**Palacký University Olomouc**

Faculty of Arts

Department of English and American Studies

# Imitation of English Coda-Voicing-Induced Vowel Duration Variability by Czech Learners

(Master thesis)

**Daniel Kopecký**

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

Olomouc 2023

## 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, 11 May 2023 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 the experiment.

## 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 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 examining changes that take place over long periods of time, continues with short-term shifts, and introduces factors that modulate these changes. Additionally, this chapter provides an overview of theoretical frameworks that seek to account for convergence. After that, the nature of the primary focus of this thesis, varying duration of vowels depending on the voicing of the following coda, is explained, and the chapter culminates in research questions and hypotheses.

Chapter two of this thesis introduces the methodology of the experimental part of the thesis. Characteristics of the stimuli are outlined along with the description of the way the model speakers' recordings were elicited, annotated, and phonetically manipulated, which is exemplified in figures with spectrograms and waveforms. The subjects who participated in the study are also described. In the last part of the section, the whole procedure of the experiment, which included three elicitation parts and four groups of subjects, is explained in detail.

In the third chapter, analyses of the data applying linear mixed-effects statistical models are introduced. The results are interpreted and discussed with relation to the research questions and the reviewed literature, aiming to provide an understanding of the statistical outcomes.

The fourth and final chapter attempts to integrate the theoretical accounts with the findings and reach a summarising conclusion about phonetic convergence.

### *1.1.1 Long-term shifts in L1*

It is not unfamiliar to many people 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, which is associated 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 or RP vowels from the 1980s. During the thirty-year period between 1950s and 1980s, the Queen's prestigious sounding vowels thus became similar to the more common Standard Southern British English pronunciation, while still remaining distinct to some extent (i.e., her productions did not became entirely like those of the Standard Southern British English speakers).

To give another example of pronunciation changes, 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 like the Standard Southern British English accent spoken in the university environment.

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, wherein 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 the 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 smaller gradual changes. The nature of these changes is addressed below in section 1.1.3 Short-term shifts in L1. Before that, however, an overview of shifts from a long-term perspective is presented.

### *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. None of them reported any significant exposure to the language prior to taking part in 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. Although the terms phonetic imitation or accommodation 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. Other terms, perhaps trying to convey more nuanced distinctions and theoretical approaches, include alignment (Pickering and Garrod 2004) or entrainment (Menshikova et al. 2020).

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.

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 pertaining to 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, showing that cognitive and personality traits must also be considered as factors of imitation.

Attractiveness and likability were also found to influence imitation of pitch over the course of a conversation (Michalsky and Schoormann 2017). Ten female and ten male previously unacquainted heterosexual students participated in a speed dating setting. Each participant evaluated their interlocutor in terms of visual attractiveness and likability, and these positively predicted the imitation. Attractiveness measured by rating of photographs of male model talkers was also observed 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 each heard word leaves a trace or echo in the memory. A newly heard trace influences the phonological representation and the subsequent retrieval and production of words. Imitation is more likely for low-frequency words as these are represented with fewer traces and are thus more prone to influence. Since it is an exemplar-based account, the effect should be cumulative, i.e., it should be more prominent after hearing several repetitions of the same token.

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.

Linguistic distance between interlocutors was observed to facilitate imitation also in Lin et al.'s (2021) study. They investigated an ongoing trend of merging two tones in Cantonese Chinese. 63 native speakers were first recorded producing words in isolation. They were then exposed to model speakers who clearly distinguished the tones. Participants who exhibited more merging in their baseline producing (i.e., less distinction between the two tones), were found to imitate the model speakers to a greater extent than did speakers with less merging. A larger phonetic distance in a shadowing task thus promoted the extent of accommodation.

A cross-dialectal study by Walker and Campbell-Kibler (2015) seems to support the effect of language distance on imitation. Twenty English speakers from New Zealand and 16 Midland American English speakers were recruited to participate in a shadowing task concerned with vowel formants. They were exposed to stimuli recorded by speakers of the same dialect (NZ or US Midland), speakers having a dialect relatively close to theirs (Australian or US Inland North), and by speakers of different dialects (NZ for the American speakers, US Midland for New Zealanders).

Convergence was observed with speakers whose baseline productions were furthest away from the model speakers. For the same or close dialect condition, maintenance of spectral values was observed.

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 appear 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 that 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 of the limitation to the language space claim. 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.

Not only has imitation been observed in socially minimal conditions, but it has also been reported to occur after exposure to synthesised voices in a sentence-shadowing task in German (Gessinger et al. 2021) or in Polish (Jankowska et al. 2020). In the latter study, they used five sets of sentences with various phonetic phenomena recorded by twenty people. Phonetic imitation was observed upon exposure to both natural and synthetic voices. Convergence to the human voice, however, was higher than to the computer-generated speech.

Some studies (Zellou et al. 2021) on human–computer interaction report the influence of conversational role and the status of the interlocutor (human vs computer). In a word list task, the participants in the role of information giver instructed the interlocutor on which of two lists a word belongs to. In a more collaborative map task, the interlocutors worked together on completing a worksheet. The productions were judged through similarity ratings. In the first less collaborative task, subjects were reported to imitate the human interlocutor to greater extent than the computer voice regardless of their conversational role. In the significantly more collaborative task, participants who were in the role of providing information exhibited greater imitation, especially towards the human voice. Convergence of role-givers is consistent with Pardo's (2006) findings in human dyads.

### *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, akin to 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.

As shown by Kim et al. (2011) above, language distance is an important factor to consider in imitation. Liu's (2017) study with Mandarin speakers of English, too, investigated its effect. The non-native subjects participated in a shadowing task in which they were repeating individual English words after a model speaker with the same L1. They reliably imitated the non-native model in terms of vowel formants and durations. Greater imitation was found for speakers whose baseline vowel duration productions were further away from the model speaker's productions. There was also evidence for selectivity in terms of the individual speech sounds since different vowels were imitated to different degrees.

Interactions between L1 and L2 have also been studied. Lewandowski and Jilka (2019) conducted a research study 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. Cognitive and personality factors such as 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.

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.

In a study by Olmstead et al. (2021), interlocutors in dyadic interaction tasks resorted to different imitation strategies based on whether they were assigned to matched or mismatched pairs in terms of L1. The authors investigated the realisation of English /i/ and /ɪ/ vowels and the voicing of the following codas in 34 Mandarin L1 and 10 native English speakers. Mandarin Chinese speakers mainly distinguish these vowels in terms of duration while native English speakers mainly contrast duration as a cue to the voicing of the following coda and otherwise use different spectral realisations for the two vowels. In a collaborative matching task, a speaker produced the words *bit*, *beat*, *bid*, and *bead*, while the other partner indicated which word they think has just been produced. The realisations were compared across pre-test and post-test conditions. Matched non-native dyads generally increased the vowel duration (i.e., resorted to their L1 pattern) in the post-test while Mandarin speakers in the unmatched native–non-native pairs accentuated the formant differences along with the contrasting durations. The native speakers mainly maintained their spectral distinction between the two vowels.

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 entirely 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.

It is known that native speakers judge non-native speech as less credible (Lev-Ari and Keysar 2010). In their paper, twenty-eight native speakers of American English reacted to trivia statements by three native English speakers, three non-native speakers with a mild accent, and three speakers with a heavy accent. Each participant listened to 15 statements by each of the three speakers. The veracity of the statements was measured by an indication of a point on a line. Accented speech was found to negatively influence truthfulness of the statements.

Speakers of a second language, too, appear to assign lower status to non-native speakers in terms of their pronunciation as opposed to native accents. 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. Furthermore, this preference was enhanced by previous familiarity with the accent.

On the other hand, although the target-language model speaker 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.

Jiang and Kennison (2022) showed that the belief about the nativeness status of an interlocutor is in itself sufficient to affect imitation. Twenty Mandarin Chinese speakers of English were involved in the experiment on vowel formants in short picture-description conversations with a native speaker of American English. The speaker was introduced to half of the participants as a native speaker, and as a non-native speaker to the other half. The first group of participants significantly adjusted their productions in the direction of the native speaker. The majority of the speakers in the other group failed to accommodate to the native speaker (introduced as a non-native), even showing slight diverging patterns in their productions.

### *1.1.5 Theoretical accounts of changes in pronunciation*

There are two major opposing theoretical frameworks seeking to account for inter-speaker accommodation. One of them, the interactive alignment model (IAM), claims that there is a parity between perception and production, and the process of adjusting one's speech upon exposure to another talker is thus automatic (Pickering and Garrod 2004). The other framework, the communication accommodation theory (CAT), argues that imitation is mediated by social factors (Giles 2016).

The IAM assumes that interlocutors cooperate in order to understand each other. It tries to account for alignment on all levels of dyadic communication via priming. A speaker using a particular lexical item, for example, prompts their conversational partner to use an item that is consistent with the representation that has just been activated. Speakers thus mediate an implicit common ground throughout the conversation, priming each other and making choices compatible with the shared representations. The alignment happens on an abstract situational level but also on semantic, syntactical, lexical, or phonological levels. When speakers align on a representation, they are then able to produce an item compatible with said representation, so there seems to be a perception–production link on all levels of linguistic representation.

Another account assuming automaticity, the exemplar-based theory by Goldinger (1998), has already been discussed. By extension, the link between perception and production of IAM and the exemplar-based episodic theory are consistent with the findings that convergence happens in both social and non-interactive laboratory settings or even with synthesised voices. They, however, fail to explain the situation in which speakers become less like one another or diverge since exposure should automatically lead to the production of more similar speech.

The need for a social explanation thus arises along with findings about the effect of social context. According to 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), perhaps even consciously to some extent. Speakers converge when they adjust their speech to become more similar to one another. To accentuate the differences or to distance oneself from the interlocutor, a speaker may diverge in their speech. If the interlocutors do not adjust their style after mutual exposure, the term maintenance may be used to describe the lack of convergence or divergence. Since speakers have different attitude towards different speakers, and different communicative goals in different social situations, they use different strategies to negotiate relationships.

Although CAT explains the fact that convergence should be more prominent in more social settings, neither CAT nor IAM nor the exemplar theory provide an explanation for the social-psychological variables such as gender (Babel et al. 2014), attitude (Yu et al. 2013), or attractiveness (Babel et al. 2014); linguistic aspects like language distance (Babel 2012) or word frequency (Goldinger 1998); or cognitive factors (Lewandowski and Jilka 2019).

Babel (2012), then, provides an integrated explanation, claiming that the process of accommodation is automatic at the low level but that it may be modulated by other factors, especially social ones. This could explain why imitation occurs in socially minimal conditions on the one hand while taking into account the vast array of factors that seem to have a bearing on phonetic imitation.

From the perception–production point of view, Pardo and Remez (2021) claim that there indeed is a relationship between perception and production, but the two processes are not symmetrical. A listener may distinguish certain features but will not converge fully. Furthermore, the process is selective, that is, the speaker will not converge in all areas.

## 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 the 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 is created in such a way that there are two versions of both models' recordings. One of them will feature systematic variability in vowel duration due to coda voicing and the other one's vowel durations will be invariable for the two voicing conditions. 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 questions are (1) whether Czech learners of English exhibit CVIVDV and whether it is influenced by their proficiency in English, and (2) whether and how they differ in the direction and degree of imitation of vowel duration variability induced by coda voicing due to language background of model speakers or due to the target-language-like durations present in the stimuli.

Although not part of this thesis, another variable could be included in future research. 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). Twenty-five 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.

Skarnitzl and Šturm (2016), too, used speakers with a relatively strong Czech accent in English. Their study with 10 speakers reading a set of carrier sentences concludes that the difference between vowel duration in voiced vs voiceless contexts is not significant in their participants' English productions. Accentedness here seems to preclude the effect of varying vowel duration based on the voicing of the following coda, which is a salient phenomenon in English.

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 speakers of English exhibit CVIVDV?**

$H_1$: As reported by Podlipský and Chládková (2007), Czech speakers of English are expected to produce minimal distinctions between vowel durations conditioned by the following voiced vs voiceless coda. However, since Fejlová (2013) suggests the influence of accentedness on the employment of CVIVDV, more proficient participants are, by extension, expected to produce more distinct values.

**$Q_2$: Do Czech learners of English exhibit imitation?**

$H_2$: It is believed that Czech speakers of English will not differ from other learners (such as Polish as in Zając and Rojczyk' 2014 study) and will adjust their production in their second language upon exposure to another speaker.

**$Q_3$: Are Czech learners of English more likely to imitate a native or a non-native model with naturalistic CVIVDV values?**

$H_3$: 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. Because non-natives have been reported to prefer native accents (Dalton-Puffer et al. 1997), and specifically because Czech speakers were shown to be critical of Czech accented English (Šimáčková and Podlipský's 2012), converging to the Czech speaker of English is not likely. In fact, the subjects are likely to diverge from the Czech model to distance themselves from undesirable foreign-accented speech, in line with Zając and Rojczyk's (2014) Polish participants.

**$Q_4$: Do Czech learners of English imitate a native and a non-native model even with non-naturalistic values?**

$H_4$: The participants are expected to imitate the native model even if his speech shows non-native-like (i.e., invariable) CVIVDV values. For the non-native model speaker, the language background status 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

This thesis expands on a previous bachelor thesis, which was originally formulated as a research proposal. Neither the experiment itself nor the statistical analysis of the data was conducted at that time. The current thesis thus takes these ideas and develops them further, implementing the proposed experiment and analysing the results to address the research questions and hypotheses that were put forward.

## 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 that follow the CVC structure. The target minimal pairs differ in the voicing status of the final obstruent. The filler minimal pairs, used to conceal the objective of the study, 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 also contain diphthongs. See Table 4 and Table 5 in Appendix A: Lists of stimuli for the complete lists of stimuli used in the experiment.

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 B: 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 Czech model speaker was purposefully selected to have a relatively strong accent in English. Both male 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 (Boersma and Weenink 2022), a speech analysis software. 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 model speakers were allowed to pause at any time during the recording session and they were repeatedly being offered water.

In Praat (Boersma and Weenink 2022), the durations of vowels and of the constriction intervals in coda obstruents were measured with the boundaries labelled manually, following the guidelines of phonetic segmentation by Machač and Skarnitzl (2009). 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 created in Praat below for the segmentation in waveforms and spectrograms.

As expected, the native model speaker provided longer vowel durations when they were 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 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.

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).

## 2.2 Participants

A total of 24 subjects participated in the study. All of them were Czech speakers of English studying a bachelor's or master's degree at Palacký University in Olomouc at the time of the experiment. Their estimated proficiency ranged from B2 to C2 according to the Common European Framework of Reference for Languages (CEFR; Council of Europe 2020). See below for more details of the participants' proficiency.

There were both women (n = 17) and men (n = 7) present in the sample, and their average age was 22.2 (min = 19, $Q_1$ = 21, median = 22.5, $Q_3$ = 24, max = 25). Out of the 24 participants, 19 were enrolled in an English-related programme (English philology or English for translating and interpreting, either major or minor).

No participants reported any hearing or serious vision problems. One participant reported a neurological problem (multiple sclerosis) but was still included in the analysis. All were recruited using the opportunity sampling method along with the snowball method, taking advantage of acquaintances and their recommendations to take part in the study.

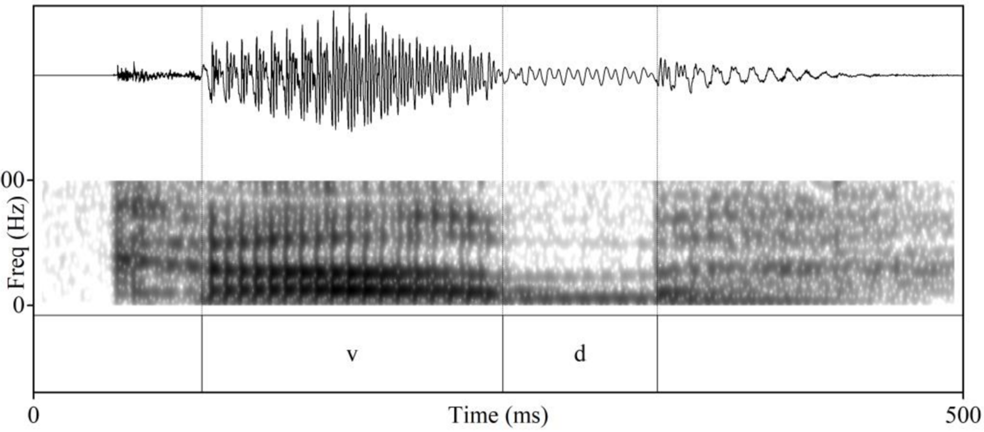


**Figure 1: Segmented realisation of the target word *cub* by the native model speaker. v = vowel /ʌ/, lasting for 162 ms, d = voiced consonant /b/.**

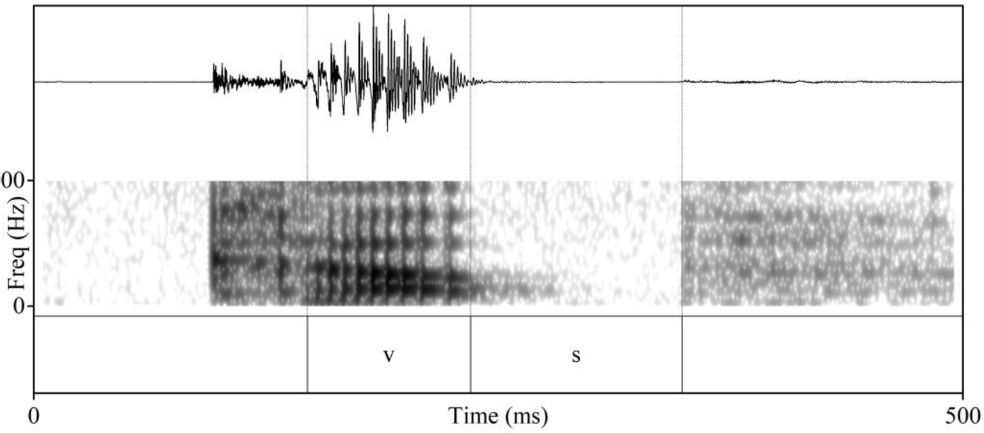


**Figure 2: Segmented realisation of the target word *cup* by the native model speaker. v = vowel /ʌ/, lasting for 86 ms, s = voiceless consonant /p/.**

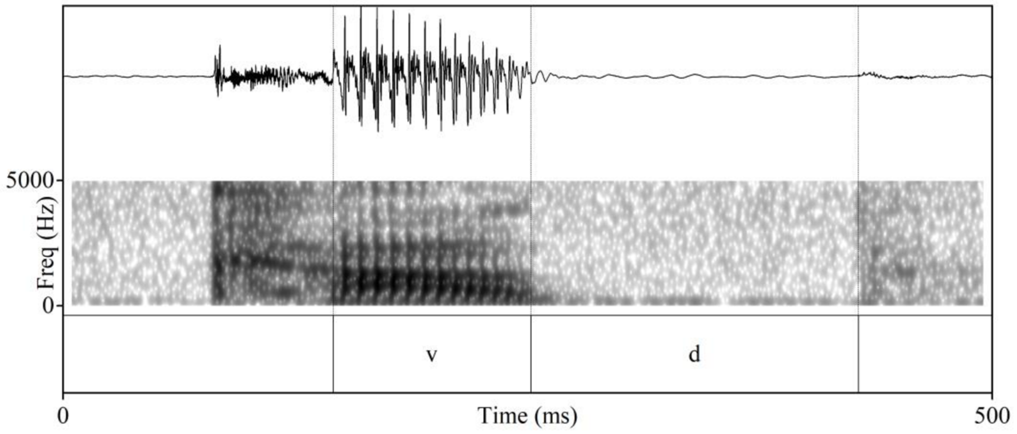


**Figure 3: Segmented realisation of the target word *cub* by the non-native speaker. v = vowel /ʌ/, lasting for 107 ms, d = voiced consonant /b/.**

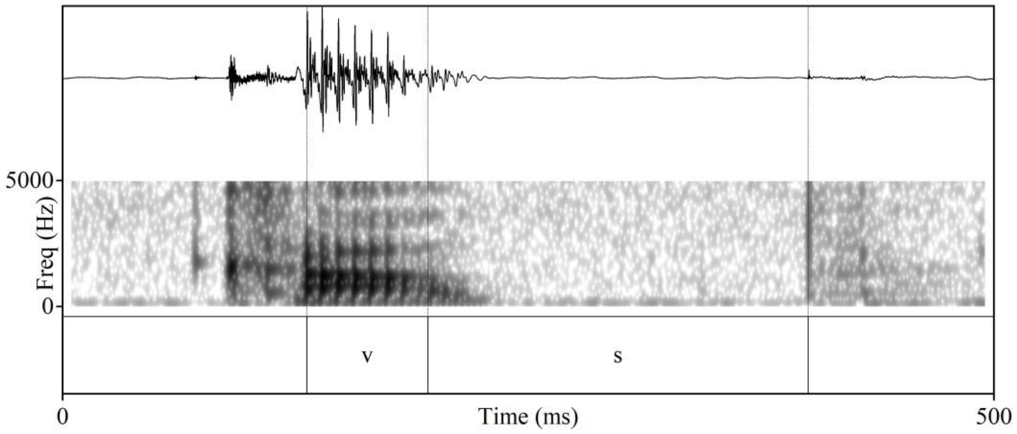


**Figure 4: Segmented realisation of the target word *cup* by the non-native speaker. v = vowel /ʌ/, lasting for 65 ms, s = voiceless consonant /p/.**

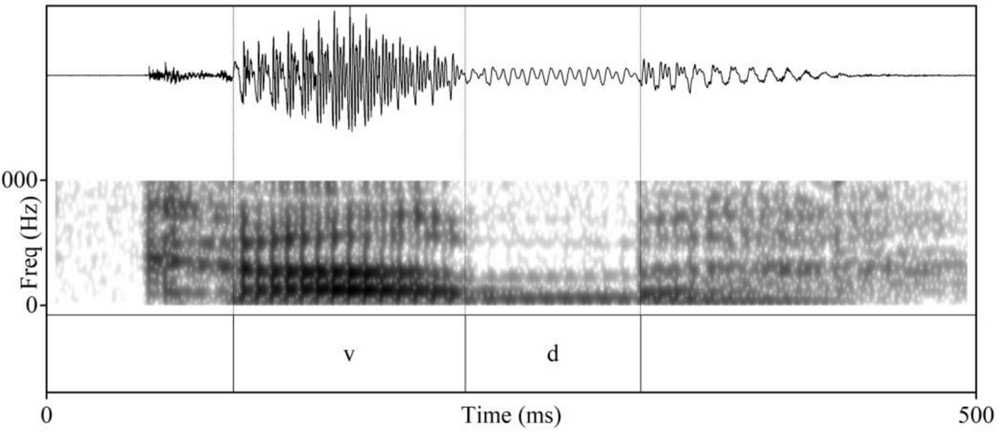


**Figure 5: Segmented realisation of the manipulated neutralised target word *cub* by the native speaker. v = vowel /ʌ/, lasting for 125 ms, d = voiced consonant /b/.**

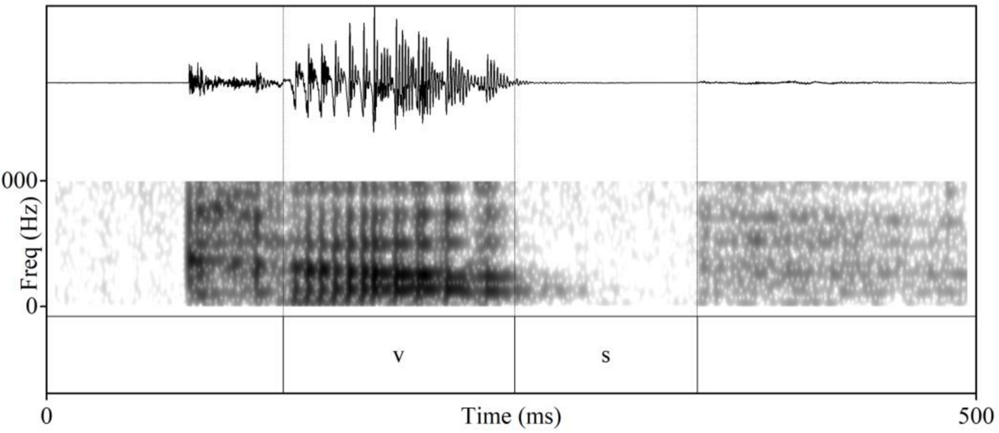


**Figure 6: Segmented realisation of the manipulated neutralised target word *cup* by the native speaker. v = vowel /ʌ/, lasting for 125 ms, s = voiceless consonant /p/.**

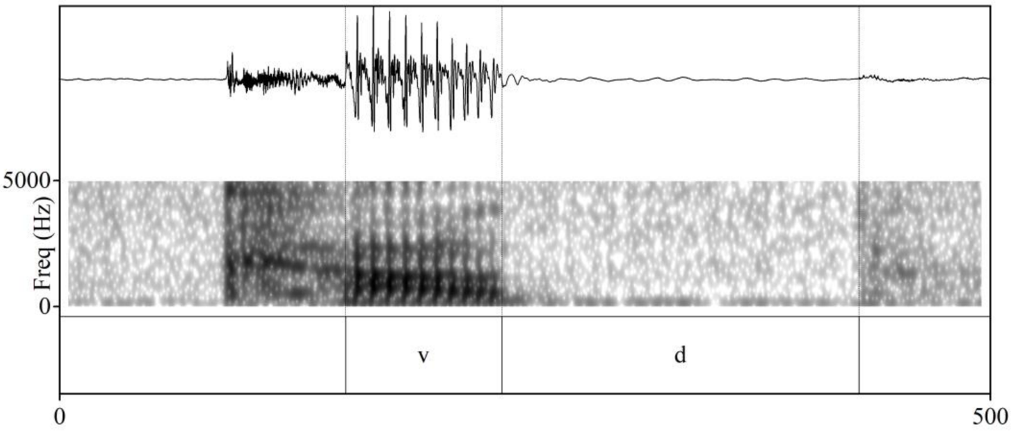


**Figure 7: Segmented realisation of the manipulated neutralised target word *cub* by the non-native speaker. v = vowel /ʌ/, lasting for 86 ms, d = voiced consonant /b/.**

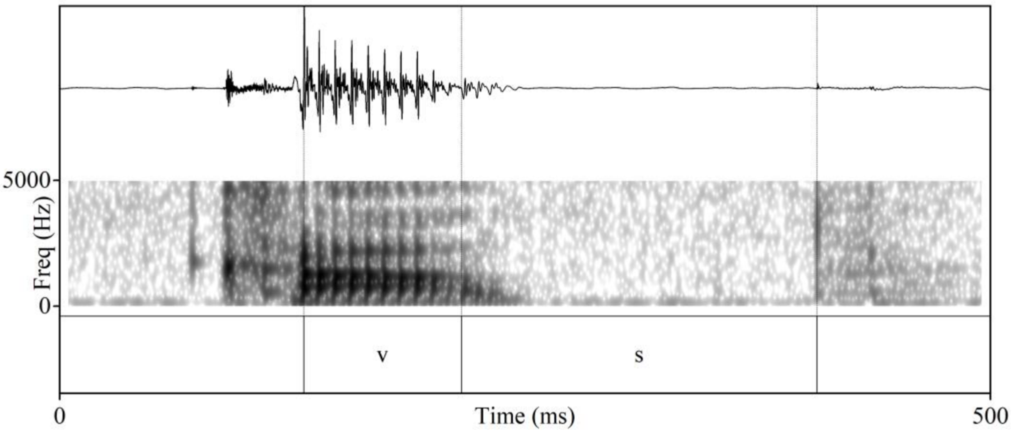


**Figure 8: Segmented realisation of the manipulated neutralised target word *cup* by the non-native speaker. v = vowel /ʌ/, lasting for 86 ms, s = voiceless consonant /p/.**

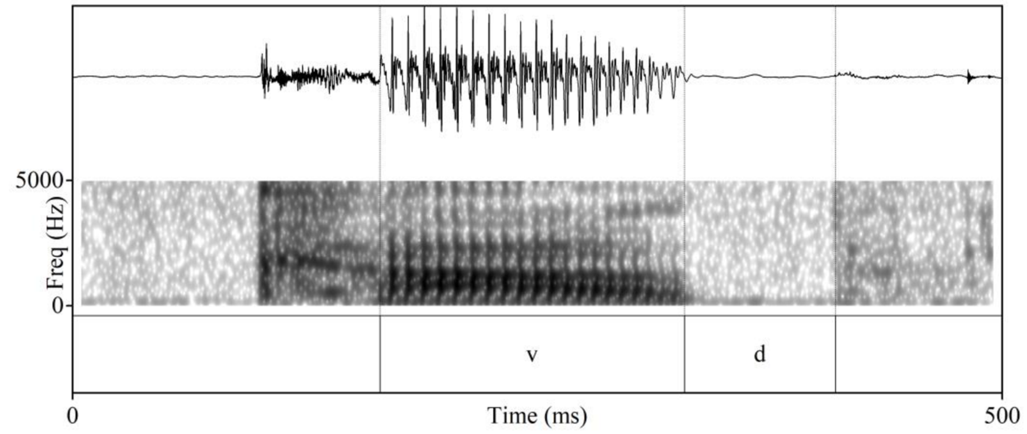


**Figure 9: Segmented realisation of the manipulated native-like target word *cub* by the non-native model speaker. v = vowel /ʌ/, lasting for 162 ms, d = voiced consonant /b/.**

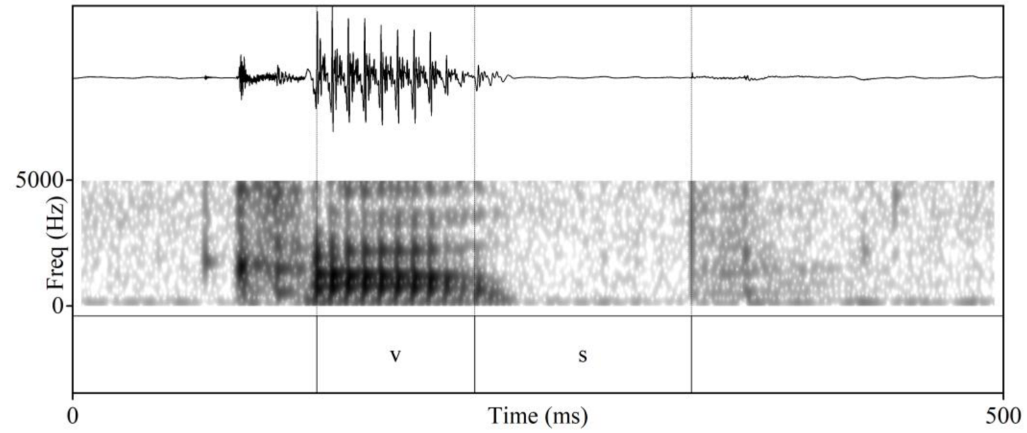


**Figure 10: Segmented realisation of the manipulated native-like target word *cup* by the non-native model speaker. v = vowel /ʌ/, lasting for 88 ms, s = voiceless consonant /p/.**

## 2.3 Procedure

The procedure of the experiment was first explained to the participants, who were answered all of their eventual questions, and they signed an informed consent form (see Appendix C: Informed consent for the exact wording of the consent). They were then seated in front of a laptop in a soundproof booth. The experimental session took place in a speech laboratory at the Faculty of Arts, Palacký University in Olomouc. The recordings were obtained using a Zoom H4n Handy Recorder microphone (recorded directly on the computer with a sampling rate of 44.1 kHz and a with a furry windscreen placed on the microphone). The experimental stimuli were played through a set of noise-cancelling headphones (Bose Noise Cancelling 700), with the noise cancellation function set to 0, so that the participants would be able to hear themselves. The administrator of the experiment was seated next to the participant. This was done in order to allow for control of the produced recordings. Occasionally, in the case of an inadequate production, the experiment allowed pausing of the recording session and repeated recording of the item.

With the administrator's help, the experiment started in a Praat script (Boersma and Weenink 2022) on the screen. Each participant was assigned an anonymous ID and supplied their age and gender. The participants then clicked through the instructions themselves at their own pace. There were three elicitation parts for each participant – baseline, shadowing 1, and shadowing 2 (see Table 1 for the whole design of the procedure).

In the baseline part, similarly to the model speakers' recording sessions, the participants first familiarised themselves with all of the 48 words on a screen sorted into an 8x5 grid. When unsure about the meaning, they were instructed to consult a wordlist on paper in front of them (see Appendix D: Wordlist). The words then started to appear automatically on the screen one by one in random order, which was different for each participant. The timing of each trial was constant to account for speech tempo influence. They were instructed to say the word as soon as it appears on the screen. The baseline productions were thus elicited to establish the subjects' pronunciation prior to the models' exposure.

In the shadowing 1 phase, one of the two model-L1-inducing recordings was played to the participants to reinforce the L1 status impression of the model speaker. The participants were moreover informed that they were about to listen to a native or a non-native speaker. This was done because preliminary tests showed some participants were not able to discern the nativeness status of the speakers. An image with the words "RESTAURANT MENU" was visible on the screen during the recording. After that, the words again started randomly appearing, this time accompanied by the model speaker's recording of that particular item. The model speaker's nativeness status corresponded to the previous model-L1-inducing recording. In other words, if the participant heard a menu recording by a native speaker, they would now hear the stimuli pronounced by the native speaker.

To mitigate the effect of the recently heard words on the next elicitation part, a short pause followed. During this pause, the participants completed a Lexical Test for Advanced Learners of English (LexTALE) to assess their proficiency. This is a test of vocabulary knowledge based on a lexical decision task, in which the participants are shown sixty lexical items one by one and are asked to judge whether they believe these to be existing words in English or not. Although primarily a lexical knowledge test, it has been shown to correlate with general language proficiency in English and to outperform proficiency self-ratings (Lemhöfer and Broersma 2012). Moreover, it is quick and easy to administer in a laboratory setting, taking no more than four minutes. At the end of the test, each participant receives a score from 0 to 100. LexTALE scores from 80 to 100 roughly correspond to C1 and C2 CEFR levels, scores from 60 to 80 are estimated by the authors to correspond to B2 level, and scores below 59 are predicted to correlate with proficiency level B1 and lower. The participants scored an average of 80 (min = 61, $Q_1$ = 76, median = 79, $Q_3$ = 88, max = 100).

After the LexTALE test, the participants were asked to watch the first eight minutes of a relaxing BBC video depicting various wildlife shots. It is titled "Unwind with 20 minutes in nature | Springwatch – BBC" and can be found online at https://www.youtube.com/watch?v=VTsz_tO3iSc. The first few seconds showing the title of the video have been trimmed so as not to confuse the participants about the amount of time they would be asked to watch it. The video contains no speech but there are different names of locations appearing on the screen throughout the video.

After the pause, the last elicitation part (shadowing 2) followed, again preceded by a model-L1-inducing recording and information about the model speaker's L1 status. The order of the stimuli across shadowing 1 and shadowing 2 was counterbalanced within the naturalistic (group 1 and 2) and non-naturalistic groups (group 3 and 4). Groups 1 and 2, and groups 3 and 4 were thus exposed to the same stimuli but in the opposite order. The naturalistic group heard a native model speaker featuring CVIVDV and a non-native model with neutralised CVIVDV values (in this order for group 1, in the opposite order for group 2). The non-naturalistic groups stimuli did not match in the CVIVDV values expected from either the native or the non-native speaker. The participants listened to a native model with neutralised CVIVDV, or non-native like, and to a non-native model with CVIVDV values that were taken from the original native model recordings (in this order for group 3, in the opposite order for group 4). See Table 1 below for an overview of the procedure.

The participants were then asked to complete a questionnaire in Google Forms about their basic personal data (age, gender), native language, study programme, impairments (visual, hearing, neurological), attitudes to English accents, and attitudes towards non-native accentedness. The last set of attitudinal questions or statements (see Appendix E: Openness to non-native accentedness  for their wording) was used to calculate the openness to non-native accentedness score. The participants expressed their attitude towards the statements on a 5-point Likert scale (1 for "Strongly agree" on one end and 5 for "Strongly disagree" on the opposite end). One statement (numbered 4 in the Appendix) was excluded from the analysis since participants found it confusing and not entirely clear. Scores from the first three statements, which were in favour of openness, were added to the score while answers to statements 5 to 7 were subtracted from the final score. The scores were then rescaled to 0–100.

After all of the elicitation parts and the questionnaire, a short debriefing followed. Most of the participants have reported that they have not recognised CVIVDV in the stimuli. When asked directly, though, the majority reported that they knew about the existence of the phenomenon as a result of their phonetics courses. At the end, the purpose of the experiment was explained to them.

The whole experimental session lasted for no more than 40 minutes. The participants received a book voucher worth CZK 100 as an incentive. Data collection took place during the faculty exam period in January and February 2023.

| **order** | **naturalistic** | | **non-naturalistic** | |
|---|---|---|---|---|
| | **group 1** | **group 2** | **group 3** | **group 4** |
| 1 | *baseline* | *baseline* | *baseline* | *baseline* |
| 2 | [+native] recording | [–native] recording | [+native] recording | [–native] recording |
| 3 | *shadowing 1* | *shadowing 1* | *shadowing 1* | *shadowing 1* |
| | [+native] voice | [–native] voice | [+native] voice | [–native] voice |
| | [+native] CVIVDV | [–native] CVIVDV | [–native] CVIVDV | [+native] CVIVDV |
| 4 | LexTALE | LexTALE | LexTALE | LexTALE |
| 5 | video | video | video | video |
| 6 | [–native] recording | [+native] recording | [–native] recording | [+native] recording |
| 7 | *shadowing 2* | *shadowing 2* | *shadowing 2* | *shadowing 2* |
| | [–native] voice | [+native] voice | [–native] voice | [+native] voice |
| | [–native] CVIVDV | [+native] CVIVDV | [+native] CVIVDV | [–native] CVIVDV |
| 8 | questionnaire | questionnaire | questionnaire | questionnaire |

**Table 1: Procedure of the experiment displaying the different stimuli between naturalistic and non-naturalistic groups, and the different order of stimuli within these groups. Native features are in turquoise, non-native in pink.**

# 3 Data analysis

The segmentation, annotation, and measurements of the recordings were done in Praat (Boersma and Weenink 2022) observing the same criteria as with the model speakers' productions described in section 2.1. When a boundary could not be determined or when the recording was corrupted, it was excluded from the analysis. Following Winter (2020), the subsequent data manipulation and statistical analyses were carried out using R 4.1.0 (R Core Team 2022) and especially the lme4 (Bates et al. 2015), tidyverse (Wickham et al. 2019), ggeffects (Lüdecke 2018), emmeans (Lenth 2023), and afex (Singmann et al. 2023) packages. The complete R script is supplemented in Appendix F: R script.

## 3.1 Raw data and its discussion

The raw distribution of the target vowel durations is plotted in Figure 11 and that of the target vowel / coda constriction duration ratio is shown in Figure 12. This data is pooled across all participants and words, so some caution is necessary when interpreting these plots. Still, a preliminary evaluation of the research questions can be attempted.

Figure 11 shows the log-transformed target vowel durations split by task, presence of the vowel and coda duration cue vs lack thereof, and the model speaker's nativeness status. Figure 12 uses the same plot design but plots instead the ratio between vowel duration and coda constriction duration (V/C ratio). The ratio was computed as the log-transformed vowel duration divided by the log-transformed consonant duration for each individual word of each participant (see section 1.1 in Appendix F: R script for the calculation). Notice that in each figure, there are always two quadrants that show the same baseline data distribution. This is the result of the design of the experiment wherein each participant completed one baseline test and then two shadowing tests. Also note that obviously the model speaker and cue presence variables do not apply to the baseline data itself (elicited using a reading task without any auditory stimuli), which is included in the figure for comparison, always shown alongside the shadowing data of the respective participant.

Although the interpretation of Figure 11, i.e., a plot with (log-transformed) vowel durations, would be more intuitive, the measure of the ratio between vowel and coda (log-transformed) durations (as in Figure 12) is considered to be superior. This is because, although the experiment was designed in such a way that speech tempo was constant across baseline and shadowing, there could be small local differences of speech tempo and using the ratio between vowel and coda durations for each word helps factor out these differences.

The research questions are now addressed preliminarily:

**Q$_1$: Do Czech speakers of English exhibit CVIVDV?**

Looking at the baseline productions of vowels in voiced contexts in Figure 11 and comparing them to the baseline values in voiceless contexts within the same quadrants, it can be observed that in each case the distributions are shifted towards longer values when they are preceding voiced consonants than when they are preceding voiceless consonants. There seems to be great variety in the distribution of the values. The same pattern can be observed in Figure 12 for the V/C ratio, tentatively suggesting that speakers indeed exhibit CVIVDV already in their baseline productions.

**Q$_2$: Do Czech learners of English exhibit imitation?**

To allow assessing the potential effects of exposure during shadowing, the violin plots in Figure 11 and Figure 12 show the mean model-speaker values as the horizontal lines in each panel (the very top line is the model speaker's voiced value, the line below is for the voiceless contexts, and the single lines for the removed-cue conditions are the means of the values used in these conditions). Within individual quadrants in both figures, the distribution of the values in the baseline vs shadowing values suggests that there are indeed shifts towards model speakers' values. The more evenly distributed values in the baseline seem to be more grouped around the heard values in the shadowing. Therefore, based on the violin plots alone, there are reasons to believe that the Czech learners indeed exhibited phonetic imitation in their L2 English.

**Q$_3$: Are Czech learners of English more likely to imitate a native or a non-native model with naturalistic CVIVDV values?**

For this question, the values measured in the condition of the English model speaker and the V/C duration cue present must be compared to the values elicited using the

Czech model speaker with V/C duration cue removed. These conditions correspond to the lower-left vs the upper-right panel in both Figure 11 and Figure 12. Notice that there is a suggestion of a larger shift from baseline to shadowed values for the English model speaker on the right than there is for the Czech model on the left. Moreover, the distribution of the shadowing values upon exposure to the English model speaker seems to be more compact than for the lower-left Czech model panel. There thus appears to be a greater differentiation between voiced and voiceless distributions in the shadowing for the English model speaker.

**Q4: Do Czech learners of English imitate a native and a non-native model even with non-naturalistic values?**

For the participants who were exposed to the English model with cue removed, there appears to be a bimodal distribution in the shadowing values in voiceless contexts. Some durations were thus shifted towards the non-naturalistic values upon exposure to the stimuli. It isn't clear from the pooled distribution, however, whether these were values from the same speakers or whether only some words were imitated. In other words, it cannot be decided whether some speakers systematically imitated the non-naturalistic model, while other did not, or whether all speakers imitated only some words, but not others. For the voiced conditions, the shadowed values seem to be more centred around the model speaker's value, which is more pronounced with the English model as opposed to the Czech one, where the values are more evenly distributed. The voiceless shadowing values are also less clearly bimodal for the Czech model with cue present. It is thus suggested that participants imitate both model speakers even with non-naturalistic values, but the effect seems to be stronger with the English model.

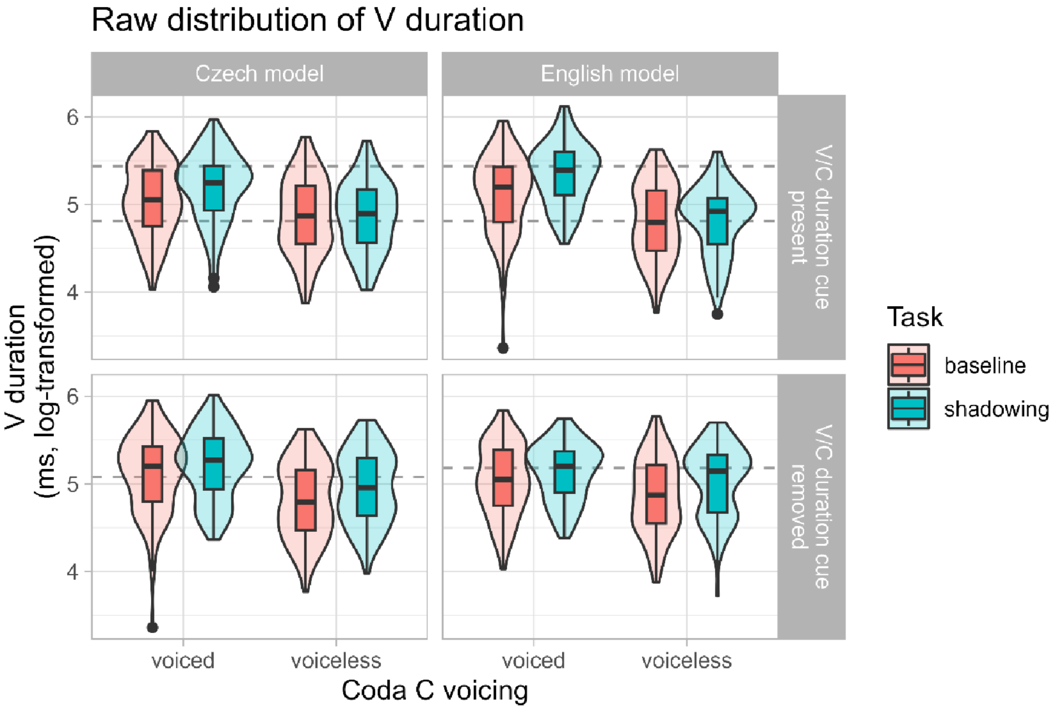


**Figure 11: Violin plot of raw distribution of vowel duration values. Plot split by accent of the model speaker, presence or lack of cue, task, and coda voicing. The models' values are shown for comparison as the horizontal dashed lines.**

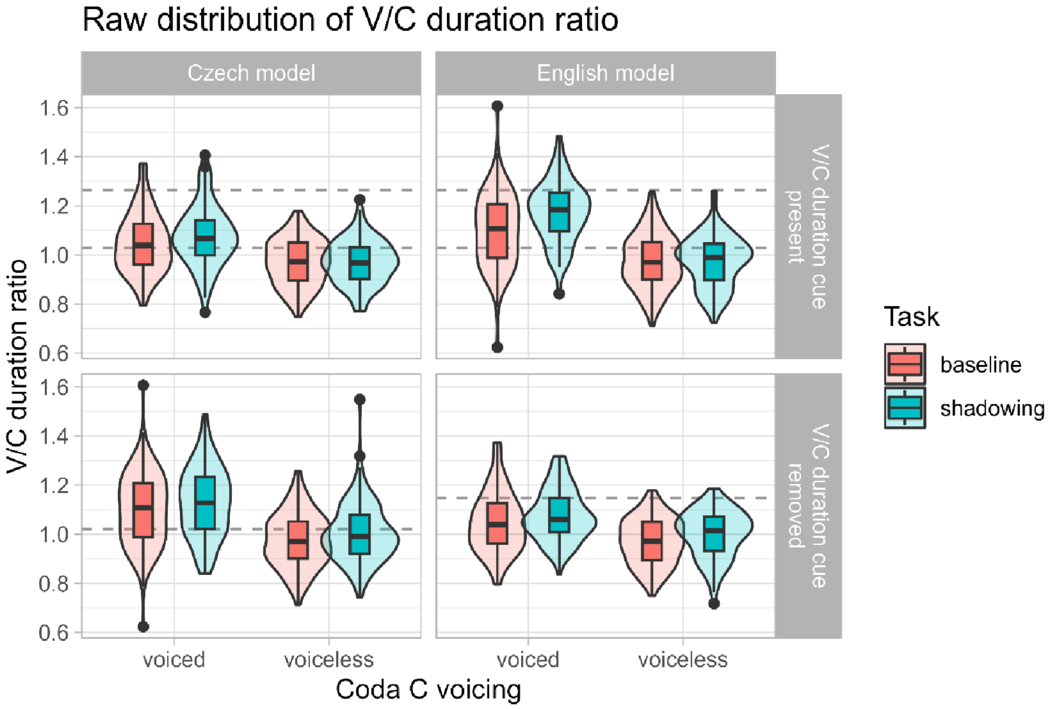


**Figure 12: Violin plot of raw distribution of V/C ratio. Plot split by accent of the model speaker, presence or lack of cue, task, and coda voicing. The models' values are shown for comparison as the horizontal dashed lines.**

## 3.2 Baseline model results and discussion

The first model addresses the participants' baseline performance with respect to proficiency and attitudinal factors. As described in the Methodology section, the participants took a LexTALE test and completed a questionnaire, which included questions used to compute a score of openness to non-native accentedness. Both these variables are hypothesised to predict the baseline differences in the realisation of the target vowel duration in voiced vs voiceless coda contexts. Speakers with a low score of openness to non-native accentedness might be more motivated to minimise their accentedness and hence be more likely to exhibit native-like CVIVDV. At the same time, participants with a higher LexTALE scores might have acquired English to a higher level, including the phenomenon in question.

As explained in section 3.1 above, the V/C ratio is believed to factor out any possible speech tempo influence. Henceforth, for the sake of limiting the number of models on data elicited from the same groups of participants, only the V/C ratio will be modelled and not V durations alone.

The baseline data were fitted to a linear mixed-effects model (estimated using REML and the BOBYQA optimiser) to predict the ratio between log-transformed target vowel duration and log-transformed coda constriction duration as a function of coda voicing, openness score, and LexTALE score, including all the possible interactions between the predictors (see section 3.1 in the R script supplemented in Appendix F: R script). As for these fixed effects, coda voicing is a sum-coded two-level categorical predictor (voiceless -1 vs voiced 1) while LexTALE and openness are continuous and mean-centred. Regarding the random effects, the model estimated by-word varying intercepts and by-participant varying intercepts and slopes for voicing with the slope/intercept correlation included. The formula in R for this model was thus as follows: V/C ratio ~ voicing * LexTALE * openness + (1 + voicing | participant) + (1 | word). The total explanatory power of the model is substantial (conditional $R^2$ = 0.71), and the part related to the fixed effects alone (marginal $R^2$) is of 0.19. 95% Confidence Intervals (*CI*) and *p*-values were computed using a Wald *t*-distribution approximation.

After looking at the coefficients of the fixed effects estimated by the model, no significant effect was found for the openness score factor alone nor for any interaction including the openness score. On the other hand, significant effect was found for the interaction between the LexTALE score and voicing.

There was thus reason to omit the openness factor from the model. Before that, an ANOVA comparison of the model just described was made with a model differing only in omitting the openness factor. The models were fitted without REML to allow for the statistical comparison. A likelihood ratio test of the model including the openness factor against the model without the openness effect did not reveal a significant difference between the two models [$\chi^2$ (4) = 3.026, p = 0.553].

For this reason, the openness factor was dropped from the model and the data were fitted to a new model, whose design and predictors remain the same, except for the missing predictor of openness (see section 3.2 in the R script supplemented in Appendix F: R script). The reduced model had the following formula: V/C ratio ~ underlying coda voicing * LexTALE + (1 + underlying coda voicing | participant) + (1 | word). The total explanatory power of this model is substantial (conditional $R^2$ = 0.70), and the part related to the fixed effects alone (marginal $R^2$) is of 0.18, suggesting that the random effects have a significant bearing on the outcome.

A histogram, Q–Q plot, and residual plot of the model were assessed visually, and it was judged that the assumptions of linear regression (normality and constant variance or homoscedasticity) have been met. In other words, the model's residuals are approximately normally distributed and their spread across the range of fitted values is approximately equal.

| Predictor | Estimate | *SE* | *df* | *CI* | *t*-value | *p*-value |
|---|---|---|---|---|---|---|
| (Intercept) | 1.024 | 0.021 | 30.08 | [0.98, 1.06] | 49.027 | **<0.001** |
| voicing1 | 0.051 | 0.018 | 20.74 | [0.01, 0.09] | 2.747 | **0.012** |
| LexTALE | 0.002 | 0.001 | 22.09 | [-8.63e-04, 4.09e-03] | 1.281 | 0.214 |
| voicing1:LexTALE | 0.002 | 0.001 | 21.75 | [1.16e-03, 3.78e-03] | 3.714 | **<0.001** |

**Table 2: Coefficients of the fixed effects estimated by the baseline model. *p*-values lower than 0.05 are in bold.**

As shown in Table 2, the model's intercept is at 1.024 (*SE* = 0.021, *CI* [0.98, 1.06], *t* = 49.03, *p* < 0.001). This corresponds to the estimated grand mean of V/C ratio value between the two voicing conditions and LexTALE. Within this model, the main effect of voicing1 (present) is positive (*β* = 0.051, *CI* [0.01, 0.09], *t* = 2.75, *p* = 0.012). This estimate relates to the difference between the mean of the estimate V/C ratio in voiced contexts and the intercept. When multiplied by 2, this gives the average distance between the voiced and voiceless conditions. The slope of the LexTALE score is positive but statistically non-significant (*β* = 0.002, *CI* [-8.63e-04, 4.09e-03], *t* = 1.28, *p* = 0.214). This is the average effect across the voiced and voiceless contexts. When the LexTALE is added in interaction with voicing1 though, it becomes highly significant (*β* = 0.002, *CI* [1.16e-03, 3.78e-03], *t* = 3.71, *p* < 0.001).

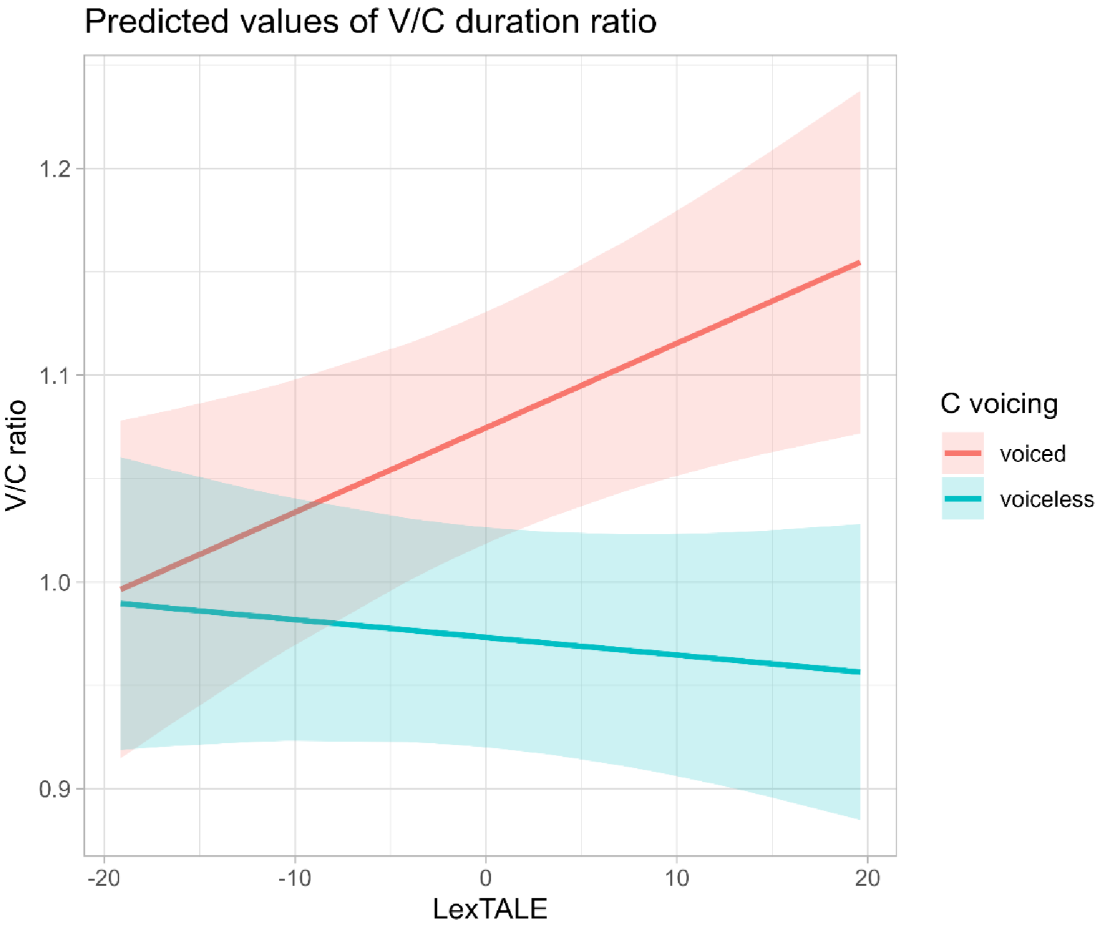


**Figure 13: Plot of predicted values of V/C duration ratio as a function of LexTALE and voicing.**

Figure 13 above visualizes this interaction. It can be seen that while for voiceless coda contexts, increasing LexTALE scores are associated with only very mildly decreasing V/C ration values, there is a clear increase for the voiced contexts.

Since the model only estimates the baseline values, it is only capable of addressing $Q_1$, repeated below for convenience.

**$Q_1$: Do Czech speakers of English exhibit CVIVDV?**

The model seems to confirm the preliminary observations from the raw V/C ratio data shown in Figure 12. The effect of coda voicing context is significant and in the expected direction, confirming that the advanced Czech learners of English participating in this study exhibit CVIVDV. Their vowels preceding voiced consonants are reliably longer than those preceding voiceless ones. At the same time, there is a clear interaction between coda voicing and the participants' LexTALE scores. In their baseline productions, then, Czech speakers of English are observed to exhibit CVIVDV but its degree is greater in speakers with higher LexTALE scores (i.e., more proficient speakers), whereas less proficient speakers seem to make little or no difference between the two contexts.

This is reminiscent of the results by Fejlová (2013), who found that Czech speakers with a strong accent in English employ CVIVDV to a lesser degree than speakers with a less strong accent. It is also in line with Skarnitzl and Šturm's (2016) study, wherein strong-accented Czech speakers failed to produce significant vowel duration differences between the two coda voicing contexts. It might be the case that accentedness correlates negatively with general proficiency, here measured by LexTALE. As speakers are becoming more proficient in English, they are also gradually eliminating the interference of Czech pronunciation rules from their L2 productions. The LexTALE test, primarily a test of lexical knowledge, is thus useful not only for the assessment of lexical knowledge specifically but also for assessing the learners' general second language proficiency and may be reflected in the degree of acquisition of L2 pronunciation.

## 3.3 Shadowing model results and discussion

Next, the V/C ratio values in the shadowing conditions were modelled. There are predictors (such as model speaker accent and cue presence) that only make sense for the shadowing, so fitting the baseline values along with the shadowing values would not be logical. At the same time, ignoring the baseline durations completely would not be reasonable either since participants have different input behaviours (as seen in the baseline model, proficient speakers already exhibit CVIVDV and therefore have less space for exposure-induced shifts). To account for the baseline values without having them included in the model explicitly as individual datapoints, for each measured V/C ratio the baseline V/C ratio measure for that specific word and speaker was subtracted from the shadowing V/C ratio value. In other words, a difference between the baseline and shadowing session 1, and between baseline and shadowing session 2 was computed per speaker and per word. Since some values were missing, there was a total number of 921 datapoints submitted to the modelling.

### *3.3.1 The effect of openness on shadowing productions*

The predictors included in the model (see section 3.3 in the R script supplemented in Appendix F: R script) are model speaker accent (sum-coded, English -1 vs Czech 1), coda voicing cue (sum-coded, present -1 vs removed 1), phonological coda voicing (sum-coded, voiceless -1 vs voiced 1 speech sound). Since the LexTALE score was found to be an important predictor in the baseline productions, it is also to be included in the model. Although openness to non-nativeness was not found to have a bearing on CVIVDV baseline production, it is still reasonable to considered it as a predictor in this shadowing model since it could potentially influence the participants' imitation behaviour, especially towards the Czech non-native model. For this reason, a comparison of two models was conducted. One model included the predictors of cue presence, voicing, accent, and LexTALE, and the other model included all of the above plus openness (each model including all the possible interactions).

The models were fitted without REML to allow for the statistical comparison. A likelihood ratio test of the model including the openness factor [specified in R using the formula V/C ratio (difference from baseline) ~ cue presence * underlying coda voicing * accent * LexTALE * openness + (1 + coda voicing | participant) + (1 | word)] against the model without the openness effect [V/C ratio (difference from baseline) ~ cue presence * underlying coda voicing * accent * LexTALE + (1 + coda voicing | participant) + (1 | word)] did not reveal a significant difference between the two models [$\chi^2$ (16) = 22.102, $p$ = 0.140].

Furthermore, the summary of the model including the openness factor showed that openness alone had no significance, neither was there an interaction that would involve the accent and openness together, which was the expected influence. The openness was a relatively crude measure aggregating the answers to several questions on the questionnaire and possibly it was not a very reliable measure of the participants' complex attitudes to accentedness, native speech and native speakers, and so it is not too surprising that was not found to predict reliably the V/C ratios in the shadowed production or interact with the predictor of accent. Since there is no significant improvement of goodness of fit if the openness predictor is included, it is dropped from consideration and the following model is only modelled with the LexTALE score.

### *3.3.2 Final shadowing model results and discussion*

The same data as in the previous shadowing models were thus fitted to a linear mixed-effects model (estimated using REML and the BOBYQA optimiser) to predict differences between participants' baseline and shadowing productions in the values of the ratio of the target vowel and the coda constriction duration (i.e., the same response variable as in the two previous models). These were modelled as a function of model speaker accent (sum-coded, English -1 vs Czech 1), cue presence (sum-coded, present -1 vs removed 1), underlying coda voicing (sum-coded, voiceless -1 vs voiced 1), and the LexTALE scores (mean-centred). Interactions were included between all the predictors. To account for the variation between participants and individual words and for the fact that the design of the experiment employs repeated measures, incorporated in the model were also random effects – by-word varying intercepts and by-participant varying intercepts and slopes for coda voicing with the slope/intercept correlation also included.

The formula for this final model has the following syntax in R: V/C ratio (difference from baseline) ~ cue presence * coda voicing * model speaker accent * LexTALE + (1 + coda voicing | participant) + (1 | word). The total explanatory power of the model is moderate (conditional $R^2$ = 0.23), and the part related to the fixed effects alone (marginal $R^2$) is of 0.07. 95% confidence intervals and p-values were computed using a Wald *t*-distribution approximation. The relatively low conditional $R^2$ could be explained by the fact that the data was considerably aggregated. Since there is always some degree of noise present in each measure, and the fitted values underwent several calculations (first dividing V duration by C duration, then subtracting the value of this ratio for the baseline condition from the shadowing ratio value), the noise could have added up, making the model less capable of explaining the variance of the residuals.

A histogram, Q–Q plot, and residual plot of the model were assessed visually, and it was judged that the assumptions of linear regression (normality and constant variance or homoscedasticity) have been met. In other words, the model's residuals are approximately normally distributed and their spread across the range of fitted values is approximately equal.

Table 3 below shows the coefficient estimates for the fixed effects, standard errors, degrees of freedom, confidence intervals, *t*-values, and *p*-values. Figure 14 displays the fitted values.

| Predictor | Estimate | *SE* | *df* | *CI* | *t*-value | *p*-value |
|---|---:|---:|---:|---:|---:|---:|
| (Intercept) | 2.74E-02 | 8.64E-03 | 21.9 | [0.01, 0.04] | 3.172 | **0.004** |
| cue1 | -5.77E-04 | 3.07E-03 | 848.0 | [-6.60e-03, 5.44e-03] | -0.188 | 0.851 |
| voicing1 | 1.27E-02 | 5.41E-03 | 21.4 | [2.04e-03, 0.02] | 2.339 | **0.029** |
| accent1 | -5.50E-03 | 3.07E-03 | 848.0 | [-0.01, 5.22e-04] | -1.792 | 0.073 |
| LexTALE | -5.86E-05 | 9.48E-04 | 20.1 | [-1.92e-03, 1.80e-03] | -0.062 | 0.951 |
| cue1:voicing1 | -1.46E-02 | 3.07E-03 | 848.2 | [-0.02, -8.60e-03] | -4.767 | **<0.001** |
| cue1:accent1 | 5.16E-03 | 8.37E-03 | 20.0 | [-0.01, 0.02] | 0.616 | 0.545 |
| voicing1:accent1 | -7.78E-03 | 3.07E-03 | 848.1 | [-0.01, -1.77e-03] | -2.538 | **0.011** |
| cue1:LexTALE | -3.28E-05 | 3.52E-04 | 848.4 | [-7.23e-04, 6.57e-04] | -0.093 | 0.926 |
| voicing1:LexTALE | -9.39E-04 | 5.64E-04 | 19.9 | [-2.05e-03, 1.68e-04] | -1.665 | 0.112 |
| accent1:LexTALE | 7.53E-04 | 3.52E-04 | 848.6 | [6.32e-05, 1.44e-03] | 2.142 | **0.032** |
| cue1:voicing1:accent1 | 9.78E-03 | 4.96E-03 | 19.5 | [4.17e-05, 0.02] | 1.971 | 0.063 |
| cue1:voicing1:LexTALE | 7.83E-04 | 3.52E-04 | 849.1 | [9.28e-05, 1.47e-03] | 2.227 | **0.026** |
| cue1:accent1:LexTALE | -4.40E-04 | 9.48E-04 | 20.1 | [-2.30e-03, 1.42e-03] | -0.464 | 0.648 |
| voicing1:accent1:LexTALE | 5.84E-04 | 3.52E-04 | 849.2 | [-1.06e-04, 1.27e-03] | 1.662 | 0.097 |
| cue1:voicing1:accent1:LexTALE | -9.76E-05 | 5.64E-04 | 19.9 | [-1.20e-03, 1.01e-03] | -0.173 | 0.864 |

**Table 3: Coefficients of the fixed effects estimated by the final shadowing model. *p*-values lower than 0.05 are in bold.**

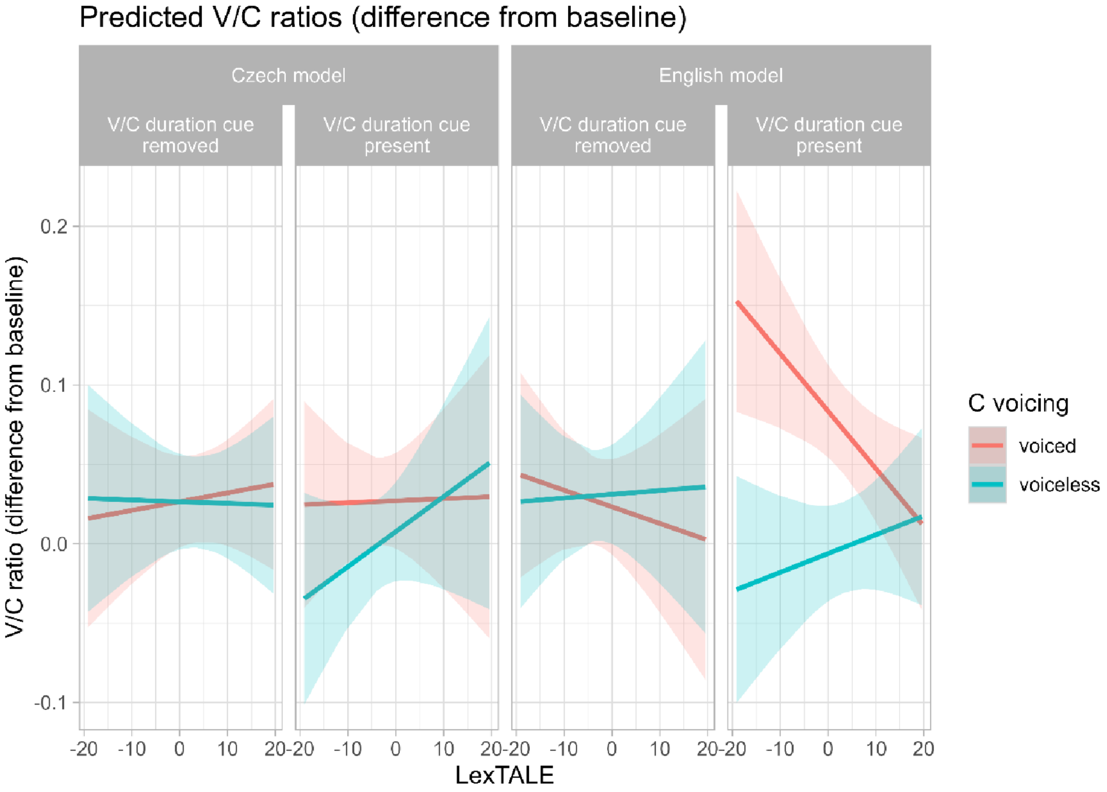


**Figure 14: Plot of predicted values of the V/C ratio difference between baseline and shadowing. Plot split by accent of the model speaker, presence or lack of cue, and coda voicing.**

Looking at the model's summary in Table 3, we can see that the intercept (i.e., mean of means) is significantly higher than 0 (*β* = 2.74E-02, *CI* [0.01, 0.04], *t* = 3.17, *p* = 0.004), meaning that on average, the ratio of vowel to final coda constriction duration was larger in shadowing than in the baseline, indicating an overall lengthening of the V duration in proportion to the duration of the coda. Within this model, the slope for voicing (present) is positive (*β* = 1.27E-02, *CI* [2.04e-03, 0.02], *t* = 2.34, *p* = 0.029). This relates to the difference between the mean of the estimated V/C ratio differences in voiced contexts and the intercept, indicating an overall larger V/C ratio in the voiced than in the voiceless coda contexts. The interaction term between cue presence (removed) and coda voicing (present) is significant (*β* = -1.46E-02, *CI* [-0.02, -8.60e-03], *t* = -4.77, *p* < 0.001), indicating that while the V/C ratio values in the removed-cue conditions largely overlap for voicing, those in the present-cue conditions show a separation between the lower V/C values for voiceless and the higher for voiced codas (see Figure 14). The interaction between voicing (present) and accent (Czech) is also significant (*β* = -7.78e-03, *CI* [-0.01, -1.77e-03], *t* = -2.54, *p* = 0.011), indicating a greater separation of the V/C ratio values between the voiced and voiceless contexts in the expected direction for the English than for the Czech model speaker. The interaction between accent (Czech) and LexTALE is positive (*β* = 7.53e-04, *CI* [6.32e-05, 1.44e-03], *t* = 2.14, *p* = 0.032), which suggests that on average, the effect of LexTALE depends on the accent the participants were exposed to (cf. Figure 14). Finally, the three-way interaction between cue (removed), voicing (present), and LexTALE is positive (*β* = 7.83e-04, *CI* [9.28e-05, 1.47e-03], *t* = 2.23, *p* = 0.026). While the values predicted for voiced vs voiceless coda context for the cue-removed conditions mostly overlap and show little or no effect of LexTALE, those in the cue-present condition show an expectable effect of LexTALE: lower LexTALE scores predict greater baseline-to-shadowing shifts and higher scores smaller shifts (since higher LexTALE scores were associated with baseline production closer to native values giving less space for imitation-induced shifts.

With the help of Figure 14 and the coefficients in Table 3, the research questions 2 to 4 can be addressed.

**Q$_2$: Do Czech learners of English exhibit imitation of CVIVDV?**

Overall, participants seem to have adjusted their V/C ratio values between the baseline and shadowing elicitation parts, as evidenced by the intercept being significantly different from zero. The strongest effect is observed for the voiced contexts, especially in the English accent cue present condition with low LexTALE scores (cf. Figure 14). Participants with lower LexTALE scores changed their V/C productions from baseline to shadowing (thus enlarging the differences between voiced and voiceless contexts) more than did participants with higher LexTALE scores. As discussed in section 3.2, these more proficient participants already exhibit CVIVDV in the baseline to begin with, resulting in reliably longer vowels before voiced codas as opposed to voiceless ones. High LexTALE participants were thus already close to the models' productions in their baseline, and so they had less space and/or reason to shift their productions towards the direction of the model speaker.

**Q$_3$: Are Czech learners of English more likely to imitate a native or a non-native model with naturalistic values?**

For this question, the English model speaker with the CVIVDV cue present must be compared to the Czech model with the cue removed (i.e., the last and first facet of Figure 14, respectively). There is an interaction of accent and coda voicing, and the subjects seem to produce significantly distinct V/C ratio values for the two voicing conditions when exposed to the English model as opposed to the Czech one. This is especially evident for the values preceding underlyingly voiced obstruents. The native English model speaker with native CVIVDV values is thus preferred by Czech speakers of English in terms of imitation to the Czech-accented speaker. This result, to some extent, replicates that of Zając and Rojczyk (2014), whose participants converged with the native model speaker and diverged from the non-native one.

**Q$_4$: Do Czech learners of English imitate a native and a non-native model even with non-naturalistic values?**

The non-naturalistic values correspond to the panels of the English model speaker with cue removed and the Czech model with cue present (i.e., the third and second panel of Figure 14, respectively). The participants seem to have imitated the English model speaker without the cue significantly less than when they were exposed to the same speaker with the V/C cue present. As for the Czech model with cue present, low-LexTALE speakers (as opposed to higher-scoring participants) seem to make a difference between the two voicing conditions, which is similar to speakers who were exposed to the English model speaker with cue present. The effect for the Czech speaker, however, is small and not significant. Since the model shows the mean values, it is possible that imitation could have occurred for some participants, as indicated by the distribution of raw values in Figure 11 and Figure 12.

# 4 General discussion and conclusion

The thesis expands on a previous bachelor thesis, originally formed as a research proposal. The literature review shows that pronunciation changes over the course of one's lifetime in the direction of the surrounding speech. Phonetic adjustments have also been observed over the course of several months, during a single conversation, but also in non-interactive conditions after the immediate exposure to stimuli, wherein subjects are observed to imitate the model speaker. Social-psychological and linguistic factors modulate the imitation.

The specific phenomenon that is being investigated is the varying vowel duration based on the voicing of the syllable coda (coda-voicing-induced vowel duration variability or CVIVDV). It seems that the contrasting durations are a language universal to some extent, although in English the effect is much larger, perhaps even phonologised and used as a reliable cue to the voicing of the coda. The effect in the Czech language is either negligible or absent, which might be the result of it potentially clashing with the Czech phonemic vowel length and, for word-final obstruents, of final devoicing.

The study reported here expands on the work of Zając and Rojczyk (2014), who investigated Polish learners of English imitating a native and a non-native model speaker. The participants preferred the native speaker and diverged from the non-native one with neutralised vowel durations. Here, a new set of stimuli was included in the experiment. Along with the native model and his naturalistic durations, and the non-native model and his neutralised (but still naturalistic, i.e., Czech-accented) durations, the new set also included a native model speaker whose CVIVDV values were neutralised, and a non-native model who featured native-like CVIDVD values obtained from the native speaker.

The primary objectives were twofold: (1) to examine whether Czech speakers employ CVIVDV in their English speech in the first place and whether this is conditioned by their L2 proficiency, and (2) to examine whether they will imitate the model speakers due to their nativeness status or because of the naturalistic (target-language-like) pronunciation pattern in the model speech. In a socially minimal non-interactive study, 24 Czech speakers of English undertook a shadowing experiment.

First, the ratios of each target vowel duration to the duration of the constriction of the coda obstruent (V/C ratios) produced by participants in the baseline reading task were analysed. After fitting the data to a linear mixed-effects model, it was found that the values can be reliably predicted by the underlying voicing of the coda, and by the interaction between the voicing and the LexTALE scores. Openness to non-native accentedness, on the other hand, was not found to have a significant effect. Perhaps the individual statements from the questionnaire that the openness score was comprised of failed to reflect the crucial attitudes. Another explanation might be that there was too much variability between the participants for a clear pattern to emerge. Non-native speakers of English in this study thus vary their vowel duration based on the voicing of the following coda in their productions. The effect is the strongest in speakers with a high LexTALE score (i.e., more proficient L2 speakers), while less proficient speakers exhibit little to no CVIVDV. It was important that the pool of participants included those who showed little CVIVDV as well as those who showed native-like CVIVDV so, that the imitation of both native-like CVIVDV differentiation and of its absence, respectively, could be studied.

Next, a model predicting the V/C ratio differences between the baseline and the shadowing tasks was fitted to the data. Again, the effect of openness to non-native accentedness, which was hypothesised to facilitate imitation especially after the exposure to the non-native model speaker, was not significant. It was thus not included in the final model. Overall, the Czech speakers of English who participated in the experiment were found to adjust their production after the exposure to both model speakers, which could be explained by the automaticity of accommodation. However, the participants were found to imitate the English model speaker more than the Czech model. They imitated the naturalistic English model with the CVIVDV cue present to the greatest degree, which seems to be consistent with Zając and Rojczyk (2014). In their study, however, participants diverged from the non-native model speaker while here they still displayed some degree of imitation. More proficient speakers (as measured by LexTALE) adjusted their productions between the baseline and shadowing parts minimally, most probably due to them already being close to the model speech, which has been shown to preclude imitation (Babel 2012). Overall, this behaviour might be interpreted while referring to the CAT (Giles 2016). Speakers with invariable CVIVDV values might have wanted to approximate their target language model while proficient speakers no longer felt the need as they were already close. The inclusion of the introductory text recording before the model speakers' stimuli could have induced more social attitudes, thus facilitating accommodation. In any case, the present study replicates the finding that there is no need for the task to be interactive for social selectiveness to be manifested (Babel 2010).

As for the behaviour of the participants after the exposure to non-naturalistic values, a weak indication of greater imitation in the expected direction was present with the Czech model with the cue present, though this was not statistically significant. The distribution of the raw durations, however, suggests that for some participants the effect indeed might have been present.

The strong imitation of the native model speaker with CVIVDV cue present (facet 2 in Figure 14) as opposed to the Czech model (facet 4) can only be explained by the participants' social selectivity (i.e., the model speakers' status of native vs non-native speaker).

When participants' productions are compared for the two model speakers in the cue removed conditions, there seems to be very little shadowing-induced changes for the non-native model while the fitted values of the English model suggest a slight forking between voiced and voiceless V/C values in the higher LexTALE participants. Moreover, the violin plots in Figure 11 show a binomial distribution. This indicates that some participants imitated even the wrong direction to somewhat greater extent after exposure to the native voice than the non-native.

Because there is significant imitation of the native model with naturalistic values (facet 4) but little imitation for the same model with non-naturalistic values (facet 3), participants also seem to be selective in terms of the linguistic patterns they are exposed to. In other words, contrary to our expectations, the nativeness status of the model speaker did not override fully the non-native language patterns and imitation failed to appear to a significant degree. The nativeness status and phonetic values must be matched for the effect to be strong.

The results of the experiment underline the importance of proficiency in L2 studies in phonetics as Czech speakers of English seem to have gradually acquired the native-like CVIVDV values with increasing proficiency. It is remarkable that the LexTALE task, a short online lexical decision task, i.e., an estimate of L2 proficiency based on lexical knowledge, predicts quite clearly a very specific feature of second langauge pronunciation. This often-used measure is thus a good proxy of L2 proficiency not only in the lexical domain. The main and new finding is that in a shadowing task, speakers preferred to converge with a native model speaker, but the native-language-like speech patterns and the phonetic space also need to be present for significant imitation to occur. Phonetic imitation is, therefore, a multifaceted phenomenon influenced by both linguistic and social-psychological factors, and an integrated approach along the lines of Babel (2012) must be pursued in order to further our understanding of this phenomenon.

## 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.

Bates, Douglas, Martin Maechler, Ben Bolker, and Steve Walker. 2015. “Fitting Linear Mixed-Effects Models Using lme4.” *Journal of Statistical Software*, 67(1):1–48. doi:10.18637/jss.v067.i01.

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.

Boersma, Paul and David Weenink. 2022. Praat: doing phonetics by computer [Computer program]. Version 6.3.03, available at http://www.praat.org/.

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.

Council of Europe. 2020. *Common European Framework of Reference for Languages: Learning, teaching, assessment – Companion volume*. Strasbourg: Council of Europe Publishing, available at www.coe.int/lang-cefr.

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.

Edlund, Jens, Mattias Heldner, and Julia Hirschberg. 2009. “Pause and gap length in face-to-face interaction. *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*, 2009:2779–2782. doi:10.21437/Interspeech.2009-710.

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.

Gessinger, Iona, Eran Raveh, Ingmar Steiner, and Bernd Möbius. 2021. “Phonetic accommodation to natural and synthetic voices: Behavior of groups and individuals in speech shadowing.” *Speech Communication*, 127:43–63. doi:10.1016/j.specom.2020.12.004.

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.

Jankowska, Karolina, Tomasz Kuczmarski, and Grażyna Demenko. “Phonetic convergence in the shadowing for natural and synthesized speech in Polish.” *Lingua Posnaniensis*, 62(2):7–17. doi:10.2478/linpo-2020-0008.

Jiang, Fan, and Shelia Kennison. 2022. “The Impact of L2 English Learners’ Belief about an Interlocutor's English Proficiency on L2 Phonetic Accommodation.” *Journal of Psycholinguistic Research*, 51(4). doi:10.1007/s10936-021-09835-7.

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.

Lemhöfer, Kristin, and Mirjam Broersma. 2012. “Introducing LexTALE: A quick and valid Lexical Test for Advanced Learners of English.” *Behavior Research Methods*, 44:325–343. doi:10.3758/s13428-011-0146-0.

Lenth, Russell V. 2023. “emmeans: Estimated Marginal Means, aka Least-Squares Means. R package version 1.8.5. https://CRAN.R-project.org/package=emmeans.

Lev-Ari, Shiri, and Boaz Keysar. 2010. “Why don't we believe non-native speakers? The influence of accent on credibility.” *Journal of Experimental Social Psychology*, 46(6):1093–1096. doi:10.1016/j.jesp.2010.05.025.

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.

Lin, Yuhan, Yao Yao, and Jin Luo. 2021. “Phonetic accommodation of tone: Reversing a tone merger-in-progress via imitation.” *Journal of Phonetics*, 87(5):101060. doi:10.1016/j.wocn.2021.101060.

Liu, Qiu Ting. 2017. “Phonetic Accommodation to Non-Native English Speech.” *UC Berkeley PhonLab Annual Report*, 13(1):108–140. doi:10.5070/P7131040749.

Lüdecke, Daniel. 2018. “ggeffects: Tidy Data Frames of Marginal Effects from Regression Models.” *Journal of Open Source Software*, 3(26):772. doi:10.21105/joss.00772.

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

Machač, Pavel, and Radek Skarnitzl. 2009. *Principles of Phonetic Segmentation*. Praha: Epocha.

Menshikova, Alla, Daniil Kocharov, and Tatiana Kachkovskaia. 2020. "Phonetic entrainment in cooperative dialogues: A case of Russian." *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*, 2020:4148–4152. doi: 10.21437/Interspeech.2020-2696.

Michalsky, Jan, and Heike Schoormann. 2017. "Pitch convergence as an effect of perceived attractiveness and likability." *Interspeech 2017*, 2253–2256. doi:10.21437/Interspeech.2017-1520

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, Annie 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.

Olmstead, Annie J., Navin Viswanathan, Tiana Cowan, and Kunning Yang. 2021. "Phonetic adaptation in interlocutors with mismatched language backgrounds: A case for a phonetic synergy account." *Journal of Phonetics*, 87(3):101054. doi:10.1016/j.wocn.2021.101054.

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.

Pardo, Jennifer S., and Robert E. Remez. 2021. "On the Relation between Speech Perception and Speech Production." *The Handbook of Speech Perception*, edited by Jennifer S. Pardo, Lynne C. Nygaard, Robert E. Remez, and David B. Pisoni. doi:10.1002/9781119184096.ch23.

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

Pickering, Martin J., and Simon Garrod. 2004. "Toward a mechanistic psychology of dialogue." *Behavioral and Brain Sciences*, 27(2):169–190. doi:10.1017/s0140525x04000056.

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.

R Core Team. 2022. "R: A language and environment for statistical computing." R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

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.

Singmann, Henrik, Ben Bolker, Jake Westfall, Frederik Aust, and Mattan S. Ben-Shachar. 2023. „afex: Analysis of Factorial Experiments. R package version 1.3-0. https://CRAN.R-project.org/package=afex.

Skarnitzl, Radek, and Pavel Šturm. 2016. "Pre-Fortis Shortening in Czech English: A Production and Reaction-Time Study." *Research in Language*, 14(1):1–14. doi:10.1515/rela-2016-0005.

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.

Walker, Abby, and Kathryn Campbell-Kibler. 2015. "Repeat what after whom? Exploring variable selectivity in a cross-dialectal shadowing task." *Frontiers in Psychology*, 6:546. doi:10.3389/fpsyg.2015.00546.

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.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, Alex Hayes, Lionel Henry, Jim Hester, Max Kuhn, Thomas Lin Pedersen, Evan Miller, Stephan Milton Bache, Kirill Müller, Jeroen Ooms, David Robinson, Dana Paige Seidel, Vitalie Spinu, Kohske Takashi, Davis Vaughan, Claus Wilke, Kara Woo, and Hiroaki Yutani. 2019. "Welcome to the tidyverse." *Journal of Open Source Software*, 4(43):1686. doi:10.21105/joss.01686.

Winter, Bodo. 2020. *Statistics for Linguists: An Introduction Using R*. New York: Routledge.

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.

Zellou, Georgia, Michelle Cohn, and Tyler Kline. 2021. "The influence of conversational role on phonetic alignment toward voice-AI and human interlocutors." *Language Cognition and Neuroscience*, 36(3):1–15. doi:10.1080/23273798.2021.1931372.

# Appendices

## Appendix A: 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 B: Model-L1-inducing text

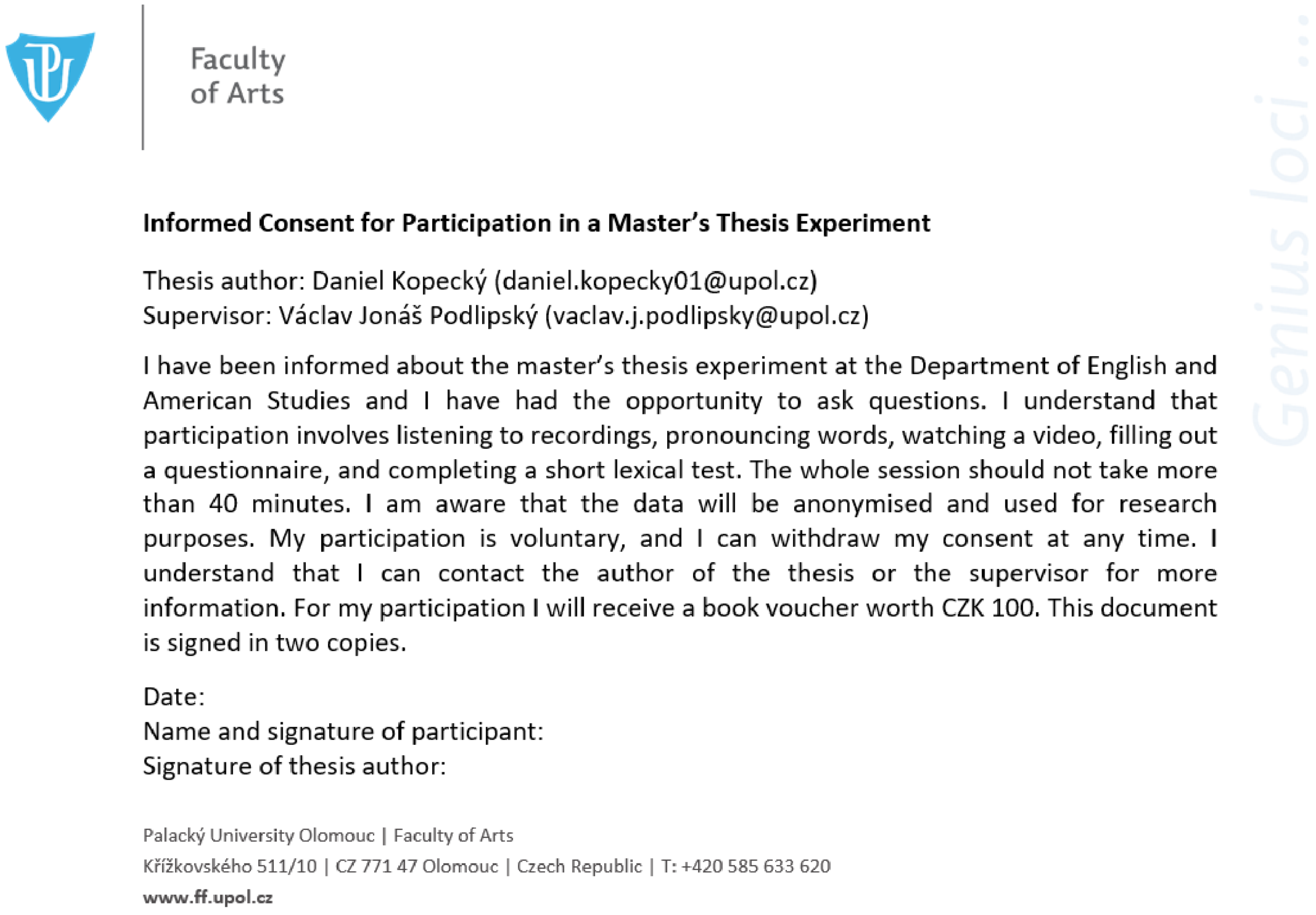
### The hidden agenda of a restaurant menu

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.

## Appendix C: Informed consent



Faculty
of Arts

**Informed Consent for Participation in a Master's Thesis Experiment**

Thesis author: Daniel Kopecký (daniel.kopecky01@upol.cz)
Supervisor: Václav Jonáš Podlipský (vaclav.j.podlipsky@upol.cz)

I have been informed about the master's thesis experiment at the Department of English and American Studies and I have had the opportunity to ask questions. I understand that participation involves listening to recordings, pronouncing words, watching a video, filling out a questionnaire, and completing a short lexical test. The whole session should not take more than 40 minutes. I am aware that the data will be anonymised and used for research purposes. My participation is voluntary, and I can withdraw my consent at any time. I understand that I can contact the author of the thesis or the supervisor for more information. For my participation I will receive a book voucher worth CZK 100. This document is signed in two copies.

Date:
Name and signature of participant:
Signature of thesis author:

Palacký University Olomouc | Faculty of Arts
Křížkovského 511/10 | CZ 771 47 Olomouc | Czech Republic | T: +420 585 633 620
**www.ff.upol.cz**

## Appendix D: Wordlist

| | |
|---|---|
| **bad** | not pleasant or enjoyable |
| **ball** | a round object used in games and sports |
| **bat** | a small animal that flies at night and looks like a mouse with large wings |
| **bed** | a piece of furniture for sleeping on |
| **bell** | a metal object shaped like a cup that makes a noise when its sides are hit by a piece inside it |
| **bet** | to risk money on a race or an event by trying to predict the results |
| **bin** | a container that you put waste in |
| **calf** | the back part of the leg between the ankle and the knee |
| **calve** | to give birth to a calf (baby cow) |
| **come** | to move or travel to the place where you are |
| **cub** | a young bear, lion, fox, wolf, or other wild animal |
| **cup** | a small round container for a drink, usually with a handle |
| **dock** | a part of a port where ships are repaired, or where goods are put onto them |
| **dog** | an animal kept as a pet, for guarding buildings, or for hunting |
| **fan** | a person who admires somebody very much |
| **fang** | one of the long, pointed teeth that dogs have |
| **fill** | to make something full of something |
| **fin** | a thin flat part that sticks out from the body of a fish, used for swimming and keeping balance |
| **gab** | to talk a lot and for a long time about unimportant things |
| **gap** | a space where something is missing |
| **hid** | the past tense of hide |
| **hill** | an area of land that is higher than the land around it, but not as high as a mountain |
| **hit** | to bring your hand or an object against somebody/something quickly and with force |
| **keel** | a long piece of wood or metal along the bottom of a boat that helps it to balance in the water |
| **keen** | wanting to do something or wanting something to happen very much |
| **lame** | unable to walk well because of an injury to the leg or foot |
| **long** | measuring or covering a great length or distance |
| **meal** | an occasion when you eat, especially breakfast, lunch, or dinner |
| **mean** | to have something as a meaning; to represent something |
| **mill** | a building where grain is made into flour |
| **name** | a word that a particular person is known by |
| **numb** | a part of your body that is numb has no feeling |
| **peck** | to move the beak forward quickly and hit or bite something |
| **peg** | a wooden or plastic object used for fastening wet clothes onto a line so that they will dry |
| **pin** | a small thin piece of metal with a sharp point, used for holding cloth in place while are sewing |
| **seat** | something you can sit on |
| **seed** | the small hard part produced by a plant, from which a new plant can grow |
| **sell** | to exchange something for money |
| **some** | an unspecified amount of something |
| **sun** | the star that shines in the sky during the day |
| **tab** | an additional document or page that can be opened on computer software |
| **tall** | having a greater than average height |
| **tap** | to hit somebody/something quickly and lightly |
| **teal** | a colour between blue and green |
| **team** | to touch someone or something gently |
| **thin** | not covered with much fat or muscle |
| **thing** | an object, or an item |
| **wrong** | not right or correct |

## Appendix E: Openness to non-native accentedness statements

1. I don’t care about my accent in English.
2. If I understand somebody, it doesn’t matter that they have a foreign accent.
3. When I speak English, I’m happy to be identified as a Czech speaker.
4. Teachers of English should present both the accent of native and of non-native speakers in lessons.
5. Having a non-native accent is bad.
6. I aim for native-like English pronunciation.
7. It’s important to me to sound like an English native speaker.

## Appendix F: R script

```
library(tidyverse)
library(broom)
library(lme4)
library(ggeffects)
library(emmeans)
library(afex)
library(effects)
library(optimx)
library(gridExtra)
library(readxl)
library(RColorBrewer)
library(report)
library(MuMIn)
library(ggh4x)

# 1. LOADING DATA
# 1.1 main data sheet
data <- read_excel("data/data.xlsx")
data <- rename(data, cue = codaVoi)
data <- mutate(data, session = as.factor(session),
          task = as.factor(task),
          cue = as.factor(cue),
          accent = as.factor(accent),
          wordVoice = as.factor(wordVoice),
          sex = as.factor(sex),
          log.vDur = log(vDur*1000),
          log.cDur = log(cDur*1000),
          vc.ratio = log(vDur*1000)/log(cDur*1000))

# 1.2 questionnaire data
quest <- read_excel("data/questionnaire_data.xlsx")

# 1.3 data sheet with model speakers' durations
model <- read_excel("data/model copy.xlsx")
model <- mutate(model, log.vDur = log(vDur*1000),
          log.cDur = log(cDur*1000),
          vc.ratio = log(vDur*1000)/log(cDur*1000))

# compute mean durations split by the factors
model.grp <- group_by(model, panel, x, facet)
mod.means <- summarise(model.grp, log.vDur = mean(log.vDur), log.cDur =
mean(log.cDur), vc.ratio = mean(vc.ratio))
mod.means <- mutate(mod.means, facet = as.factor(facet))
```

```
# rename the levels of the cue factor
levels(mod.means$facet) <- c("V/C duration cue\npresent","V/C duration
cue\nremoved")

# 2. GETTING THE DATA SHEETS INTO SHAPE
# 2.1 add to the main data sheet the participants' questionnaire openness, and lexTale
scores
data <- mutate(data, openness = quest$openness[data$sbjNo],
        prefer.AmE = as.factor(quest$prefer.AmE[data$sbjNo]),
        adopt.AmE = as.factor(quest$adopt.AmE[data$sbjNo]),
        lexTale = quest$lextale[data$sbjNo])

# 2.2 rename levels and sum-code factors
levels(data$accent) <- c("Czech model","English model")
levels(data$cue) <- c("V/C duration cue\nremoved","V/C duration cue\npresent")
contrasts(data$task) <- contr.sum(2)
contrasts(data$cue) <- contr.sum(2)
contrasts(data$accent) <- contr.sum(2)
contrasts(data$wordVoice) <- contr.sum(2)
contrasts(data$prefer.AmE) <- contr.sum(2)
contrasts(data$adopt.AmE) <- contr.sum(2)

# 2.3 centre openness and lexTale to mean
data <- mutate(data, openness_c = openness - mean(openness),
        lexTale_c = lexTale - mean(lexTale))

# 2.4 add log vowel and consonant duration differences, and vowel/consonant ratio
differences between
# baseline and shadowing for each word and participant
data <- data %>% mutate(log.vDifBS = "", log.cDifBS = "",  vc.ratioDifBS = "")
data$log.vDifBS <- as.numeric(data$log.vDifBS)
data$log.cDifBS <- as.numeric(data$log.cDifBS)
data$vc.ratioDifBS <- as.numeric(data$vc.ratioDifBS)

for (i in seq(nrow(data))) {
  pairRow <- filter(data, sbjNo == sbjNo[i] & word == word[i] & task == "baseline")
  if (nrow(pairRow) == 1 && data$task[i] != "baseline") {
    data$log.vDifBS[i] <- data$log.vDur[i] - pairRow$log.vDur
    data$log.cDifBS[i] <- data$log.cDur[i] - pairRow$log.cDur
    data$vc.ratioDifBS[i] <- data$vc.ratio[i] - pairRow$vc.ratio
  }
}

# 3. MODELLING
# 3.1 baseline with openness
baseLexOpenVC.mdl <- lmer(vc.ratio ~ wordVoice * lexTale_c * openness_c + (1 +
wordVoice|sbjNo) + (1|word),
                  data = filter(data, task == "baseline"),
                  control = lmerControl(optimizer = "bobyqa"))
r.squaredGLMM(baseLexOpenVC.mdl)
summary(baseLexOpenVC.mdl)
write.csv(round(summary(baseLexOpenVC.mdl)$coefficients, 8),
"coefs/baseLexOpenVC.csv")

# 3.2 baseline without openness
baseLexVC.mdl <- lmer(vc.ratio ~ wordVoice * lexTale_c + (1 + wordVoice|sbjNo) +
(1|word),
                data = filter(data, task == "baseline"),
                control = lmerControl(optimizer = "bobyqa"))

r.squaredGLMM(baseLexVC.mdl)
summary(baseLexVC.mdl)
write.csv(round(summary(baseLexVC.mdl)$coefficients, 8), "coefs/baseLexVC.csv")
baseLexVC.estim <- ggemmeans(baseLexVC.mdl, type = "fixed",
                     terms = c("lexTale_c", "wordVoice"), ci.lvl = 0.95)
baseLexVC.fit <- ggplot(data = baseLexVC.estim, aes(x = x, y = predicted, color =
group, group = group)) +
  geom_line(linewidth = 1) +
  geom_ribbon(aes(ymin = conf.low, ymax = conf.high, fill = group), alpha = 0.2,
linetype = 0) +
  ggtitle("Predicted values of V/C duration ratio") +
  xlab("LexTALE") +
  ylab("V/C ratio") +
  labs(fill="C voicing") +
  labs(color="C voicing") +
  theme_light()

baseLexVC.fit

ggsave("figs/baseLexVC.fit.png", plot = baseLexVC.fit, width = 6, height = 5, units =
"in",
      dpi = 600)

# plotting histogram, Q-Q plot, residual plot
par(mfrow = c(1, 3))
hist(residuals(baseLexVC.mdl))
qqnorm(residuals(baseLexVC.mdl))
qqline(residuals(baseLexVC.mdl))
plot(fitted(baseLexVC.mdl), residuals(baseLexVC.mdl))

# anova
openBaseREML.mdl <- lmer(vc.ratio ~ wordVoice * lexTale_c * openness_c + (1 +
wordVoice|sbjNo) + (1|word),
                    data = filter(data, task == "baseline"),
                         REML = FALSE,
                         control = lmerControl(optimizer = "bobyqa"))

baseREML.mdl <- lmer(vc.ratio ~ wordVoice * lexTale_c + (1 + wordVoice|sbjNo) +
(1|word),
                     data = filter(data, task == "baseline"),
                     REML = FALSE,
                     control = lmerControl(optimizer = "bobyqa"))

anova(openBaseREML.mdl, baseREML.mdl, test = "Chisq")

# 3.3 main model (vc.ratioDifBS)
vc.ratioDifBS.mdl <- lmer(vc.ratioDifBS ~ cue * wordVoice * accent * lexTale_c + (1
+ wordVoice|sbjNo) + (1|word),
                          data = data,
                          control = lmerControl(optimizer = "bobyqa"))

vc.ratioDifBS.estim <- ggemmeans(vc.ratioDifBS.mdl, type = "fixed",
                                 terms = c("lexTale_c", "wordVoice", "cue", "accent"), ci.lvl = 0.95)

r.squaredGLMM(vc.ratioDifBS.mdl)
summary(vc.ratioDifBS.mdl)

write.csv(round(summary(vc.ratioDifBS.mdl)$coefficients, 8),
"coefs/vc.ratioDifBS.csv")

vc.ratioDifBS.fit <- ggplot(vc.ratioDifBS.estim, aes(x = x, y = predicted, colour =
group, group = group)) +
  geom_smooth() +
  geom_ribbon(aes(ymin = conf.low, ymax = conf.high, fill = group), alpha = 0.2,
linetype = 0) +
  facet_nested(~panel + facet) +
  ggtitle("Predicted V/C ratios (difference from baseline)") +
  xlab("LexTALE") +
  ylab("V/C ratio (difference from baseline)") +
  labs(fill="C voicing") +
  labs(color="C voicing") +
  theme_light()
vc.ratioDifBS.fit
ggsave("figs/vc.ratioDifBS.fit.png", plot = vc.ratioDifBS.fit, width = 7, height = 5, units
= "in",
       dpi = 600)

# plotting histogram, Q-Q plot, residual plot
par(mfrow = c(1, 3))
hist(residuals(vc.ratioDifBS.mdl))
qqnorm(residuals(vc.ratioDifBS.mdl))
```

```
qqline(residuals(vc.ratioDifBS.mdl))
plot(fitted(vc.ratioDifBS.mdl), residuals(vc.ratioDifBS.mdl))

# anova
openREML.mdl <- lmer(vc.ratioDifBS ~ cue * wordVoice * accent * lexTale_c *
openness_c + (1 + wordVoice|sbjNo) + (1|word),
                     data = data,
                     REML = FALSE,
                     control = lmerControl(optimizer = "bobyqa"))

REML.mdl <- lmer(vc.ratioDifBS ~ cue * wordVoice * accent * lexTale_c + (1 +
wordVoice|sbjNo) + (1|word),
                     data = data,
                     REML = FALSE,
                     control = lmerControl(optimizer = "bobyqa"))

anova(openREML.mdl, REML.mdl, test = "Chisq")

# 4. RAW DATA PLOTS
# getting data into shape for violin plots
write.csv(data, "data/olddata.csv")

#  a copy of the data tibble was made and in excel the following changes were made:
# for each baseline production in each participant, the baseline row
# was copied and the values in columns codaVoi, accent, and session were changed
# to correspond to the values in the row shadowing session 2 of the same word,
# so that the baseline productions could be plotted alongside both
# shadowing session 1 and shadowing session 2
datavio <- read_excel("data/datavio.xlsx")
datavio <- rename(datavio, cue = codaVoi)
datavio <- mutate(datavio, cue = as.factor(cue))
levels(datavio$cue) <- c("V/C duration cue\npresent","V/C duration cue\nremoved")

mod.means2 = mod.means
mod.means2 <- rename(mod.means2, wordVoice = x, accent = panel, cue = facet)

# 4.1 raw V dur violin plots
dodge <- position_dodge(width = 0.8)
vDur.vio <- ggplot(datavio, aes(y=log.vDur, x=wordVoice, fill=task)) +
  geom_hline(data = mod.means2, aes(yintercept = mod.means2$log.vDur),
             linetype = 2, linewidth = 0.5, color = '#999999') +
  geom_violin(width=1, alpha=0.25, position = dodge) +
  geom_boxplot(width=0.25, position = dodge) +
  facet_grid(cue~accent) +
  labs(fill = "Task") +
  ggtitle("Raw distribution of V duration") +
  xlab("Coda C voicing") +
  ylab("V duration\n(ms, log-transformed)") +
  theme_light()
vDur.vio
ggsave("figs/vDur_vio.png", plot = vDur.vio, width = 6, height = 4, units = "in",
      dpi = 600)

# 4.2 raw V/C dur violin plots
vc.ratio.vio <- ggplot(datavio, aes(y=vc.ratio, x=wordVoice, fill=task)) +
  geom_hline(data = mod.means2, aes(yintercept = mod.means2$vc.ratio),
             linetype = 2, linewidth = 0.5, color = '#999999') +
  geom_violin(width=1, alpha=0.25, position = dodge) +
  geom_boxplot(width=0.25, position = dodge) +
  facet_grid(cue~accent) +
  labs(fill = "Task") +
  ggtitle("Raw distribution of V/C duration ratio") +
  xlab("Coda C voicing") +
  ylab("V/C duration ratio") +
  theme_light()
vc.ratio.vio
ggsave("figs/vc.ratio_vio.png", plot = vc.ratio.vio, width = 6, height = 4, units = "in",
      dpi = 600)
```

## Annotation

**Author:** Daniel Kopecký

**Field of study:** English Philology & General Linguistics

**Title:** Imitation of English Coda-Voicing-Induced Vowel Duration Variability by Czech Learners

**Type**: Master thesis

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

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

**Number of pages:** 72

**Number of characters:** 108,051

**Description:** This thesis focuses on changes in the pronunciation of L2 speakers that happen after immediate exposure to the speech of others. In minimal pairs such as *bat* and *bad*, native speakers systematically produce a vowel that is longer in the latter word due to the following underlyingly voiced obstruent. First it was found that, akin to native English speakers, proficient Czech speakers vary their vowels in the two voicing conditions, but less proficient speakers fail to do so. In a shadowing experiment, Czech speakers were exposed to natural and manipulated stimuli from a native and a non-native speaker of English to investigate whether they vary their speech because of these models' language background or because of the target-language-like pattern in the model speech. In the experiment, the native voice was imitated to a greater extent than the non-native one, but only when the duration values matched the native English values.

**Keywords:** phonetic imitation, phonetic convergence, phonetic accommodation, pre-fortis clipping, vowel length as a cue to coda voicing, L2 speakers of English

# Anotace

**Autor:** Daniel Kopecký

**Studijní obor:** Anglická filologie & Obecná lingvistika

**Název:** Imitation of English Coda-Voicing-Induced Vowel Duration Variability by Czech Learners [Imitace variability délky vokálů vyvolané znělostí kody v angličtině českými mluvčími]

**Typ práce:** diplomová práce

**Fakulta a katedra**: Filozofická fakulta, Katedra anglistiky a amerikanistiky

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

**Počet stran:** 72

**Počet znaků:** 108,051

**Charakteristika:** Tato práce se zaměřuje na změny ve výslovnosti mluvčích druhého jazyka, ke kterým dochází po bezprostředním kontaktu s řečí ostatních jedinců. V minimálních párech jako například *bat* a *bad* rodilí mluvčí systematicky vyslovují samohlásku, která je v druhém slově delší kvůli následující fonologicky znělé obstruentě. Nejprve bylo zjištěno, že čeští mluvčí s vysokou úrovní angličtiny variují své hlásky v obou hláskových kontextech podobně jako rodilí mluvčí angličtiny, kdežto méně zdatní mluvčí tak nečiní. V rámci experimentu byli čeští mluvčí vystaveni přirozeným a modifikovaným stimulům od rodilého a nerodilého mluvčího angličtiny, aby se zjistilo, zda se jejich řeč mění kvůli jazykovému pozadí těchto modelových mluvčích, nebo kvůli vzorcům v modelové řeči, které se podobají cílovým jazykům. V experimentu byl hlas rodilého mluvčího napodobován ve větší míře než hlas nerodilý, ale to pouze v tom případě, kdy hodnoty trvání hlásek odpovídaly hodnotám rodilé angličtiny.

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