Univerzita Palackého v Olomouci

Filozofická fakulta

Katedra anglistiky a amerikanistiky

Imitation-based visual-feedback training of English vowels for Czech learners of English

A research proposal

(Bakalářská práce)

Autor: Kateřina Plachká

Vedoucí práce: Mgr. Václav Jonáš Podlipský, Ph.D.

Olomouc 2020

Prohlašuji, že jsem tuto bakalářskou práci vypracovala samostatně a uvedla plný seznam citované a použité literatury

V Olomouci dne 14.12.2020 ....

I would like to thank my supervisor Mgr. Václav Jonáš Podlipský, Ph.D. for his help and support throughout the process of writing this thesis.

Table of contents

1. Introduction

The aim of this thesis is to make a research proposal that examines whether articulatory training with visual feedback has an influence on the final quality of production of chosen English vowels by Czech learners of English.

L2 learners all around the world experience difficulties with learning specific attributes of the languages they are learning and one of the greatest problems will always be the one with the acquisition of the native accent by a non-native speaker. A specific problem of L2 learners is the acquisition of pronunciation of the L2 sounds. L2 experienced adult learners often learn how to properly use an L2 vocabulary and a large portion of grammatical rules, but they usually experience difficulties when it comes to eliminating their non-native accent (Scovel 1969). There are many other studies that support a similar idea and claim for example that even heavily accented non-native speakers "may have nevertheless mastered grammar and idioms of English as well as native speakers" (Moyer 2013, 133). Thus, the proper pronunciation of foreign sounds remains one of the most enormous difficulties L2 learners experience and there are many ways which try to facilitate this issue for L2 learners for example by different kinds of training.

In this thesis, I would like to investigate firstly the general problems L2 learners may encounter while acquiring L2 native-like pronunciation. Then I would like to focus on the Czech L2 learners of English in particular and present a training program which might improve the pronunciation of problematic English vowels to sound more native-like. The thesis contains a proposal of an empirical test of the efficiency of the training.

There has been much research dedicated to comparisons of different L1 speakers, the characteristics of their native L1 speech, and its reflection in the pronunciation of an L2. According to the Speech Learning Model proposed by Flege (1987, 1992, 1995), after reaching the age of six years, children have already finished the process of creation of their L1 phonetic categories. This theory thus states that if a new L2 sound is similar (for example close in the phonetic space) to a different sound in an L1, the L1 speaker will have difficulty with creating a new category for such sound and instead what they are going to do is assimilate this sound to already existing sound in their native language. On the other hand, a sound that is completely unknown and not similar to any speaker's native sound category, the new category of this sound should be easier for them to acquire. The SLM theory will further be discussed in the chapter 7.

Languages as such are very diverse all around the world and just by listening to a person from a foreign country speak in their native language, we may notice sounds that are not involved in our own native language. As we have already stated for the SLM theory, after 6 years of age people have created their native language sound categories and similar sounds from an L2 then assimilate to these categories. Because different languages differ not only from English but also from each other, we may assume that the acquired categories and the subsequent assimilation of L2 sounds will also be different for different L1 speakers. The current paper is concerned with vowels so we should provide examples from different L1s and their difficulties with English vowels.

One of the languages that have been studied with respect to the way of their perception and production of English vowels is Japanese. A study by Ohata (2004) made a thorough comparison of Japanese and English sounds, mentioning that the two factors that usually make the major difference between two language vowels are the differences between the number of vowels in their vowel inventory and the distinction between lax and tense vowels, which are differentiated by the extent of muscle involvement while producing a certain vowel (Ladefoged 1982). If the muscles are moving more during the production of a vowel, the vowel is said to be tense. On the other hand, when the muscle involvement in a vowel production is weak, the vowel is said to be lax. The difficulty of English lax and tense vowels differentiation for Japanese speakers is also observable in a study made by Grenon et al. (2019), which only proved that Japanese speakers do have difficulties with the qualitative perception of this contrast in the vowel pair /i/ and /ɪ/ and rather rely on the durational difference between these two vowels even after a phonetic training. Ladefoged & Johnson (2014), however, describe that the main difference between these two vowels lies in the position of the tongue and the lips. The vowel /i/ in *beet* would then be a high front vowel and the vowel /ɪ/ in *bit* would be a mid-high vowel.

Ohata (2004) further states there are only five vowels in Japanese compared to a large vowel inventory of English which contains 15 vowel monophthongs. There is no differentiation between lax and tense vowels in Japanese and no central vowels are included in its vowel inventory. Ohata states that the English central vowels would cause difficulties in the production for Japanese L1 speakers because Japanese are only used to differentiate between front and back vowels. However, as we have already mentioned for the SLM theory, the vowels which are further from their L1 vowel counterpart are likely to be perceived better. Furthermore, because of no distinction between lax and tense vowels in Japanese, the Japanese learners of English are also expected to have difficulties with vowel pairs containing such differentiation as in /i/ and /ɪ/ or /ʊ/ and /u/.

If we talk about vowels, Spanish is in some way very similar to Japanese, and as its vowel inventory also contains only five vowels (del Puerto et al. 2008), Spanish and Japanese both seem like languages adequate for a comparison with Czech. Spanish L1 speakers are as well as Japanese L1 speakers also expected to have trouble differentiating English lax and tense vowels. According to Finch et al. (1982), the vowel space of a Spanish speaker also lacks two English low back vowels /ɒ/ and /ɑ/ which is also the same for Czech native speakers. As well as for Japanese, Spanish speakers of English often experience difficulties recognizing lax and tense contrast, and the vowel pairs /i/ and /ɪ/ or /ʊ/ and /u/ are recognized as /i/ and /u/, respectively. (Finch et al.1982). As we have already mentioned that there are certain similarities between the Spanish, Japanese, and Czech vowel inventories and both Spanish and Japanese speakers often do not perceive the difference between the lax and tense vowels, we might expect that also in this study this distinction will be hard for our participants. The Spanish interpretation of English central vowels is also considered to be difficult, as they tend to have difficulties perceiving all of the English central vowels incorrectly. As we will see in the chapter 9. which compares Czech and English vowels, in the Czech vowel inventory there are no central vowels as well as in Japanese and Spanish. In our study, we will examine only the production of a central vowel /ʌ/ which as mentioned by Finch et al. (1982) is by Spanish speakers of English widely assimilated to the Spanish low central vowel /a/.

Non-native speakers of English can often be recognized by native speakers according to several signs of their speech (Bradlow 1999). That could be a fact related to a wrong application of an L1 speakers' existing sounds into the pronunciation of a given L2. As we have already mentioned above, different L1 speakers can have difficulties with different L2 sounds, depending on various factors, some of them being unfamiliar or familiar sounds (SLM theory), the age of the language acquisition, lax and tense vowels differentiation, or the number of vowels in a vowel inventory among others. The main goal of this thesis is to try to eliminate the difficulty in English vowel production by a proposing training method.

For this thesis, I was inspired by a study by Kartushina et al. (2015) on the effect of indirect articulatory feedback training of four Danish vowels by French speakers who have no experience with the Danish language. This study revealed that an hour of training with trial-by-trial corrective feedback per each tested vowel improves the production of the vowels by 17 % on average. The accuracy differed mainly in the difficulty of specific vowels. The study by Kartushina is a default source for the present paper and will be therefore described in more detail in chapter 5.

My study will be different from Kartushina et al. (2015) in several aspects. First of all, the speakers will be Czech and they will already have had considerable experience with English. Also, they will not be training isolated vowels, the trained vowels will rather be a part of whole existing English words. I will focus on the contrasting pairs of vowels /ɛ/ and /æ/, /ʌ/ and /ɑ/ and /ɔ/ and /ɒ/ and /u/ and /ʊ/ and the corrective feedback will be both visual and auditory meaning that the participants will be able to replay their own recording after producing a given word.

I am going to try to answer the question of whether an approach that would be similar to the one used in the Danish study on French speakers of Danish vowels is beneficial for the production of chosen English vowels for Czech speakers of English.

This paper will be divided into two parts. I will dedicate the first part of this thesis to the theoretical aspects of the current topic by first briefly discussing the general issue of the acquisition of L2 speech sound patterns. We will discuss difficulties experienced by L2 learners, focusing on the pronunciation aspects of this L2. First, there will be listed the two ways of language training, according to whether a particular study is focused rather on the aspect of how a learner correctly perceives the native sounds or how they can produce these sounds in a native-like way. We will discuss the main feedback approaches used for these two training methods, as feedback is one of the most important learning tools in any kind of training. We will try to demonstrate the basic principles of these feedbacks, their advantages, and disadvantages. We will discuss models that were created to predict difficulties in L2 acquistion. Subsequently, we will introduce the two vowel systems that are the subject of this study, thus we will list the English and Czech monophthongs describing their basic properties and differences between them. By comparing similar existing studies and the main ideas of models predicting difficulties we will try to hypothesize which English vowels will be more difficult for Czech speakers to achieve in a native-like way.

The second part of this paper will be dedicated to a proposal of methodology, the description of vowels chosen for this study and the description of program which could be used for training.

2. L2 acquisition importance and problems

In today’s world, second language learning is becoming a part of almost every person’s life. There are already many good teaching ways for teaching grammar, vocabulary, but still, I think there should be more focus on the pronunciation of L2 sounds. As there are still not many ways to improve this language aspect, this thesis has an aim to test one of the possible training methods. It is well known that learning a foreign language in general may cause many problems because of differences such as those in the grammatical systems and vocabulary of an L2 and L1 which usually form the first and very apparent difficulty encountered by any L2 learner. However, even after acquiring mastery of a given language in terms of its grammatical system, syntactic structure, and vocabulary, one can experience further troubles with the pronunciation. This can be caused, among other things, by the lacking exposure to the native environment. A specific problem may sometimes emerge when a learned L2 includes sounds that are not in the inventory of one's L1. Because this study is observing exclusively vowels, we talk about the L2 and L1 vowel inventory. It seems intuitive that if an L2 has a larger vowel inventory than L1, meaning that it consists of vowel categories that are not even included in the L2 learner's native language, the pronunciation will be even more difficult to achieve. However, because of the increasing interest in language studies and innovations in technology, there have risen some new methods to train L2 pronunciation using computers. We need to further divide this training into production training and perception training, according to what is being trained. Production training focuses on the learner's ability to produce the target L2 sounds in a native-like way while perception training is concerned with the correct recognition of a given sound of an L2. What these two training methods have in common is the need for certain feedback offered to the participant after their attempt of correct sound production or perception. If the feedback allows a more precise way of determining the correctness of the participant's try, and as for the production feedback they can also be able to see how far away their attempt was from the native version of this sound. In the next section, I review the types of feedback used in previous studies and try to demonstrate their advantages and disadvantages.

3. Training paradigms and types of feedback

As mentioned above, there are several types of feedback used for L2 training studies. Some of these are prototypically used for production training and others for the training of perception.

3.1. Perception feedback paradigms

If we start with the perception feedback, one of the typical types of feedback is a simple correct-incorrect type of feedback which as the name suggests only admits two answers, namely correct and incorrect. This type of feedback makes use of its simplicity and when a participant seems to be having problems with identifying a sound, the correct/incorrect type of feedback might be a good way to practice the problematic sound by simply including more words with this particular sound throughout the training so that it is easier to remember it. With this training paradigm, a participant can be asked for example to identify a single sound (e.g., a single vowel) or to identify a whole word including the sound or sounds that are targetted by the study.

3.2. Production feedback paradigms

The simple correct/incorrect feedback might be partly applicable also on the production training for example by means of immediate evaluation by a native speaker, but it is important to mention that this approach would probably not be very reliable. Moreover, it is possible that the participant would need more detailed feedback about their production accuracy than a simple statement whether their trial was right or wrong. For a non-native inexperienced speaker, this approach might also be overly strict and the probability for a learner’s sound to be considered simply as correct might be very low, perhaps causing demotivation of the participant. Therefore, production training feedbacks are focused rather on the closeness of each sound trial of an L2 learner to the native speaker’s production of the same word or sound. In this way, the speaker can observe their mistakes more clearly and can see how far they are from the native speaker’s speech and how much they need to improve. One of the most innovative approaches in the language acquisition field is an immediate ultrasound feedback paradigm.

Ultrasound imaging can be used to monitor the speaker's articulatory gestures of the tongue during speech and can provide the learner with immediate feedback. Thanks to this procedure, the speaker is able to see the exact position of their articulators and they can observe details and differences in the tongue position in problematic sounds which could possibly facilitate learning foreign language pronunciation acquisition. One of the successful phonetic training studies using ultrasound feedback on Japanese learners of French made by Pillot-Loiseau et al. (2015) revealed a significant improvement of vowels considered to be very problematic for the Japanese participants. The study and its results will be further discussed below. Ultrasound imaging is, however, especially for its cost, used mainly in speech therapies and medical research. Bernhardt et al. (2005) made a medical-phonetic study focusing not only on training for people with hearing impairment or speech impairment but also on people with accented English. In both the medical and accent training, ultrasound imaging proved to be beneficial when improving speech production for people with hearing impairment or accented speech. Bernhardt et al. (2005) also demonstrate several advantages of the used ultrasound feedback, one of them being the possibility of the view of the tongue from various perspectives. Bernhardt also states that unlike other types of technologies used nowadays to display the articulators' movements during the speech for the participant, the ultrasound feedback is not so invasive when it comes to the human body. However, as for every type of training feedback, ultrasound imaging also has its major flaws, among the most striking the time inefficiency (compared e.g. with the correct/incorrect feedback discussed above) or the inability to give the "tongue-palate contact information [or] acoustic information" (Bernhardt et al. 2005, 615) as well as the tip of the tongue which is not very visible. The limitation of the tongue-palate difficult view could be eliminated by providing additional information about acoustics and the tongue-palate position on a separate screen coming from palatography.

A more typical and less expensive approach when training production of an L2 is visual feedback given to a participant that includes the acoustic information of their attempt and the same information about the sound that is imitated. This type of feedback usually analyses the sound properties as the resonating frequencies and displays them on a screen. This approach was used in the study by Kartushina et al. 2015 and is the closest to the method used for this thesis.

4. Production and perception training studies

There are many existing studies that could help us get an overview of the most and least successful methods used in L2 pronunciation training. In this section we will try to review some of these studies and demonstrate what they had in common, the methods they used especially in terms of the type of feedback, and most importantly, what findings they brought. Perhaps by comparing the advantages and disadvantages of these studies, we can predict the results of our study and discover what difficulties our participants might have. We can assume that speakers with different L1s will have problems with different sounds according to many factors, such as to what extent the L1 and L2 sound inventories differ and to what extent their L1 and L2 target sounds overlap. These factors and their eventual advantages and disadvantages will be discussed in chapter 7. By discussing the model theories predicting difficulties we will be able to evaluate what English vowels will be expected to be difficult for Czech learners of English.

5. Kartushina et al. (2015)

As Kartushina et. al’s (2015) study is the main inspiration for the present paper, the description of its procedure and findings requires a more detailed focus. As mentioned above, Kartushina et al. (2015) focused primarily on the production training and also on the production of vowels, which is very relevant for our study too. The participants of the study were French speakers who had no experience with the Danish language, which will differ from our study in the sense that our participants will be advanced learners of English. However, Kartushina et al. mention that Danish vowel inventory is much larger (specifically 16 monophthongs, moreover with a very specific distribution) while French has only 10 vowels. The French speakers were trained on those vowels which are not included in their vowel inventory. Apart from the participants’ proficiency in the trained language, we can therefore say that the situation for our participants will be similar because they will be training vowels of a language with a larger vowel inventory than their native one.

The participants of Kartushina et al's study underwent a production training of 4 Danish vowels with a trial by trial visual articulatory feedback of their own F1 and F2 information based on the acoustic analysis of a given vowel alongside with the same parameters of the native speaker of the same sex as theirs. The focus of the training was isolated vowels only. These vowels were extracted from native Danish speakers' real and non-existing words which were in an ideal environment necessary for vowel clarity. There are two main differences between our and the French-Danish study recordings. The recordings we used are downloaded from an online dictionary Outlook.com and the tokens that we are training are whole words, not isolated vowels only as in Kartushina et al's study. As the study admits, however, regardless of the final results of the training, the usage of the trained vowels in real life and context (in real words, in the presence of other consonants and vowels), might not be as good as the imitation of isolated vowels. As Kartushina et al. (2015) also include at the end of their study as a possible disadvantage of their training procedure, the isolated vowels training, however successful, does not show its effect in the subsequent real speech. And as isolated vowels in general hardly ever appear in real speech, we consider training the vowels in the context of real words more beneficial for our participants. Another reason for this choice is that English vowels are subordinated to various allophonic processes and therefore might behave differently in each given context and can be for example lengthened when in an open syllable or closed by a voiced consonant or shortened when closed by a voiceless consonant (Ladefoged & Johnson 2014). There are also several other differences depending for example on the stressed or unstressed position of a syllable containing a given vowel, however, our study eliminates all words containing more than one syllable.

Kartushina et al's study also use a control group apart from the experimental group. This procedure is important to see if there are any differences between the two groups in the final test. The control group received a similar training after the pre-test, but unlike the experimental group, only received an NS's parameters and after each block, they were able to see a percentage of the accuracy only of their F1. The control group did not benefit from this impoverished training and therefore confirmed that only trial by trial corrective feedback with both F1 and F2 position information of a native speaker as well as of the participant is crucial for non-native vowels production learning. The production training was divided into five 45-minute sessions, each divided into 5 blocks and these blocks were further divided into 5 mini blocks.

Another difference between our study and the study by Kartushina et. al (2015) is the experience of the participants with the language they are being trained in. Kartushina et al. (2015) used exclusively participants who had had no experience at all with the Danish language. However, our participants will be students of English enrolled in the Bachelor program and therefore are expected to have gained a certain level of proficiency in English. Although Kartushina et al. (2015) state that some of the vowels trained in their study were misinterpreted as their closest French counterparts, they still were training a completely unfamiliar language which might be an advantage if some of the vowels were also unfamiliar to them. What this study might therefore also bring as an adjacent benefit is to determine whether this kind of training helps students of English with the production of vowels they have been taught for many years and that are therefore probably longer in their memory and more difficult to improve any more.

The research found that the accuracy of the participants improved on average by 17 %. These results were based on the pre-test and post-test outcomes which were measured by the Mahalanobis distance from the position of the parameters of a native speaker. The Mahalanobis distance was used instead of the usual Euclidean distance because it should take into account also the natural variable of native speech. Despite the general percentual improvement in this study, the results of each individual speaker also differed a lot and a participant who improved in the production of one vowel was not necessarily the one who improved in another. It was also discovered that the vowels which were found to be the most difficult for the participants in the pre-test experienced the biggest improvement, while the vowels which were closer to the NS did not benefit from the training as relevantly. These results were not the same for the Danish vowel /y/ which was produced poorly in the pre-test and did not experience such major improvement. This was explained utilizing the SLM as the Danish vowel /y/ had been poorly categorized as its French equivalent /y/ which is however phonetically different from the Danish one.

What also differentiates Kartushina et al's study and our present research is that they have also included an ABX discrimination task as a part of the pre-test and post-test to demonstrate whether this production training also influences the perception of the trained vowels. The perception of the Danish vowels also perceptually improved, but not so significantly as the production. This proves that the focus of the training is crucial and one will probably not benefit from production training if their goal is gaining proficiency in the perception. The vowels which had been poorly produced before and after the training (as in the example of Danish vowel /y/) were also not necessarily those which were poorly perceived. The correlation between the improvement in perception and production was therefore not very significant in this study.

Despite the relevant improvement in the production of Danish vowels, Kartushina et al's study admit that there is still some space for further investigation since the production accuracy in the post-test is still not native-like. As one of the possible explanation for this, they proposed that the recordings used for this study fall into the category of Low Variability Phonetic feedback as opposed to High Variability Phonetic feedback which has been proved to be more beneficial for the production accuracy learning (e.g. Bradlow et al. 1997).

5.1. Research questions

As the training method used in the study by Kartushina et al. (2015) had many similarities with the training method proposed in this study, we could apply their findings to create the main research questions for the current thesis. The following research questions were made:

1. Does this kind of training work for Czech speakers of English, specifically on the English vowel pairs /u/- /ʊ/, /æ/-/ɛ/, /ɒ/- /ɔ/, /ɑ/-/ʌ/?
2. Does any potential improvement during the training itself transfer to subsequent speech production?

6. Other production and perception training studies

In the following literature review, I would like to start by mentioning firstly several perceptual training studies. Because our study focuses on the production training, we will further move on to that, trying to go from the perceptual training studies to examples from production training studies that are more related to our study.

6.1. Perceptual training

A study by Bradlow et al (1997) focused on the perception of the /l/ and /r/ sounds by Japanese speakers of English and showed that there is a significant correlation between the perception auditory training and the improvement of words that contain English /l/ and /r/ sounds.

Bradlow et al (1997) used the High Variability Phonetic Training (HVPT) paradigm for the English /l/ and /r/ sounds perception training which means that more native speakers were recorded for the study in order to expose the participants to a more naturalistically varied material and improve the training effect. This study is therefore significant for our paper because we are also using the HVPT paradigm. The HVPT has been already tested in more studies (e.g. Lively et al. 1993) and usually brings better results with respect to the trained subject than training with materials from a single model speaker. The HVPT usually consists of training based on the application of the trained sound or sounds in different phonetic environments and focuses on the different perceptual cues of a given sound in order to teach the participant to differentiate the L2 sounds in a native-like manner and not based on their L1 categories or cues (Lively et al. 1993, Bradlow et al. 1999). The results of the improved production in Bradlow et al.'s study differed for every speaker. The study included an examination of two tests of generalization. The first of these tests of generalization concentrated on novel words in minimal pairs spoken by a new speaker (a speaker that the participants had not heard in the stimuli before) and the second test used words that were produced by a speaker the participants had heard in the stimuli. The pretest of both the control and the trained group revealed a similar level of perception accuracy but the post-test was significantly better in terms of the perception accuracy of minimal pairs in the trained group than the control group in all three tests (the main test and both generalizations). The tests of generalization will not be included in our study although they might be very useful in putting into further practice what has been learned during the training by the participants. However, as the computer in our study will monitor the imitation accuracy trial by trial, the tests of generalization are not necessary.

The /r/ and /l/ contrast was not investigated and tested only in the study by Bradlow et al. (1997) but has been further investigated in other studies (Logan et al. 1991, Lively et al. 1993) because the liquid consonants /r/ and /l/ have proved highly problematic for Japanese speakers of English especially in terms of its differential cues which are by Japanese speakers weighed respect to mainly concerning its F2 quality while the most distinctive cue for the native speakers would be the F3 formant.

6.2. Production training

One of the studies that focused on the production training of vowels was made by Pillot-Loiseau et al. (2015). This study trained seven Japanese speaking participants to produce the French /y/-/u/ contrast, vowel /y/ being more of a closed front vowel while /u/ a back vowel (Pillot-Loiseau et al 2015). These two French vowels represent a phonemic contrast causing trouble to many Japanese speakers of French, especially in terms of the front-back position represented by F2 (Pillot-Loiseau et al 2015). The training consisted of a twelve-week-long vowel training and for the experimental group, it was accompanied by three training sessions with immediate ultrasound visual feedback. Each of these additional training sessions was 45 minutes long. The results of the study were highly variable with some surprising conclusions. The French vowel /u/ had been considered at first to be highly problematic to acquire for Japanese speakers of French but eventually the same vowel was approached to the NS´s (native speaker) accuracy more than the front vowel /y/. The three additional ultrasound training sessions were focused more on the /u/ vowel concerning this assumption and the time spent on its training might have had a direct influence on the results. Based on this study, regardless of how difficult the English vowels may be for the Czech speakers of English, especially considering the difference of the quantity of English vowel inventory, we may not be so sure in predicting the final results of our study because as the study by Pillot-Loiseau et al. (2015) shows, there is always more room for improvement especially as for the vowels considered to be difficult at first. This supports the idea of SLM (Flege 1987, 1992, 1995) which states that unfamiliar sounds are easier to learn while sounds that are close to their L1 equivalent are usually wrongly assimilated to the L1 existing category.

The best results of Pillot-Loiseau et al.'s study in terms of achieving native-like pronunciation of French /u-y/ contrast were observed on isolated vowels, and were better in monosyllabic words than disyllabic ones and also revealed that non-words were the most difficult to produce by the participants. Because in our study the participants will not be training isolated vowels so that they would be given the HVPT paradigm, the expected results would not have to be so successful. However, we are not going to test non-words but only real English words in order to train the participants English vowels in their natural context and we are also training exclusively monosyllabic words, no disyllabic ones as to avoid differences between the vowel quality due to the position in a stressed or unstressed syllable. Based on Pillot-Loiseau et al.'s study and also based on considering the fact that in monosyllabic words there is still enough room to concentrate on a restricted number of elements, we may predict the results of our study may also be quite successful.

As will be discussed in the section 9 concerning the description of Czech and English vowel inventories, we will know that probably the most relevant difference between the English vowel inventory and Czech vowel inventory is the number of vowels contained in them. Czech has comparatively fewer vowels, specifically only 5 vowels which are differentiated from the other set of 5 vowels by a length with the exception of the vowels /i/ and /ɪ/ which have been proved to differ mostly qualitatively rather than by means of duration (Podlipský et al. 2009). Skarnitzl & Volín (2012) also proved a different quality between the back vowels /u/ and /u:/, however not so significant as for the front vowels /i/ and /ɪ/ which are for the same reason represented by different IPA symbols. As for the number of vowels, we can notice that for example in Spanish there is a similar situation as Spanish has fewer vowels than English. There are also only 5 vowels in the Spanish vowel inventory which makes Spanish quite similar to Czech. Because Spanish production training studies have been made more widely than Czech production training studies, I would now like to review some of them.

A study by Flege (1988) focused on one Spanish participant with very little experience with English who trained the production of English /i/ and /ɪ/ and /æ/ and /ɑ/ vowels differences using visual feedback which focused more on the visualization of the tongue to palate position and the target places of the tongue positions needed in order to produce the /i/-/ɪ/ and /æ/-/ɑ/ vowel contrasts. The results of a pre-test revealed that the Spanish participant neutralized the differences between the contrasting vowels targetted in this paper. Only after ten minutes of training using visual feedback and a device called glossometer which should measure and show you a location of the tongue on a computer, there was a noticeable improvement in the subject's differentiation of the /i/ and /ɪ/ vowel contrasts, however, the results in the production of /æ/ and /ɑ/ contrasts were still neutralized. The author speculates whether the unsatisfying results of the /æ/ and /ɑ/ vowels were caused by a limited time of training or because this kind of visual feedback does not allow the subject to see properly the difference in the pharyngeal width.

As mentioned, Flege (1988) trained a Spanish subject. There are many other studies that concentrate on the same vowels as Flege's study for Spanish speakers of English (Aliaga-García 2009, Iverson et al. 2009) because of the large difference between the number of vowels in Spanish and English vowel inventories. Also, based on the SLM the perception of the vowels which have an equivalent in L1 could, unlike those which are completely unknown, cause difficulties for L1 speakers to perceive. As for production training, SLM states that if an L1 speaker perceives L2 sounds correctly, it does not necessarily mean that they will produce these sounds in a native-like way. However, if the production of L2 sounds is correct, the perception is necessarily correct as well.

If we take into consideration a low front vowel /æ/ which is neither in Spanish nor in Czech vowel inventory it can be predicted that it would be more difficult for Czech or Spanish speakers to produce it. There are studies that focus on wrong identification of this vowel (e.g. Šimáčková 2003) and thus prove that it may be a source of difficulties for learners who do not have low front vowel /æ/ in their vowel inventory. The perception of the low front vowel /æ/ by a Spanish L1 speaker and on the other hand by a Czech L1 speaker might also differ a lot, according to which vowel is the front low vowel /æ/ closest to and thus to which L1 category it will assimilate. The perception of a given sound is influenced by many factors, however, among one of the most obvious the time of the participant's interaction with a given language and also for example the frequency of the learned sound in the L2 lexicon (Best & Tyler 2007).

We could predict that both Spanish and Czech L1 speakers might have difficulties with the production of the vowels /ɛ/ and /æ/. The Czech production should be closer to the English /ɛ/, considering for example the fact that most of originally English words including front low /æ/ are usually taken over to the Czech language creating words that actually do consist of Czech vowel /ɛ/, rather than their original sound /æ/ as in the words *'match'* or *'backhand'*, originally pronounced as /*mætʃ*/ and /*bækhænd*/, but when taken over to Czech, usually pronounced in Czech as /*mɛtʃ*/ and /*bɛkhɛnt*/.

A major discussion has been made trying to answer the question about what effect a different L1 vowel inventory has on L2 sound learning. One of the most arguable is the question about whether the fact of not having an overlapping counterpart of a vowel in an L1 vowel inventory and the L2 which is being learned is an advantage or a disadvantage in learning a foreign language pattern. There are several basic theories that take into consideration this issue.

7. Models predicting the difficulty of L2 acquisition

7.1. Speech Learning Model (SLM)

Flege (1995, 2003) and his Speech Learning Model (SLM) claims that there is evidence that not having an equivalent counterpart in a vowel category in an L2 is an advantage in learning new vowel categories. According to Flege (1995, 2003), the whole vowel system works for all languages based on one common phonological space which includes all the speech sounds from an L1 inventory as well as the learner's L2. When an L1 sound is close in the learner's perceptual space to an L2 sound, it can be predicted to be more difficult for an L1 speaker to learn because the learner is more likely to equate the L2 sound with the corresponding L1 sound. This could lead to the question of whether a small L1 vowel inventory is an advantage in learning an L2 because if the L1 vowel inventory is small, there are fewer possible counterparts of vowels in L2 and the acquisition of the new categories should be easier to learn. The SLM makes predictions about the ultimate attainment, i.e. an experienced learner's final achieved level, not initial stages, in comparison with another difficulty predicting model discussed below, the Perceptual Assimilation Model (PAM) made by Best (1993, 1995) which is by contrast better applicable during the initial stages of L2 acquisition and where a larger L1 inventory predicts fewer problems.

The SLM takes into consideration how L2 learners perceive the phonological aspects of a sound rather than its articulatory quality while in contrast, PAM (Best 1993, 1995) focuses on the articulation. Also, there is a very important factor that might make the SLM useful for this study more than other models and that is the fact that it takes into consideration speakers of an L2 who have a large experience in this language. SLM does not concentrate on novice speakers of an L2. Because in our study the participants are highly experienced students of English who actively work on their language abilities, we can assume that the rules defined in the SLM will work. On the other hand, SLM is primarily about L2 learners in natural settings, i.e. the country where the L2 is spoken, nevertheless, our participants are experienced but have learned their English via formal instruction and will have only limited experience of living in English-speaking countries. These factors might actually fit the best into a Perceptual Assimilation Model (PAM-L2, Best & Tyler 2007) which is discussed in the section 7.3. Another issue that is discussed in the SLM is the age of an L2 learner. Essentially, the older the participant is, the more problems we could anticipate in their L2 pronunciation learning. The participants in this study are university students so essentially we will only recruit young adults.

7.2. Perceptual Assimilation Model (PAM)

Another model predicting difficulties in L2 sounds acquisition is the Perceptual Assimilation Model (PAM) made by Best (1993, 1995). The PAM is also focused on the cross-language perception and agrees with the SLM (Flege 1995, 2003) in many ways, for example in terms of the easier acquisition of a new L2 sound which is not close to any L1 vowel category. Only PAM is not applied to speakers who have already managed to acquire proficiency in an L2 but to speakers who have very little or no experience at all with the new sounds that they are learning or with the language itself. Unlike SLM, PAM (1993, 1995) besides the phonological aspects of sounds takes into consideration also their articulatory qualities by promoting the theory of speech perception called Direct Realism. This theory, discussed e.g. by Fowler (1986), posits that when listeners are perceiving speech, what their brains perceive are not the sounds defining particular phonological categories but the gesture needed to pronounce them. Based on this perception theory, a listener may need to imagine how their own articulators would need to move to produce the encountered sound in order to identify a phoneme represented by this sound. The content of the phonemic representation in the mind is therefore defined in terms of the articulatory gestures needed to realize the sound in speech.

PAM (1993, 1995) defines specifically how the efficiency in production or perception of a new sound could be expected to be assimilated to an L2 learner's L1 categories. According to PAM, when two different new sounds assimilate (i.e. become categorized by the learners as belonging) to only one speaker's L1 vowel category, the production and perception of these two new sounds will be likely to cause problems. When two different new sounds assimilate to different L1 vowel categories, the perceptual discrimination of the L2 sounds should be very easy, however, their production may still sound accented since the learner essentially transfers the L1 contrast into the L2 resulting in pronouncing the sounds in an L1-like way.

PAM (1993, 1995) also suggests only one category assimilation in which case one category is correctly perceived but the other one is uncategorized. In this case, the results of correct discrimination of the L2 sounds are predicted to be good. However, both categories can also be unassimilated and that essentially leads to a non-predictable discrimination accuracy that could range from very good to very bad. A quite special case would emerge if the new sounds are not even recognized as speech and therefore can not be categorized.

7.3. Perceptual Assimilation Model-L2 (PAM-L2)

The Perceptual Assimilation Model-L2 (PAM-L2, Best & Tyler 2007) is a difficulty predicting model made as a continuation of the earlier PAM discussed above. The PAM-L2 is also a cross-language perception study. The first major difference between the earlier PAM and the later PAM-L2 is their application on different types of learners. While PAM studies a cross-language perception of sounds of a foreign, unknown language or at least a language of which the speaker is not very proficient, the PAM-L2 is its extension for later stages of L2 speech learning especially in terms of the broadened knowledge in the phonetics and phonology of a given L2. This condition might naturally bring better results for the L2 speakers' perceptual discrimination of L2 categories as they have already gained a certain level of proficiency in the particular L2 since they have probably already managed to create new categories within their L1 while learning the L2. As mentioned above, PAM-L2 is based on the previous PAM, and thus it uses the basic perception assimilation categorization described in the PAM as a tool for predicting difficulties of L2 learners in the perception of L2 sounds. If an L1 category of these speakers works well enough to enable them to recognize the L2 category contrast, they may not need any further training. However, if the speaker's L1 category does not suffice the category contrast in a given L2, further perceptual training might be needed.

While both models discussed in this paper so far focused prevailingly on one specific framework, either phonetic, phonological or articulatory, PAM-L2, by stating that "L1–L2 differences at a gestural, phonetic, or phonological level may each influence the L2 learner's discrimination abilities, separately or together, depending on the context or the perceiver's goals" (Best & Tyler 2007, 25) clearly admits there is a correlation between all of these frameworks and each has its importance in perceptual determination studies depending on what specific contrast is the goal of a given perceptual study. However, it is implied that when it comes to determining the category of contrasting pair of sounds the phonological information appears to be the most relevant.

8. General characteristics of vowels

In the following section of this paper, I would like to dedicate a few words to the comparison of Czech and English vowels. Before that it seems appropriate to briefly discuss general characteristics of vowel sounds in general.

Vowels are by Ladefoged & Johnson (2014) described as sounds that are produced using vocal cords. In comparison with most consonants, when producing vowels, the vocal tract is not obstructed in any way and the airflow is, therefore, consistent. Because the vocal cords are always vibrating when a vowel is being produced, there is no distinction between voiced and voiceless vowels, as it is for consonants, and vowels are always voiced. Given all these descriptions, it is clear that vowels are the highest on the sonority scale, followed by semi-vowels and consonants.

The quality of a given vowel is always determined by two basic factors. First of them being the tongue position which divides vowels into front and back and the openness of the mouth which would divide vowels into open and closed. The minor articulatory difference lies also in the position of the lips which usually divides vowels into rounded and unrounded.

9. Comparison of Czech and English vowels

Because this paper is a study focused on the training of English vowels on Czech speakers of English, we will first have to describe the basic properties of both Czech and English vowel inventories and demonstrate their differences. We will also try to predict which vowels will be the most difficult to acquire for the Czech participants based on previous studies.

Cruttenden (2014) describes that Standard Southern British English consists of 12 monophthongs. Based on the position of the tongue we can divide English vowels into front, central and back vowels. There are four following front vowels in English: /i/, /ɪ/, /ɛ/ and /æ/ as in the words /*neat*/, /*hit*/, /*bed*/ and /*bad*/, respectively. Back vowels /u/, /ʊ/, /ɔ/, /ɒ/ and /ɑ/ could be found in the words /*mood*/, /*bush*/, /*fork*/, /*lot*/ and /*car*/, respectively. English central vowels, namely /ɜ/, /ʌ/ and /ə/ appear in the words /*word*/, /*duck*/ and /*alone*/. The following table presents a vowel chart of English vowels.

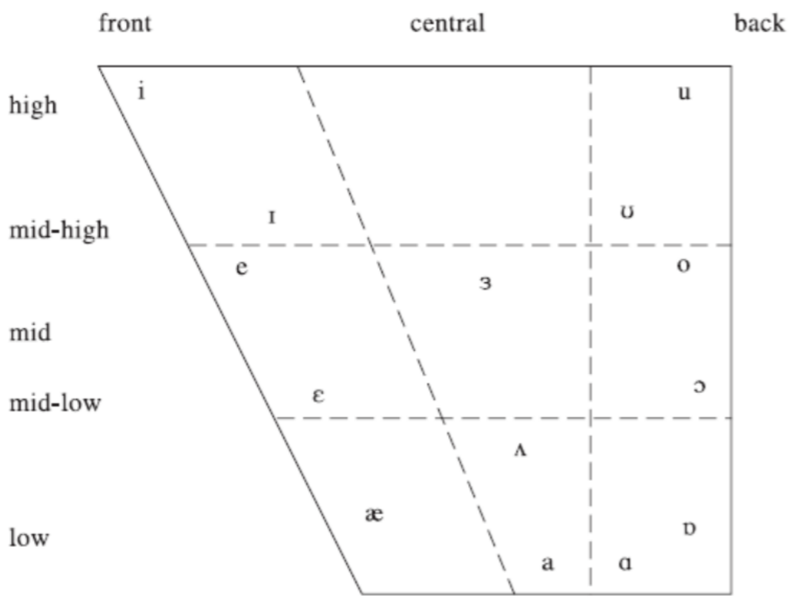

Table 1:English vowel chart (Ladefoged & Johnson 2011)

As for the Czech vowel inventory, there are five vowels mostly differentiated by means of their duration. There are therefore ten different vowel phonemes. (Skarnitzl et al. 2016). The phonemes /ɛː/, /ɛ/, /aː/, /a/, /oː/, /o/, /uː/, /u/ are differentiated by their duration and they can be found in the words /*lék*/, /*med*/, /*náš*/, /*lak*/, /*tón*/, /*vosk*/, /*úl*/ and /*už*/, respectively. The Czech vowels /ɪ/ and /i:/ are an exception when it comes to the usage of different IPA symbols as they have been proved to differ mainly in their quality and not like the rest of the Czech long-short vowel pairs in terms of their duration (Podlipský et al. 2009).

The following table shows average formant values of the selected English vowels and their Czech closest counterparts. The data were borrowed from Skarnitzl & Volín (2012) for the Czech vowels and from Hawkins & Midgley (2005) for the English vowels. The latter were chosen from the data produced by speakers between 20-25 years of age as our participants will also be young adults.

	English			Czech	
vowel	F1 (Hz)	F2 (Hz)	vowel	F1 (Hz)	F2 (Hz)
/ɛ/	600	1914	/ɛ/	566	1519
/æ/	917	1473	/ɛ:/	576	1578
/ʊ/	413	1285	/u/	359	937
/u/	289	1616	/u:/	304	769
/ɒ/	484	865	/o/	458	1055
/ɔ/	392	630	/o:/	484	1028
/ɑ/	604	1040	/a:/	699	1206
/ʌ/	658	1208	/a/	649	1286

Figure 1:F1 and F2 values of the target English vowels (Hawkins & Midgley 2005) and their Czech counterparts (Skarnitzl & Volín 2012)

The Czech and English vowel inventories differ not only in the number of vowel categories but in terms of the features which are crucial and distinctive for each language. Although in some particular cases it may seem that a vowel from one language has an overlapping equivalent in the second language, they may be actually different and we will hardly ever encounter one vowel from Czech vowel inventory identical in its all acoustic properties to another vowel from the English vowel inventory. There are also different cues according to which a certain vowel is differentiated from another one. There have been many studies that proved that quantity (duration) is not a phonologically relevant feature for distinguishing vowels in English language vowel inventory. A study by McAllister et al. (2002) claims that many languages including English rely much more on the qualitative features rather than the quantitative features. The general distinction of English vowels into peripheral /tense/ and non-peripheral /lax/ is mainly a spectral differentiation, however, there can also some durational difference (Ladefoged & Johnson 2014). Furthermore, the lax-tense differentiation is said to be based more on the historical development of English (Ladefoged 2014). The durational differentiation seems to be even more irrelevant considering that there are differences even within the lax-tense groups as for the lax vowel /æ/ which tends to be longer than its other lax members (Wells 1982).

As the durational feature is used in Czech for making vowel quantity, we might assume that Czech L1 speakers will probably assimilate the lax-tense differentiation of English vowels into their L1 short-long vowel pairs. Given this information, although we have already mentioned that the lax-tense vowel distinction is mainly spectral rather than durational, the fact that the English lax /æ/ has some level of durational difference with other lax vowels, we may assume this might be an advantage for Czech learners of English when training the vowel pair /æ/ and /ɛ/ since the latter is a member of other lax vowels. As lax vowels have their own specific distribution and do not appear for example in an open syllable, in this paper we have chosen only words that contain coda (that are closed by a consonant).

Standard Southern British English used and practiced in this paper contains eleven vowel monophthongs (Cruttenden 2014). Although there is a certain differentiation which divides English vowels into two groups, namely lax and tense, lax vowels being /ɛ, ɪ, ʊ, ʌ, ɒ, æ / and tense being /i, u, ɑ, ɔ/. However, as mentioned above, duration as a differential cue for English vowels is not the main cue for native speakers because its usage is varied depending on many factors. Klatt (1976) mentions many contexts in which the duration of vowel changes, the position before a voiced or a voiceless consonant being one of them, followed by a position of a vowel in secondary or primary lexical stress or lengthening of a vowel when at a word boundary position. Volín (2002) supports the idea of unstable stress by saying that "The length of a vowel in English depends on whether it is followed by a fortis or lenis consonant, whether it is in a stressed or unstressed syllable, whether it is in a monosyllabic or polysyllabic word and also, whether it is in a sentence-final stress group or not." (Volín 2002, 14). This statement makes it clear that the same English vowel can have different duration in different contexts. If we consider the distinction between the fortis (voiceless consonant) and lenis (voiced consonant) in the example such as *bat* and *bad*, we can see that the vowel in both of the words is the same, however, when followed by a fortis the vowel is pronounced more briefly and is lengthened in *bad* when followed by a lenis. This example is sufficient if we want to demonstrate that duration cue is in most cases not a distinctive feature of vowels in English.

The Czech vowel inventory includes ten vowel monophthongs forming five pairs contrasted by quantity (Skarnitzl et al. 2016) with the exception of the vowel pair /i/- /ɪ/ which is differentiated by its quality (Podlipský 2009). This information is supported by the findings made by many studies (Šimáčková & Podlipský 2018, Šimáčková 2003) which demonstrated that Czech speakers of English vowels often rely on the durational differentiation where native speaker´s cue on the differentiation would be based on the vowel quality. The quality of two Czech contrasting vowels as for example in the vowels /a/ and /a:/ is mainly represented in its duration, however, Skarnitzl & Volín (2012) found that the long /a:/ has changed for young speakers of Czech also in quality as it has become more fronted. Hence, we might expect that for example, the pronunciation of the vowel pair in the words *bet* and *bat* will be produced as [bɛt] an [bɛːt] and not as [bɛt] an [bæt] as it would be in the speech of a native speaker of English or similarly, for example, the difference between the pronunciation of the vowels in the back vowel pair /u/- /ʊ/ will also be most likely differentiated in terms of the vowel length rather than its frontness and backness. There are studies which still state that the vowels in both of the examples of vowel pairs mentioned above /æ, ɛ/ and /ʊ, u/ are even by native speakers of English sometimes differentiated not only by means of their quality but also quantity (duration) (Ladefoged 2014, Wells 1982, Volín 2002).

Based on research that has been done (e.g. Šimáčková 2003) we may assume that one of the most problematic vowels in our study for the Czech participants will be the low front vowel /æ/. As mentioned above, not only is this vowel an exception between English lax vowels being differentiated by its duration (which might be a help for the Czech L1 speakers but also a little confusing when compared to other lax vowels) but also it does not have any real correspondent in Czech vowel inventory. The closest vowel quality for the English low front /æ/ is could be the Czech the mid-low front /ɛ/ (Volín 2002) which means it may cause difficulties when produced by non-native speakers of English as they could classify the L2 /æ/ as a counterpart of the L1 /ɛ/.

As for the vowels that we expect to cause troubles to Czech participants producing English vowels, we might assume not specifically /æ/ itself, but all the vowels occurring in the open region of English vowels which is not filled in the Czech vowel inventory and therefore the vowels /ʌ, ɑ, ɒ/, for which we would not find equivalents in the Czech vowel inventory (Skarnitzl & Rumlová 2019), might also be a source of problems. The latter study also confirmed results of the study by Šimáčková & Podlipský (2018) with the open front vowel /æ/ as being the problematic vowel for Czech speakers and another vowel, also from the open region /ɒ/ being realised by a majority of the speakers as the Czech counterpart /o/.

9.1. English problematic vowels for Czech speakers of English in other studies

A study by Šimáčková & Podlipský (2018) recorded Czech advanced English L2 learners and examined vowels that are said to be problematic even for these speakers. Because the participants were all university students of English motivated to get better in the field they study and want to work in and the participants of this study are also enrolled in English language program we can expect similarities in the results as they are also active learners of English. The participants of our study are also surrounded by

Czech speakers but they are constantly working on improving their English proficiency at university.

Šimáčková & Podlipský (2018) focus on the limitations on L2 native-like production caused by daily exposure to the Czech environment and the fact that the participants started learning the English language in a later phase of their childhood which increases the likelihood of L1 transfer. The paper also suggests that the participants are very proficient in English language production and perception due to their long-term previous and current study of English at university. The question of whether or not it is an advantage to study English at university at the time of training the native-like pronunciation is mentioned not only in study by Šimáčková & Podlipský (2018) but also for example in the study by Šturm & Skarnitzl (2011), a paper which concentrates on the English low front vowel /æ/ by Czech learners of English, although not in terms of its production as much as its perception. However, Šturm & Skarnitzl (2011) revealed that more experienced Czech learners of English (in particular students of English who had had a basic introduction to the field of phonetics) showed better results in the perception accuracy than those who did not have such experience.

The study by Šimáčková & Podlipský (2018) examined three English contrasting pairs of British vowels, two of which (/æ/ vs /ɛ/ and /u/ vs /ʊ/ are included in the present study. In addition, Šimáčková & Podlipský (2018) also examine the contrasting vowel pair /i/ and /ɪ/. The recordings made by Šimáčková & Podlipský consisted of whole sentences with the target word with the required vowel at the beginning of the sentence and then in its end. An interesting observation was made on the vowel pair /æ/and /ɛ/ where even highly experienced university L2 learners of English often interchanged the vowels /æ/and /ɛ/ in the words containing them. The front mid-low vowel /ɛ/ was often produced with a higher F1 typical for the low front /æ/ and the vowel /æ/ was realized with a lower F1 more typical for the native /ɛ/. These vowels were also the most difficult for the learners in terms of native-like production. As expected for example by the study by Šimáčková (2003), the vowels /æ/and /ɛ/ were mostly differentiated by their duration where the vowel /æ/ was presented as the longer vowel. However, the durational differentiation of vowels was not only observed on the vowel pair /æ/and /ɛ/. A similar situation was also present for the back vowel pair /u/ and /ʊ/ which was also contrasted by its duration rather than its quality, with better results for /ʊ/ in terms of native-like production. Therefore, as our participants will be also highly experienced in English as well as in the study by Šimáčková & Podlipský, we might expect better results for the high back vowel pair /u/ and /ʊ/ than for the low front /æ/ and /ɛ/ which is problematic even for advanced Czech learners of English and that the participants will rely on their L1 vowel categories in production of new English L2 vowel categories.

10. Vowel contrasts selected for training

In the training part of this study, the participants will be training the following pairs of English vowels. Front vowel pair /æ/-/ɛ/ spectrally differentiated by means of their height, back vowel pair /ɒ/- /ɔ/ also differentiated by its height, high front /ʊ/- /u/ contrast differentiated by the contrast lax-tense and a pair of a central /ʌ/ with low back /ɑ/.

The vowels have been chosen mostly according to the assumptions about their difficulty for Czech learners of English with the focus on the vowel pair /ɛ/ and /æ/ as being the vowel pair that proved challenging to differentiate correctly in other studies (e.g. Šimáčková 2003) which showed that while the perception of the vowel /æ/ is relatively correct and so is the vowel /ɛ/ except when in the position before final voiced obstruent, the production of the vowel pair /æ/ by Czech learners of English could qualitatively be mistaken by a native listener with the vowel /ɛ/, as the English front low /æ/ is with reference to its F1 and F2 values somewhere between the Czech vowel /ɛ/ and /a/.

For the vowel pair /ʌ/-/ɑ/, the central /ʌ/ is in terms of its F1 and F2 values closest to the Czech /a/ which is also created with the tongue in the central position. It is possible that the Czech learners of English have perceived this vowel as its Czech counterpart /a/ and therefore its production might be closer to the Czech /a/ rather than the native central /ʌ/. However, if we compare the spectral values of the low back English /ɑ/, we probably would not find a closer Czech counterpart than the Czech /a/ as well.

The vowel pair /ʊ/ and /u/ that could also be considered problematic for Czech learners of English as the lax vowel /ʊ/ is not in the Czech vowel inventory and the two Czech vowels /u/ and /u:/ are differentiated by their duration. Šimáčková & Podlipský (2017) revealed, however, that Czech speakers of English are able to approach the native-like pronunciation by a certain level of fronting. For the vowel pair /ʊ/-/u/ we are going to investigate based on the SLM whether the participants are going to rely on their existing L1 distinctive feature of short and long vowels rather than their spectral differentiation.

Concerning the vowel pair /ɒ/-/ɔ/, it is possible that the back vowel /ɔ/ will be assimilated to an L1 close counterpart /o:/ and as the Czech vowel space is empty around the L2 low back /ɒ/, we suppose there is a possibility of creating a new category and the subsequent eventual native-like production of this vowel.

11. Methodology

11.1. Participants

All the participants in this study will be young adults at the time of the training enrolled in the university English philology program. In the study by Kartushina et al. (2015) the participants were given recordings produced by a native speaker of the same sex as each participant, however, for this study, there will be no such distinction. Instead, the program will use a normalized vowel space which will eliminate differences created by participants' different sizes and lengths of vocal tracts (Ménard et al. 2002).

11.2. Stimuli and procedure

For the training itself, the following words were chosen to practice by the participants. For the vowel pair /ɒ/-/ɔ/ there are the words *bought, caught, cause, core, fork, fought, fraught, lord, ought, pause, short, thought, book, bond, boss, box, chop, cop, cost, cough, dot, job, knot, not, pod, pot, top.* The vowel pair /ʊ/-/u/ is represented in the words *book, bush, cook, could, foot, good, hook, push, put, shook, should, took, wood, boot, booze,* *choose, goof, juice, poo, shoe, two.* For the front vowel pair /æ/-/ɛ/ the words *sad, apt, ash, bag, bass, bat, cap, cat, dad, fact, fat, pass, pat, trap, bed, beg, deck, egg, get, jet, peg, sec, set, sex, step, test, vet, web* were chosen. Lastly, the vowel pair /ʌ/- /ɑ/ is included in the words *buck, bug, bus, chuck, cuss, cut, duck, hut, such, touch, tuck, arse, barn, calf, cart, část, chart, fat, half, last, part, pass, past, path, shaft, tart* and *vast.* All of the words were chosen according to several criteria to receive high-quality recordings. The chosen words are always monophthongal in order to avoid a position of a vowel in a different stress position. In order to create an advantageous position (for example not distracted by noise) of a vowel, the onset of the syllables is always composed of obstruents as well as the coda.

The program named "Listen-and-repeat vowel production trainer. Version 1.2." (Podlipský 2015) will be used for the purposes of this study. There is a folder named 'vowel trainers' attached to this thesis which includes four other folders, each for one vowel pair. This folder is to be used for the training itself. Each of the four folders contains another folder called 'monophthongs' including all the words for a given vowel pair, and a file called 'v prod trainer 2 (monophthongs)- update'. By clicking this file a participant opens the software Praat (Paul Boersma & David Weenink) and the program for the vowel training and subsequently uploads all the words from the folder 'monophthongs'. After clicking the option 'run', the participant will be able to start the training.

There are different procedures that could be used in order to achieve good results in terms of the native-like pronunciation of the target vowels. The participants could be training only one vowel contrast or multiple contrasts. The training could be divided into training blocks, each block representing a specific vowel contrast, or the order of the vowel contrasts could be randomized. In this study, we will divide the vowels into blocks of vowel contrasts with the option of subsequent randomized order training.

During the actual training, the participants will get the instructions before the imitation of the first word. After that, they will hear a word, click "record" and pronounce the imitation of the word they have just heard by a native speaker. They should be able to see visual feedback of their imitation after clicking “play yours” and at the same time listen to their own imitation. They can also listen to the model recording and their imitation right after each other. After the participants’imitation it might be beneficial to let them play their recording once again to be able to realize the possible mistake immediately.

The imitation will be immediately measured and the mean fundamental frequency (F0) of the imitated vowel and the F1, F2, and F3 values of both the original vowel and the imitated one together with the number of trial will be saved in a table. The computation of the distance of the imitated vowel quality in the normalized space from the original quality in the normalized space will subsequently be done. This distance will be measured by Euclidean distance and then in case that the number of trials predicts reduced distance between the imitation and the original vowel, we will find out whether there is a correlation between the time spent on the training and getting closer to the model vowel using regression analysis. Figure 2, Figure 3 and Figure 4 show an example of visual feedback on the word “*chart*“ provided to a participant before any recording, after hearing a native speaker, and after their trial, respectively.

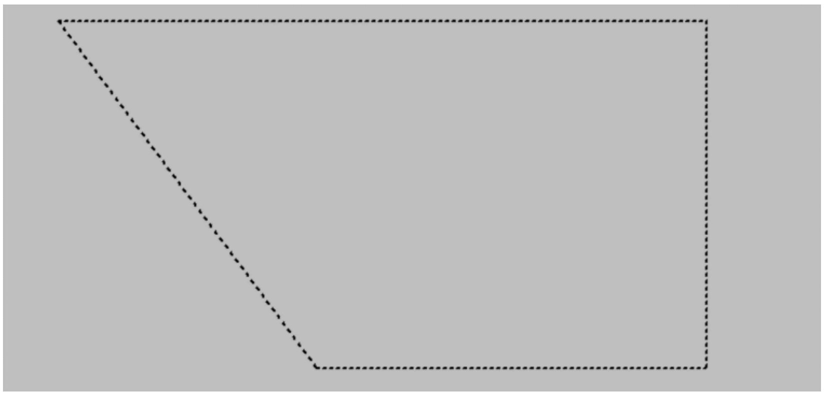

Figure 2: Visual feedback provided to participants before the NS's or the participants' trial

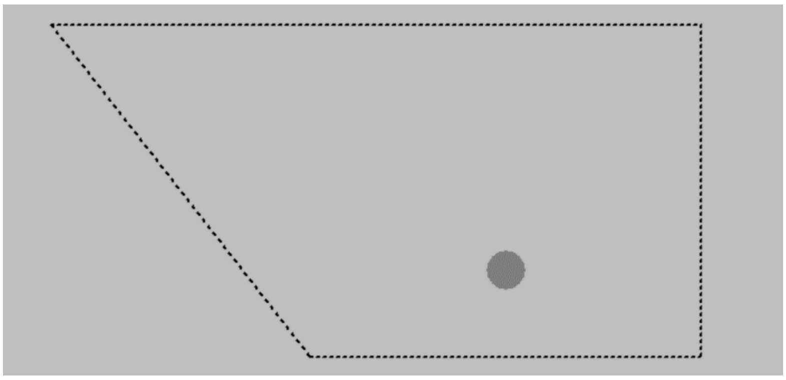

Figure 3: Visual feedback provided to participants after NS's recording

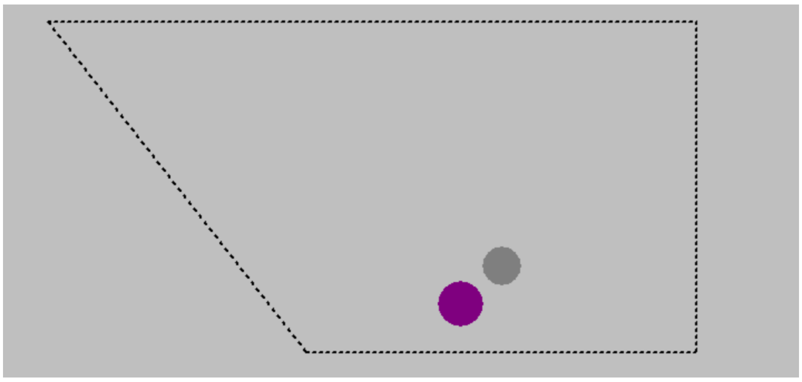

Figure 4: Visual feedback provided to participants after NS's recording and participant's trial

11.3. Measurement

The program will not measure the mean fundamental frequency and formant frequencies in hertz but in a logarithmic scale modeling, human perception of frequency called Bark scale. To compare the data with the mean frequencies of English and Czech vowels subsequent conversion to hertz will be needed. As mentioned above, the feedback uses a normalized vowel space. That means that the x-axis which measures vowel height shows F1 in Bark minus F0 in Bark. The y-axis which measures the vowel frontness or backness shows F2 in Bark minus F0 in Bark. The x-axis shows therefore the distance of the first formant from the fundamental frequency, while the y-axis shows the distance of the second formant from the fundamental frequency. This normalized vowel space is used to eliminate the difference between the speakers' pitch and the length and size of their vocal tract which affects the height of the resonating frequencies. If the participants’s F1 and F2 values are close to each other, the normalized vowel space should highlight the differences (Ménard et al. 2002). The normalized vowel space thus only displays the target feature, namely vowel quality and tries to eliminate differences between men and women's voice.

The script skips the initial 15 % of the vowel's duration and the final 15 % too, and measures the mean F0, F1, F2, F3 for the middle 70 % of the vowel, as the values F0 and the formants are usually not stable during the vowel production and for example, the F0 tends to gradually drop (e.g. Gerfen & Baker 2005). This approach was chosen because of the changes of the frequencies in the vocal tract, namely formant transitions, which occur when vowels are surrounded by consonants depending on the place of articulation of these consonants. The effect of the F0 drop caused by consonants surrounding the vowel has been showed for example in the study by Vilkman et al. (1989). The script finds the beginning and end of each vowel by first finding out which part of the signal is periodic, meaning that it has the F0 and is voiced, and then also by using Praat's automatic annotation of silent intervals. The size of the circle depicted depends on the measured intensity of the vowel in the decibel.

12. Conclusion

This thesis aimed to make a research proposal on whether a phonetic training of Czech speakers of English has an impact on their native-like pronunciation achievement of selected English vowels, namely the vowel pairs /u/- /ʊ/, /æ/-/ɛ/, /ɒ/- /ɔ/ and /ɑ/-/ʌ/.

In the theoretical part, the author discussed models which predict difficulties for L2 learners, training methods to overcome these difficulties, and then listed examples of existing studies dealing with similar training methods to compare their effectiveness to finally be able to make a hypothesis about the results for future studies. Also, a detailed description of the main source for this thesis which presents a similar training method is made altogether with its limitations and possible ways to overcome some of them.

The discussion and comparison of Czech and English vowel inventories were made and existing studies and models predicting difficulties were applied to predict which vowels could be problematic for Czech speakers of English and should therefore be trained to achieve native-like pronunciation.

The methodology part proposes a training program created for English L2 learners to be able to train the problematic vowels by means of imitation after hearing a native speaker. This part is also dedicated to the description of the vowel training program which is attached as a part of this thesis. Each vowel pair has its own folder.

Based on the literature review and comparison of existing training studies we can hypothesize that there is a possibility for the participants to benefit from the phonetic training, however, considering that the participants would be only advanced learners of English, it is also likely that the results will differ for every speaker as each of them may have got differently stable L2 categories. Moreover, the improvements may vary between the vowel pairs too. A little unpredictability arises when considering the SLM and PAM theories state that the further an L2 sound is from its closest L1 counterpart, the more precise should the perception of such vowel be, however, other studies show that even advanced Czech speakers of English experience difficulties when producing for example English low front /æ/ which does not have any real counterpart in Czech vowel inventory and therefore contradict the theories. Compared to the study by Kartushina (2015), we might expect that with both auditory and visual feedback the participants are likely to improve their native-like pronunciation more than the French participants did in the Kartushina et al.'s, as in the latter study only a visual feedback was provided to the participants. It seems that the possibility of hearing one's own voice and the imitation of the native word right after hearing the native speaker might be good when it comes to realizing the mistake and distance from the target vowel immediately.

The worst results are most likely expected for a problematic vowel pair /æ/-/ɛ/, which has proved to be differentiated in terms of its duration even by experienced Czech learners of English (Šimáčková & Podlipský 2018), as the L2 low front /æ/ is in terms of its quality in between the Czech vowels /a/ and /ɛ/. It is very likely that the misinterpretation in the production of these two vowels might change the meaning of the whole word and thus become confusing for a NS to understand. Hence, we would recommend dedicating more time for the training of this vowel pair, hopefully with the result of NS's production accuracy achievement. On the other hand, better results are expected for example for the vowel pair /ʊ/-/u/. While it might be expected that the Czech speakers would differentiate these two vowels by their duration, Šimáčková & Podlipský (2017) revealed that they are able to achieve the differentiation by fronting, which would be a success.

As for the vowel pair /ʌ/-/ɑ/, we have found out that the L2 central /ʌ/ is closest to the Czech /a/ but the same Czech vowel is the closest for the contrasting English /ɑ/. It is either possible that the vowel pair /ʌ/-/ɑ/ will be contrasted by duration, meaning that their final production by Czech speakers would be /a/ and /a:/, or a case based on PAM-L2 called One-Category assimilation might arise, meaning that both vowels from the vowel pair /ʌ/-/ɑ/ will be assimilated to the closest Czech counterpart and they will both be interpreted and possibly produced as Czech /a/.

For the vowel pair /ɒ/- /ɔ/ it is also possible that a Two-Category assimilation will take place, where the two L2 vowels will assimilate to their closest Czech counterparts, namely /o/ and /o:/.

Because the focus of our training are not isolated vowels but real English words, we can assume that the participants will be able to use the learned vowels better in real contexts which seems to be another practical advantage compared to some of the other mentioned studies. Also, some studies (e.g. Pillot-Loiseau et al. 2015) revealed that training real words is easier than training non-words. That is why we have only chosen real English words for the current training.

To answer the research question which asks whether the training transfers to subsequent speech production, an additional pretest and posttest would be required in order to investigate whether a participant is able to read the words without hearing the recording of a native speaker.

For better effectiveness of the training itself, the HVPT paradigm is recommended to use as well as more time spent on the actual training.

This research proposal is to serve as a theoretical basis for future phonetic training studies.

13. Resumé

Tato bakalářská práce si klade za cíl vytvořit hypotézu na základě existujících studií, zda imitační trénování anglických samohlásek českými studenty angličtiny s vizuální zpětnou vazbou má kladný vliv na přiblížení výslovnosti těchto samohlásek výslovnosti rodilého mluvčího.

Největší inspirací pro tuto práci byla následující studie: Kartushina et al. (2015), která se zabývala fonetickým trénováním vybraných dánských samohlásek francouzskými mluvčími bez jakékoli zkušenosti s dánštinou. Jako primární rozdíl mezi současnou prací a dánským výzkumem je tedy zkušenost účastníků výzkumu v cílovém jazyce, protože tato bakalářská práce má za cíl poskytnout hypotézu pro podobný trénink na již pokročilejší studenty angličtiny.

Teoretická část této práce obsahuje podrobný výklad modelů, které byly vytvořeny pro předpověď obtížnosti osvojení hlásek v cílovém nerodilém jazyce založených na srovnání hláskového spektra tohoto jazyka a jazyka mateřského. Dále byly popsány typy tréninkových modelů, které mají za úkol těmto obtížím předcházet na základě imitačního trénování hlásek v cílovém jazyce.

Srovnáním českých a anglických samohlásek, jejich vlastnostmi a přibližnou vzdáleností mezi nimi, byly určeny některé samohlásky, které by mohly způsobovat problémy při rodilé výslovnosti pro české mluvčí a byly tak navrženy k fonetickému tréninku. Určení těchto problematických samohlásek bylo také navrženo na základě předchozích studií, které se zabývaly podobnou tematikou a zkoumaly například, které anglické samohlásky působí největší potíže pro české mluvčí.

Součástí práce je příloha se samotným programem pro následné fonetické trénování daných samohlásek pro studenty angličtiny.

14. Works cited

Aliaga-García, Cristina, and Joan C. Mora. "Assessing the effects of phonetic training on L2 sound perception and production." Recent research in second language phonetics/phonology: Perception and production 231 (2009).

Bernhardt, Barbara, Bryan Gick, Penelope Bacsfalvi, and Marcy Adler-Bock. "Ultrasound in speech therapy with adolescents and adults." Clinical Linguistics & Phonetics 19, no. 6-7 (2005): 605-617.

Best, Catherine T. "Emergence of language-specific constraints in perception of non-native speech: A window on early phonological development." In *Developmental neurocognition: Speech and face processing in the first year of life*, pp. 289-304. Springer, Dordrecht, 1993.

Best, Catherine T. "Learning to perceive the sound pattern of English." *Advances in infancy research* 9 (1995): 217-217.

Best, Catherine T., and Michael Tyler. "Nonnative and second-language speech perception." Language experience in second language speech learning: In honour of James Emil Flege (2007): 13-34.

Bradlow, Ann R., and David B. Pisoni. "Recognition of spoken words by native and non-native listeners: Talker-, listener-, and item-related factors." The Journal of the Acoustical Society of America 106, no. 4 (1999): 2074-2085.

Bradlow, Ann R., David B. Pisoni, Reiko Akahane-Yamada, and Yoh'ichi Tohkura. "Training Japanese listeners to identify English/r/and/l: IV. Some effects of perceptual learning on speech production." The Journal of the Acoustical Society of America 101, no. 4 (1997): 2299-2310.

Clink, Kellian.,"OneLook: Dictionary Search", Emerald Group Publishing Limited, Reference Reviews, Vol. 21 No. 8,2007: pp. 35-35. https://doi.org/10.1108/09504120710838912

Cruttenden, Alan. Gimson's pronunciation of English. Routledge, 2014.

del Puerto, Francisco Gallardo, and Esther Gómez Lacabex. "La enseñanza de las vocales inglesas a los hablantes de español." *PULSO. Revista de Educación* 31 (2008): 37-66.

Finch, Diana F., and Héctor Ortiz Lira. *A course in English phonetics for Sanish speakers*. Heinemann Educational Books, 1982.

Flege, James E. "Second language speech learning: Theory, findings, and problems." Speech perception and linguistic experience: Issues in cross-language research 92 (1995): 233-277.

Flege, James E., Carlo Schirru, and Ian RA MacKay. "Interaction between the native and second language phonetic subsystems." Speech communication 40, no. 4 (2003): 467-491.

Flege, James Emil, and Kathryn L. Fletcher. "Talker and listener effects on degree of perceived foreign accent." The Journal of the Acoustical Society of America 91, no. 1 (1992): 370-389.

Flege, James Emil. "The production of “new” and “similar” phones in a foreign language: Evidence for the effect of equivalence classification." Journal of phonetics 15, no. 1 (1987): 47-65.

Flege, James Emil. "Using Visual Information to Train Foreign-Language Vowel Production." *Language Learning* 38, no. 3 (1988): 365-407.

Fowler, Carol A. "An event approach to the study of speech perception from a direct–realist perspective." Journal of phonetics 14, no. 1 (1986): 3-28.

Gerfen, Chip, and Kirk Baker. "The production and perception of laryngealized vowels in Coatzospan Mixtec." Journal of Phonetics 33, no. 3 (2005): 311-334.

Grenon, Izabelle, Mikio Kubota, and Chris Sheppard. "The creation of a new vowel category by adult learners after adaptive phonetic training." Journal of Phonetics 72 (2019): 17-34.

Hawkins, Sarah, and Jonathan Midgley. "Formant frequencies of RP monophthongs in four age groups of speakers." Journal of the International Phonetic Association (2005): 183-199.

Iverson, Paul, and Bronwen G. Evans. "Learning English vowels with different first-language vowel systems II: Auditory training for native Spanish and German speakers." The Journal of the Acoustical Society of America 126, no. 2 (2009): 866-877.

Kartushina, Natalia, Alexis Hervais-Adelman, Ulrich Hans Frauenfelder, and Narly Golestani. "The effect of phonetic production training with visual feedback on the perception and production of foreign speech sounds." The journal of the acoustical society of America 138, no. 2 (2015): 817-832.

Klatt, Dennis H. "Linguistic uses of segmental duration in English: Acoustic and perceptual evidence." The Journal of the Acoustical Society of America 59, no. 5 (1976): 1208-1221.

Ladefoged, P. "A Course in Phonetics. San Diego, New York, Chicago, Atlanta." (1982).

Ladefoged, P., and K. Johnson. "A course in phonetics Wadsworth." Cengage Learning, Scarborough (2011)

Ladefoged, Peter, and Keith Johnson. A course in phonetics. Nelson Education, 2014.

Lively, Scott E., John S. Logan, and David B. Pisoni. "Training Japanese listeners to identify English/r/and/l/. II: The role of phonetic environment and talker variability in learning new perceptual categories." The Journal of the acoustical society of America 94, no. 3 (1993): 1242-1255.

Logan, John S., Scott E. Lively, and David B. Pisoni. "Training Japanese listeners to identify English/r/and/l: A first report." The Journal of the Acoustical Society of America 89, no. 2 (1991): 874-886.

McAllister, Robert, James E. Flege, and Thorsten Piske. "The influence of L1 on the acquisition of Swedish quantity by native speakers of Spanish, English and Estonian." Journal of phonetics 30, no. 2 (2002): 229-258.

Ménard, Lucie, Jean-Luc Schwartz, Louis-Jean Boë, Sonia Kandel, and Nathalie Vallée. "Auditory normalization of French vowels synthesized by an articulatory model simulating growth from birth to adulthood." The Journal of the Acoustical Society of America 111, no. 4 (2002): 1892-1905.

Moyer, Alene. *Foreign accent: The phenomenon of non-native speech*. Cambridge University Press, 2013.

Ohata, Kota. "Phonological differences between Japanese and English: Several potentially problematic." *Language learning* 22 (2004): 29-41.

Pillot-Loiseau, Claire, Takeki Kamiyama, and Tanja Kocjančič Antolík. "French/y/-/u/contrast in Japanese learners with/without ultrasound feedback: vowels, non-words and words." 2015.

Podlipský, "Václav Jonáš, Listen-and-repeat vowel production trainer." Version 1.2., 2015.

Podlipský, Václav Jonáš, Radek Skarnitzl, and Jan Volín. "High front vowels in Czech: A contrast in quantity or quality?" In Tenth Annual Conference of the International Speech Communication Association. 2009.

Scovel, Tom. "Foreign accents, language acquisition, and cerebral dominance 1." Language learning 19, no. 3-4 (1969): 245-253.

Skarnitzl, Radek, and Jan Volín. "Referenční hodnoty vokalických formantů pro mladé dospělé mluvčí standardní češtiny." Akustické listy 18, no. 1 (2012): 7-11.

SKARNITZL, RADEK, and JANA RUMLOVÁ. "Phonetic aspects of strongly-accented Czech speakers of English." AUC Philologica 2, no. 2019 (2019): 109-128.

Skarnitzl, Radek, Pavel Šturm, and Jan Volín. Zvuková báze řečové komunikace: fonetický a fonologický popis řeči. Charles University in Prague, Karolinum Press, 2016.

Šimáčková, Šárka, and Václav Jonáš Podlipský. "Foreign language learners acquire L2 phonetic detail: Goose and Foot fronting in non-native English." Research in Language 15, no. 4 (2017): 385-403.

Šimáčková, Šárka, and Václav Jonáš Podlipský. "Production accuracy of L2 vowels: Phonological parsimony and phonetic flexibility." Research in Language 16, no. 2 (2018): 169-191.

Šimáčková, Šárka. "Engela's Fleshes": Cross-linguistic perception and production of English [æ] and [ɛ] by Czech EFL learners trained in phonetics." In Proceedings of the 15th International Congress of Phonetic Sciences (ICPhS 2003), Barcelona, pp. 3-9. 2003.

Šturm, Pavel, and Radek Skarnitzl. "The open front vowel/æ/in the production and perception of Czech students of English." In Twelfth Annual Conference of the International Speech Communication Association. 2011.

Vilkman, Erkki, Olli Aaltonen, Ilkka Raimo, Paula Arajärvi, and Hanna Oksanen. "Articulatory hyoid-laryngeal changes vs. cricothyroid muscle activity in the control of intrinsic F0 of vowels." Journal of Phonetics 17, no. 3 (1989): 193-203.

Volín, Jan. IPA-Based Transcription for Czech Students of English. Karolinum, 2002.

Wells, John C., and John Corson Wells. Accents of English: Volume 1. Vol. 1.

Cambridge University Press, 1982.

15. Anotace

Autor: Kateřina Plachká

Název katedry a fakulty: Katedra anglistiky a amerikanistiky, Filozofická fakulta UP

Název bakalářské práce: Imitation-based visual-feedback training of English vowels for Czech learners of English

Název bakalářské práce česky: Imitační trénování anglických samohlásek s vizuální zpětnou vazbou

Vedoucí práce: Mgr. Václav Jonáš Podlipský, Ph.D.

Počet znaků: 73729

Počet příloh: 1

Počet titulů použité literatury: 46

Klíčová slova: české samohlásky, anglické samohlásky, fonetické trénování, návrh výzkumu, fonetika

Klíčová slova v angličtině: Czech vowels, English vowels, phonetic training, research proposal, phonetics

Anotace česky: Tato práce si klade za cíl vytvořit hypotézu, zda má fonetické trénování problematických anglických samohlásek českými studenty angličtiny vliv na jejich následné přiblížení rodilé výslovnosti. Důraz je kladen na přehled literatury s podobnou tematikou a jejich úspěšnost, samotné rozdíly mezi samohláskami českými a anglickými a následné shrnutí s předpokládanými problémy s osvojováním rodilého přízvuku u anglických samohlásek při samotném tréninku.

Anotace anglicky: This thesis aims to make a hypothesis about whether a phonetic training of English problematic vowels for Czech L1 speakers influences their final native-like pronunciation. The emphasis is put on the literature review concerning similar topics and their success, the differences between Czech and English vowels, and a subsequent summary with predictable problems when achieving native-like pronunciation of English vowels.