.2478/rela-2018-0009

Šimáčková, Šárka, Kateřina Kolářová and Václav J. Podlipský. 2014. Tempo and connectedness of Czech-accented English speech. *Concordia Working Papers in Applied Linguistics* 5, 667-77.

Skarnitzl, Radek and Pavel Šturm. 2016. Pre-fortis shortening in Czech English: A production and reaction-time study. *Research in Language* 14 (1), 1-14. https://doi.org/10.1515/rela-2016-0005

Skarnitzl, Radek and Pavel Šturm. 2017. Voicing assimilation in Czech and Slovak speakers of English: Interactions of segmental context, language and strength of foreign accent. *Language and speech* 60 (3), 427-453. https://doi.org/10.1177/0023830916654509

Skarnitzl, Radek and Jan Volín. 2012. Referenční hodnoty vokalických formantů pro mladé dospělé mluvčí standardní češtiny. *Akustické listy* 18 (1), 7-11.

Verhelst, Norman, et al. 2009. *Common European Framework of Reference for Languages: learning, teaching, assessment*. Cambridge: Cambridge University Press.

Volín, Jan and Radek Skarnitzl. 2010. The strength of foreign accent in Czech English under adverse listening conditions. *Speech Communication* 52 (11-12), 1010-1021. https://doi.org/10.1016/j.specom.2010.06.009

Volín, Jan, Lenka Weingartová and Radek Skarnitzl. 2013. Spectral characteristics of schwa in Czech accented English. *Research in Language* 11 (1), 31-39. https://doi.org/10.2478/v10015-012-0008-6

Wanrooij, Karin, Paul Boersma and Titia L. van Zuijen. 2014. Distributional vowel training is less effective for adults than for infants. A study using the mismatch response. *PloS One* 9 (10), e109806. https://doi.org/10.1371/journal.pone.0109806

Wanrooij, Karin, Paul Boersma and Titia Benders. 2015. Observed effects of "distributional learning" may not relate to the number of peaks. A test of "dispersion" as a confounding factor. *Frontiers in Psychology* 6. Article 1341. https://doi.org/10.3389/fpsyg.2015.01341

Wanrooij, Karin, Johanna De Vos and Paul Boersma. 2015. Distributional vowel training may not be effective for Dutch adults. In: *18th International Congress of Phonetic Sciences (ICPhS 2015)*. University of Glasgow.

Wells, J. C. 1982. *Accents of English*. Volume 1. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511611759

Appendix

Table 4. List of settings that can be changed in the English Vowel Trainer script.

Variable name in the script	Value of the variable in the current setting	Description of the variable	Script line
addNoise =	= 0 (in Session 4 the setting was 1)	Should noise be added to the signal? 1 is yes.	69
snr =	= 40	If addNoise = 1, setting the speech-to-noise ratio.	70
noiseOnset =	= 20	If addNoise = 1, after which trial should noise be added?	72
dynamic.noise =	= 1	If addNoise = 1, is noise level dynamic (with a gradually decreasing SNR)? 1 is yes.	74
finalSnr =	= 20	Signal-to-noise ratio of the last trial. The value defined in the variable snr is then the initial SNR.	75
n.initial.repetitions =	= 3	Diagnostic phase: number of initial repetitions of all 11 SSBE vowels	77
n.criterion =	= 4	How many last occurrences of a vowel does the script look at to determine how many times the vowel will be repeated in the next round of trials?	81
drop.trial =	= 60	After how many trials should well-answered vowels start dropping out?	84
cutoff2 = cutoff1 =	= 2/3 = 1/5	What is the mean occurrence of errors in the last relevant trials (whose number is defined in the variable n.criterion), i.e. levels for requiring at least 2, 1, and 0 tokens in the upcoming round?	87
lastTrial =	= 200	Stop the training session after this many trials.	96
pauseAfter =	= 25	Pause the training session after this many trial. The participant can resume by clicking.	99
maxReplays =	= 2	Maximum number of times one stimulus can be replayed	102
repeatWrong =	= 1	The number of times the user can repeat their attempt after a wrong answer.	105
showOrthography =	= 1	After the users response the word is shown in orthography along with correct/incorrect answer feedback, if set to 1.	108