

UNIVERZITA PALACKÉHO V OLOMOUCI
FILOZOFICKÁ FAKULTA

KATEDRA ANGLISTIKY A AMERIKANISTIKY



**Voicing Assimilation Across Word
Boundaries in Czech and English**

Bakalářská práce

Autor: Lenka Pštrossová

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

Olomouc 2025

Prohlašuji, že jsem bakalářskou práci na téma Asimilace znělosti na hranicích slov v češtině a angličtině vypracovala samostatně, a že jsem uvedla úplný seznam citované a použité literatury.

V Olomouci dne

Podpis

Chtěla bych poděkovat panu Mgr. Václavovi Jonáši Podlipskému, Ph.D. za odborné vedení práce, ochotu a cenné rady, které mi pomohly tuto práci zkompletovat. Mé poděkování patří též účastníkům výzkumu za spolupráci při získávání údajů pro výzkumnou část práce.

Dále bych chtěla poděkovat OpenAI ChatGPT za pomoc s jazykovou korekturou a stylistickou úpravou textu. Veškeré myšlenky, závěry a odborný obsah práce jsou mé vlastní.

Anotace

Jméno a Příjmení: Lenka Pštrosová

Katedra: Katedra anglistiky a amerikanistiky

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

Rok obhajoby: 2025

Název práce: Asimilace znělosti na hranicích slov v češtině a angličtině

Název práce v angličtině: Voicing Assimilation Across Word Boundaries in Czech and English

Přílohy vázané v práci: 8

Počet stran: 40

Počet znaků: 73 266

Jazyk práce: angličtina

Abstrakt v jazyce práce

This thesis examines voicing assimilation in the English pronunciation of Czech learners. It contrasts the systematic voicing assimilation in Czech with the more variable patterns in English and analyzes their impact on second-language pronunciation. A pilot study was conducted using a custom phonetic training tool in Praat. Results showed limited improvement after training, highlighting the difficulty of modifying L1 phonological habits. Longer-term training is recommended to achieve lasting pronunciation changes.

Klíčová slova v jazyce práce

assimilation, voicing, phonetics, training, Czech, English, visual feedback

Table of Contents

Anotace	IV
Table of Contents	V
1 Introduction	1
2 Theoretical Background	3
2.1 Voicing Assimilation	3
2.1.1 Cross-Linguistic Variability	4
2.1.2 Phonetic and Phonological Voicing Distinctions	5
2.2 Czech Voicing Assimilation	6
2.2.1 Voiced Labiodental	7
2.2.2 Differences Between Bohemian and Moravian Phonetics	9
2.3 English Assimilation	11
2.3.1 Voicing Assimilation of English as L2	11
2.4 Phonetic Training	13
2.4.1 Introduction to Phonetic Training	13
2.4.2 Approaches to Phonetic Training	14
2.4.2.1 Auditory Training	14
2.4.2.2 Articulatory Training	14
2.4.2.3 Multimodal Feedback and Computer-Assisted Pronunciation Training	15
2.4.2.4 High Variability Phonetic Training	15
2.4.3 Factors Influencing Phonetic Training Outcomes	17
2.4.4 Challenges and Limitations of Phonetic Training	17
2.4.5 Future Directions in Phonetic Training Research	18
2.4.6 Conclusion	18
3 Practical Part	19
3.1 Research Design	19

3.2	Participants.....	20
3.3	Materials and Stimuli.....	20
3.4	Procedure	23
3.4.1	Pretest.....	23
3.4.2	Training session	24
3.5	Data Collection and Analysis.....	25
3.5.1	Audio recordings and waveforms	26
3.5.1.1	Participant-Specific Results.....	28
3.5.1.2	Effects of Speech Rate and L1 Transfer on Training Outcomes	29
3.5.1.3	Conclusion.....	29
3.5.2	Questionnaire analysis	30
3.6	Limitations and Lessons Learned	31
4	Conclusion	33
	Resumé	35
	References	36
	Appendix	39

1 Introduction

Although much research in English phonology focuses on assimilation processes affecting place of articulation, to the best of the author's knowledge, relatively little detailed attention appears to have been given to voicing assimilation in English. This relative gap in literature motivated the focus of the present study. By investigating voicing assimilation among Czech learners of English, this thesis aims to contribute to an underexplored area and to provide insights into how native language interference can affect connected speech production in English.

Recent research suggests that targeted phonetic training, especially methods combining auditory exposure, articulatory guidance, and visual feedback, can significantly support second-language learners in adjusting their pronunciation patterns. Training approaches such as Computer-Assisted Pronunciation Training (CAPT), ultrasound tongue imaging (UTI), and High Variability Phonetic Training (HVPT) offer promising tools for raising learners' awareness of phonetic detail and supporting more native-like speech production.

The aim of this thesis is twofold. First, it provides a theoretical overview of voicing assimilation in both Czech and English, focusing on cross-linguistic differences and their impact on second-language acquisition. Second, it presents a pilot study testing a customized phonetic training script developed in Praat, designed to assist Czech learners of English in mastering final obstruent voicing across word boundaries. The study evaluates whether short-term training combining auditory and visual-acoustic feedback can improve learners' pronunciation, and it reflects critically on the effectiveness and limitations of the intervention.

Chapter 2 reviews the theoretical background of voicing assimilation, covering its typological variability across languages, the specific patterns found in Czech and English, and the challenges these differences pose for second-language learners.

Chapter 3 discusses current approaches to phonetic training, including auditory, articulatory, and computer-assisted methods, with particular focus on techniques relevant to voicing contrasts.

Chapter 4 presents the practical part of the thesis, describing the design of the phonetic training tool, the pilot study procedure, the analysis of pretest and post-training recordings, and the results of participant questionnaires.

Chapter 5 offers a discussion of the study's findings, identifies limitations, and proposes directions for future research in pronunciation training for Czech learners of English.

By integrating theoretical insights with practical experimentation, this thesis contributes to the broader discussion on pronunciation pedagogy for learners from typologically different language backgrounds. It highlights the importance of tailored phonetic training in second-language development.

2 Theoretical Background

2.1 Voicing Assimilation

Voicing assimilation is a process that occurs across various languages but is implemented differently depending on phonological and phonetic constraints. At its core, it refers to how adjacent consonants influence each other's voicing properties, either progressively (from left to right) or regressively (from right to left) (Simon 2010). In native English speech, prosodic structure—such as word boundaries and phonological phrasing—has been shown to influence the application of assimilation rules (Gow 2002).

English exemplifies a system where voicing assimilation is typically gradient and phonetic rather than categorical. In connected speech, word-final voiced obstruents may undergo partial devoicing before voiceless segments, but the underlying voicing contrast is usually preserved (Myers 2010). In contrast, Czech applies voicing assimilation as a categorical phonological rule, resulting in somewhat complete neutralization of voicing distinctions across word boundaries (Šturm and Tylečková 2019).

One key factor in voicing assimilation is laryngeal coarticulation, where the overlap of vocal fold gestures influences the voicing of neighboring sounds. Dmitrieva, Jongman, and Sereno (2020) found that the strict voicing system of Russian influenced even the native-language phonetics of learners, supporting the idea that voicing assimilation in Russian is phonologically categorical, unlike the more flexible, phonetic pattern in English. In contrast to the more variable, phonetic patterns found in English.

The extent of voicing assimilation is also influenced by speech rate and prosodic structure. Faster speech often increases assimilation effects as speakers reduce articulatory effort, whereas clear and careful speech minimizes assimilation (Agwuele, Sussman, and Lindblom 2009). Additionally, prosodic boundaries—such as phrase breaks or pauses—can either enhance or block assimilation, depending on the language (Gow 2002).

Voicing assimilation may develop diachronically from gradient phonetic processes into categorical phonological rules. Studies in phonetically based phonology suggest that such patterns often originate as coarticulatory effects, which, through repeated exposure and perceptual reinterpretation, may become phonologized (Hume and Johnson 2001; Ohala 1993). This perspective highlights the role of articulatory and perceptual factors in shaping phonological systems over time.

2.1.1 Cross-Linguistic Variability

Cross-linguistic studies show that not all languages treat voicing assimilation equally. For instance, in English, assimilation is largely phonetic and does not affect the underlying phonological representation of words. In contrast, languages such as Russian, exhibit categorical phonological assimilation, where voicing features are systematically altered (Myers 2010).

Olson (2019) found that Spanish–English bilinguals maintained separate voicing assimilation rules for each language, but when switching between languages, assimilation in Spanish was weakened while English remained unaffected. This suggests that even when phonological rules are language-specific, bilinguals may still experience asymmetrical cross-linguistic influence, particularly in boundary contexts.

In bilingual speakers, the ability to process assimilation varies depending on language dominance, with second-language learners often applying their native assimilation rules to a second language (Simon 2010).

Broersma (2005) found that Dutch learners of English were often able to perceive final voiced and voiceless obstruents with near-native accuracy, even in the absence of certain acoustic cues such as final release bursts. Although the study did not explicitly investigate voicing assimilation, it demonstrates that L2 learners can, under controlled conditions, accurately categorize difficult phonemic contrasts. However, Broersma also concluded that native-like use of perceptual cues for familiar contrasts in unfamiliar positions may hardly ever be attained, suggesting that perceptual confusion could still arise when available cues are limited or ambiguous.

Flege and Bohn (2021) revised Speech Learning Model (SLM-r) suggests that difficulties in acquiring L2 phonetic contrasts stem from the automatic association of new L2 sounds with pre-existing L1 categories. Rather than forming entirely new categories for each L2 sound, learners often link incoming L2 input to the closest L1 equivalent, which can block or interfere with accurate category formation.

Over time, this process may lead to the development of merged or composite categories that reflect features of both the L1 and L2 sound systems. Furthermore, L2 learners are exposed to input conditions that differ from those of monolingual native speakers, which further complicates acquisition. Despite these challenges, the model emphasizes that adult learners retain the capacity to detect subtle L1–L2 phonetic

differences, store them, and use these perceptual distinctions to guide their own production.

2.1.2 *Phonetic and Phonological Voicing Distinctions*

It is essential to distinguish between phonological voicing assimilation, which is dictated by the phonological system of a language, and phonetic voicing assimilation, which occurs at a surface level due to articulatory constraints (Skarnitzl and Šturm 2014).

Jansen (2007) shows that in English, voicing assimilation is not categorical but occurs to a limited extent at the phonetic level. Voiced obstruents at the end of a word may be partially devoiced before voiceless consonants, though the underlying voicing contrast is typically preserved. For example, in the phrase *bad time*, the final /d/ in *bad* might be slightly devoiced but remains distinguishable from a voiceless [t]. This suggests that assimilation in English functions more as a surface-level phonetic adjustment rather than a systematic phonological process.

In contrast, Czech exhibits a categorical, phonological rule of regressive voicing assimilation. Final voiced obstruents are systematically devoiced before voiceless consonants, resulting in a fully voiceless realization, such as [t] instead of /d/ (Skarnitzl and Šturm 2014). This difference has important implications for L2 pronunciation, as Czech learners often transfer their native assimilation patterns into English speech, leading to non-native-like voicing patterns.

Moreover, phonetic assimilation is often less perceptually salient than phonological assimilation, making it harder for learners to perceive and internalize. The revised Speech Learning Model (SLM-r) proposes that adult learners retain the ability to form new L2 phonetic categories, but this process depends on factors such as the degree of phonetic dissimilarity from L1 categories, the structure of L1 categories, and the amount and variability of perceptual input they receive (Flege and Bohn 2021). Since phonetic-level assimilation in English is subtle and gradient, Czech learners may have difficulty recognizing when it occurs and may instead rely on more categorical L1-based assimilation patterns. This highlights the importance of training that raises awareness of these cross-linguistic differences in assimilation behavior.

2.2 Czech Voicing Assimilation

Cvrček (2015) note that in Czech the process of *spodoba* (assimilation) involves transferring a phonetic feature from one sound to an adjacent sound, thereby changing the identity of the affected.

Unlike English, where phonetic assimilation is variable, Czech assimilation generally follows strict phonological rules, making it a key feature of connected speech, particularly in obstruents. Word-final obstruents typically adjust their voicing to match the following sound through regressive assimilation, which applies both within words and across word boundaries (Skarnitzl and Šturm 2014). However, some exceptions exist, notably in cases involving the consonant /v/, where voicing assimilation may be less consistent (Volín and Skarnitzl 2006).

In Czech, regressive voicing assimilation applies systematically across word boundaries. Voicing of obstruents is determined by the voicing of the following sound, regardless of whether the obstruent is word-final or word-internal (Šturm and Tylečková 2019). For example, the word *pes* (“dog”) ends in the voiceless phoneme /s/, which is realized as [s] in *Byl tam pes* (“There was a dog”), but assimilates to [z] in *Pes byl doma* (“The dog was at home”), where the initial /b/ of *byl* triggers voicing. This illustrates that voicing assimilation in Czech is categorical and phonological in nature, affecting obstruents both within words and across word boundaries. In contrast, English voicing assimilation is more limited and occurs only at the phonetic level, typically in fast or casual speech (Jansen 2007).

As previously mentioned, Czech phonology follows a reliable pattern of regressive voicing assimilation. There is a strong correspondence between phonological voicing categories and their phonetic realization: phonologically voiced obstruents are generally produced with substantial, though not necessarily continuous, vocal fold vibration, while voiceless obstruents are generally produced without vocal fold vibration (Skarnitzl and Šturm 2014).

Beyond regressive assimilation, another defining feature of Czech phonology is final obstruent devoicing, which occurs when a voiced obstruent appears at the end of a word without a voiced sound following it. In such cases, the obstruent is automatically pronounced as voiceless, neutralizing the distinction between voiced and voiceless consonants (Skarnitzl and Šturm 2014). For example, the words *led* (“ice”) and *let* (“flight”) are both pronounced as [let], even though one is historically voiced and the

other voiceless. A similar process can be observed with the word *provaz* (“rope”), which ends in the phoneme /z/. When uttered phrase-finally or before a voiceless consonant, such as in *Mám provaz* (“I have a rope”), the final /z/ is devoiced to [s]. However, when followed by a voiced consonant, as in *Dám provaz dolů* (“I will lower the rope”), the underlying voicing is preserved. This interaction between final devoicing and assimilation demonstrates the regular and context-sensitive nature of Czech obstruent voicing patterns.

This rigid assimilation process in Czech creates challenges for Czech speakers learning English. Since voicing assimilation does not typically occur across word boundaries in English, Czech speakers may unintentionally apply their native rules, leading to non-native pronunciation patterns (Skarnitzl and Šturm 2014). For instance, in Czech, a phrase like *opět byl* (“again he was”) is typically pronounced as [opjɛd bɪl], with the final /t/ in *opět* voiced to [d] due to the following voiced /b/ in *byl*. When this native rule is transferred to English, learners may similarly voice final voiceless obstruents inappropriately, producing forms like *black bird* as [blæg bɜ:d].

The role of sonorants in Czech voicing assimilation is also noteworthy. Unlike in English, where sonorants generally do not influence the voicing of preceding obstruents, in Czech, some dialects exhibit voicing assimilation before sonorants (Šturm and Tylečková 2019). For instance, in Moravian Czech, a phrase like *pět nocí* („five nights“) may be pronounced as [pjɛd notsi:], showing that even sonorants can trigger voicing changes. This feature is more characteristic of Slovak and some Polish dialects than of standard Bohemian Czech, which retains *pět nocí* without voicing the final /t/.

These systematic assimilation patterns, while essential to Czech phonology, present challenges in second-language acquisition. Understanding these differences not only aids in mastering foreign pronunciation but also highlights the structural uniqueness of Czech in contrast to languages like English.

2.2.1 Voiced Labiodental

While Czech obstruents follow somewhat reliable voicing assimilation, /v/ behaves differently. One particularly unusual aspect of the Czech phoneme /v/ is that it does not always follow the same rules as other voiced consonants. Typically, in Czech, voiced consonants like /b/, /d/, and /g/ lose their voicing when they appear at the end of a word, but /v/ does not. This makes it act more like a sonorant (such as /m/, /l/, or /r/) than an obstruent (like /b/ or /d/). However, in some cases, /v/ still follows the typical voicing

assimilation rules, meaning it can lose its voicing when placed before a voiceless consonant. This inconsistency has led some researchers to argue that /v/ does not fully belong to either group—it has characteristics of both sonorants and obstruents (Skarnitzl, Šturm, and Volín 2016).

The Czech voiced labiodental /v/ is quite different from its English counterpart in pronunciation and how it behaves in speech. In Czech, /v/ is often realized as an approximant [v] with much weaker frication than the English [v], essentially behaving more like a sonorant (Volín and Skarnitzl 2006). This difference poses a challenge for Czech learners of English, who must learn to produce /v/ as a true fricative rather than as a glide.

Studies have shown that Czech speakers differ in their pronunciation of /v/: some realize it as a fricative [v], while others produce it more like an approximant [v] (Volín and Skarnitzl 2006). However, the manner of articulation is distinct from the phonological behavior of /v/ in assimilation processes. Some speakers allow /v/ to trigger regressive voicing assimilation, treating it like an obstruent, while others do not, treating it like a sonorant (Volín and Skarnitzl 2006).

Volín and Skarnitzl (2006) in their study of radio broadcasters and students showed that about half of the speakers applied voicing assimilation to obstruents before /v/, while the other half did not. This suggests that the phonetic realization of /v/ appears to be undergoing gradual change, and people may be shifting toward treating it more like an obstruent.

Czech /v/ has long presented a challenge for phonological classification. While it is now generally considered an obstruent, it still retains some sonorant-like properties (Volín and Skarnitzl 2006). Its phonological behavior often differs from that of other obstruents, especially in assimilation contexts (Skarnitzl, Šturm, and Volín 2016).

In addition to these classification challenges, /v/ shows atypical voicing patterns depending on its phonetic environment. For example, if /v/ follows a voiceless velar fricative /x/ (as in *abych viděla*), it is more likely to lose voicing, while in other cases, it tends to stay voiced (Skarnitzl, Šturm, and Volín 2016).

For Czech speakers learning English, this difference in pronunciation can be a challenge. Many Czech speakers tend to pronounce /v/ more like [v], which can sound unusual to English speakers. They may also mistakenly devoice final /v/, pronouncing words like *have* as [hæf], i.e. replacing the final [v] with a [f]. (Volín and Skarnitzl 2006).

The status of /v/ in Czech is not fixed, and its pronunciation can vary depending on dialect, speaker, and speech situation (Skarnitzl, Šturm, and Volín 2016). Because of this, /v/ remains one of the most complex consonants in Czech, and researchers are still studying how it fits into the language. Understanding these pronunciation differences is especially important for teaching Czech speakers how to pronounce /v/ correctly in English, as well as improving our knowledge of Czech phonetics and phonology (Volín and Skarnitzl 2006).

2.2.2 *Differences Between Bohemian and Moravian Phonetics*

The distinction between Bohemian Czech and Moravian Czech is particularly evident in phonetics and pronunciation patterns, with key differences in voicing assimilation, vowel length, and segmental articulation. These variations also influence second-language pronunciation, particularly for Czech speakers acquiring English (Šturm and Tylečková 2019).

One of Bohemian and Moravian Czech's most striking phonological differences is how they handle voicing assimilation. Bohemian Czech follows standard assimilation rules, where obstruents change voicing only before another obstruent. For example, *dost peněz* ('enough money') is pronounced as [dost pɛɲɛs], while *dost bodů* ('enough points') follows regressive assimilation as [dozd bodu:] (Šturm and Tylečková 2019)

However, in Moravian Czech, voicing assimilation extends to sonorants as well. This means that in Moravian Czech, a phrase like *dost masa* (enough meat) can be pronounced as [dozd masa], a pattern also found in Slovak but not in Bohemian Czech (Šturm and Tylečková 2019). This difference plays a role in second-language English pronunciation. Moravian speakers tend to carry over their native assimilation patterns into English, leading to voicing errors in connected speech (Šturm and Tylečková 2019). Bohemian Czech speakers, by contrast, are less likely to apply native assimilation in inappropriate English contexts (Volín and Skarnitzl 2006).

Labiodental and dental sounds also show regional variation. Bohemian Czech speakers generally follow standard Czech articulation, while Moravian Czech speakers exhibit more significant variability in the realization of /v/. Research shows that in some Moravian dialects, /v/ behaves more like an approximant [ʋ], similar to Slovak, whereas Bohemian Czech maintains a more consistent fricative [v] (Volín and Skarnitzl 2006).

This results in Moravian speakers producing English /v/ with less friction, making it sound softer or closer to [w] in some cases (Volín and Skarnitzl 2006)

Similarly, Moravian Czech shows greater lenition of certain consonants, which can lead to non-standard pronunciations when learning English. For instance, the realization of word-final voiced obstruents in Moravian Czech can sometimes remain fully voiced, whereas Bohemian Czech speakers often apply devoicing, aligning more closely with English phonetics (Šturm and Tylečková 2019).

Prosody and stress patterns also differ between the two dialects. Bohemian Czech follows more standardized Czech intonation, whereas Moravian Czech tends to have flatter pitch contours, sometimes resembling Slovak intonation patterns (Volín and Skarnitzl 2006). This prosodic difference can influence the way Moravian Czech speakers apply stress and intonation in English, leading to monotonous speech patterns that do not align with the rhythm of English (Skarnitzl and Šturm 2014).

The phonetic differences between Bohemian and Moravian Czech have important implications not only for native pronunciation but also for English language acquisition (Šturm and Tylečková 2019). The most critical distinction is voicing assimilation, where Moravian Czech aligns more with Slovak than Bohemian Czech. These findings underscore the necessity of dialect-sensitive pronunciation training for Czech speakers learning English (Šturm and Tylečková 2019).

2.3 English Assimilation

In English, phonological assimilation primarily affects place of articulation rather than voicing and occurs more frequently within words than across word boundaries. The most common form is regressive assimilation, where a sound is influenced by the following phoneme, causing it to shift its place of articulation (Jansen 2007). For example, in casual speech, the phoneme /n/ in *input* may be realized as [m] before bilabial /p/, due to regressive assimilation.

English also exhibits place assimilation at word boundaries, but it is optional and context-dependent. For instance, in fast speech, a phrase like *green park* may be pronounced as [gri:m p^hɑ:k], where the alveolar nasal /n/ assimilates to the bilabial /m/ before the bilabial stop /p/. However, these changes are not strictly required by English phonological rules, and speakers may still pronounce them carefully as separate sounds (Jansen 2007).

Unlike Czech, where voicing assimilation operates as a systematic phonological rule, English assimilation processes tend to be more variable and speaker-dependent. Research shows that English does not strictly enforce cross-word voicing assimilation, meaning that final voiced obstruents do not necessarily take on the voicing of the following consonant. Instead, English speakers may produce a partially devoiced obstruent, but complete assimilation does not typically occur (Jansen 2007).

Overall, assimilation in English is more limited and optional, and it primarily affects the place of articulation, whereas voicing assimilation is not obligatory. This is a key difference from languages like Czech and Dutch, where voicing assimilation is phonological and applies across word boundaries. These distinctions create challenges for learners whose native languages have strict assimilation rules, leading to potential pronunciation difficulties in English (Jansen 2007).

2.3.1 Voicing Assimilation of English as L2

Adjusting to the more relaxed English system can be challenging for learners of English from languages with strict assimilation rules. Research on Dutch learners of English by Simon (2010) showed that they often applied Dutch assimilation rules to English, incorrectly changing the voicing of final obstruents based on the next sound. This led to mispronunciations that made speech less clear (Simon 2010).

Similarly, Czech speakers of English tend to assimilate final voiced consonants in ways that native English speakers do not, which affects their pronunciation. In the phrase *bad time*, Czech learners may devoice the final /d/ in *bad*, resulting in a production closer to [bæt taim]. In contrast, native English speakers generally preserve the voicing of the /d/ (Skarnitzl and Šturm 2014).

Myers (2010) explains that languages like Czech use voicing assimilation as a reliable phonological rule, meaning it usually happens under specific conditions. In English, voicing assimilation is optional and depends on the speaker. It is also gradual, meaning sounds may be partially affected but not completely changed. This explains why English voicing assimilation is less noticeable, while in Czech, it is an essential part of pronunciation.

L1 phonological rules can influence L2 speech production, as observed in Dutch-English bilinguals (Simon 2010). Flege and Bohn (2021) also argued that L1 categories exert a strong influence on the acquisition of L2 phonetic categories, especially when the L2 sound is similar but not identical to an L1 counterpart.

Research by Šturm and Tylečková (2019) underscores the significant implications of voicing assimilation for language learning. Czech learners of English frequently transfer their native voicing assimilation patterns into English, leading to systematic pronunciation errors. This type of cross-linguistic interference highlights the importance of explicit phonetic training to help learners internalize English phonotactic constraints, making the study of voicing assimilation a crucial aspect of language acquisition (Thomson and Derwing 2015).

Dmitrieva, Jongman, and Sereno (2020) found that learning a language with strict voicing rules can even change how people speak their first language. For example, American English speakers learning Russian started to shorten their VOT in English, showing that second-language learning can affect native pronunciation. This further supports the idea that English voicing rules are more flexible, while languages like Czech and Russian follow more strict phonological patterns.

The difference between voicing assimilation in English and Czech highlights a broader contrast between phonetic and phonological rules. While Czech follows more strict, predictable assimilation patterns, English allows for variation, with assimilation happening in some cases but not all. This difference makes it difficult for Czech and Dutch speakers learning English, as they must learn to keep final obstruents stable instead of changing them based on the following word (Skarnitzl and Šturm 2014). These

findings highlight the need for phonetic training to help learners adapt to English pronunciation rules .

2.4 Phonetic Training

2.4.1 Introduction to Phonetic Training

The present study focuses on the role of phonetic training in helping Czech learners of English overcome challenges related to voicing assimilation, which differs significantly between Czech and English. In Czech, assimilation is somewhat phonologically obligatory, while in English, it remains largely phonetic and gradient. Consequently, Czech learners often misapply L1 voicing rules to English, leading to non-native pronunciation patterns.

Studies suggest that combining visual and auditory feedback can improve phonetic accuracy in second language learning (Thomson and Derwing 2015). For instance, Suemitsu et al. (2015) demonstrated significant improvement in Japanese learners' production of English vowels after real-time articulatory feedback. Also, Kartushina et al. (2015) demonstrated that visual articulatory feedback significantly improved Danish learners' vowel production, highlighting the value of multimodal training.

Olson (2021) examined how learners acquire voice onset time (VOT) contrasts in L2 stop consonants and found that while training improved production of voiceless stops, voiced stops showed less improvement. This suggests that phonetic contrasts relying on subtle timing cues, such as glottal vibration and VOT, may be more difficult to acquire. For Czech learners, whose native phonology lacks final obstruent voicing, this poses a significant challenge. The present study builds on Olson's findings by testing whether targeted phonetic training can improve English voicing assimilation in such learners.

Beyond SLA, phonetic training is widely used in speech therapy to assist individuals with articulation disorders, such as lisps. Since both voicing errors in L2 speech and articulation disorders involve difficulties in precise phonemic control, similar phonetic training principles—focused auditory input, articulatory precision, and feedback-based learning—apply across both domains (Lipetz and Bernhardt 2013). Understanding how phonetic training improves voicing assimilation not only contributes to language learning research but also has implications for speech therapy interventions.

2.4.2 *Approaches to Phonetic Training*

Various phonetic training techniques have been developed, each targeting different aspects of speech perception and production. These approaches can be broadly categorized into auditory training, articulatory training, multimodal feedback, and computer-assisted pronunciation training (CAPT) (Bogach et al. 2021).

2.4.2.1 *Auditory Training*

Auditory training has been shown to improve learners' ability to distinguish minimal pairs that are phonemically contrastive in English but neutralized in their native language. For Czech learners, this includes contrasts like /b/ vs. /v/ or voicing in final obstruents. In English, final obstruents maintain their voicing distinction, as in *bad dog* vs. *bat dog*, whereas in Czech, such contrasts are neutralized through final devoicing (Cvrček 2015). This means that both *bad* and *bat* would be pronounced [bet] in Czech, making this distinction particularly difficult for learners without targeted training.

Theoretical models suggest that articulatory and perceptual feedback interact dynamically in shaping phonological categories (Kröger, Kopp, and Lowit 2009). Additionally, for individuals with articulation disorders, auditory training can enhance their ability to differentiate and produce sounds more accurately, which is essential for speech correction (Lipetz and Bernhardt 2013).

2.4.2.2 *Articulatory Training*

Articulatory training involves explicit instruction on the positioning and movement of speech organs, often using real-time visual feedback. One effective approach is **ultrasound tongue imaging (UTI)**, which allows learners to visualize their tongue shape and position during speech in real-time. UTI is a non-invasive method that provides immediate visual information about tongue articulation, making it easier for learners to understand and correct articulatory errors (Kocjančič, Bořil, and Hofmann 2024).

Kocjančič, Bořil, and Hormann (2024) demonstrated that UTI can be effectively used in L2 pronunciation training in classroom settings. Their study showed that Czech learners of Swedish vowels, a particularly challenging set of targets due to large cross-linguistic differences, were able to improve their production after only a few short training sessions. Although both ultrasound tongue imaging and acoustic formant visualization were beneficial, UTI led to a greater number of improved vowel targets, suggesting its particular strength in promoting articulatory awareness and adjustment.

Suemitsu et al. (2015) showed that real-time articulatory feedback improved learners' production of a difficult English vowel, suggesting this approach can help with sounds requiring precise tongue positions. In clinical applications, this method is also used to treat speech disorders such as lisps, where individuals may struggle with proper tongue placement for producing fricatives (Lipetz and Bernhardt 2013).

In addition to assisting with segmental features such as vowel quality and fricative production, ultrasound tongue imaging could also be adapted to support learners in managing voicing contrasts by helping them better coordinate laryngeal and supralaryngeal articulation during connected speech.

Overall, the use of articulatory visual feedback tools, especially ultrasound imaging, provides L2 learners with a clearer, more tangible understanding of speech production mechanisms, facilitating more rapid and accurate pronunciation development.

2.4.2.3 Multimodal Feedback and Computer-Assisted Pronunciation Training

Technological advancements have led to the integration of multimodal feedback in phonetic training. Computer-Assisted Pronunciation Training (CAPT) systems use automatic speech recognition to provide learners with visual and auditory feedback on their pronunciation. Research indicates that multimodal training, which combines auditory, visual, and tactile feedback, enhances both perception and production skills. Modern CAPT systems often display acoustic visualizations (pitch curves, spectrograms, etc.) to help learners compare their pronunciation with a model (Bogach et al. 2021)

The present study incorporates CAPT tools to help Czech learners visualize how their pronunciation differs from native English speakers, focusing on maintaining voicing in final obstruents.

2.4.2.4 High Variability Phonetic Training

High variability phonetic training (HVPT) has been shown to improve learners' perception of difficult second-language contrasts by exposing them to multiple speakers and varied phonetic contexts (Bradlow et al. 1997). Although this method was originally developed for /r-l/ training in Japanese learners, it has been extended to other contrasts and learner groups. For Czech learners of English, HVPT offers promising potential in addressing challenges with voicing assimilation—particularly in final obstruents, where Czech phonology neutralizes voicing distinctions maintained in English.

Thomson (2011) demonstrated that HVPT can enhance both perception and production of segmental contrasts, especially when variability is paired with focused input. This aligns with the Revised Speech Learning Model (SLM-r), which emphasizes that successful L2 phonetic learning—even in adults—depends on exposure to rich input and continued perceptual learning (Flege and Bohn 2021).

High variability phonetic training (HVPT) has been shown to be effective in helping second-language learners acquire difficult phonemic contrasts. Research indicates that training with multiple speakers improves learners' ability to distinguish sounds that do not exist in their native language by promoting more generalized phonemic representations. This approach gradually reduces reliance on speaker-specific acoustic cues and strengthens category formation (Bradlow et al. 1997). Additionally, visual feedback techniques, such as those used by Kartushina et al. (2015), have demonstrated improvements in vowel production, suggesting that multimodal input can further support phonetic learning.

Research has also highlighted the importance of integrating HVPT into long-term phonetic training programs. While short-term HVPT sessions improve learners' ability to distinguish non-native contrasts, continuous exposure and reinforcement are necessary for these improvements to transfer to spontaneous speech (Bogach et al. 2021; Thomson and Derwing 2015). Combining HVPT with real-time feedback and interactive pronunciation exercises enhances learners' ability to both perceive and produce target phonemes correctly.

The success of HVPT suggests that variability is a crucial factor in effective phonetic training. While traditional methods focus on repeated exposure to a single pronunciation model, HVPT's use of diverse stimuli strengthens learners' adaptability to real-world language use. This approach is particularly beneficial for the participants of this study, as it equips them with the perceptual flexibility needed to navigate the variable pronunciation patterns found in native English speech.

For Czech learners of English, HVPT could be useful in addressing difficulties with English voicing assimilation. HVPT could train Czech learners to recognize and maintain voicing in final obstruents where English maintains the voicing contrast. By exposing learners to numerous speech samples, HVPT might encourage them to internalize the correct phonemic boundaries of the target language.

2.4.3 Factors Influencing Phonetic Training Outcomes

Several factors affect the success of phonetic training, including age, linguistic background, training intensity, and individual cognitive abilities. One of the most significant influences is the learner's exposure to the target language and their ability to integrate new phonetic patterns into their speech (Thomson and Derwing 2015).

According to the Speech Learning Model–revised (SLM-r), the development of new phonetic categories depends not only on the perceptual similarity between L1 and L2 sounds but also on the quantity and quality of input (Flege and Bohn 2021).

The effectiveness of phonetic training also depends on the learner's ability to perceive phonetic contrasts. Some learners may struggle with voicing distinctions due to perceptual biases formed by their native language. Olson (2021) demonstrates that feature size, such as the difference in voice onset time (VOT), significantly affects learnability, with smaller or gradient contrasts being more difficult to acquire. For Czech learners, this means that training must explicitly address the differences in voicing assimilation rules between Czech and English to prevent the automatic transfer of Czech assimilation patterns to English pronunciation.

Additionally, motivation and cognitive engagement play an essential role in training success. Learners who actively participate in phonetic exercises and receive immediate feedback tend to achieve more lasting improvements. Research indicates that self-monitoring techniques, such as recording and analyzing one's own pronunciation, can enhance learning outcomes (Thomson and Derwing 2015).

2.4.4 Challenges and Limitations of Phonetic Training

Despite its benefits, phonetic training presents several challenges. One of the most persistent issues is the difficulty of transferring trained pronunciation skills to spontaneous speech. Learners often perform well in controlled exercises but struggle to apply these skills in real-world interactions (Thomson and Derwing 2015). To address this, phonetic training should integrate conversational practice and contextualized speech exercises, ensuring that new pronunciation habits extend beyond isolated drills.

Another challenge is the reliance on visual feedback, particularly in CAPT-based training. While visual feedback is valuable for demonstrating articulatory targets, over-reliance on visual cues can prevent learners from developing strong auditory perception skills (Thomson and Derwing 2015). Effective training programs should balance visual,

auditory, and physical feedback, such as sensations of movement and positioning, to create a well-rounded learning experience.

2.4.5 Future Directions in Phonetic Training Research

As phonetic training continues to evolve, new research areas are emerging. One promising development is the use of neurofeedback and EEG-based training, which monitors brain activity in response to phonetic stimuli and provides real-time adjustments to training exercises (Kim, Emory, and Choi 2021). This approach could enhance learners' ability to internalize new phonetic contrasts by directly targeting neural processing mechanisms.

Additionally, future research might explore the use of Virtual Reality (VR) as a tool for pronunciation training, as immersive technologies could potentially support more engaging and individualized learning environments.

Ultimately, these advancements suggest a shift towards personalized and adaptive phonetic training programs, capable of specifically addressing the unique phonetic and phonological challenges faced by Czech learners of English. Such approaches could significantly enhance pronunciation training outcomes, facilitating more efficient mastery of English voicing assimilation.

2.4.6 Conclusion

Phonetic training is an essential component of second language acquisition, aiding learners in overcoming phonetic challenges posed by their native language. It is also an effective tool in speech therapy, helping individuals with articulation disorders, such as lisps, improve their pronunciation accuracy (Lipetz and Bernhardt 2013), and while traditional auditory and articulatory training remain effective, technological advancements in CAPT and multimodal feedback provide promising new avenues for improving pronunciation.

3 Practical Part

3.1 Research Design

The aim of the practical part of this study is to pilot test the effectiveness of a phonetic training program designed specifically to support Czech learners of English in mastering English voicing assimilation.

The practical work did not intend to provide a large-scale empirical study but rather served as initial pilot testing of a targeted training script developed in **Praat** (Boersma and Weenink 2024), a phonetic analysis and training software. The script was designed to guide learners through perception and production tasks, specifically targeting final obstruent voicing in contexts where English retains the voicing contrast across word boundaries—a contrast often neutralized in Czech. The training incorporated both auditory input and visual-acoustic feedback. These included waveforms and spectrograms, enabling participants to see their own speech output in real time and make targeted articulatory adjustments.

Each participant completed a **pretest**, which consisted of recording a list of phrases involving final obstruents followed by voiced or voiceless consonants. This served as a baseline for later comparison.

The individual training session that followed used a Praat script (Boersma and Weenink 2024) developed for this study. In each trial, participants first saw the target phrase and heard a recording by a native speaker. They then attempted to repeat the phrase, after which the script displayed both their own and the native speaker's waveforms and spectrograms for comparison. This visual-acoustic feedback aimed to raise awareness of voicing differences and support more accurate production.

To gain additional insights into the participants' background, attitudes, and perceived difficulties with English pronunciation, a short questionnaire was administered after the training. The questionnaire included both closed and open-ended questions and aimed to gather qualitative feedback on the participants' awareness of voicing distinctions, their exposure to English, and their impressions of the training session.

While this pilot study, due to its limited scope, cannot yield statistically robust conclusions, it provides valuable information for refining the training procedure and for considering the feasibility of future research with a larger sample size. It also offers

practical insights into how Czech learners respond to phonetic training grounded in visual-acoustic feedback.

3.2 Participants

To protect participants' privacy, their identities have been anonymized using initials. No personally identifiable information was collected, and participation in the study was voluntary.

The participants in this pilot study were seven Czech university students. All participants were either majoring or minoring in English Philology at Palacky University; no distinction between major and minor programs was made for the purposes of this study. All participants, except for one, were native speakers of Czech and they all had an advanced level of English proficiency (at least B2 according to the CEFR scale), as required by their study program.

Participants were selected on the basis of convenience sampling, as they were accessible and willing to participate in the pilot training session. While this limits the generalizability of the findings, it was sufficient for the exploratory nature of this pilot study, whose primary aim was to assess the feasibility and clarity of the training procedure rather than to make generalizable claims.

None of the participants reported any speech or hearing disorders. Prior to the training, all were given an explanation of the aim of the study. Specifically, they were informed that the training focused on improving pronunciation of English voicing contrasts and that they would receive individual feedback in the form of visual-acoustic representations of their speech.

To better understand the background of each participant and their subjective experiences with English pronunciation, a brief follow-up questionnaire was administered after the training. This included questions about their awareness of voicing assimilation, perceived pronunciation difficulties, and general impressions of the training method. These insights helped complement the recorded data with qualitative reflections and were useful for evaluating learner awareness and engagement.

3.3 Materials and Stimuli

Table 1: Stimulus Phrases by Voicing Context

The following table (Table 1) summarizes the 54 stimulus phrases used in the phonetic training program, grouped by type of word-boundary voicing context, including a separate analysis of the typical manner of articulation involved in the examples.

Category	Number of Phrases	Example Phrases	Manner of Articulation (Examples)
Final voiced obstruent before voiceless consonant	12	brave person, good time, large field	fricative–stop, stop–stop, fricative–fricative
Final voiceless obstruent before voiced consonant	18	help them, ask them, back there	stop–fricative, stop–fricative, stop–fricative
Final obstruent before sonorant	13	nice meeting, sharp rock, good morning	fricative–nasal, fricative–stop, stop–nasal
Fillers (voiced–voiced X voiceless–voiceless)	11	good day, hot soup, fresh start	stop–stop, stop–fricative, fricative–fricative

The stimuli used in this pilot phonetic training program were carefully selected to reflect real-world examples of English voicing assimilation across word boundaries. A total of 54 short phrases were compiled, systematically grouped into three main categories: (1) phrases with final voiced obstruents followed by initial voiceless consonants (e.g., *brave person, good time*), (2) phrases with final voiceless obstruents followed by initial voiced consonants (e.g., *let's go, this day*), and (3) phrases with final obstruents preceding sonorants (e.g., *nice meeting, sharp rock*), the latter reflecting dialectal tendencies found in Moravian Czech and Slovak, where assimilation may occur even before sonorants. The target phrases were selected with the aim of eliciting voicing assimilation challenges typical for Czech learners of English.

In addition to these, a few filler phrases were included in the pretest to help balance learners' attention across sentence types. These fillers did not contain contexts where voicing assimilation would typically occur, thereby serving for minimizing response bias.

To ensure naturalness and representativeness, the **British National Corpus (BNC)** (Davies 2004) was consulted during phrase selection. Words were searched individually, and collocations were examined based on their frequency of occurrence, with attention

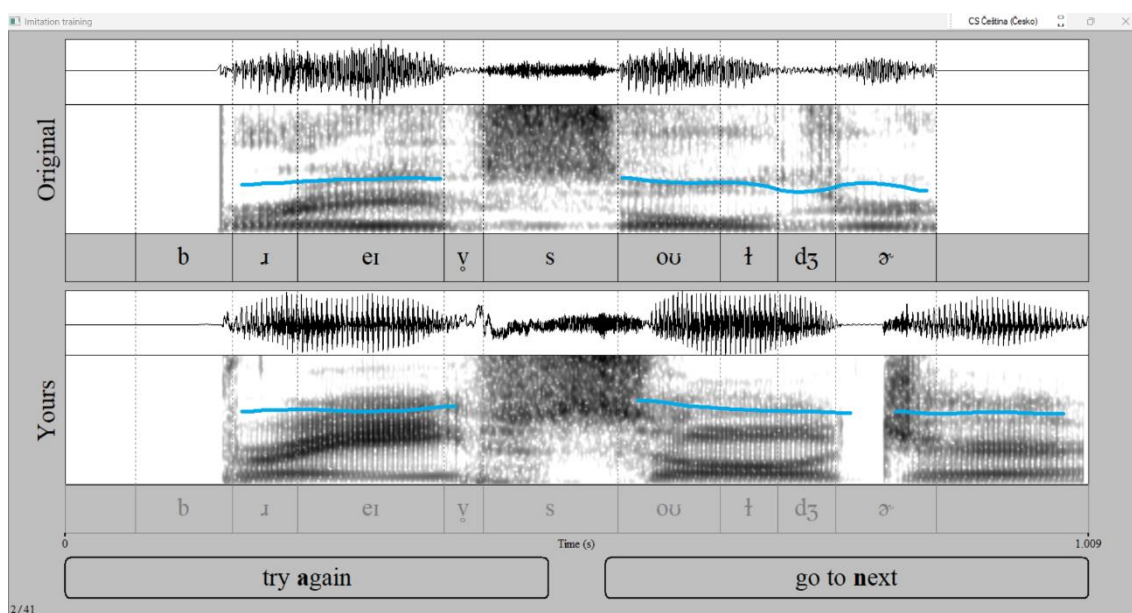
focused on combinations in which the second word ranked among the top 100 most frequent co-occurring items. Preference was given to phrases that appeared frequently in spontaneous speech and reflected common usage patterns. Although the process was not formally quantified, the reliance on high-frequency collocational data helped ensure that the selected items were both authentic and representative of naturally spoken English.

Corresponding recordings were sourced via **YouGlish**, prioritizing clarity and variation. All native speaker recordings were saved using **Audacity**, an open-source digital audio editor and recording application. This ensured clarity and consistency in the audio stimuli presented during training (“Audacity: Computer Software” 2024). Importantly, a range of native speaker accents (including both regional and standard British varieties) was included to implement **High Variability Phonetic Training (HVPT)**. This exposure to diverse pronunciation models aimed to promote learners’ generalization of voicing patterns beyond a single speaker or dialect. Native pronunciation was verified manually by checking the phonetic realization of the final obstruents and their assimilation patterns.

To prepare the training stimuli, each target phrase was phonetically transcribed and segmented using the *TextGrid* feature in Praat (Boersma and Weenink 2024). Segmentation was guided by established principles of phonetic boundary marking, particularly those focusing on the identification of articulatory transitions between consonants (Macháč and Skarnitzl 2010). The alignment of these boundaries was crucial in providing precise and reliable visual feedback to learners. Both the model (native speaker) and the learner’s own productions were segmented and annotated, allowing for direct comparison in waveform and spectrogram displays.

The training script in Praat (Boersma and Weenink 2024), custom-built for this project, presented the target phrase visually and aurally. It then prompted the participant to repeat the phrase and displayed both the waveform and spectrogram of their recording alongside the native speaker model. This visual-acoustic feedback was designed to enhance learners' perceptual awareness of voicing distinctions and to encourage self-correction based on phonetic cues.

Figure 1: Screenshot of the Praat training interface displaying waveform and spectrogram.



Although brief, this type of training introduces learners to the use of instrumental phonetic tools in a pedagogical context and opens up possibilities for further refinement of their pronunciation habits through data-driven self-monitoring. The integration of authentic language input, precise phonetic segmentation, and interactive visual feedback makes this pilot training program a promising step toward more evidence-based and personalized approaches in second language pronunciation instruction.

3.4 Procedure

The practical part of this study aimed to pilot test a phonetic training script developed in *Praat* (Boersma and Weenink 2024) to support Czech learners of English in improving their pronunciation of English voicing assimilation. The procedure consisted of several stages: selecting materials, preparing the training script, conducting the pretest, guiding participants through the training, and administering a follow-up questionnaire.

3.4.1 Pretest

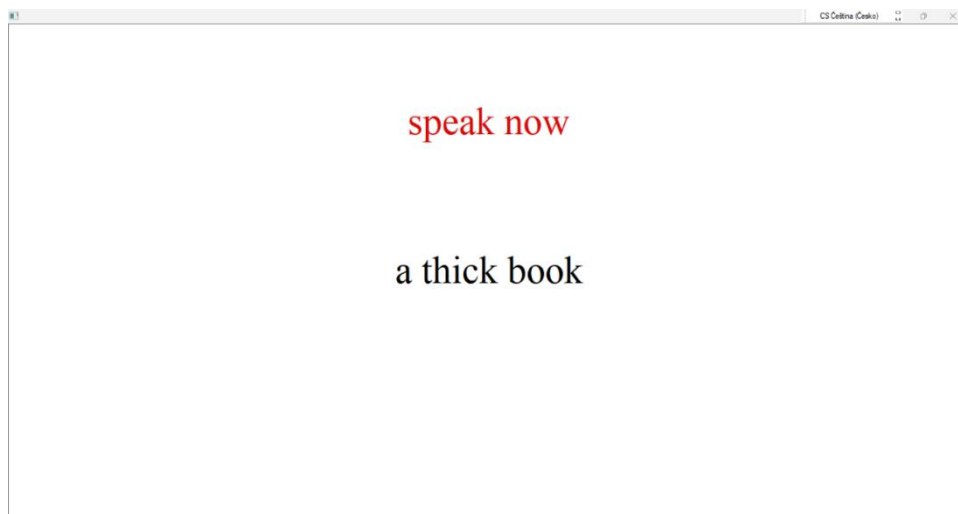
Prior to the training, participants completed an extended pretest designed to establish a detailed baseline for later comparison. This task involved reading a list of approximately 50 phrases, which included both test phrases specifically targeting voicing assimilation (e.g., “back there“ vs. “large field“) and filler phrases to reduce predictability and maintain natural reading rhythm (e.g., “good day”).

The recordings were conducted in a quiet room using a high quality microphone and the *Praat* software (Boersma and Weenink 2024). Participants were instructed to read

each phrase naturally at a conversational pace. Importantly, no corrective feedback or additional instructions on pronunciation were provided during this stage to ensure the responses reflected spontaneous production.

At the start of the pretest session, the researcher clearly explained how the program functioned, demonstrating how the recordings were made using five initial practice examples. During the pretest, phrases appeared sequentially on the screen accompanied by a red prompt that read "speak now" (see Figure 2) clearly indicating the precise moment when participants should begin speaking.

Figure 2: Screenshot of the Praat pretest interface showing the 'Speak now' prompt



Although it was technically possible to pause the program between individual phrases, the researcher intentionally avoided doing so, believing interruptions would disrupt the natural flow and potentially increase errors when restarting. However, this continuous approach led to participants reporting that the number of phrases made the pretest somewhat exhausting, highlighting the need to adjust the procedure in future implementations.

3.4.2 Training session

Following the pretest, each participant completed one individual training session, lasting approximately 10 to 15 minutes. The training was conducted using a custom Praat script, with the researcher providing support throughout. The goal of the session was to raise participants' awareness of English voicing assimilation and to allow them to practice producing target phrases with real-time visual-acoustic feedback.

Each exercise (e.g., *this guy*) began with the visual representation of the phrase, listening to a native speaker's recording of the same phrase, and then attempting to

replicate it. Participants were allowed to replay the native speaker's recording multiple times to familiarize themselves with the target pronunciation before attempting their own production.

After their recording, the Praat script displayed both the waveform and spectrogram of the native speaker model and the participant's production side by side. To assist interpretation, the script was designed to highlight periods of voicing with a blue line, making it visually clear where vocal fold vibration occurred (see Figure 1).

The researcher provided additional support by guiding participants in reading and interpreting the spectrograms. This was done in a consistent way across participants. For example, they were shown where to look for the key consonants between words and were encouraged to pay close attention to voicing cues, such as the presence of a blue line indicating vocal fold vibration. While feedback wasn't given after every phrase, the researcher would occasionally point out whether assimilation had occurred or not, especially during the first few examples. The goal was to help participants learn how to recognize these patterns on their own over time. This semi-guided approach gave participants a clearer sense of what to look for and helped them become more aware of how their own pronunciation compared to the native model.

However, several limitations were encountered during the session. The native recordings were extracted from naturally occurring speech via YouGlish and were therefore relatively fast-paced. This required participants to produce their utterances at a similarly high speech rate, which added to the complexity of the task. Additionally, technical issues with the Praat script occurred, which occasionally disrupted the session flow and prevented most participants from completing all 41 phrases. Nevertheless, each participant succeeded in recording at least 20 target phrases, providing a sufficient dataset for observation and informal comparison.

3.5 Data Collection and Analysis

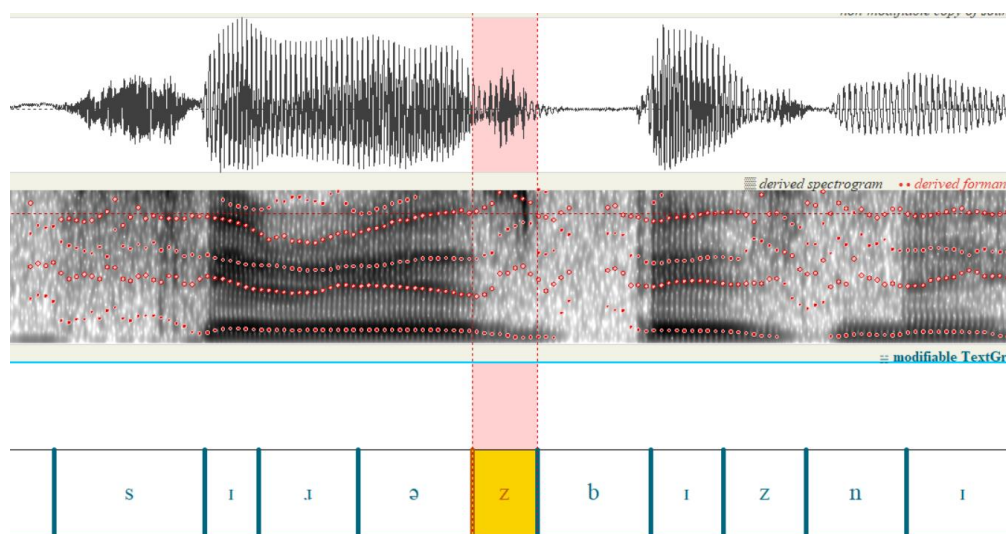
The data collected for this pilot project included two main types: audio recordings (from both the pretest and the training) along with the visual representations of these recordings (waveforms and spectrograms), and written responses from a post-training questionnaire.

3.5.1 Audio recordings and waveforms

The primary dataset consisted of audio recordings collected during both the pretest and training sessions. Each participant recorded a series of English phrases containing final obstruents in contexts that, for Czech speakers, typically trigger voicing assimilation. The primary aim was to investigate whether a short phonetic training intervention could help Czech learners improve the production of voicing in connected speech, aligning their pronunciation more closely with English phonotactic patterns.

The analysis of recordings was conducted through a combination of auditory perceptual judgment and instrumental inspection using Praat. The results were recorded in a spreadsheet. Each entry included the target consonant, the participant's realization in the pretest and post-training recordings, any observed assimilation behavior, and an overall evaluation of change (see Figure 3)

Figure 3: Example of Speaker Evaluation



Importantly, recordings were assessed by **direct comparison between each participant's pretest and corresponding post-training realizations**, without reference to an idealized native-speaker model. The analysis focused on whether the participant's pronunciation improved relative to their own baseline performance.

Categories of evaluation were defined as follows:

Improvement: clear reduction of Czech-like regressive assimilation patterns.

Partial improvement: some reduction in assimilation behaviors, though not complete.

No improvement: no observable change between pretest and training realizations.

Deterioration: increase in Czech-like regressive assimilation after training.

It is important to note that "no improvement" does not necessarily indicate that the participant produced incorrect voicing or that assimilation was present; it simply reflects the absence of change relative to the pretest production. Similarly, minor phonetic assimilation, which is common in casual native English speech, was not automatically treated as an error. Only clear Czech-style regressive assimilation across word boundaries was marked as deterioration.

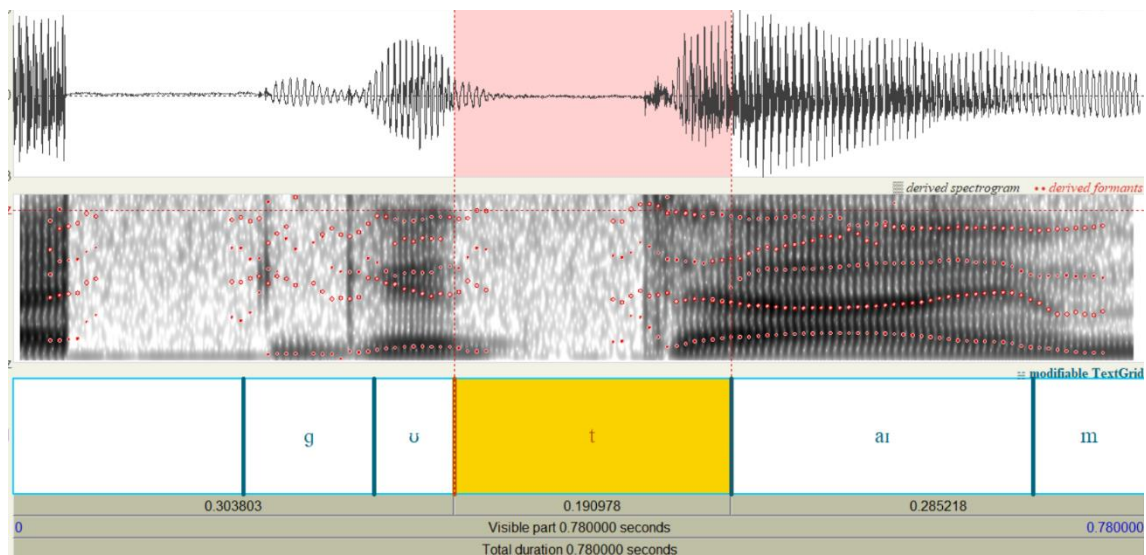
Table 2: Summary of Participant Outcomes Based on Audio Recordings

Category	Number of Tokens	Percentage
No improvement	132	63.8%
Deterioration	26	12.6%
Improvement	10	4.8%
Partial improvement	2	1.0%
Inconclusive	36	17.4%
Total	207	100%

A total of 207 phrase recordings were collected across seven participants. Of these, 132 (63.8%) showed no improvement, 26 (12.6%) were marked as deterioration, 10 (4.8%) were classified as improvement, and 2 (1.0%) were evaluated as partial improvement. Thirty-six tokens (17.5%) were deemed inconclusive due to missing or ambiguous voicing realization.

Most inconclusive cases involved final stops where the closure was unreleased or elided entirely, leaving no acoustic material on which to assess voicing. Since the consonant in question was effectively absent from the signal, these tokens were excluded from analysis. In a few additional instances, technical issues during recording resulted in missing or unreadable data, also leading to an inconclusive classification (see Figure 4).

Figure 4: Example of an unreleased final stop. No clear release burst is present, making it impossible to assess assimilation acoustically. As such, this token was classified as inconclusive.



The breakdown by target consonant revealed additional patterns. Final voiceless consonants like /p/, /t/, and /k/ tended to remain stable, whereas obstruents like final /s/, /v/, and /dʒ/—showed high instability. Final /s/ was the most problematic, with 33 tokens showing no improvement and 10 showing deterioration. This pattern reflects Czech speakers' tendency to simplify final fricatives, either through full assimilation or incomplete articulation.

3.5.1.1 Participant-Specific Results

Participant-specific analysis revealed that the majority of learners demonstrated limited progress. For instance, Participants TH, LP, and EV had the highest number of tokens rated as "no improvement," while Participants AV and LB each showed isolated cases of improvement, primarily in easier target contexts such as final /p/ or /f/. Nonetheless, across all participants, there was no meaningful shift suggesting consistent learning or generalization of voicing maintenance rules between the pretest and post-training phases.

Interestingly, one participant underwent a striking transformation between the pretest and training. During the pretest, the participant spoke with a strong Czech accent, characterized by clear Czech intonation patterns and stress placed in a regular, Czech-like pattern. However, in the training phase, their speech suddenly mirrored the intonation and rhythm of the native recordings to such an extent that the researcher initially suspected a file mix-up. Upon careful checking, it became clear that the recordings were indeed from the same participant.

This phenomenon highlights how closely participants may mimic models they are exposed to when visual-acoustic feedback is combined with auditory input. However, the

extent to which such short-term adjustments reflect stable pronunciation acquisition remains uncertain and suggests that future research should examine long-term retention effects.

3.5.1.2 Effects of Speech Rate and L1 Transfer on Training Outcomes

Another major factor influencing voicing behavior was speech rate. During training, participants noticeably increased their speech tempo in an effort to match the native speaker recordings. This faster speech led to greater coarticulatory effects, increasing the likelihood of Czech-style regressive assimilation. When participants were allowed to speak at their preferred, natural pace as in the pretest, the rate of assimilation was considerably lower (Agwuele, Sussman, and Lindblom 2009). This observation supports earlier findings that faster articulation generally enhances coarticulatory effects, while careful speech tends to reduce them.

Thus, faster speech during training recordings often led to greater assimilation but did not necessarily progress toward more native-like English voicing patterns. In many cases, it represented a reversion to L1 phonological strategies.

This discrepancy suggests that future training procedures would benefit from including slowed-down native speaker recordings, allowing learners to focus more precisely on the intended voicing contrasts before gradually increasing speech rate. Slower input models could reduce the cognitive burden on learners, facilitating better acquisition of subtle phonetic cues that are otherwise masked during rapid articulation.

3.5.1.3 Conclusion

Despite isolated examples of improvement and some increase in phonetic awareness (confirmed by questionnaire responses), there was no significant overall difference between the pretest and post-training recordings. Most participants continued to exhibit Czech voicing patterns even after the training session. This confirms that a single, short intervention is insufficient for reorganizing deeply entrenched L1 phonological habits.

In conclusion, while the training program showed promise as a method for raising phonological awareness, the limited duration, fast-paced input, and strong L1 transfer prevented significant short-term changes in pronunciation. These findings reinforce the view that pronunciation improvement, particularly in subtle features such as English voicing assimilation, requires sustained, carefully structured practice over an extended period rather than a single session.

Recordings are available as supplementary digital material submitted along with the thesis.

3.5.2 *Questionnaire analysis*

A comprehensive post-training questionnaire was administered to students to supplement the phonetic training session and gain detailed insights into participant backgrounds and experiences. The questionnaire received responses from six participants, providing valuable qualitative and quantitative data.

All respondents except one were Czech native speakers enrolled in English Philology at Palacký University. Specifically, five respondents were in their 2nd year of studies, one in the 1st year. Regarding accent, four participants (66.7%) identified their accent as Moravian, while one (16.7%) identified as Bohemian and one (16.7%) as Slovak.

When rating their pronunciation proficiency on a scale from 1 (below average) to 5 (excellent, near-native), five participants (83.3%) rated themselves at level 3, indicating average proficiency, and one participant (16.7%) rated themselves at level 4, indicating good proficiency. This distribution suggests that the participants generally perceive their pronunciation as adequate but see room for improvement.

All six participants (100%) reported studying English phonetics and phonology during their coursework. However, notably, none (0%) of the participants had practical pronunciation training integrated into their previous phonetics coursework. This highlights a critical gap between theoretical knowledge and practical application, underscoring the need for hands-on phonetic training modules.

Regarding awareness of voicing assimilation before the training, six participants (100%) indicated that they had studied it previously in their phonetics or phonology courses. Opinions varied on whether Czech speakers naturally use voicing assimilation when speaking English: three participants (50%) answered "Yes," demonstrating awareness of linguistic interference, while three participants (50%) answered "No," indicating less awareness of native language influences.

All respondents (100%) found listening to native speakers beneficial for their pronunciation development, citing enhanced awareness and improved auditory discrimination skills. Additionally, four participants (66.7%) found phonetic transcription (IPA) helpful, while two (33.3%) did not. These findings suggest variability in

participants' comfort and familiarity with phonetic transcription, highlighting the need for additional or simplified instruction in using IPA.

Participants reported increased confidence in their pronunciation after training, with four participants (66.7%) rating their confidence at level 4 (high confidence) and two participants (33.3%) rating it at level 3 (moderate confidence). All respondents (100%) would recommend this type of training to other students, reflecting high perceived effectiveness and overall satisfaction with the training approach.

Participants reported several challenges during the training: two mentioned difficulties related to listening, particularly due to the high speech rates of native speakers. One participant specifically mentioned difficulty pronouncing certain unfamiliar words without audio recording in the pretest, while another participant struggled to fit the activities within the allotted time. These responses suggest that future training should include longer sessions or slow down speech samples initially.

Additionally, some participants expressed difficulty with technical aspects, such as interpreting spectrograms and waveforms, suggesting that future training should include a preliminary session to explain these acoustic visualization tools.

All participants (100%) expressed strong interest in more pronunciation-focused instruction as part of their regular academic program. Several mentioned (during post training conversations) that ongoing structured training could significantly enhance their proficiency and confidence in speaking contexts, particularly in academic and professional environments. This unanimous response underscores the participants' recognition of the critical role pronunciation plays in effective communication and language mastery, thus advocating for integrating sustained phonetic practice within their curricular activities.

3.6 Limitations and Lessons Learned

Based on the results and participant feedback from this pilot testing, several limitations were identified, and corresponding improvements for future implementations can be proposed. A primary limitation was the relatively small sample size, consisting of only seven participants, which restricts the generalizability of the findings. Future studies could address this limitation by recruiting a larger and more diverse group of participants, potentially providing more robust and generalizable conclusions.

Additionally, the single-session format limited participants' exposure and practice opportunities. Some learners reported difficulty matching their pronunciation to the native model, mainly due to the rapid speech rate of the native recordings. To mitigate this challenge, future training could integrate multiple sessions over an extended period or include slowed-down versions of native recordings alongside the original-speed audio, enabling learners to more clearly perceive and imitate subtle voicing contrasts.

Another issue that emerged during the training sessions was the participants' initial difficulty with technical components, particularly interpreting spectrograms and waveforms. While visual feedback was beneficial overall, some learners found the interface unintuitive at first. Addressing this in future revisions might involve providing an introductory tutorial or demonstration session that clearly explains these acoustic visualization tools. Additional refinements such as simplified visual layouts, color-coded cues, or automated alignment of native and learner recordings could significantly enhance usability and accessibility.

A notable technical limitation encountered during the pilot was related to the functionality of the phonetic training script and issues with the recording device settings. Due to these technical difficulties, the script malfunctioned repeatedly, causing interruptions and requiring frequent restarts of the training sessions. Consequently, fewer successful recordings were captured than initially planned, and multiple restarts were necessary to achieve at least 20 original phrases per participant. These technical challenges were subsequently resolved, ensuring smoother operations in future implementations.

Further improvements could also involve integrating perception tests to evaluate learners' abilities to identify voicing contrasts explicitly. This addition would allow for a more comprehensive assessment of both production and perception skills, offering valuable insights into learners' phonetic awareness. Incorporating longitudinal tracking could also provide valuable data on long-term retention and the practical efficacy of phonetic training sessions over time.

Overall, while this pilot study successfully raised participants' awareness and highlighted the potential effectiveness of targeted phonetic training, it clearly identified critical areas for refinement. These insights emphasize the importance of ongoing refinement in phonetic training methodologies, aiming for enhanced pedagogical effectiveness and learner engagement in future studies.

4 Conclusion

This thesis set out to investigate voicing assimilation across word boundaries, focusing specifically on the challenges Czech learners of English face due to cross-linguistic phonological differences. While English voicing assimilation is largely phonetic and optional, Czech voicing assimilation is generally systematic but subject to certain exceptions, creating frequent transfer issues for Czech speakers when producing connected speech in English.

The theoretical part of the thesis provided a detailed overview of voicing assimilation from both cross-linguistic and language-specific perspectives. It outlined the predominantly systematic nature of voicing assimilation in Czech, contrasted with the more variable and speaker-dependent patterns observed in English. This contrast laid the foundation for understanding why Czech learners often misapply their native phonological rules when speaking English.

The practical part presented a pilot study testing a phonetic training program developed in Praat. The training combined auditory input and visual-acoustic feedback with the aim of raising learners' awareness of English voicing patterns and encouraging more native-like speech production. Pretest and post-training recordings were analyzed to assess improvement, and participants' perceptions were gathered through a post-training questionnaire.

The findings suggest that while participants displayed increased phonetic awareness after the training, no significant systematic improvement in pronunciation was observed. Instead, many participants showed either no change or even deterioration in voicing patterns, often linked to an increased speech rate during training. These results highlight the difficulty of modifying deeply rooted first-language phonological habits through a single short intervention.

Nevertheless, the study revealed important insights into learners' responses to phonetic feedback, the challenges of interpreting acoustic visualizations, and the influence of speech rate on voicing assimilation. It also demonstrated strong participant interest in further pronunciation training, underlining the practical relevance of this research area.

Future research should aim to implement longer, multi-session training programs with a greater focus on perception tasks, slower model recordings, and clearer guidance in interpreting acoustic feedback. Incorporating longitudinal studies would also allow for

the assessment of long-term improvements and retention of newly acquired pronunciation habits.

In conclusion, while this pilot study did not yield dramatic short-term improvements, it contributed valuable knowledge to the relatively underexplored topic of voicing assimilation in English as a second language and highlighted the necessity for tailored, long-term phonetic training approaches for learners from typologically different language backgrounds.

Resumé

Tato bakalářská práce se zabývá problematikou asimilace znělosti na hranicích slov v angličtině, se zvláštním zaměřením na výslovnost českých studentů anglického jazyka. Zatímco v češtině je asimilace znělosti většinou systematickým fonologickým procesem, v angličtině se jedná spíše o jev fonetický, variabilní a často závislý na tempu a plynulosti řeči. Tyto rozdíly mezi jazyky představují pro české mluvčí významnou výzvu, zejména při produkci plynulé angličtiny.

Teoretická část práce shrnuje poznatky o asimilaci znělosti v češtině a angličtině, popisuje jejich hlavní rozdíly a zaměřuje se na specifické potíže, které tyto rozdíly způsobují při osvojování anglické výslovnosti. Zvláštní důraz je kladen na roli fonetického tréninku, zejména na možnosti využití sluchového vstupu, artikulační zpětné vazby a počítačem asistovaného tréninku výslovnosti jako nástrojů pro podporu nácviku výslovnosti.

Praktická část představuje pilotní studii, ve které byl testován fonetický trénink využívající kombinaci sluchového vstupu a vizuálně-akustické zpětné vazby prostřednictvím skriptu v programu Praat. Sedm účastníků nejprve absolvovalo pretest a následně tréninkovou intervenci, během níž byly pořizovány záznamy pro srovnání. Výsledky ukázaly, že u většiny účastníků nedošlo ke statisticky významnému zlepšení ve výslovnosti cílových hláskových kombinací. Naopak, v některých případech došlo k vyšší míře asimilace znělosti v důsledku rychlejšího tempa řeči během tréninku.

Závěry práce naznačují, že krátkodobý trénink není dostatečný k překonání hluboce zakořeněných fonologických návyků mateřského jazyka. Budoucí výzkum by měl zahrnovat delší tréninkové programy s důrazem na postupné navyšování obtížnosti a podporu vnímání jemných fonetických rozdílů.

Tato práce přispívá k poměrně málo prozkoumané oblasti asimilace znělosti v angličtině jako cizím jazyce a zároveň poukazuje na nutnost systematické a dlouhodobé fonetické výuky při zvládnutí výslovnostních obtíží českých studentů angličtiny.

References

- Agwuele, Augustine, Harvey M. Sussman, and Björn Lindblom. 2009. "The Effect of Speaking Rate on Consonant Vowel Coarticulation." *Phonetica* 65 (4): 194–209.
- Audacity Team. 2024. "Audacity: Computer Software." <https://www.audacityteam.org/>.
- Boersma, Paul, and David Weenink. 2024. "Praat: doing phonetics by computer: [Computer program]." <https://www.praat.org>.
- Bogach, Natalia, Elena Boitsova, Sergey Chernonog, Anton Lamtev, Maria Lesnichaya, Iurii Lezhenin, Andrey Novopashenny, et al. 2021. "Speech Processing for Language Learning: A Practical Approach to Computer-Assisted Pronunciation Teaching." *Electronics* 10 (3): 1–22.
- Bradlow, Ann R., David B. Pisoni, Reiko Akahane-Yamada, and Yoh'ichi Tohkura. 1997. "Training Japanese listeners to identify English /r/ and /l/: IV. Some effects of perceptual learning on speech production." *J Acoust Soc Am.* 101 (4): 2299–2310.
- Broersma, Mirjam. 2005. "Perception of familiar contrasts in unfamiliar positions." *J Acoust Soc Am.* 117 (6): 3890–3901.
- Cvrček, Václav. 2015. *Mluvnice současné češtiny: Jak se píše a jak se mluví*. Praha: Karolinum.
- Davies, Mark. 2004. "British National Corpus." <https://www.english-corpora.org/bnc/>.
- Dmitrieva, Olga, Allard Jongman, and Joan A. Sereno. 2020. "The Effect of Instructed Second Language Learning on the Acoustic Properties of First Language Speech." *Languages* 5 (4): 1–33.
- Flege, James E., and Ocke-Schwen Bohn. 2021. "The Revised Speech Learning Model (SLM-r)." In *Second Language Speech Learning*, edited by Ratrete Wayland, 3–83. Cambridge University Press. <https://doi.org/10.1017/9781108886901.002>.
- Gow, D. W. 2002. "Does English coronal place assimilation create lexical ambiguity?" *Journal of Experimental Psychology: Human Perception and Performance* 28 (1): 163–179.
- Hume, Elizabeth, and Keith Johnson. 2001. "The Role of Speech Perception in Phonology." In *The Role of Speech Perception in Phonology*, Elizabeth Hume, edited by Elizabeth Hume, 3–26. Brill Academic Pub.

Jansen, Wouter. 2007. "Phonological 'voicing', phonetic voicing, and assimilation in English." *Language Sciences* 29 (2–3): 270–293.

Kartushina, Natalia, Alexis Hervais-Adelman, Ulrich Hans Frauenfelder, and Narly Golestani. 2015. "The effect of phonetic production training with visual feedback on the perception and production of foreign speech sounds." *Journal of the Acoustical Society of America* 138 (2): 817–832.

Kim, Subong, Caroline Emory, and Inyong Choi. 2021. "Neurofeedback Training of Auditory Selective Attention Enhances Speech-In-Noise Perception." *Front. Hum. Neurosci.* 15: 1–10.

Kocjančič, Tanja, Tomáš Bořil, and Susanna Hofmann. 2024. "Acoustic and Articulatory Visual Feedback in Classroom L2 Vowel Remediation." *Language and Speech*: 1–22.

Kröger, Bernd J., Stefan Kopp, and Anja Lowit. 2009. "A model for production, perception, and acquisition of actions in face-to-face communication." *Cogn Process* 11: 187–205.

Lipetz, Heidi Massel, and B. May Bernhardt. 2013. "A multi-modal approach to intervention for one adolescent's frontal lisp." *Clinical Linguistics and Phonetics* 27 (1): 1–17.

Macháč, Pavel, and Radek Skarnitzl. 2010. *Principles of phonetic segmentation*. Vol. 14. Edice Erudica. Nakladatelství Epoque.

Myers, Scott. 2010. "Regressive voicing assimilation: Production and perception studies." *Journal of the International Phonetic Association* 40 (2): 163–179.

Ohala, John. 1993. "The Phonetics of Sound Change." In *Historical Linguistics Problems and Perspectives*, Charles Jones, 237–278. London: Longman.

Olson, Daniel J. 2019. "Phonological processes across word and language boundaries: Evidence from code-switching." *Journal of Phonetics* 77 (100937): 1–16.

Olson, Daniel J. 2021. "Phonetic feature size in second language acquisition: Examining VOT in voiceless and voiced stops." *Second Language Research* 38 (4): 913–940.

Simon, Ellen. 2010. "Phonological transfer of voicing and devoicing rules: evidence from L1 Dutch and L2 English conversational speech." *Language Sciences* 32 (1): 63–86.

Skarnitzl, Radek, and Pavel Šturm. 2014. "Assimilation of Voicing in Czech Speakers of English: The Effect of the Degree of Accentedness." *Research in Language* 12 (2): 199–208.

Skarnitzl, Radek, Pavel Šturm, and Jan Volín. 2016. *Zvuková báze řečové komunikace: Fonetický a fonologický popis řeči*. Univerzita Karlova v Praze: Nakladatelství Karolinum.

Suemitsu, Atsuo, Jianwu Dang, Takayuki Ito, and Mark Tiede. 2015. "A real-time articulatory visual feedback approach with target presentation for second language pronunciation learning." *The Journal of the Acoustical Society of America* 138 (4): EL382–EL387.

Šturm, Pavel, and Lea Tylečková. 2019. "Dialectal differences in voicing assimilation patterns: The case of Moravian Czech English." *Acta Universitatis Carolinae Philologica* 2019 (2): 129–143.

Thomson, Ron I. 2011. "Computer Assisted Pronunciation Training: Targeting Second Language Vowel Perception Improves Pronunciation." *CALICO Journal* 28 (3): 744–765.

Thomson, Ron I., and Tracey M. Derwing. 2015. "The Effectiveness of L2 Pronunciation Instruction: A Narrative Review." *Applied Linguistics* 36 (3): 326–344.

Volín, Jan, and Radek Skarnitzl. 2006. "Fonologická výjimečnost české znělé labiodentály." In *Kapitoly z fonetiky a fonologie slovanských jazyků*, Jarmila Janoušková, edited by Zdena Palková, 251–264. Praha: Filozofická fakulta Univerzity Karlovy.

Appendix

Appendix A	Stimuli Spreadsheet
Appendix B	Native speakers recordings with TextGrids
Appendix C	Praat script pretest
Appendix D	Praat script training
Appendix E	Participants' recordings pretest
Appendix F	Participants' recordings training
Appendix G	Questionnaire replies
Appendix H	Analysis of Participant Recordings Spreadsheet

(See attached files for Appendices A–H)