Palacký University Olomouc
Faculty of Arts
Department of English and American Studies

Imitation of English Coda-Voicing-Induced Vowel Duration Variability by Czech Learners: A Research Proposal

(Bachelor thesis)

Daniel Kopecký

Supervisor: Mgr. Václav Jonáš Podlipský, Ph.D.

Olomouc 2020

Declaration

I hereby declare that this thesis has been composed by myself under the supervision of Mgr. Václav Jonáš Podlipský, Ph.D., and that I have provided a complete list of the literature used.

Olomouc, 15 August 2020 Daniel Kopecký

Acknowledgements

I wish to express my deepest gratitude to my supervisor, Mgr. Václav Jonáš Podlipský, Ph.D., for his invaluable help, advice, guidance, and encouragement throughout the writing of the thesis, which would not have otherwise been completed. I would also like to thank those who participated in recording of the stimuli.

Table of Contents

1 Introduction

1.1 Dynamic changes of pronunciation

It is a relatively well-known fact that language, including pronunciation, is not invariable, but instead is constantly changing over long periods of time (Aitchison 2001). This development of pronunciation occurs on the level of a language community as the generations of speakers exchange and it is thus a continuous process covering centuries. Similar changes, however, also happen within individuals across their lifespan. In fact, even within a much shorter period, speakers are known to adjust their speech in reaction to the pronunciation features encountered in their environment. During a single conversation, measurable changes of the speakers' accents take place. This effect, also known as phonetic convergence, is the main focus of this thesis. Specifically, it aims to explore phonetic convergence in a second language.

This chapter first reviews the literature concerned with changes in pronunciation. It starts with changes taking place over long periods of time, continues with short-term shifts and introduces factors that modulate these changes. After that, the nature of the main focus of this thesis is explained and the chapter culminates in research questions and hypotheses.

The second chapter introduces the methodology of the proposed experiment. Characteristics of the stimuli are provided along with a description of the process of model speakers' recordings elicitation. The recordings are then analysed, and their measured values reported. They are further segmented and manipulated, which is exemplified in figures with spectrograms and waveforms. In the last part, the different conditions and subject groups of the proposed procedure of the experiment is explained.

In the third and final chapter of the thesis a summarising conclusion based on the discussed literature and the proposed experiment is reached.

1.1.1 Long-term shifts in L1

It is not unfamiliar to many that pronunciation is changing throughout time. The vowel qualities of Chaucer, for example, were radically different to the vowels of Shakespeare's time and to those of today's English speakers (Wolfe 1972).

However, changes in pronunciation have been observed even within the lifetime of a single adult speaker. To illustrate this point, Harrington et al. (2000) compared recordings of Queen Elizabeth II's annual Christmas Messages from the 1950s and 1980s with recordings of BBC broadcasters from the 1980s. They spoke Standard Southern British English, associated more with younger or lower social class speakers with respect to the Queen. Using acoustic analysis, the authors measured the first two resonant frequencies of the vocal tract (F1 and F2 formants). Significant shifts in one or both formants were found, although the shift was not complete. In the formant space, the Queen's 1980s vowels were midway between her 1950s realisations and the Standard Southern British English vowels from the 1980s. During the thirty-year period, her prestigious sounding vowels thus became similar to the more common Standard Southern British English pronunciation, while still remaining distinct to some extent.

To give another example, the acquisition of a second dialect of one's first language (L1) in an adult person who moves to a different region or becomes a member of a different speech community also occurs commonly. It has been studied by several authors including Munro et al. (1999), who found that the accent of Canadians who had been living in the USA was rated as intermediate between Canadian and American accent by listeners from both Canada and the USA, showing that the shift was salient. Evan and Iverson (2007) investigated Northern English students attending a university in the south of England over a period of two years. They were recorded before the beginning of their studies, after three months, and before the end of the first and second years attending the university. During their study in the south, the northern speakers were found to change their pronunciation of vowels to become more similar to the Standard Southern British English accent spoken at the university.

A similar change within an even shorter period of time was reported in a study by Pardo et al. (2012). They recorded five pairs of previously unacquainted American college roommates. At four times during the academic year, the students provided

words with different vowels and two sentences. The recordings were judged in an AXB perceptual similarity test, where independent listeners judged which of A or B was more similar to X. On each trial either A or B was the speaker's baseline recording of the particular target word from the first session, before the roommates had met, and the other recording (A or B) was from a later session. X was one of the roommate's productions of that word. In other words, the listeners had to decide which of a speaker's productions of a word, A or B (baseline or a later production), was more similar to the speaker's roommate's production from a later session after the speaker and their roommate had been exposed to each other's accents for some time. Furthermore, the speakers completed a questionnaire assessing their relationship with their roommates based on the amount in common and the closeness they felt towards each other. It was shown using perceptual and vowel spectra measurements that the change throughout a single academic year was facilitated by the perceived closeness of the subjects' relationships.

Although the existence of pronunciation changes over long periods of time is generally well known and intuitively logical, they obviously do not happen from one day to the next. Instead, as is logical, they must consist of an accumulation of small gradual changes. The nature of these changes is addressed below in section 1.1.3 Short-term shifts in L1. Before that, however, shifts in a long-term frame will be reviewed.

1.1.2 Long-term shifts in L1 induced by L2

As shown above, phonetic changes happen because of exposure to another speaker's accent. Besides that, however, the changes may be a result of the influence of another language. The interfering influence of a speaker's native language (L1) on the speaker's second language (L2) is not surprising as it accounts for accentedness, which is a noticeable feature of L2 learners. However, researchers have also found evidence for influence in the opposite direction where L1 of an individual is influenced by a relatively long-term exposure to the speaker's L2. This effect is known as phonetic drift (Chang 2019) and it has been found to take place after years, months, or even weeks of exposure.

Phonetic drift was studied by Sancier and Fowler (1997) with a Brazilian-Portuguese speaker. She was found to shift in the voice onset time (VOT), a phonetic correlate of the phonological voice feature of stop consonants, towards the ambient language, be it her native language or her L2 English. She was recorded translating sentences on three occasions after several-months-lasting trips: once after a stay in the USA and before leaving for Brazil, then after returning to the USA, and finally right before leaving for Brazil again. The VOT of stops in her native language, Brazilian-Portuguese, is inherently shorter than that of English. The participant shortened her VOT after staying in Brazil not only in Brazilian-Portuguese but also in English. In contrast, after staying in the USA she used longer VOT in English but also in Brazilian-Portuguese. Brazilian-Portuguese listeners judged which one of the two sentences in their L1, one from the second and one from the third recording session, they perceived as more foreign accented, and they were more likely to choose the sample recorded after the participant's stay in Brazil. Native English listeners, however, were not able to discriminate between the two versions in English. The author explains this by the claim that while Brazilian-Portuguese listeners distinguished between a normal-accented and foreign-accented speech, American speakers had to distinguish between two foreign-accented versions differing in the degree of accentedness. The study shows that phonological repertoires are interconnected cross-linguistically. Phonetic drift thus entails that the repertoires are not rigid but malleable and the direction of the change is predicted by the phonetic features the speaker encounters in their language environment.

Tobin et al. (2017) tried to replicate the findings of Sancier and Fowler (1997) with Spanish-English late bilinguals. They were recorded reading English (their L2) and Spanish (their L1) sentences after spending a few months in an English-speaking country and then after a few weeks in a Spanish-speaking country. Unlike the results of Sancier and Fowler's study, drifting towards the ambient language was found only for English VOTs, and not for Spanish ones. The authors argue that the shorter realisation of Spanish VOTs as opposed to English could account for this asymmetry, as short realisations are more resistant to change. The effect might also be explained by using linear measures as opposed to logarithmic ones, since human perception of differences

in stimuli duration is proportionally related to the magnitude of the stimuli. For short realisations a smaller difference seems bigger. For example, if 10 ms is increased by 5 ms, it would be perceived as an increase of a certain extent. In the case of 100 ms, however, the same increase in duration (5 ms) would not yield the same degree of difference in perception as it is proportionally a much smaller part of the original duration.

Chang (2011) found that in order to influence L1 realisations, the speakers need not be fluent in the L2. In the study, English adults started to attend a six-week course of Korean in South Korea. All of them reported no significant exposure to the language prior to the course. Furthermore, they spoke mainly English outside of class. Each week recordings of the subjects reading English and Korean monosyllabic words in isolation were elicited. An acoustic analysis of VOT, F1, F2, and F0 at the onset of the following vowel was then conducted for stop consonants and vowels. The results showed that novice learners of Korean drifted towards the language's phonetic properties in their native English productions, and this drift occurred as soon as after the first week of L2 learning.

1.1.3 Short-term shifts in L1

Furthermore, changes have been found to occur even in a more short-term shifts, during one dialogue or after non-interactive immediate exposure to stimuli known as shadowing. With respect to the direction, two terms are used by researchers. Phonetic convergence, according to Nguyen and Delvaux (2015), is the tendency wherein interlocutors approximate each other's speech during the course of conversation. Phonetic divergence, then, is the tendency to sound more distinct from each other. If the interlocutors do not adjust their style after exposure to each other, another term, maintenance, may be used to describe the lack of convergence or divergence (Giles 2016). Although the terms phonetic imitation, accommodation, and alignment can all be used interchangeably to refer to the same phenomenon, i.e. modifying one's speech after exposure to the speech of others, some authors make distinctions between them. According to some (Zając and Rojczyk 2014), imitation may be seen as a result of adjusting one's speech in a strict laboratory setting or when explicitly instructed to

imitate, while accommodation may imply obtaining the relevant data from more natural circumstances. Alignment might also be used in a broad sense.

A number of social-psychological factors, such as gender or attitude, affect the degree and direction of accommodation. Besides these, the magnitude of phonetic imitation may also be influenced by linguistic factors. Most literature is concerned with the influence of L1 which sounds atypical to other speakers in some regards or which is a different variety than that of the individual.

According to the communication accommodation theory (CAT), originally referred to as speech accommodation theory (Giles 1973), convergence or divergence is used to manage social distances between interlocutors and to facilitate communication (Giles et al. 1991, Giles 2016).

Among social-psychological factors taking place within a single conversation, Pardo (2006) and Pardo et al. (2010) found that gender influences convergence. In their studies, male speakers converged more than female speakers in same-sex pair conversational tasks.

Pardo (2006) also studied the effect of roles in a conversation. The creation of uneven roles of a giver and a receiver was enabled by the nature of the elicitation method used. In the map task, one of the participants in a pair is given a map including a path around various points, and the other participant receives a map without this path. Without looking at each other's maps, the receiver must complete the path on their map with the help of their partner. The target utterances are thus elicited naturally. Overall, the participants in giver roles converged to a greater extent to the receivers than receivers did to the givers.

The effect of social-psychological factors is present even in immediate situational imitation devoid of social interactions. Unlike the findings of Pardo (2006) and Pardo et al. (2010) on the role of gender above, in Namy et al. (2002) women have been found to converge more than men in a shadowing task. Greater imitation with women was also found in a study by Babel et al. (2014) concerned with voice attractiveness. In this study, women were more likely to imitate both male and female models previously rated as attractive than men were. The ratings had been performed by independent

participants, who had listened to male and female voice recordings and then rated the attractiveness of their voice on a scale.

In the same study by Babel et al. (2014), it was also found that gender atypical voices facilitate imitation. The ratings were based on how quickly independent participants assigned the models their gender upon exposure to their voices. They heard a word and were asked to judge whether the voice was either “male” or “female” as quickly as possible. The quicker they made their decision, the more gender typical the voice was regarded.

Another social-psychological factor, the effect of attitude towards a nationality on accommodation, was investigated by Babel (2010). In a shadowing task concerning vowels, New Zealand participants were presented with an Australian model. Participants with an implicit positive bias to Australia were more likely to converge to the model. The attempt to create an immediate attitude towards the model by adding an insulting anti-New Zealand comment before the task was not found to have any significance on convergence. Not only is imitation influenced by attitude towards nationality, but the facilitating nature of positive subjective attitude towards the model speaker was reported as well by Yu et al. (2013). In their study with extended VOTs, baseline productions were elicited from the subjects before they were played a story containing the manipulated VOT values. Then a post-exposure test was conducted to examine whether any influence took place. Besides attitude, the role of openness, attention switching, and the outcome of the story was found to influence convergence.

Attractiveness, measured by rating of photographs of male model talkers, was found to facilitate imitation for female participants in a study by Babel (2012). Male speakers, on the other hand, diverged from the models they had rated as attractive. A possible explanation might be that men might have seen the speaker as a “threat”, therefore diverging from him. The awareness of the model’s sexual orientation should perhaps remove this difference with male gay model talkers. Yu’s study (2013), however, did not find an influence of perceived models’ sexual orientation on convergence.

Other studies focus on linguistic factors. One of these factors in terms of word frequency has been suggested by Goldinger (1998). He proposed an episodic theory of

the mental lexicon, in which he claims that each heard word leaves a trace in the memory. A newly heard trace influences phonological representation and subsequent production of words. Imitation is less likely for low-frequency words as they are represented with fewer traces and are thus more prone to be influenced.

Babel (2012) implemented Goldinger's findings into her study by using low-frequency lexical words as the shadowing stimuli. Open vowels were imitated to a greater extent than close ones, possibly due to larger differences in the participants' dialects with respect to the realisation of those vowels. As Babel puts it, "there is more imitation when there is the phonetic space to do so" (2012, 188). Tobin's already-mentioned study (2017), in which participants with longer English VOTs shifted towards Spanish VOTs to the largest extent, is consistent with this claim.

As shown in Kim et al. (2011), however, it seems that the facilitating feature of the language space for imitation is limited by size. Very large differences in the interlocutors' phonetic repertoires seem to inhibit imitation. In this study, it was hypothesised that phonetic convergence is facilitated by interlocutors' close language distance. Indeed, non-native speakers converged to native speakers to a lesser degree than did interlocutors of the same language with the same dialect. They participated in a so-called diapix task, where each member of a same-sex pair was given a picture which differed in ten details from a picture of the other interlocutor. They were seated so that they would not see one another, and they were asked to identify these differences by conversing. The names for the three language distance conditions were labelled "close", "intermediate", and "far", corresponding to same-L1 and same-dialect pairs at the same time, same-L1 but different-dialect pairs, and different-L1 pairs, respectively. The participants in the same-dialect and different-dialect pairs were both speaking either American English or Korean. In different-L1 pairs, English was the language used. They were pairs consisting of a native American English and native Korean speaker, and pairs consisting of a native English and native Chinese speaker. To measure the imitation, an independent group of people judged from natural conversations where repetition was not common. The results showed that close language distance between interlocutors facilitates phonetic convergence to a greater extent than do the intermediate or far distances. In addition, no significant difference was found between

different-dialect and different-L1 interlocutors. Having the same L1 but a different dialect is therefore not a facilitating factor as opposed to not having the same L1.

The results of Olmstead et al. (2013) also seem to support this hypothesis. In their study, English and Spanish speakers imitated VOT durations only within their phonetic inventories. A continuum varying in VOT duration was used in the experiment, in which the participants were explicitly told to imitate the stimuli they hear. Native Spanish subjects imitated only short VOTs, which is the natural realisation in Spanish. Similarly, L1 English subjects imitated long VOTs, consistently with their natural pronunciation.

Similar findings have been reported by Lev-Ari and Peperkamp (2014). In this perceptual learning study, interlocutors failed to adapt to a linguistically distant speaker. Native speakers of French adjusted their phonological representations of VOT in bilabial stops only upon the exposure to a native speaker, as opposed to a non-native one, and this change was later extended to other speakers as well. It does not provide answers to whether the influence relies on their knowledge of the model's language background or on the native features of the voice.

Finally, the results of another study (Nielsen 2011) concerned with shadowing support the phonetic space limitation. A large difference that would interfere with the speaker's phonological repertoire, was not imitated. The author manipulated VOT duration of the voiceless stop /p/. The English-speaking subjects were first recorded reading the stimuli, then they heard the model pronouncing all of the words, and finally they read them again. The extended VOT of /p/ was imitated and the effect was also extended for previously unheard words including a new phoneme, /k/. Reduced VOT, on the other hand, was not imitated by the participants. The answer may lie in the speakers' phonological categories. While extended VOT would not affect the English category boundary, decreasing VOT might clash with the realisation of voiced stops, where aspiration is often the discriminatory factor between voiceless and voiced sounds.

1.1.4 Short-term shifts in L2

Little research has been conducted on L2 speakers' convergence in L2. It might be assumed that non-native speakers will be influenced by atypical realisations, like L1

speakers. Trofimovich et al. (2014) studied convergence between interlocutors of the same L2. They were 41 students with different L1s enrolled in an English for academic purposes class. Three- and four-syllable words stressed on the second syllable were embedded in four collaborative pair tasks during one semester. The students exchanged information and discussed some topics using supplementary information containing the target words. Recordings of these classroom-setting interactions were analysed for cases when one student produced a correct stress pattern following the other student's correct production, not necessarily of the same word. The results showed that non-native students of English did converge in word stress.

Interactions between L1 and L2 have also been studied. Lewandowski and Jilka (2019) conducted a research with 20 German speakers of English involved in conversational tasks with native English speakers. The non-native speakers were all highly proficient and selected from a previous study so that they could be divided in a group of 10 phonetically talented speakers and 10 less talented speakers. The phonetic talent measures included a variety of tests in speech perception, production, and also imitation. They participated in a diapix task described above in Kim et al.'s study (2011). Convergence was measured acoustically by comparing the amplitudes of target words at different frequencies from the two speakers. The similarity value of the native and non-native speaker productions from an early time during the dialogue was compared to the similarity value of the early native production and a non-native's production from a later point in the conversation. Phonetic talent, openness to experience, mental flexibility, but also neuroticism (as measured by the Neuroticism Extraversion Openness Five Factor Inventory, or NEO-FFI, and a so-called Simon Test) in non-native speakers significantly facilitated convergence in non-native speakers. The authors explain the counter-intuitive influence of neuroticism by referring to the speaker's intensified need for social approval, with accordance to the CAT.

In a communication task, Enzinna (2018) investigated whether accommodation is affected by language background, i.e. being an English monolingual or a Spanish-English bilingual, and by long-term exposure to a monolingual or bilingual community. For the bilinguals, Spanish was their L1 and English their L2. The focus was on the duration of VOT in word-initial voiceless stops, which are inherently longer in English

than in Spanish. A total of 20 self-reported monolinguals and late fluent bilinguals from either a monolingual or a bilingual community participated in the experiment. They engaged in a referential communication task with a pre-recorded voice belonging either to a monolingual English speaker or a bilingual Spanish-English speaker. The participants were given a board with words and were asked by the recorded voices to identify some of these words. Long-term exposure to a monolingual or bilingual community showed an effect. Speakers (both mono- and bilingual) who had spent at least a year in a predominantly monolingual community had overall longer VOTs, i.e. more monolingual-like, than speakers after spending a year in the bilingual community. The influence of language background was also significant. Spanish-English bilinguals were influenced more by long-term exposure than English monolingual speakers. The bilinguals diverged from the speaker who belonged to the minority in their community either by increasing or decreasing their VOTs. Bilinguals from a monolingual community diverged from a bilingual speaker and bilinguals from a bilingual community diverged when listening to a monolingual speaker. It seems that interlocutors' linguistic closeness was, at least in the latter group, preferred over the desire to sound native-like.

Because non-native learners are less proficient than native speakers, they may be more resilient to changes in pronunciation and therefore less likely to converge to their interlocutors. However, since their phonological repertoire is represented by less traces in their minds, they may be influenced by recently heard speech to a greater extent than native speakers, consistent with Goldinger's (1998) episodic theory.

It is not clear whether L2 learners are equally likely to adjust to native and non-native interlocutors. It has been shown that social factors influence convergence, and the interlocutor's native status may perhaps affect this degree. Indeed, native speakers seem to be assigned higher status by L2 speakers in terms of their pronunciation than non-native accents do. In Dalton-Puffer et al.'s study (1997), 132 Austrian students of English evaluated three different native and two Austrian accents in English. Overall, the native accents seemed to be preferred by the non-native speakers in a scalar subjective evaluation of qualities such as likeability, honesty, education, or suitability

for being a radio presenter. Also, this preference was enhanced by previous familiarity with the accent.

On the other hand, although the target-language model might facilitate convergence because of the positive attitude towards it and because of greater linguistic distance, the same may apply for the non-native interlocutors. The subjects may identify with the same-L1 speaker as their peer and converge to them. However, the results of Šimáčková and Podlipský's study (2012) of Czechs' attitudes towards Czech-accented English might imply that this is unlikely. Among other objectives of their paper, they tried to examine the neutrality of an interpreter in terms of his accent, i.e. whether he is efficient in not drawing attention away from the topic or not creating unnecessary attitudes. They played a 1-minute English recording of a Czech-accented interpreter to 60 listeners of several different L1s. One of the questions of a subsequent questionnaire asked the subjects to assess the degree of selected pronunciation qualities on a scale. These were meant to show to what extent the interpreter's pronunciation is perceived as neutral. Czech L1 speakers tended to judge the Czech interpreter more negatively than other L1 speakers as they attached more value to descriptors such as odd, unpleasant, or irritating whereas the others tended to evaluate him as educated.

1.2 Current thesis

Following a study by Zając (2013), where imitation was not systematic and the participants converged towards the native model in some cases and diverged from the non-native in other words, Zając and Rojczyk (2014) aimed to investigate whether the models' native or non-native status influences the extent of imitation. The vowels /ɪ/, /æ/, and /ɛ/ were placed into seven minimal pairs consisting of word-initial /b/, /m/, or /s/. The contrasting voicing contexts were provided by /d/ and /t/. Forty first-year Polish students of English with B2 proficiency participated in the study of vowel length as a cue to the voicing of following consonant. Their English phonetics course had not yet covered durational variability of English vowels, therefore they probably had no conscious knowledge of this phenomenon. However, their baseline productions imply that it already functions as a feature in their interlanguage to some extent.

The model recordings were made by a native speaker of English and a phonetician imitating Polish accent. The native model produced longer vowels before voiced stops than before voiceless ones. The non-native model, on the other hand, produced similar durations in both contexts. The subtle length differences were equalised because vowel length in Polish does not contribute to the perception of voicing of the following final consonant, which is always realised as voiceless.

The participants were found to converge to the native model and diverge from the non-native one. Longer vowels were thus produced before voiced stops and shorter vowels before voiceless stops upon the exposure to both model speakers. The extent of imitation was greater when imitating (converging to) the native model than (diverging from) the non-native one.

Unlike Zając and Rojczyk, where vowel length durational variability was present in the native model's speech but not in the non-native's, in this thesis the stimuli will be created in such a way that there will be two versions of both models' recordings. One of them will feature systematic variability in vowel duration due to coda voicing and the other one will not. That way, it will be possible to test whether the participants converge with the model speaker because of their language background or because of the target-language-like pattern in the model speech.

The main question is whether and how Czech learners of English differ in the direction and degree of imitation of vowel duration variability induced by coda voicing due to language background of model speakers.

Possibly, another variable could be included. A non-native model with a different L1 to the L1 of the subjects, i.e. a non-Czech learner of English. The reason for this can be found in what has been described as the interlanguage speech intelligibility benefit (Bent et al., 2003). For native listeners, native speech is the most intelligible. Proficient non-native speech, however, seems to be intelligible to the same extent for the same-L1 non-native listeners as native speech does. This is true for same-L1 non-native subjects as well as different-L1 speakers, for which the intelligibility can even be greater than native speech. It is probably due to the shared phonology and L2 learning strategies.

1.2.1 Coda-Voicing-Induced Vowel Duration Variability

The term Coda-Voicing-Induced Vowel Duration Variability (CVIVDV) is used for the purpose of this thesis to combine two parts of a phenomenon in English speech, namely pre-fortis clipping, or shortening, and pre-lenis lengthening.

The fact that there is a difference in vowel length before voiced (lenis) and voiceless (fortis) codas has been observed as early as in Heffner (1937). House and Fairbanks (1953) later measured the difference in English to be relatively consistent across the contrasting consonant pairs. They reported an average vowel duration to be 174 ms before voiceless consonants and 253 ms before the corresponding voiced consonants. The ratio of voiceless to voiced thus being approximately 2:3, or 0.69. The absolute length values of each pair differed with respect to the manner of articulation. The longest vowels appeared when preceding fricatives and the shortest vowels preceded stops.

These results were subsequently reproduced by Peterson and Lehiste (1960), whose values were 197 ms for the voiceless consonants and 297 for the voiced consonants, a ratio of 0.66. The authors furthermore calculated intrinsic durational values of individual vowels by averaging their length before voiced and voiceless codas. This resulted in them categorising vowels into short and long. Finally, they observed that preceding consonants, unlike following consonants, do not influence the vowel length.

The above-mentioned studies on CVIVDV, however, were applicable to English only. In Chen (1970) it was discovered that as well as in English, vowels vary they length as a result of coda voicing value also in French, Korean, and Russian. The highest degree was reported for English (146 ms for voiceless-consonant-preceding vowels vs 238 ms for vowels preceding voiced consonants, a ratio of 0.61). French speakers' productions, on the other hand, had a ratio of 0.87 (354 vs 407 ms for the voiceless-context and voiced-context vowels, respectively). The author concluded that the phenomenon itself is likely a language-universal, but the degree thereof is language-specific.

Moreover, evidence against a number of hypotheses which tried account for the phenomenon was presented. In particular, against a so-called compensatory temporal

adjustment. Because the closure of voiceless obstruents is inherently longer than that of voiced ones, it was hypothesised that the vowel would compensate for this difference so that the total duration of the syllable would be identical for both contexts, but this was disproved by Chen. He assumed that the best plausible explanation is the rate of closure transition. When producing speech sounds, the energy needed for voiceless obstruents is greater than for their voiced counterparts. Because of the anticipatory effect, the time needed for the transition from a vowel to a voiceless obstruent should be shorter.

Thanks to this study, CVIVDV was thus proven to function in speech in at least some languages other than English. Other authors studied the phenomenon from a perception point of view to determine whether listeners employ vowel length as a cue to the following coda voicing. Denes (1955) observed than not only are the differences between the vowel durations depending on the voicing feature of the following consonants, but the consonants themselves systematically vary in duration. The vowel before a voiced coda is longer, and the coda itself is shorter, whereas the vowel preceding a voiceless consonant is shorter and the voiceless coda seems to compensate for this by being slightly longer. In the study, Denes selected a minimal pair consisting of the words use (noun) and use (verb), which differ in the voicing status of the last speech sound. Then, recordings of the words were manipulated in such a way that the /s/ in the noun was shortened and inserted in the place of the /z/ in the verb. Similarly, the /z/ was lengthened and it replaced the /s/ in the noun. The shortened /s/ sounded like /z/, and vice versa. The words were synthesised, except for the final fricative, which was realised as voiceless using a human recording. Next, four different durations of the vowel and five durations of the coda were determined, creating vowel and consonant continua. They were combined, making a total of 20 different items. 33 listeners were instructed to judge the sequence of words and determine whether they had heard the noun or the verb. The results showed that perception of final consonant voicing is stronger when the durational ratio of the coda consonant to the nucleus vowel is reduced. The vowel and final consonant duration are thus a clear cue to the perception of final consonant voicing.

Evidence with more minimal pairs was needed to confirm this, though. An experiment with a variety of synthetic vowels and final consonants was conducted by

Raphael (1972). 25 listeners were supposed to determine which word of a contrasting minimal pair they had heard. Final consonants were perceived as voiced when preceded by a long vowel, and as voiceless when following a vowel with short duration. Raphael concluded that vowel duration is a sufficient cue to the voicing of the following coda in English.

Many years later, Tanner et al. (2020) conducted a large-scale study utilising a number of different spoken-language corpora. The magnitude of CVIVDV was analysed across different speakers and across a total of 30 dialects. It was found that the degree of the effect was substantially less visible in spontaneous speech than in the literature concerned with speech from laboratory settings. Furthermore, it differed significantly between different English dialects. The highest ratio of the vowel durations was reported for US dialects. In dialects of Scottish English, the magnitude of the phenomenon was minimal, or even non-existent. The variation across speakers within the different dialects was rather small. The variations, however, were found to be affected by a number of factors. Greater CVIVDV was measured in low frequency words, slow speech rate, and lexical words as opposed to function words.

Unlike many studies on CVIVDV in English, little research has focused on the effect in the Czech language, and the results are not conclusive. Keating (1984) reported three Czech speakers reading words with phonemically short or long vowels followed by either a voiceless or a voiced stop. Although there was a tendency in speakers to apply CVIVDV, the results were not statistically significant, the ratio of voiceless-stop-preceding vowels to vowels preceding voiced stops being 0.95, or approximately 11 ms. Similarly, Machač and Skarnitzl (2007) found longer vowels before voiced consonants than before voiceless ones, but the results lacked significance and were not manifested in all cases. Furthermore, the post-vocal consonants in question were not part of the same syllable but instead belonged to the onset of the following syllable.

Other authors, on the other hand, report significant results of CVIVDV in Czech speakers. Podlipský and Chládková (2007) created a minimal set of 3 nonsensical words differing in the coda voicing. The words contained a voiced coda, a voiceless coda, and an underlyingly voiced coda, which was devoiced due to regressive assimilation in Czech. They obtained the stimuli recordings from 9 Czech speakers. The results

indicated that although the differences were minor vowel duration does indeed vary with respect to the coda voicing in Czech. The vowel was the shortest before a voiceless obstruent, longer before an underlyingly voiced but overtly devoiced obstruent, and the longest before a voiced coda. Next, they aimed to find whether CVIVDV is utilised also by listeners. They constructed two vowel-duration continua differing in the voicing of the coda obstruent. 54 listeners judged whether they heard a long or a short vowel in nonsensical words imbedded in carrier sentences. There were no differences between a voiceless and devoiced coda, but there were differences between a coda which was voiceless and that which was voiced. Ambiguous vowels in terms of duration were perceived as short when they preceded a voiceless coda to a greater extent than when appearing before a voiced obstruent. The difference, approximately 3.3 ms, was subtle but highly significant.

Fejlová (2013) studied the effect in Czech speakers speaking English and it was found to be conditioned by different degrees of accentedness of the non-native speakers. She examined 13 Czech speakers of English and English native speakers reading news bulletins in English. The non-native speakers were placed into three groups based on their level of accented speech. The differences in vowel duration based on the voicing of the following consonant contexts were the lowest in speakers with a strong Czech accent. Overall, however, the differences were very small compared to reported literature from isolated speech.

From the published studies, it thus cannot definitely be decided whether CVIVDV works in Czech. The results indicate that the magnitude of the phenomenon from both a production and perception aspect is minimal, if existent at all. The subjects of the experiment in this thesis should thus have room for accommodation in English, where the phenomenon is of great magnitude.

In English, the larger CVIVDV differences compared to other languages may be accounted for by the feature having been phonologised in English speakers to be the cue to underlying coda voicing. Indeed, English speakers' obstruents are commonly not realised as voiced. They are voiced especially when they appear between voiced speech sounds. In pre- and post-pausal positions, they may be voiced only partially or even realised as completely voiceless (Cruttenden 2014; 164, 193). As Walsh and Parker

(1981) argue, the vowel duration is influenced by the underlying voicing feature rather than by the overt coda obstruent voicing. Klatt (1976) further suggested that this dependency on the vowel length rather than the coda voicing may be a diachronic change.

In Czech, there is word-final devoicing of obstruents, i.e. no voiced obstruents appear in the coda position. Devoicing usually fails to appear, however, when the next word begins with a voiced obstruent. In the case of Czech spoken in the Moravia region, the devoicing-inhibiting speech sound may also be a vowel or a sonorant (Šimáčková et al. 2012, Volín 2015). Moreover, unlike English, Czech is a quantity language, which contrasts short and long vowels. The duration value in Czech thus functions as a phonemic feature.

1.2.2 Research questions and hypotheses

Building on the findings of the literature discussed in the previous sections, the following research questions (Q) and their corresponding hypotheses (H) have been formulated.

Q_1: Do Czech learners of English exhibit imitation?
H_1: It is believed that Czech speakers of English will not differ from other learners and will adjust their L2 upon exposure to another speaker.

Q_2: Are Czech learners of English more likely to imitate a native model or a non-native model?
H_2: The participants are hypothesised to converge to the native speaker to a greater degree than to the non-native speaker, as a result of their intention to approximate a representative of the target language. Alternatively, they might be found to converge to the non-native, who they may identify as a member of the same social group of learners. Because non-natives have been reported to prefer native accents, and specifically because Czech speakers were shown to be critical of Czech accented English, converging to the Czech speaker of English is not likely. (linguistic distance?) In fact, the subjects are likely to diverge from the Czech model to distance themselves from undesirable foreign-accented speech.

Q_3: Do Czech learners of English imitate a native model even if accented features (no difference in vowel duration between voiced and voiceless coda contexts) are imposed onto otherwise native-like speech?
H_3: The participants are expected to imitate the native model even if the feature has non-native-like characteristics.

Q_4: Do Czech learners of English imitate a non-native model if a feature in their speech is manipulated to sound native-like?
H_4: The language background status of the model is likely to have more value for the participants who are thus not likely to imitate the non-native model even if they sound native-like regarding one feature in their speech.

2 Methodology

The original aim of this thesis was to conduct the experiment described in more detail in the following sections. Relevant literature was studied to carefully design the experiment. Once the whole scheme and the individual stimuli were devised, a script for their elicitation was written in Praat, a speech analysis software. Then, a native and a non-native speaker were recorded to serve as the speech shadowing models. The elicited recordings were segmented and labelled so that they could be subsequently manipulated in terms of duration. A script for the shadowing speakers using the prepared stimuli was in preparation and it was planned to record the non-native speakers. Due to the ensuing COVID-19 pandemic, however, the experiment itself was not conducted and the scope of this thesis was thus narrowed down to a research proposal with the stimuli recordings already prepared.

2.1 Stimuli

A total of 20 target words (10 minimal pairs) and 28 filler words (14 minimal pairs) were created for the proposed experiment. All of them are English monosyllabic words which follow the CVC structure. The target minimal pairs differ in voicing status of the final obstruent. The filler minimal pairs differ in either the initial or final consonant, the latter of which is not an obstruent in any case. The target words have only monophthongs in the middle position, whereas some of the fillers contain also diphthongs. See Table 4 and Table 5 in Appendix I: Lists of stimuli for the complete lists of stimuli.

A short recording, approximately 40 seconds long, was recorded to encourage the participants' impression of the model status so that the participants would know whether they are listening to a native or a non-native speaker. The text is an extract from an article on a BBC website by Richard Gray called "The secret tricks hidden inside restaurant menus". It was adapted so that it would not contain any of the phenomena tested in the subsequent experiment, i.e. no vowel–obstruent sequences at the end of syllables. For the full text, see Appendix II: Model-L1-inducing text.

All the stimuli were recorded by a native American speaker in his forties and a non-native Czech speaker of English in his twenties. The two models first recorded the model-L1-inducing text, which they read from a piece of paper. The rest of the recording session then continued on a computer using a script in Praat. All of the 48 words appeared on a screen sorted into an 8x5 grid, and the models were instructed to read them to themselves so that they would not hesitate in pronunciation during the ensuing recording. The target and filler words then started appearing automatically on the screen one by one in random order. Each word was visible for 2.5 s, and after 1 s of white background another word in black font appeared in the middle of the monitor. The models were allowed to pause at any time during the recording session and they were repeatedly being offered water.

vowel	**word**	**[+native]**		**[–native]**	
		original	**mean**	**original**	**mean**
æ	bad	267	229	184	152
	bat	191		120	
	calve	393	274	313	305
	calf	155		297	
	gab	263	227	201	179
	gap	190		157	
	tab	240	192	153	159
	tap	143		164	
ɛ	bed	174	143	135	132
	bet	111		129	
	peg	200	135	171	153
	peck	69		134	
ʌ	cub	162	125	107	97
	cup	88		86	
ɑ/ɒ	dog	275	208	250	205
	dock	160		159	
ɪ	hid	158	129	106	103
	hit	100		100	
i	seed	248	190	291	265
	seat	132		239	

Table 1: Vowel durations of the native and non-native models' target words in ms. The original durations of the native model and the means of both models are to be used in the experiment.

In Praat, the durations of vowels and of constriction intervals in coda obstruents were measured (see the values in Table 1 above) with the boundaries labelled manually. The vowel duration was measured as the interval following either the release of a stop or a fricative noise and consisting of the onset and offset of visible formant structure in the spectrogram. In the case of final stops, the constriction was segmented as beginning with the vowel offset and ending just before the release of the stop. For coda fricatives, the prominent friction noise following the vowel was used. See Figure 1 to Figure 10 below for the segmentation in waveforms and spectrograms.

As expected, the native model speaker provided longer vowel durations when preceding a voiced obstruent, and shorter realisations before the voiceless counterpart for the same vowel, as illustrated in Figure 1 and Figure 2. Similarly to Zając and Rojczyk's (2014) Polish participants, the non-native model also systematically exhibited the use of CVIVDV in his productions, although to a lesser extent than the native model (compare Figure 1 and Figure 2 with Figure 3 and Figure 4). The only exception is the tap/tab pair, where a longer vowel was produced before the voiceless stop than before the voiced one. Even then, the difference (11 ms) is minor. Absolute differences between the two voicing contexts ranged from 6 to 91 ms, a ratio ranging from 0.64 to 1.1 for the non-native model and from 58 to 238 ms, or from 0.35 to 0.72 for the native productions. As can be observed, the overall durational differences between the voiced and voiceless contexts of the non-native speaker are relatively short and correspond to less than 1/3 of the length differences of the native model. For the mean values, see Table 2 below.

model	[+voice]	[–voice]	difference	ratio
[+native]	238	134	104	0.56
[–native]	191	159	32	0.83

Table 2: Mean durations of vowels of the native and non-native models in voiced and voiceless coda obstruent contexts, length differences between the two contexts, and the ratio of voiceless- to voiced-context vowel duration. All values in ms.

Two variables, CVIVDV and model language background, create four sets of model recordings made by a native and a non-native Czech speaker of English. In terms of the non-native model, the recordings were manipulated in two ways using the

overlap-add method in Praat. First, the stimuli were matched with the native model's CVIVDV and final consonant constriction durations so as to make the recordings sound native-like in these regards (see Figure 9 and Figure 10). The other way of manipulating the stimuli involved neutralising these CVIVDV and consonant duration values so that the words would be perceived as non-native-like. This was performed as an average of the vowel duration before a voiced obstruent and before a voiceless obstruent, and as an average of the values of the obstruent constrictions (see Figure 7 and Figure 8). This was done in order to provide the non-native participants with space for imitation, as was described in Babel (2012). In terms of the native model, his recordings were either neutralised (see Figure 5 and Figure 6), similarly to the non-native, or kept non-manipulated so that they would retain their native CVIVDV values and sound native-like (see Figure 1 and Figure 2).

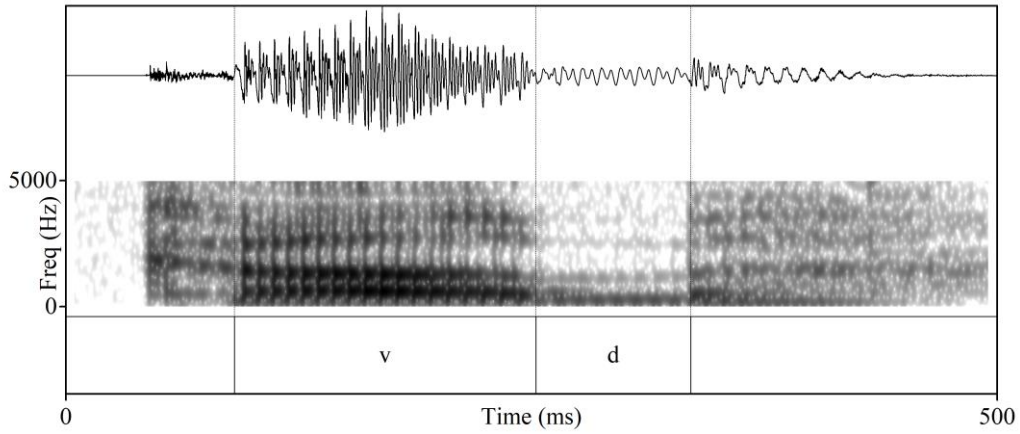

Figure 1: Segmented realisation of the target word *cub* by the native model speaker. v = vowel /ʌ/, lasting for 162 ms, d = voiced consonant /b/.

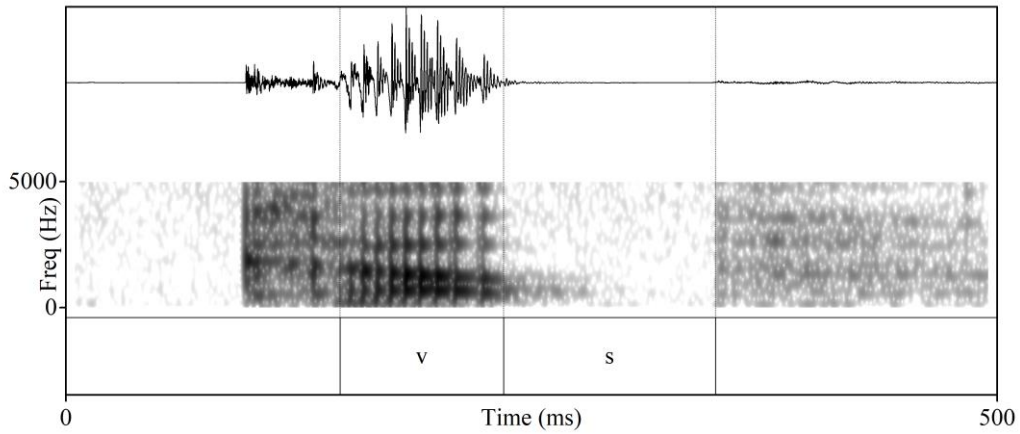

Figure 2: Segmented realisation of the target word *cup* by the native model speaker. v = vowel /ʌ/, lasting for 86 ms, s = voiceless consonant /p/.

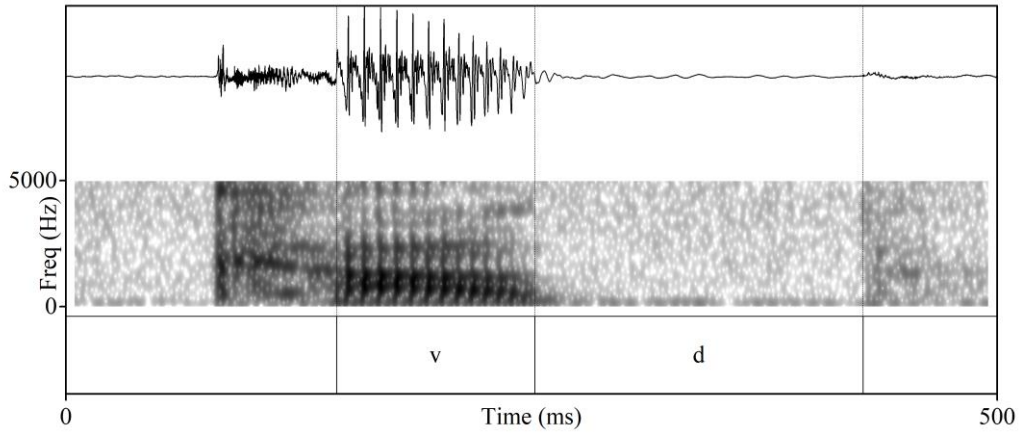

Figure 3: Segmented realisation of the target word *cub* by the non-native speaker. v = vowel /ʌ/, lasting for 107 ms, d = voiced consonant /b/.

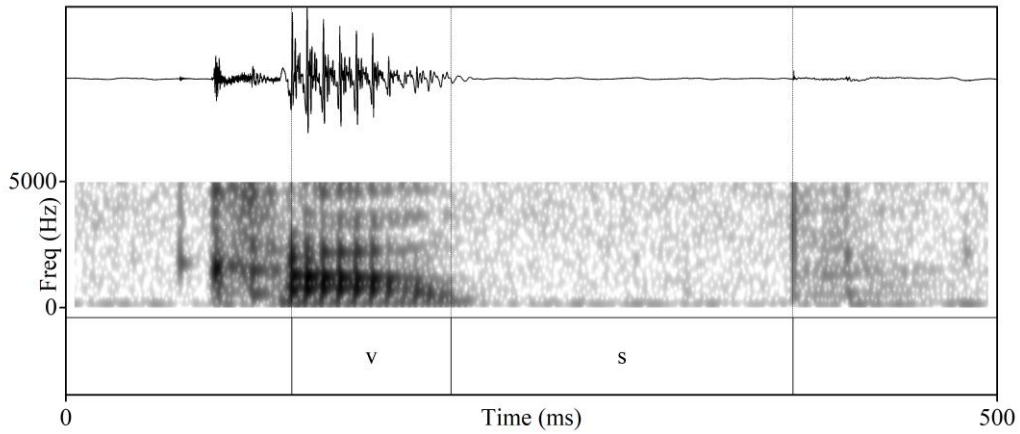

Figure 4: Segmented realisation of the target word *cup* by the non-native speaker. v = vowel /ʌ/, lasting for 88 ms, s = voiceless consonant /p/.

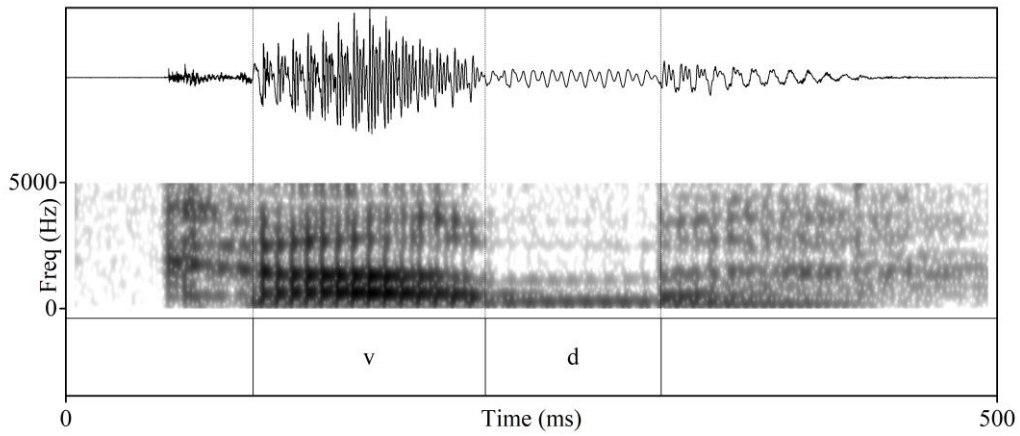

Figure 5: Segmented realisation of the manipulated neutralised target word *cub* by the native speaker. v = vowel /ʌ/, lasting for 125 ms, d = voiced consonant /b/.

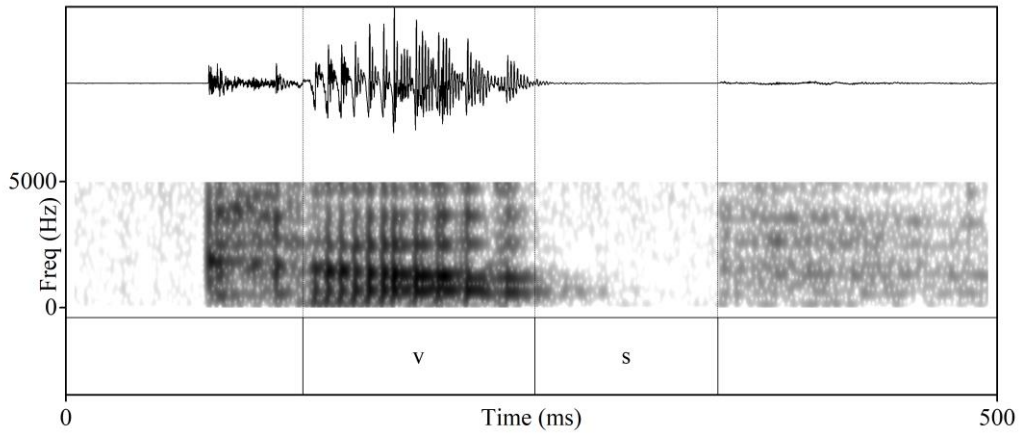

Figure 6: Segmented realisation of the manipulated neutralised target word *cup* by the native speaker. v = vowel /ʌ/, lasting for 125 ms, s = voiceless consonant /p/.

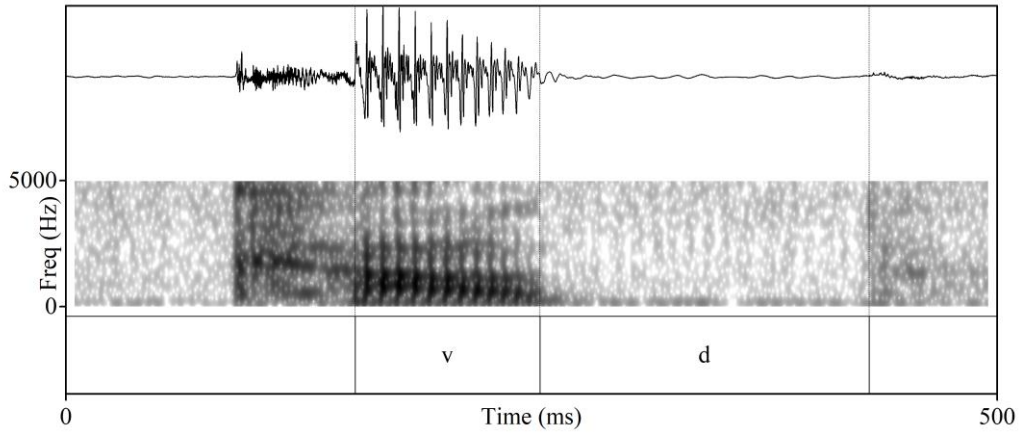

Figure 7: Segmented realisation of the manipulated neutralised target word *cub* by the non-native speaker. v = vowel /ʌ/, lasting for 97 ms, d = voiced consonant /b/.

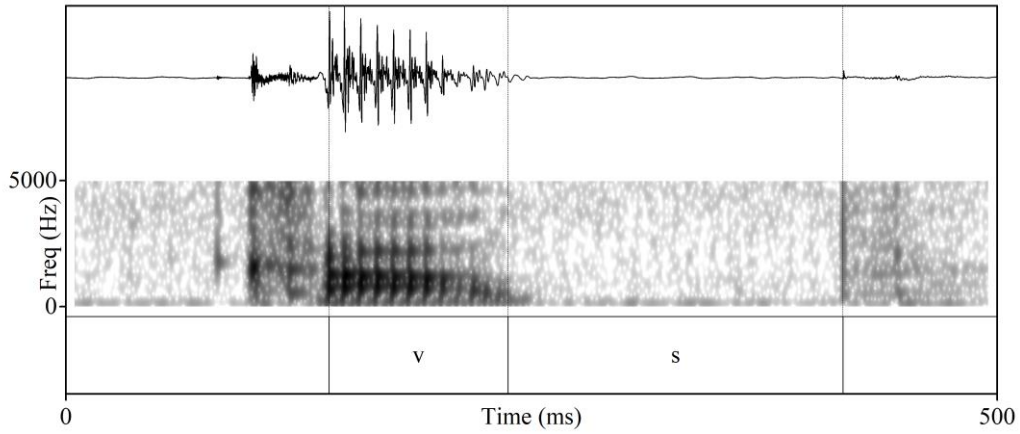

Figure 8: Segmented realisation of the manipulated neutralised target word *cup* by the non-native speaker. v = vowel /ʌ/, lasting for 97 ms, s = voiceless consonant /p/.

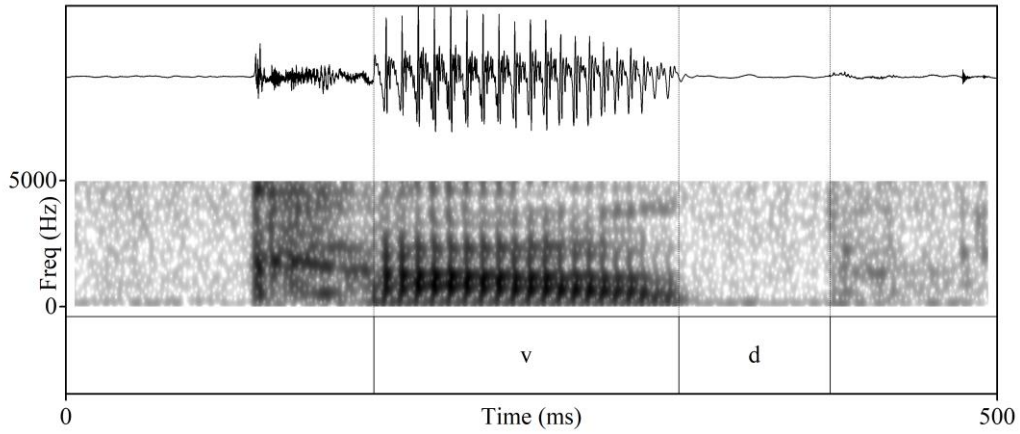

Figure 9: Segmented realisation of the manipulated native-like target word *cub* by the non-native model speaker. v = vowel /ʌ/, lasting for 162 ms, d = voiced consonant /b/.

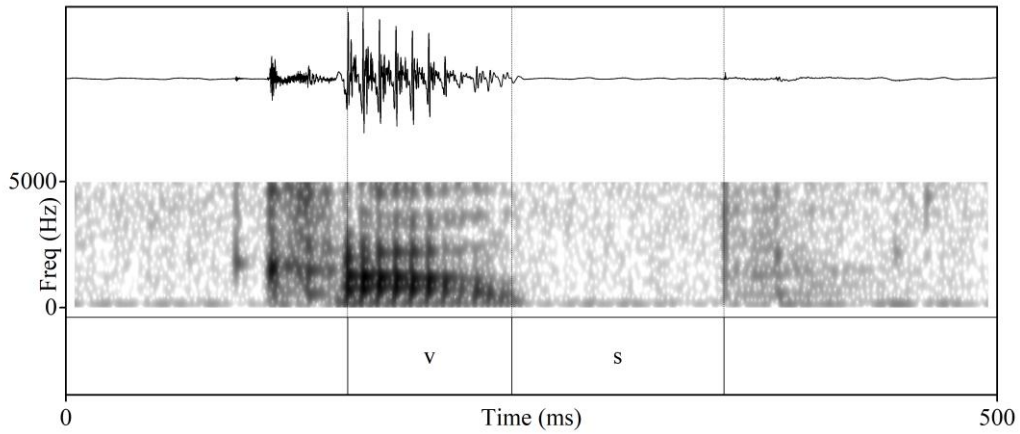

Figure 10: Segmented realisation of the manipulated native-like target word *cup* by the non-native model speaker. v = vowel /ʌ/, lasting for 88 ms, s = voiceless consonant /p/.

2.2 Procedure

The participants in the proposed experiment will be non-native Czech speakers of English. The recordings will be elicited with a microphone and a computer using a script in Praat. As with the model recordings, the participants will first familiarise themselves with all of the 48 words on a screen sorted into an 8x5 grid. The words will then start to appear automatically on the screen one by one in random order and the baseline productions will be elicited to establish the subjects' pronunciation prior to the models' exposure. Next, one of the two model-L1-inducing texts will be played to the participants to encourage the L1 status impression of the subsequent model recording. After that, the words will again start randomly appearing, this time accompanied by the model recording, whose status would correspond to the previous model-L1-inducing recording. A pause will follow to eliminate the influence of the recently heard words on the next recording part. Then another elicitation part will follow, again preceding with a model-L1-inducing recording. The order will be counterbalanced within the naturalistic and non-naturalistic groups. Group 1 and 2, and groups 2 and 3 will thus have the same stimuli but presented in the opposite order. The naturalistic group will have a native model featuring CVIVDV and a non-native model with neutralised CVIVDV values. The non-naturalistic groups stimuli will not match in the CVIVDV values expected from either a native or a non-native speaker. The participants will listen to a native model with neutralised CVIVDV, or non-native like, and to a non-native model with CVIVDV values taken from the original native model recordings. See Table 3 below for an overview of the procedure.

order	naturalistic		non-naturalistic	
	group 1	**group 2**	**group 3**	**group 4**
1	baseline test	baseline test	baseline test	baseline test
2	[+native]-inducing text	[–native]-inducing text	[+native]-inducing text	[–native]-inducing text
3	[+CVIVDV], [+native voice]	[–CVIVDV], [–native voice]	[–CVIVDV], [+native voice]	[+CVIVDV], [–native voice]
4	pause	pause	pause	pause
5	[–native]-inducing text	[+native]-inducing text	[–native]-inducing text	[+native]-inducing text
6	[–CVIVDV], [–native voice]	[+CVIVDV], [+native voice]	[+CVIVDV], [–native voice]	[–CVIVDV], [+native voice]

Table 3: Procedure of the experiment.

3 Conclusion

This thesis reviews changes in pronunciation which take place after exposure to other speakers' speech. In particular, it proposes an experiment concerned with phonetic convergence of L2 speakers of English in a non-interactive setting.

In the first chapter, changes in pronunciation are introduced from a broader perspective and an overview of literature is provided. It is shown that pronunciation changes across long periods of time within communities but also within individual speakers. The influence of L2 on L1 and other short-terms shifts in L1 during dialogues or asocial settings has been found and this convergence is influenced by social-psychological and linguistic factors, most importantly by atypicality.

Next, shifts in the speakers' L2 are described. It seems that for convergence there must be a gap between the interlanguage and the perceived L2. Imitation is inhibited, however, if this linguistic distance between interlocutors is too large. Non-native speakers prefer a native speaker's accent to the accent of a non-native speaker.

The main focus of the thesis concerns vowel duration variability induced by coda voicing, or CVIVDV. An overview of literature is provided for English and Czech from both a production and perception view. In English, the phenomenon seems to be phonologised and the magnitude of the effect is thus large. In Czech, it functions to a minimal degree, if at all, possibly due to it clashing with phonemic vowel length.

The proposed experiment builds on a study by Zając and Rojczyk (2014) with Polish learners of English who converged with a native model and diverged from a non-native one. Their study featured the native model with naturalistic CVIVDV and the non-native model with neutralised CVIVDV. This thesis added yet another set of stimuli, in which the native model features non-naturalistic CVIVDV and the non-native features naturalistic CVIVDV values taken from the native speaker. The aim is to learn whether the non-native speakers will imitate the model because of their language background or because of the native-like pattern in the model speech. It is hypothesised that they will imitate the native model even if accented features are imposed on his speech.

References

Aitchison, Jean. 2001. *Language change: progress or decay?* 3rd ed. Cambridge: Cambridge University Press.

Babel, Molly. 2010. "Dialect divergence and convergence in New Zealand English". *Language in Society*, 39(4):437–456. doi: 10.1017/S0047404510000400.

Babel, Molly. 2012. "Evidence for phonetic and social selectivity in spontaneous phonetic imitation." *Journal of Phonetics*, 40(1):177–189. doi:10.1016/j.wocn.2011.09.001.

Babel, Molly, Grant McGuire, Sophia Walters, and Alice Nicholls. 2014. "Novelty and social preference in phonetic accommodation." *Laboratory Phonology*, 5(1):123–150. doi:10.1515/lp-2014-0006.

Bent, Tessa, and Ann R. Bradlow. 2003. "The interlanguage speech intelligibility benefit." *The Journal of the Acoustical Society of America*, 114(3):1600–1610. doi:10.1121/1.1603234.

Chang, Charles B. 2011. "Rapid and multifaceted effects of second-language learning on first-language speech production." *Journal of Phonetics*, 40(2):249–268. doi:10.1016/j.wocn.2011.10.007.

Chang, Charles B. 2019. "Phonetic drift." *The Oxford Handbook of Language Attrition*, 191–203. doi:10.1093/oxfordhb/9780198793595.013.16.

Chen, Matthew. 1970. "Vowel Length Variation as a Function of the Voicing of the Consonant Environment." *Phonetica*, 22:129–159.

Cruttenden, Alan. 2014. *Gimson's Pronunciation of English*. 8th ed. London: Routledge.

Dalton-Puffer, Christiane, Gunther Kaltenboeck, and Ute Smit. 1997. "Learner attitudes and L2 pronunciation in Austria." *World Englishes*, 16(1):115–128. doi:10.1111/1467-971X.00052.

Denes, P. 1955. "Effect of Duration on the Perception of Voicing." *The Journal of the Acoustical Society of America*, 27(4):761–764. doi:10.1121/1.1908020.

Enzinna, Naomi. 2018. "The influence of language background and exposure on phonetic accommodation." *Proceedings of the Linguistic Society of America*, edited by Patrick Farrell, 41:1–15. doi:10.3765/plsa.v3i1.4333.

Evans, Bronwen G., and Paul Iverson. 2007. "Plasticity in vowel perception and production: A study of accent change in young adults." *The Journal of the Acoustical Society of America*, 121(6):3814–3826. doi:10.1121/1.2722209.

Fejlová, Dita. 2013. "Pre-fortis shortening in fluent read speech: A comparison of Czech and native speakers of English." *AUC Philologica*, 2014(1):101–112.

Giles, Howard. 1973. “Accent mobility: A model and some data.” *Anthropological Linguistics*, 15(2):87–105.

Giles, Howard. 2016. “Communication Accommodation Theory.” *The International Encyclopedia of Communication Theory and Philosophy*, 1–7. doi:10.1002/9781118766804.wbiect056.

Giles, Howard, Nikolas Coupland, and Justine Coupland. 1991. “Accommodation theory: Communication, context, and consequence.” *Contexts of Accommodation*, 1–68. doi:10.1017/CBO9780511663673.001.

Goldinger, Stephen D. 1998. “Echoes of Echoes? An Episodic Theory of Lexical Access.” *Psychological Review*, 105(2):251–279. doi:10.1037/0033-295X.105.2.251.

Harrington, Jonathan, Sallyanne Palethorpe, and Catherine I. Watson. 2000. “Does the Queen speak the Queen's English?” *Nature* 408:927–928.

Heffner, R.-M. S. 1937. “Notes on the Length of Vowels.” *American Speech*, 12(2):128–134. doi:10.2307/452621.

House, Arthur S., and Grant Fairbanks. 1953. “The Influence of Consonant Environment upon the Secondary Acoustical Characteristics of Vowels.” *The The Journal of the Acoustical Society of America*, 25(1):105–113.

Kim, Midam, William S. Horton, and Ann R. Bradlow. 2011. “Phonetic convergence in spontaneous conversations as a function of interlocutor language distance.” *Laboratory Phonology*, 2(1):125–156. doi:10.1515/labphon.2011.004.

Klatt, Dennis H. 1976. “Linguistic uses of segmental duration in English: Acoustic and perceptual evidence.” *The Journal of the Acoustical Society of America*, 59(5):1208–1221. doi:10.1121/1.380986.

Munro, Murray J., Tracey M. Derwing, and James E. Flege. 1999. “Canadians in Alabama: a perceptual study of dialect acquisition in adults.” *Journal of Phonetics*, 27:385–403. doi:10.1006/jpho.1999.0101.

Nielsen, Kuniko. 2011. “Specificity and abstractness of VOT imitation.” *Journal of Phonetics*, 39(2):132–142. doi:10.1016/j.wocn.2010.12.007.

Keating, Patricia A. 1984. “Universal phonetics and the organization of grammars.” *UCLA Working Papers in Phonetics*, 59:35–49.

Lev-Ari, Shiri, and Sharon Peperkamp. 2014. “Do people converge to the linguistic patterns of non-reliable speakers? Perceptual learning from non-native speakers.” *Proceedings of the 10th International Seminar on Speech Production*, 257–160.

Lewandowski, Natalie, and Matthias Jilka. 2019. “Phonetic Convergence, Language Talent, Personality and Attention.” *Frontiers in Communication*, 4:18. doi:10.3389/fcomm.2019.00018.

Machač, Pavel, and Radek Skarnitzl. 2007. “Temporal compensation in Czech?” *Proceedings of the 16th International Congress of Phonetic Sciences,* 537–540.

Namy, Laura L., Lynne C. Nygaard, and Denise Sauerteig. 2002. “Gender differences in vocal accommodation: The role of perception.” *Journal of Language and Social Psychology* 21:422–432. doi:10.1177/026192702237958.

Nguyen, Noël, and Véronique Delvaux. 2015. “Role of imitation in the emergence of phonological systems.” *Journal of Phonetics,* 53:46–54. doi:10.1016/j.wocn.2015.08.004.

Olmstead, Anne J., Navin Viswanathan, M. Pilar Aivar, and Sarath Manuel. 2013. “Comparison of native and non-native phone imitation by English and Spanish speakers.” *Frontiers in Psychology,* 4:475. doi:10.3389/fpsyg.2013.00475.

Pardo, Jennifer S. 2006. “On phonetic convergence during conversational interaction.” *The Journal of the Acoustical Society of America* 119:2382–2393. doi:0.1121/1.2178720.

Pardo, Jennifer S., Isabel Cajori Jay, and Robert M. Krauss. 2010. “Conversational role influences speech imitation.” *Attention, Perception, & Psychophysics* 72(8):2254–2264. doi:10.3758/APP.72.8.2254.

Pardo, Jennifer S., Rachel Gibbons, Alexandra Suppes, and Robert M. Krauss. 2012. “Phonetic convergence in college roommates.” *Journal of Phonetics,* 40:190–197. doi:10.1016/j.wocn.2011.10.001.

Peterson, Gordon, and Ilse Lehiste. 1960. “Duration of Syllable Nuclei in English.” *The Journal of the Acoustical Society of America,* 32(6):693–703.

Podlipský, Václav Jonáš, and Kateřina Chládková. 2007. “Vowel duration variation induced by coda consonant voicing and perceptual short/long vowel categorization in Czech.” *17th Czech-German Workshop – Speech Processing*, 68–74.

Raphael, Lawrence J. 1972. “Preceding Vowel Duration as a Cue to the Perception of the Voicing Characteristic of Word-Final Consonants in American English.” *The Journal of the Acoustical Society of America,* 51(4B):1296–1303. doi:10.1121/1.1912974.

Sancier, Michele L., and Carol A. Fowler. 1997. “Gestural drift in a bilingual speaker of Brazilian Portuguese and English.” *Journal of Phonetics,* 25(4):421–436. doi:10.1006/jpho.1997.0051.

Šimáčková, Šárka, and Václav Jonáš Podlipský. 2012. “Pronunciation Skills of an Interpreter.” *Teaching Translation and Interpreting Skills in the 21st Century,* edited by Jitka Zehnalová, Ondřej Molnár, and Michal Kubánek, 139–149. Olomouc: Palacký University Olomouc.

Šimáčková, Šárka, Václav Jonáš Podlipský, and Kateřina Chládková. 2012. “Czech spoken in Bohemia and Moravia.” *Journal of the International Phonetic Association*, 42(2):225–232. doi:10.1017/S0025100312000102.

Tanner, James, Morgan Sonderegger, Jane Stuart-Smith, and Josef Fruehwald. 2020. “Toward ‘English’ Phonetics: Variability in the Pre-consonantal Voicing Effect Across English Dialects and Speakers.” *Frontiers in Artificial Intelligence*, 3:38. doi:0.3389/frai.2020.00038.

Tobin, Stephen J., Hosung Nam, and Carol A. Fowler. 2017. “Phonetic drift in Spanish-English bilinguals: Experiment and a self-organizing model.” *Journal of Phonetics*, 65:45–59. doi:10.1016/j.wocn.2017.05.006.

Trofimovich, Pavel, Kim McDonough, and Jennifer A. Foote. 2014. “Interactive Alignment of Multisyllabic Stress Patterns in a Second Language Classroom.” *TESOL Quarterly*, 48(4):815–832. doi:10.1002/tesq.156.

Volín, Jan. 2015. “Fonetika a fonologie.” In *Mluvnice současné češtiny I.: Jak se píše a jak se mluví*, edited by Václav Cvrček, 43–79. Prague: Karolinum.

Walsh, Thomas, and Frank Parker. 1981. “Vowel length and ‘voicing’ in a following consonant.” *Journal of Phonetics*, 9(3):305–308. doi:10.1016/S0095-4470(19)30973-8.

Wolfe, Patricia M. 1972. *Linguistic Change and the Great Vowel Shift in English*. California: University of California Press.

Yu, Alan C. L., Carissa Abrego-Collier, and Morgan Sonderegger. 2013. “Phonetic Imitation from an Individual-Difference Perspective: Subjective Attitude, Personality and ‘Autistic’ Traits.” *PLoS ONE*, 8(9):e74746. doi:10.1371/journal.pone.0074746.

Zając, Magdalena. 2013. “Phonetic Imitation of Vowel Duration in L2 Speech.” *Research in Language*, 11(1):19–29. doi:10.2478/v10015-012-0009-5.

Zając, Magdalena, and Arkadiusz Rojczyk. 2014. “Imitation of English vowel duration upon exposure to native and non-native speech.” *Poznan Studies in Contemporary Linguistics*, 50(4):495–514. doi:10.1515/psicl-2014-0025.

Appendices

Appendix I: Lists of stimuli

[+voice]	[–voice]
bad	bat
bed	bet
calve	calf
cub	cup
dog	dock
gab	gap
hid	hit
peg	peck
seed	seat
tab	tap

Table 4: List of target word pairs in voiced and voiceless coda obstruent contexts.

ball	tall
bell	sell
come	numb
fan	fang
fin	fill
keen	keel
long	wrong
meal	mean
mill	hill
name	lame
pin	bin
sun	some
teal	team
thin	thing

Table 5: List of filler word pairs differing in either the initial or final consonant.

Appendix II: Model-L1-inducing text

The hidden agenda of a restaurant menu

Creating a restaurant menu isn't trivial whatsoever. Restaurants employ some very powerful psychology to influence their clients.

In a theatrical way, the waiter hands you a sombre, leather-bound document, the menu. When you open the menu, you mainly see small writing and your attention turns to a couple of items in flamboyant print. Then you turn to the waiter and order.

The meal is now in preparation, however do we know the reason why clients come to their particular decision? The menu probably performs a more important role than you think. A restaurant menu, an essential marketing tool, can even change the clients' thinking.

Annotation

Author: Daniel Kopecký

Field of study: English Philology – General Linguistics and Theory of Communication

Title: Imitation of English Coda-Voicing-Induced Vowel Duration Variability by Czech Learners: A Research Proposal

Type: Bachelor thesis

Faculty and Department: Faculty of Arts, Department of English and American Studies

Supervisor: Mgr. Václav Jonáš Podlipský, Ph.D.

Number of pages: 39

Number of characters: 56,391

Description: This thesis focuses on changes in the pronunciation of speakers that happen after exposure to the speech of others. It reviews the relevant literature and proposes an experiment in which Czech speakers of English are exposed to natural and manipulated stimuli from a native and a non-native speaker of English. Specifically, it is concerned with varying vowel duration with respect to the voicing of the following consonant, which is a salient phenomenon in English. The aim of the proposed experiment is to investigate whether listeners vary their speech because of the models' language background or because of the target-language-like pattern in the model speech. It is hypothesised that the native status will outweigh the native speech features.

Keywords: phonetic imitation, phonetic convergence, phonetic accommodation, vowel length as a cue to coda voicing, L2 speakers of English

Anotace

Autor: Daniel Kopecký

Studijní obor: Anglická filologie – Obecná lingvistika a teorie komunikace

Název: Imitace variability délky vokálů vyvolané znělostí kody v angličtině českými mluvčími: návrh výzkumu

Typ práce: bakalářská práce

Fakulta a katedra: Filozofická fakulta, Katedra anglistiky a amerikanistiky

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

Počet stran: 39

Počet znaků: 56 391

Charakteristika: Tato práce se zabývá změnami ve výslovnosti, které se dějí, když je mluvčí vystaven řeči druhých. Představuje relevantní literaturu a navrhuje experiment, ve kterém jsou čeští mluvčí angličtiny vystaveni přirozeným a manipulovaným stimulům rodilého a nerodilého mluvčího angličtiny. Konkrétně se jedná o odlišnou délku samohlásek s ohledem na znělost následující souhlásky, což je v angličtině nápadný jev. Cílem navrhovaného experimentu je prozkoumat, zda posluchači mění svou řeč kvůli rodilému statusu modelového mluvčího, nebo kvůli rodilým charakteristikám v řeči modelového mluvčího. Předpokládá se, že rodilý status bude mít větší váhu než charakteristicky rodilé mluvy.

Klíčová slova: fonetická imitace, fonetická konvergence, fonetická akomodace, délka vokálu jako signál ke znělosti kody, nerodilí mluvčí angličtiny