Univerzita Palackého v Olomouci

Filozofická fakulta

Katedra anglistiky a amerikanistiky

Vowel Length as a Cue to the Perception of Word-final Voicing by Czech Learners of English

Bakalářská práce

Jan Fusik

Anglická filologie

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

Olomouc 2019

Prohlašuji, že jsem tuto bakalářskou práci vypracoval samostatně a uvedl úplný seznam citované a použité literatury.

V Olomouci dne _ _ _ _ _ _ _ _ _ _ _ _ Podpis _ _ _ _ _ _ _ _ _ _ _ _

Acknowledgments

I would like to thank my supervisor, Mgr. Václav Jonáš Podlipský, PhD., for his kind support, expertise on demanding tasks as well as overall guidance during the process of writing this thesis. Additionally, my thanks belong to my mother who provided an outstanding support, vital for this thesis' completion. Participants who devoted their time to contribute to the pilot study of this thesis are also subjects of my appreciation.

Table of Contents

1. Introduction

The relation between the phonological voicing value of an obstruent consonant in a syllable coda and the duration of the vowel in the nucleus of that syllable has been a widely researched topic. While in the English language a vowel is clearly shorter when followed by a voiceless coda and longer when followed by a voiced coda, the same pattern has not been reliably observed in native Czech speakers' production or perception. In order to expand our knowledge of vowel duration followed by voiced/voiceless coda and of how Czechs perceive it, a pilot study has been conducted and will be elaborated on in this thesis.

This thesis is structured into five main sections. The Introduction establishes the main topic of this work as well as the aim of this study and reviews general knowledge about and the concept of voicing. Then, it describes the English phenomenon of vowel duration variation due to coda voicing [introduce an abbreviation?]. Previous research in this field is reviewed and significant findings of studies of the relationship between nucleus vowel duration coda obstruent voicing across multiple languages are highlighted. The two examined languages, Czech and English, are contrasted.The last subsection of the Introduction explores the perception of word-final voicing and its cues by both native and non-native speakers. The Introduction concludes by presenting research questions which the thesis aims to answer.

Following sections describe the pilot study which was conducted and serves as the main body of the thesis: Participants, research techniques, equipment, testing environment and procedures are listed and analyzed in the Methods section. All results and findings of this experiment are cited in the section labeled Results.

Main purpose of the last two sections, Discussion and Conclusion is to interpret obtained results and answer the research questions formulated at the end of Introduction. Several comments regarding used methods and overall experiment execution are made.

1.1. Voicing

As described by Baart (2010), voicing is one of the phonetic as well as phonological features of consonants. Consonants are usually described phonologically as either voiced [+voiced] or voiceless [–voiced], i.e. as bearing one of the two values of a binary feature. Voicing occurs when vocal folds vibrate during speech process. When airflow is not obstructed by vocal folds, no vibration is generated and obstruents are perceived as voiceless. When the vocal folds are pulled closer to each other, airflow creates a pressure which causes vocal folds to vibrate and articulated obstruents are then perceived as voiced (Baart 2010, 77-78; Ladefoged 2015, 4; Napoli 1996, 15-17). The phonetic reality of implementing voicing during the production of an obstruent consonant gesture, however, is relatively complex.

1.1.1. Phonation modes

Besides voicing, humans are able to control glottis voluntarily to achieve various speech sounds with with different attributes. These phonation types can be categorized into multiple degrees. For easier realization, Ladefoged (1971) introduces a continuum of phonation types with two extremities – left side symbolizing the most open aperture between arytenoid cartilages and the most closed aperture representing the right side.

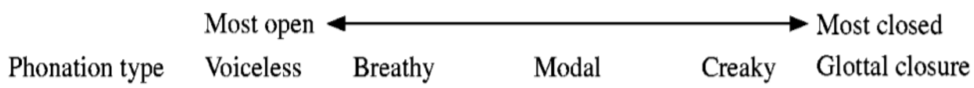


Figure 1: Continuum of phonation types (after Ladefoged, 1971; taken from Gordon & Ladefoged, 2001)

Gordon & Ladefoged (2001) make a distinction between five major phonation types:

Voiceless phonation takes place when the vocal folds are furthest apart, the airflow is not obstructed. It is most commonly found among plosives and contrasted in many languages.

Breathy voice is characterized by fairly abducted vocal folds and being longitudinaly relaxed. When the airflow goes through the glottis, airflow is turbulent and creates an impresion of voice mixed with breath. On the waveform, breathy voice can be identified by aperiodic noisy energy. Similarly, a high frequency noise is present at spectrogram images. Together with creaky voice, breathy voice is the mosst common nonmodal phonation type.

Modal voice is characterized as regular voicing, achieve by vocal folds being adducted tightier than in case of breathy voice. Modal voiced has well-defined pitch pulses on waveform representation and clear nasal-to-vowel transitions on spectrogram.

Unlike breathy voice, creaky voice is produced when vocal folds are tightly adducted but still allow for voicing. It is also sometimes refered to as "vocal fry." Irregularly spaced pitch periods and decreased acoustic intensity are characteristic for waveform of creaky voiced consonants, as well as lowered fundamental frequency.

Glottal stop is a rare occurence in English language. The vocal fold vibration is absent and English phonemic glottal stops are often realized as creaky phonation on neighbouring sounds. (Gordon & Ladefoged 2001, 383-391)

1.1.2. Multiple cues to voicing

As mentioned earlier, realization of obstruent voicing is a complex process, thus allowing for several cues based on which the voicing can be identified. Stevens (2000) provides three well-known cues.

Amplitude of low-frequency periodicity related to low-frequency amplitude of the adjacent vowel. It is stated that in case of voiced consonant the low-frequency prominence of the amplitude is below 5-15 dB than in its adjacent vowel.

Another significant cue is a change in fundamental frequency at the onset of the vowel which is preceded by a consonant. If the consonant is voiced, the fundamental frequency would be lower during the onset of the following vowel. Accordingly, if the preceding consonant is voiceless, the vowel will have higher fundamental frequency during the consonantal release.

Lastly, differences in formant transition also hint to consonantal voicing. In particular, the gap between amplitues of harmonic waves is a sign of a different degree of glottal spreading.¨

Together, these cues can be summarized as evidence for vocal fold stiffening or slackening and for the presence of glottal spreading. They partly facilitate or inhibit glottal vibration during the obstruent interval. (Stevens 2000, 141-142)

Voicing is subject to phonological rules across languages such as voice assimilation and voice neutralization.

1.2. Word-final voicing neutralization

Majority of Slavic languages—including Czech—neutralize voicing word-finally (Simonchyk 2017). For the instance of Czech language Bičan (2011) presents the voicing neutralization on a minimal pair of words *led* "ice" and *let* "flight" which are both phonetically realized as [let] by Czech native speakers. Furthermore, he lists rules concerned with coda devoicing. Firstly, if the word is utterance-final and the utterance is followed by a prosodic boundary pause then the final obstruent is devoiced. In the phonetic environment where an word-final voiced obstruent is immediately followed by a voiceless consonant, the obstruent is devoiced (cf. led taje "the ice melts", realized as [lɛt‿tajɛ]). Lastly, in context of coda followed by approximants like vibrant [r] or [l] as well as nasals and the voiced dental fricative [v] without a prosodic boundary pause, the preceding word-final obstruent will be voiceless (cf. led může "the ice can", realized as [lɛt‿muːʒɛ]). (Bičan 2011, 26)

Bičan (2010) discloses that in regards to Czech phonology, both voiced or voiceless variants of word final consonats are common but they are subject to various geographical regions. Bičan uses the term archiphoneme – a subsystem which is represented by two or more phonemes with a single distinctive feature depending on the overall context onto which they are projected (in this case voicing). For example, utterance *kupme* "let's buy" could be phonetically realized as [kupmɛ] – voiceless obstruent (typical for Bohemian region) or [kubmɛ] – voiced one (typical for Moravia). While both variants are common in everyday speech, voiced variants are not regarded as correct pronunciations. (Bičan 2010, 28-40)

Among the languages affected by coda voicing neutralization are for example Russian (Kulikov and McMurray 2017), Polish (Slowiaczek and Dinnsen 1985) from the Slavic language family and German (Charles-Luce 1985) and Dutch (Berendsen, 1983; Zonneveld, 1983) from West Germanic language family to which English also belongs.

1.3. Defining the word-final voicing rule

As was foreshadowed in section 1.1 Concept of voicing, there are certain phonological rules related to voicing. My thesis focuses on the correlation between syllable-final voicing and the duration of the preceding vowel. This phenomenon is related to voicing neutralization and was first observed and described by Denes (1955). He argues that duration of the preceding vowel plays a significant role in the perception of word-final voicing for English native speakers. (Denes 1955, 764)

Raphael (1972) extends the research of Denes and observes that vowel duration is a significant and sufficient cue to coda voicing and that this cue is more effective for stop consonants than for fricatives. (Raphael 1972, 1301). Over the years, these findings about vowel duration cue were further confirmed (Berg 1994, Chen 1970, Kessler & Treiman 1997). For the sake of simplicity, to describe the general rule for coda voicing in English, I will borrow the definition from Scheer (2017) where he argues that this pattern is common for many languages: "Vowels are longer before sonorants and voiced obstruents than before voiceless obstruents." (Scheer 2017, 116)

Now that the rule has been defined and described how it behaves in English phonology, next few chapters will explore the rule in several other languages. This will allow for comparison with English phonology and help to evaluate whether the rule applies at all.

1.3.1. Vowel duration as a cue to coda voicing in Arabic

Even though the universal is true for a range of languages including English, Mitleb (1982) claims that the voicing effect on vowel duration is not an absolute universal. The evidence is a series of spectrographic tests consisting of Arabic an English minimal pairs which were produced by eight Arabs. Results obtained from these tests show that Arabic speakers were not able to implement the temporal differences of vowels for the voicing feature in English minimal pairs. These results led to conclusion that segmental voicing feature is not governed by differences in vowel duration universally, thus must be learned by second language learners. (Mitleb 1982, 23)

Findings by Mitleb have been contradicted by a later study by Embarki (2016). Embarki performed a production experiment with 16 native Arab speakers in which 64 VC syllable sequences were analyzed. The results obtained show that Arabic language is not an exception from the coda voicing rule and coda voicing affects preceding vowel. (Embarki 2016, 2-4)

1.3.2. Vowel duration as a cue to coda voicing in Dutch

Another research that is tightly related to my thesis which established (almost) similar hypotheses was conducted by Miriam Broersma. In her work, Broersma (2005) subjected native Dutch speakers to a binary choice experiment in which participants were to distinguish final voicing contrast in English utterances. Broersma utilized manipulated stimuli comprising of two CV segments (one with phonetically long vowel and the other with phonetically short vowel to which a continuum of 11 dental fricatives–created by incrementally adding up waveforms of /f/ and /v/–was added. Her research has shown that while Dutch listeners were able to distinguish English final voicing contrast with high accuracy, they did not utilize vowel duration cue in exact manner as English listeners did. Set of stimuli with intentionally mismatched voicing information as evidence. While English participants continued using this cue to decide on voicing in word-final labiodental fricative consonants, Broersma argues that Dutch participants might have adapted more flexibly by taking into consideration aditional information in the stimulus material which could have overruled using vowel duration as the main cue. (Broersma 2008, 712)

Broersma indicated that Dutch listeners did not use vowel durationas a cue for fricative voicing simply because they did not perceive it as helpful and chose to ignore it instead. (Broersma 2008, 713) Therefore, she followed up on this in

2008 (Broersma, 2008) where she attempted to test the above-mentioned explanation by conducting a similar binary choice experiment with 28 English listeners and 28 Dutch listeners, utilizing an 11-step continuum ranging from natural /v/ to a natural /f/ which was spliced onto two carrier syllables – short-vowel token [ku] and long-vowel token [ku:]. The results showed that Dutch listeners initially used the vowel duration as a cue for final /v/-/f/ categorization but the use rapidly diminished over time. English participants, just as in the 2005 experiment, continued using the vowel duration as a word-final voicing cue.

1.3.3. Vowel duration as a cue to coda voicing in Czech

Unlike English, Czech does not seem to follow the rule for final obstruent voicing based on vowel duration. Podlipský (2008) argues that this might not be entirely true and there could be an indirect support for the hypothesis that Czech in fact follows the rule of vowel lengthening before voiced coda based on the finding that in Czech there is a higher chance that quantitatively ambiguous vowel is perceived as short when preceding phonetically voiced coda obstruent.(Podlipský 2008, 6158)

Machač and Skarnitzl carried out a study to examine how duration of vowel affects duration of adjacent plosivein Czech. They extracted 638 VCV sequences from short-text recordings of 53 university students. The duration of obstruents based on the place of articulation and contrast of voicing was analyzed in the sequences as well as duration of vowels based on vowel height. The obtained results showed that vowel tends to be shorter before a voiceless plosive than before voiced plosive. However this tendency is quite weak as from the 19 relationships in VC sequences, only three were statistically significant ($p < 0.05$) and in 4 of out of 19 the tendency was opposite. (Machač and Skarnitzl 2007, 540)

Podlipský and Chládková (2007) pointed out a lack of distinction regarding syllabicity of VC sequences in the description of their experimental material and followed up with their own study to confirm their hypothesis that tautosyllabic VC sequences are affected by the consonantal voicing more than their heterosyllabic counterparts. Their study concludes with remarks that even though the voiceless coda had no effect on distinction of short/long vowel in the perception of native Czech listeners, there was a small but statistically significant shift in perceptual boundary of vowel quantity in voiceless/voiced coda context. Therefore, the results of the study advocate dependency of vowel quantity perception on coda voicing in the minds of Czech listeners. (Podlipský and Chládková 2007, 73-74)

To further confirm the findings of the previous study, based on the data which were obtained from native Czech perception tests and used as a baseline for his 2008 study (Podlipský 2008), Podlipský points out that "(Czech) listeners indeed adjust the location of the short/long vowel boundary so that a vowel needs to last slightly longer when followed by a voiced coda to be perceived as long than it does when followed by a voiceless coda." He acknowledges that the average boundary shift is not large (only 3.93ms), nevertheless, listeners seem to be very consistent in adjusting their perception of phonemic vowel quantity. (Podlipský 2008, 6158)

There have also been attempts to improve not only perception but also production of the word-final voicing based on vowel cues of Czech learners of English. Herudková (2015) in her study used a material consisting of 58 CVC English minimal pairs differing in coda voicing. A half of these stimuli was enhanced by manipulating vowel duration, transitional duration as well as the duration of constriction. From the enhanced stimuli set, half of the stimuli containing voiced codas was further manipulated to appear more ambiguous – low frequencies for word-final obstruent were filtered out based on the speaker and obstruent type and noise intensity for fricatives was increased. The other half of the stimuli containing voiced coda, the word-final obstruents were replaced with their voiceless counterparts and the noise intensity as well as constriction duration were decreased.

The experiment consisted of a pretest and two posttests, all comprising of the 58 minimal pairs of English words in their original version. The tests were designed as binary-choice identification tasks with no feedback on correct answer. Between the pretest and first posttest, a training session took place in form of another binary-choice identification task, this time with immediate feedback on correct answer, ability to relisten the provided stimulus. The training task was also designed to provide participants with higher exposure to stimuli by repeating those words, which were answered incorrectly until participant provided correc answer.

For the training session, participants were divided into three groups. Group 1 was trained using enhanced stimuli. Group 2 continued using the original, unaltered stimuli. Group 3 was a control group and did not undergo any training.

Additionally, after the perceptual testing, a production test was conducted using a selection of six participants from each group. They were recorded individually i na soundproof room, reading out loud 21 English minimal pairs written on a sheet of paper.

Even though the study failed to confirm the hypothesis that native Czech listeners would be able to learn to use vowel duration as a cue to word-final voicing in English phonetic environment, due to high percentage of correct reponses (over 60% for all three groups) in the pretest, the findings support the claim that the degree of proficiency in L2 (English) influences the perception of word-final voicing. This hypothesis was suggested by Smith et al (2009, 272).

Moreover, the experiment was carried out using high-variability phonetic training — naturally-produced stimuli from multiple speakers in order to avoid specific phonetic patterns produced by a single individual and to introduce more variability and reduce predictability. Despite its efficiency in prior studies (Bradlow et al. 1997; Wong 2014) Results showed that the high-variability training did not improve subjects' performance which could have been caused by short period of the training task as Herudková notes. (Herudková 2015, 55-56)

The aforementioned studies by Herudková and Broersma have been described in greater detail since my own research follows up on them and is directly linked to the work of Herudková. As will be further disclosed in Methodology chapter, Herudková's material had also been used in my thesis.

1.3.4. Vowel duration as a cue to coda voicing in other Slavic languages

Research conducted by Slowiaczek and Szymanska (1989) suggests this language universal does not apply to Polish listeners. In their research, Polish and English native speakers underwent a forced-choice identification task for Polish monosyllabic minimal pairs. It was revealed that even though vowel duration seemed to play a significant role in coda voicing in a production test (Szymanska, Dinnsen; 1985) it had not been used as a primary cue for voicing in perceptual tests to diferentiate the minimal pairs. In fact, vowel duration cue is poorly acquired even by experienced Polish learners. It is one of the most difficult temporal cues for voicing for Poles and they are not able to control this vowel duration cue. (Slowiaczek and Szymanska 1989, Waniek-Klimczak 2005)

Another research concerned with acquisition of [+voice] feature in word-final position by native speakers of Russian comes from Simonchyk (2017) and indicates that "Russian learners of English did not manipulate vowel durations with respect to the voicing status of the following consonants. Instead, Russian learners used durations of closure and voicing into closure to encode voicing." (Simonchyk 2017, 1)

Judging by the diverse results of multiple studies in different languages, there is no definite resolution whether the examined phenomena of coda voicing effect on vowel duration can be applied to whole range of languages, establishing it as an absolute language universal.

1.4. Implicit versus explicit learning

In Order to asses and describe the perceptual training which the experimental group underwent, we must outline the theory behind language learning and training which will aid in describing the perceptual training methods utilized in my thesis and its overall effectivity.

As Rebuschat in his introduction to *Implicit and Explicit Learning* (2015) acknowledges, process of language acquisition can be classified into two main categories – implicit and explicit.

Implicit learning is "a fundamental feature of human cognition" (Rebuschat 2015, xiii) Rebuschat then further explains on a study by Reber (1967) that implicit learning is integral to the process of language acquisition and while person who acquires implicit knowledge is not able to describe what that knowledge is specifically, they keep utilizing it without being aware of the knowledge. However, implicit knowledge usually gives a sense of intuition to the affected person of learning something, yet the subject matter remains unspecific for the person. (Rebuschat 2015, xiii-xiv)

On the other hand, explicit knowledge, as described by DeKeyser (2008), involves instructions which direct learner's focus on specific structure. DeKeyser further distinguishes between deductive and inductive explicit learning. For example traditional methods of teaching fall into the category of deductive explicit learning whereas instructing students to find a rule in provided sample is an example of inductive explicit learning. (DeKeyser 2008, 314)

DeKeyser concludes the comparison of explicit and implicit learning with a statement based on study of Robinson (1966) that implicit learning proves to be not as effective as deductive explicit learning. Moreover, this does not apply to designed implicit learning exclusively but also to any learning in implicit condition. (DeKeyser 2008, 336) Above claim by DeKeyser is reinforced by findings of others (Alanen 1995, Baleghizadeh & Derakhshesh 2017, Ellis 2002).

1.5. High variability versus low variability phonetic training

In addition to the distinction between implicit and explicit learning, we need to define another approach to participant's training. In large volume of studies which incorporate instructional input with an aim to improve participant's knowledge, the methods utilized in such study are usually defined in terms of variability.

Bariuso & Hayes-Harb (2018) characterize high variability phonetic training as a technique which introduces variability of input into the perception practice. It is believed that higher diversity of inputs improves learner's ability to perceive non-native sounds. Over the years, HVPT became a significant factor in linguistic research, particularly in studies concerned with speech. (Bariuso & Hayes-Harb 2018, 177-180)

Unlike HVPT, Low variability phonetic training is limited in terms of input. The most usual limitation is concerned with stimuli production which includes only a single speaker or a single phonetic context.

When we compare effectivity of both phonetic trainings, large body of studies confirm that HVPT is generally more effective than LVPT. (Bradlow et al. 1997, Lively et al. 1994, Wong 2012, Logan et al. 1991) Wong argues that even though studies tend to support the above claim, they cannot support it directly because the results from such studies have not compared the effects of HVPT and LVPT simultaneously. Wong therefore conducted an experiment which utilized both approaches in production and perception tasks. Furthermore, Wong employed Cantonese low-proficiency and high-proficiency L2 speakers of English. The aim of the experiment was to improve perception and production of non-native contrast of /e/ and /æ/ vowels. While both low and high variability training modified listener's perception and production of contrast of the vowels and marked their significant improvement, HVPT proved to be more effective than LVPT. Wong concludes that not only high-proficiency learners are able to benefit from HVPT but low-proficiency learners can improve to same extent. (Wong 2014, 1-5)

1.6. Research Questions and Hypotheses

Q1: Will the perceptual training affect the ability to use vowel duration as a cue to coda voicing by improving the ratio of correct responses in the posttest as compared to the pretest?

H1: We expect that if we expose listeners to the knowledge of how duration of vowel affects following word-final consonant in English and the visual representation of vowel duration available in the trainer, they will be able to utilize the vowel duration variation to decide on coda voicing better. This should result in higher ratio of correct responses in posttest.

Q2: Will the perceptual training improve the reaction time with which listeners decide on coda voicing?

H2: As explained in first hypothesis, not only we expect the ratio of correct responses to improve but we also expect the amount of time during which listeners decide on the coda voicing to decrease.

Q3: Does vowel duration of nonsense words affect the ability to decide on coda voicing?

H3: Apart from the existing English words, we expect the listeners to be able to take into account vowel duration as a cue to word-final consonant voicing as the nonsense words avoid bias of familiarity.

Q4: Does phonological knowledge of L2 affect listener's performance?

H4: We speculate that naive listeners (i.e. listeners with no previous phonological or phonetic knowledge) will be outperformed by listeners who have been exposed to explicit knowledge of L2 phonology in terms of ratio of correct responses and reaction times.

2. Methodology

In order to achieve relevant results, set of training procedures had been established and performed. The procedures, as well as researched participants and material used in the experiment, are detailed in this chapter.

2.1. Material

Research material consisted of three main classes. First class of stimuli was—as mentioned in 1. Introduction — taken directly from the research conducted by Herudková. Specifically, set of one hundred English words from which 42 of them were minimal pairs differing in the word-final consonant. As reported by Herudková, the soundfiles were obtained from online MacMillan Dictionary using software Audacity. To avoid prosodical effect as much as possible, all words were monosyllabic with a CVC structure. It is worth noting that part of the stimuli with supposed word-final voicing was partially devoiced at the end of utterance which is absolutely natural and in accordance with phonological rules of English. Whole set of one hundred stimuli was further divided into two subclasses. The original set of stimuli that remained unchanged and manipulated set of stimuli that underwent following enhancements:

Half of the voiced stimuli (i.e. 29 stimuli) had their word-final consonant replaced by a voiceless consonant of the same consonantal type. Constriction duration of the obstruent was then decreased to exact half of duration between the boundaries of voiceless and voiced version. The noise intensity of the replacing voiceless consonant was also lowered (by 4 dB for fricative and by 6dB for stop consonant). (Herudková 2015, 39) The purpose of these manipulations was to make the stimuli more indistinguishable for Czech listeners and therefore raise the vowel duration as a cue for voicing to prominence.

The second class of the material consisted of thirty nonsense sounds. They were representation of a made-up word "yig." To prepare a set of stimuli with no actual meaning, following procedure was carried out:

Several nonsense words were proposed as candidates and from these propositions a narrower sample of seven minimal pairs was created by selecting those words that did not reflect an association with any existing meaningful word.

This proved to be a rather difficult task due to the restrictions that were established. Not much variability was allowed – only monosylabic words ending with consonantal stops were accepted. It is nearly impossible for minimal pairs of such words not to evoke a similar sounding items in both English and Czech language which carry a meaning. Therefore, it was agreed to lower the restrictions that were initially set and words that evoked as less meaning as possible were selected. These stimuli were then recorded by a native English male speaker in a controlled environment using portable digital audio recorder. After an assesment of the seven recorded minimal pairs, sound "yig" from minimal pair "yik/yig" was selected to be used and further modified for this experiment. Complete list of all candidate sounds are included in Appendix.

Next in the process of creating stimuli, recorded sound "yig" was artificially modified using PRAAT software. Modification was divided into two steps:

Voiced word-final consonant [g] was partially devoiced to create more ambiguity in the matter of voicing. The consonantal duration was manipulated creating three instances of the consonant with different durations (60ms, 85ms and 120ms). Additionally, release phase of the obstruent and fundamental frequency was also manipulated, to further impose the most [g]-like impression on one consonant, one with the most [k]-like impression and last one was left the most ambiguous, somewhat a compromise between the two.

Duration of vowel preceding the word-final obstruent was also adjusted resulting in a set of ten vowels with various vowel lengths ranging from 100ms to 180ms. Vowel length was modified using logarithmic durational interpolation. Reason for choosing natural logarithm function to modify the duration over increasing the duration incrementally was a more natural characteristic of the logarithm over simple multiplier. As Reichl (2010) states: "With the sensory system of the human body, in many cases this dependency turns out to be of logarithmic nature. Recent quantitative QoE[1] research shows that in several different scenarios a similar logarithmic relationship can be observed between the

[1] QoE = Quality of Experience

size of a certain QoS[2] parameter of the communication system and the resulting QoE on the user side as observed during appropriate user trials." (Reichl et al. 2010, 1) This behaviour is defined in Weber-Fechner's Law. It is therefore more viable to use natural logarithmic units rather than static increments when manipulating stimuli.

Last set of stimuli contained thirteen minimal pairs of English words with divergency in word-final consonantal voicing. Though comparable to the first set of stimuli, there was no overlap of same English words between the two sets. The soundfiles were obtained from online Macmillan Dictionary through online service. For each stimulus, a waveform graph with spectrogram and annotations was created for purpose of the perceptual training which will be explained in 2.3. Procedure.

For further information and full list of the material, please refer to Appendix 1 and 2 in which all stimuli are listed and categorized in detail.

2.2. Participants

Native Czech speakers who were students of Palacký University in Olomouc and University of Ostrava from various fields of study and had no previous linguistic or phonological knowledge (ie. naive speakers) have taken part in this study. All respondents had average or above average fluency in English language and had not reported any difficulties with understanding given instructions or presented stimuli. The age of respondents ranged from 18 to 26 (mean age≈21,5). None of the participants suffered from hearing or speech impairments.

Respondents were divided into two groups. Control group contained 2 males and 1 female (mean age≈22). Experimental group consisted of 6 males and 4 females (mean age≈22).

Subsequently, an additional control group consisting of 18 females and 2 males (mean age≈21) was included in the experiment due to shortcomings of the original control group. However, this new control group consisted of students of English philology at Palacký University in Olomouc who were attending atleast one course in phonetics at that time, thus being exposed to theoretical

[2] QoS = Quality of Service

phonological knowledge of English as well as having average or above average fluency in English.

2.3. Procedure

Experiment comprised of two perceptual tests and a perceptual training session.

At the start of the experiment, both control and experimental group underwent perceptual test. Test written in PRAAT (for whole script, please refer to Appendix 3), consisted of the one hundred stimuli from Herudková's thesis and to those the whole set of nonsense "yig" stimuli was added. The test was structured as follows: Instructory introduction with volume check to adjust the volume based on respondent's needs followed by listening phase divided by two short intermissions. Due to repetitive nature of the tests, participants were informed about estimate length of the test and aforementioned intermissions. This allowed respondents to have general awareness of the progress which was supposed to help avoid negative impact on the results by keeping the respondent from being impatient or fatigued.

For each question, a sound was played with no option to replay and the respondent was presented with a binary forced-choice answer in a form of an ortographic representation of the sound with both versions of the concerned minimal pair. The inability to replay the sound was implemented due to possibility of individual advantage which could have misrepresent final results.

Experimental group was then subjected to a perceptual training. The training was realized by an imitation exercise in which a respondent was presented with a sound stimulus. The respondent's task was to immitate the stimulus by pushing a recording button. After the recording was saved, the orthographic form of the current stimulus was revealed. The task then proceeded to a comparison screen in which a waveform graph and spectrogram of the original sound was compared to graphs of the recorded sound. After an inspection of the graphs, respondent moved to next stimulus. For the perceptual training, the third set of thirteen minimal pairs was used. The stimuli were subjectively ordered from easiest to hardest in terms of recognition. Training was conducted under general guidelines which were drawn on the basis of previous pilot training:

1. let the respondent solve first two minimal pairs with no input from the supervisor
2. ask them for their feedback, any noticeable differences in the minimal pairs
3. explain purpose of the task along with phenomena of voicing and its effect on vowel duration
    a. comparison of vowel length and voicing in CZ and EN
    b. phonological rule of longer preceding vowel resulting in voiced consonant and vice versa in EN
    c. demonstrate on third pair of stimuli
4. let the respondent practice on the rest of stimuli
5. be prepared for feedback and questions

The training phase was a single session with every respondent treated separately from others. Treating each individual separately helped to control the overall progress of the experiment and aided in avoiding incidental learning which could occur if two or more participants were present at the same time. The perceptual training lasted approximately one hour during which all the steps from the guidelines were performed.

After the perceptual training of experimental group was completed, both groups underwent perceptual test identical to the test at the beginning. The second perceptual test was conducted with a six-day delay from the first test.

For further information regarding the software, tools and devices used during the material preparation and the experiment procedure, please refer to Appendix 7 in which all instruments are listed and described in detail.

3. Results

Results which have been obtained during the experiment phase (described in 2.3. Procedure) will be declared in this chapter. The results were gathered from both control and experimental groups in the form of perceptual pretest and perceptual posttest output. The perceptual training for experiment group has not been recorded in any manner as it has been deemed irrelevant to the final outcome. It was expected that the training session would affect the results obtained from perceptual posttest meaning it would most likely be redundant to additionally record and analyze the imitation exercise which was part of the perceptual training. The expected outcome of the results was an improvement in percentage of correct responses and lower reaction time in perceptual posttest for experimental group.

The data were recorded for all 13 listeners which have been listed in 2.2. Paticipants. The data analysis was conducted for two dependent variables. The first dependent variable was proportion correct. The second dependent variable was reaction time.

3.1. Proportion correct

Firstly, if we compare the ratio of correct responses between all three groups in both tests (Figure 2), we see that experimental group's mean percentage of correct responses was 74% while control group A with naive speakers achieved mean 69% of correct responses, making the difference of mean correct responses between the two groups only 5%. The control group B of phonologically learned participants achieved mean 79% of correct responses. In comparison to the other two groups, there is an inclination of phonological learnedness being an advantage. The difference in mean percentage of correct responses between the three groups has not been found significant ($p>0,05$).

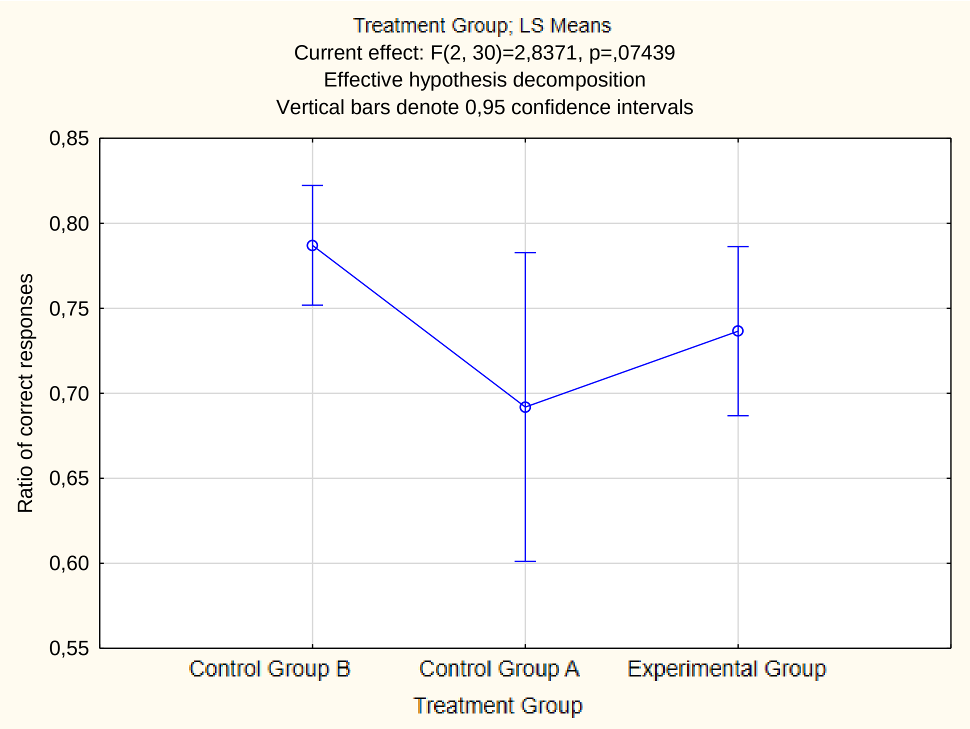


Figure 2 Percentages correct: The Main Effect of Treatment Group

Figure 3 shows a percentage correct responses based on within-subject variable of time. The mean percentage of correct responses including all three groups in perceptual pretest was approximately 72%. In the perceptual posttest, the mean percentage of correct responses was 76% which marks a genuine improvement. This main effect of test is significant (F (1, 30)=15,332, p=0,00048).

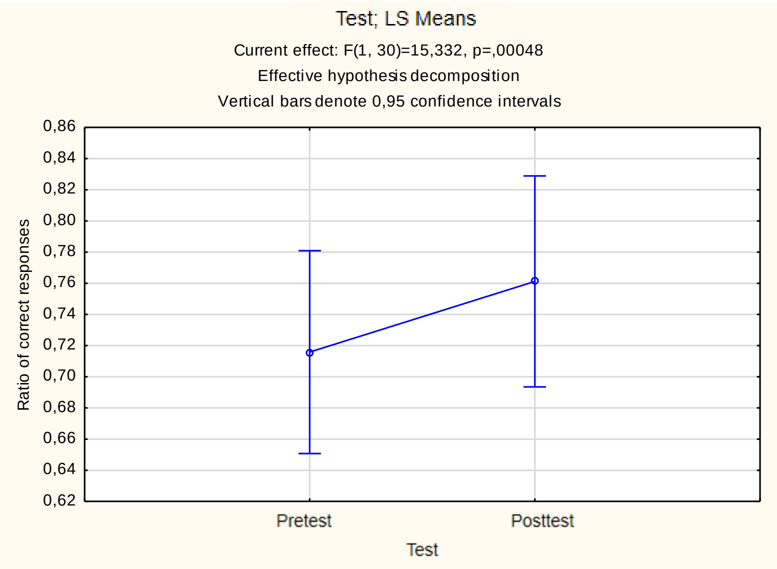


Figure 3 Percentages Correct: The Main Effect of Tests

Next, in order to assess how effective the perceptual training based on percentage of correct responses was, ANOVA for two between-subject variables of group and time was calculated (see Figure 4). From the graph, we see that mean percentage of correct responses inclines to be higher in perceptual posttest for all groups, however, the effect is not significant (p>0,05).

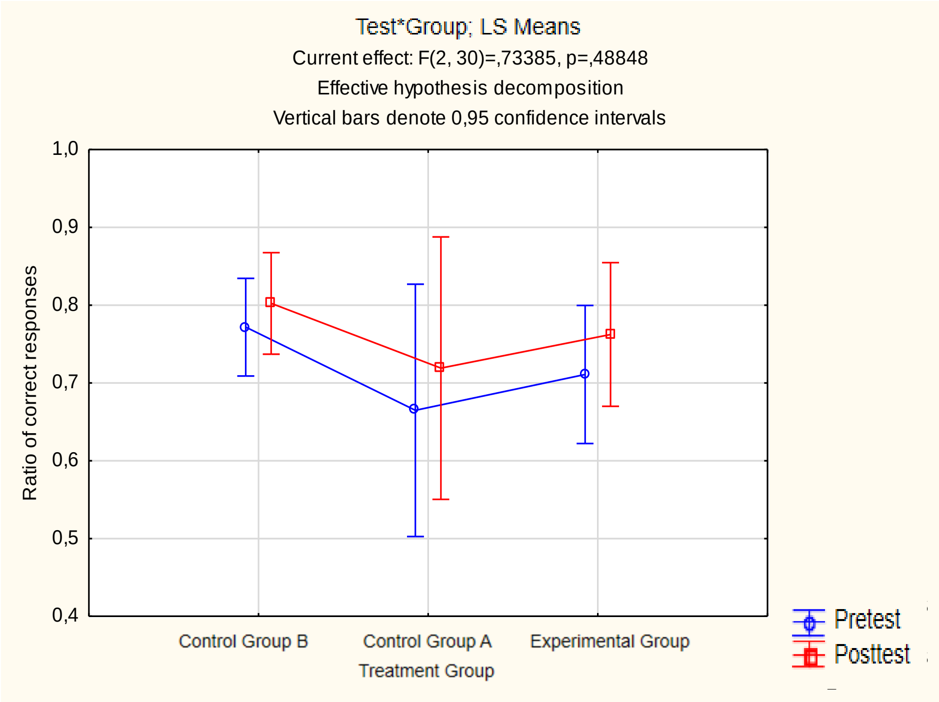


Figure 4 Percentages Correct: The Interaction Between Treatment Group and Tests

To have an estimation on how correctly were participants judging perceived word-final consonants altogether, a proportion correct ANOVA with a within-subject variable of obstruent voicing type was calculated. All three voicing types were compared — voiceless, naturally voiced and manipulated voiced obstruent.

From the graph (Figure 5), it is apparent that word-final obstruents with underlying voicing were the most easiest to identify for all three groups with mean ratio of correct responses of 87%. The amount of correct responses for voiceless obstruents were lower, averaging at 76%. The lowest mean percentage of correct responses was achieved for manipulated voice obstruents with an average 58% which is slightly above a mere chance. Significant main effect of the within-subject factor of obstruent voicing type was found for all three variants according to posthoc test. (F=(2, 60)=57,831 p=0,000117 between voiceless and naturally voiced; p=0,000117 between naturally voiced and manipulated voiced; p=0,000117 between voiceless and manipulated voiced).

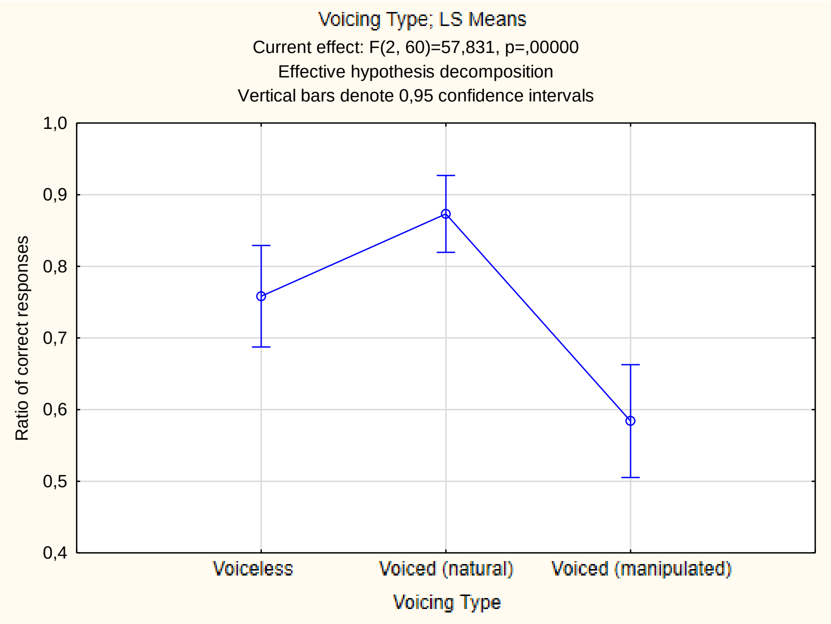


Figure 5 Percentages Correct: The Main Effect of Obstruent Voicing

With a significant main effect found for obstruent voicing type, a between-subject factor of group was added to the calculation to examine how were individual groups affected. (see Figure 6) Mean percentages of correct responses between the groups varied slightly for each respective obstruent voicing type. Although the highest difference in mean ratio of correct responses can be noted for manipulated voiced obstruents, the overall main effect is not significant ($p>0,05$).

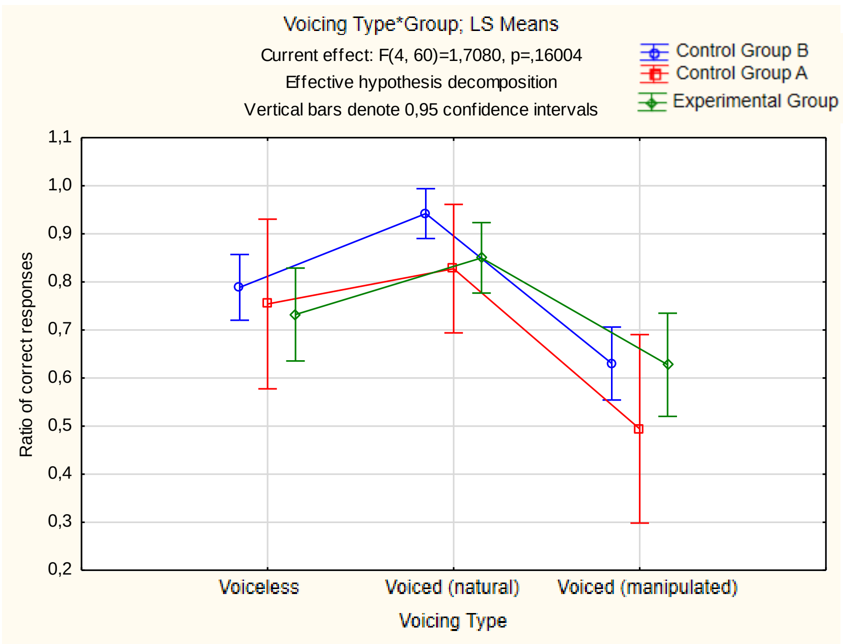


Figure 6 Percentages Correct: The Interaction Between Obstruent Voicing and Group

Last graph for dependent variable of proportion correct (Figure 7) shows two within-subject variables of obstruent voicing type and time as well as a within-subject variable of group. In the Figure 7, we see that perception of manipulated voiced obstruents in pretest with experimental and control groups A's of naive speakers mean percentage of correct responses below 60% provides the highest opportunity to improve. Even phonologically aware participants were able to achieve only mean 62% of correct responses. Nonetheless, the main effect is not significant (p>0,05).

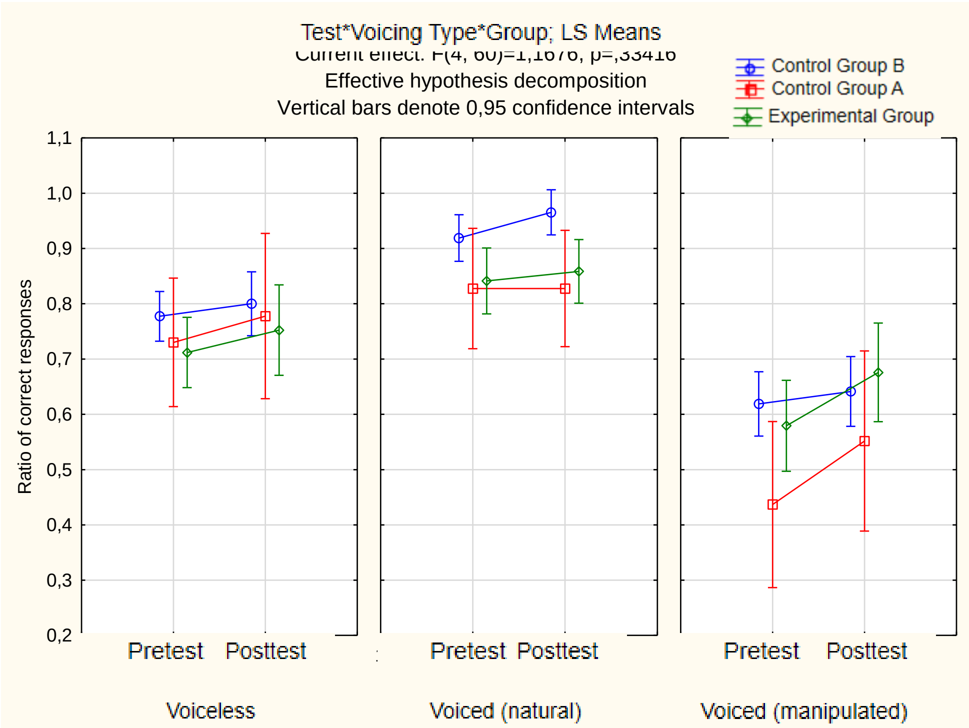


Figure 7 Percentages Correct: The Interaction Between Tests, Obstruent Voicing and Group

3.2. Reaction time

Regarding the response time need for participant's answer (either correct or incorrect) an ANOVA for a dependent variable of time and a within-subject variable of group was plotted. Upon examination of Figure 8, we notice a considerable difference in reaction times between all three groups. Lowest mean response times were achieved by control group B with phonologically aware speakers – 1,9 seconds. Next was a control group A with naive speakers with mean reaction time of 2,4 seconds. Experimental group was the slowest in reacting to stimuli, reaching approximately 2,7 seconds. The main effect was found significant (F (2, 30)=9,7442, p=0,00055).

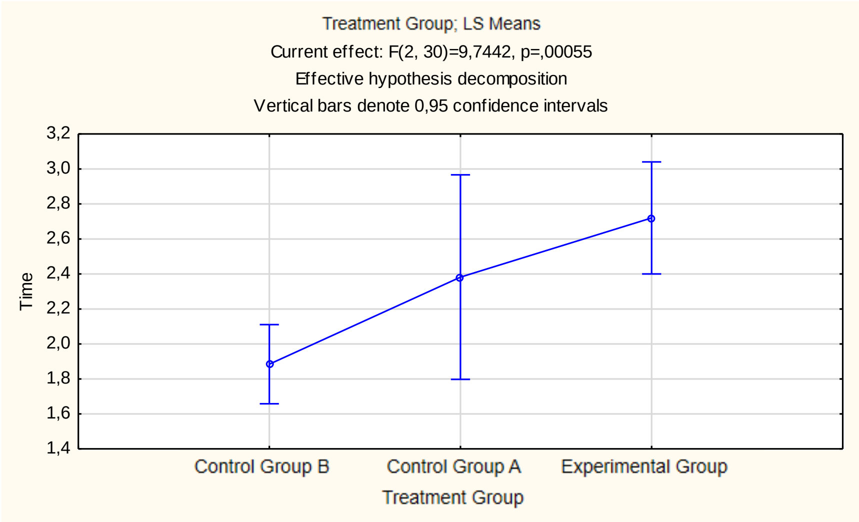


Figure 8 Reaction Times: The Main Effect of Treatment Group

ANOVA with a within-subject variable of obstruent voicing type (as seen in Figure 9) shows that all three groups were quickest to identify naturally voiced obstruent while identifying manipulated voiced obstruent took 0,36 seconds longer on average which is a 15% increase in reaction time. The average increase in response reaction time for voiceless word-final obstruents is 0,36 seconds when compared to response time of naturally voiced obstruents which was approximately 2,1 seconds. This also means an average 15% increase in response time. The main effect of obstruent voicing type is significant (F (2, 60)=12,634, p=0,00003). The findings are in agreement with the above mentioned results for a dependent variable of proportion correct in Figure 5 and further support it.

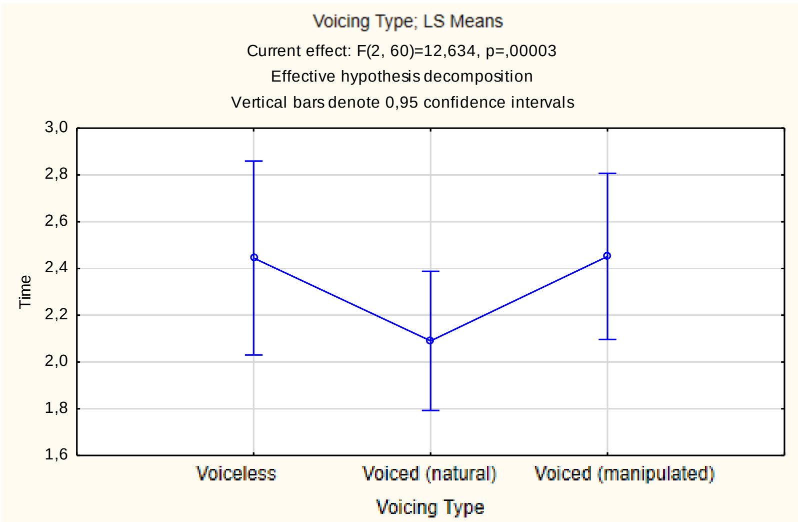


Figure 9 Reaction Times: The Main Effect of Voicing

Figure 10 shows detailed graphs for reaction times for test divided by group. There is apparent reduction of reaction time of both control groups and a visible increase of the reaction time for experimental group by average 0,39 seconds. Control group B of phonologically aware speakers achieved average decrease of 0,67 seconds. The main effect was significant (F (2, 30)=13,178, p=0,00008).

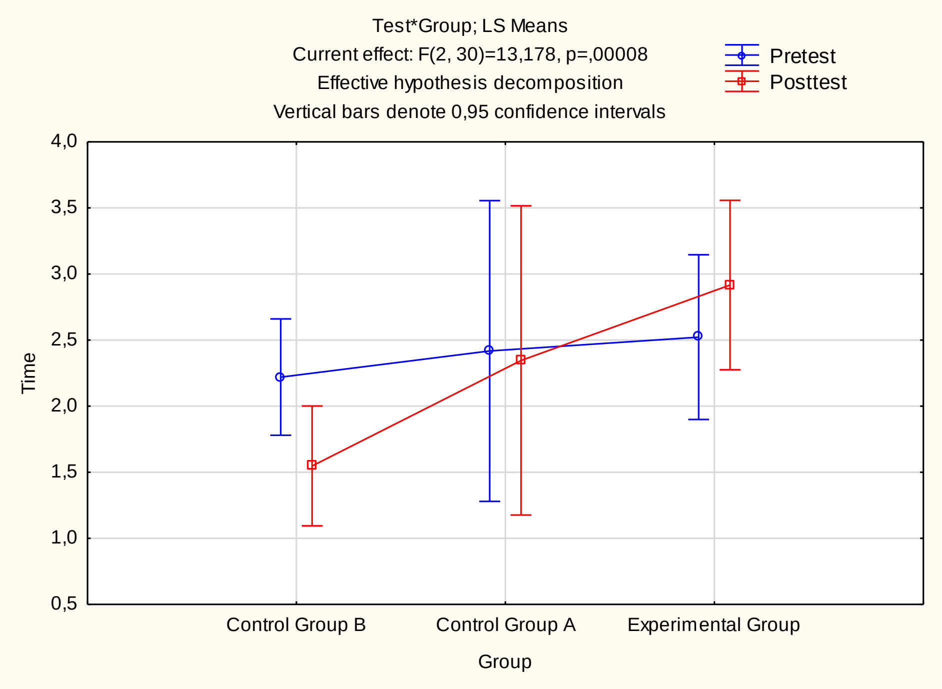


Figure 10 Reaction Times: The Interaction Between Tests and Group

3.3. Perception of coda voicing in nonsense-word continua

A two dimensional non-linear logistic regression combining independent variables of vowel duration and consonant duration is not included. Upon modelling, we have found out that duration of word-final obstruent had not affected likelihood of [yig] response significantly and has therefore been omitted.

For the [yig] continuum which has been created and described in 2.1. Material, a non-linear logistic regressions with dependent variable of response (either [yig] or [yik]) and independent within-subject variable of vowel duration were calculated.

Examining Figure 11, it is evident that the likelihood of a response being [yig] which is represented as the upper line labeled as 1,0 increases proportionally to the increased length of the vowel preceding word-final consonant. The graph shows that respondents were likely to choose [yik] (represented by the lower line labeled as 0,0) as the curve of likelihood did not exceed the threshold of 50%. However, the main effect of vowel duration was not found significant (Chi2(1)=0,27524, p=0,59984).

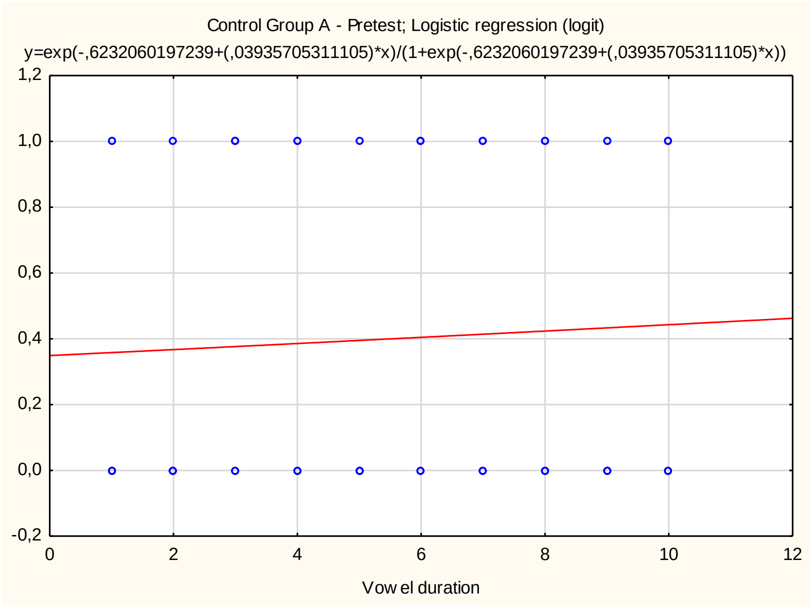


Figure 11 Control Group A – Pretest Logistic Regression: The Main Effect of Vowel Duration

Figure 12 shows same model for the same group but this time for the posttest. There is a noticeable difference in the responses which shows stronger tendency for the answer to be [yig] based on increasing vowel duration. The main effect of vowel duration is significant (Chi2(1)=4,1643, p=,04129).

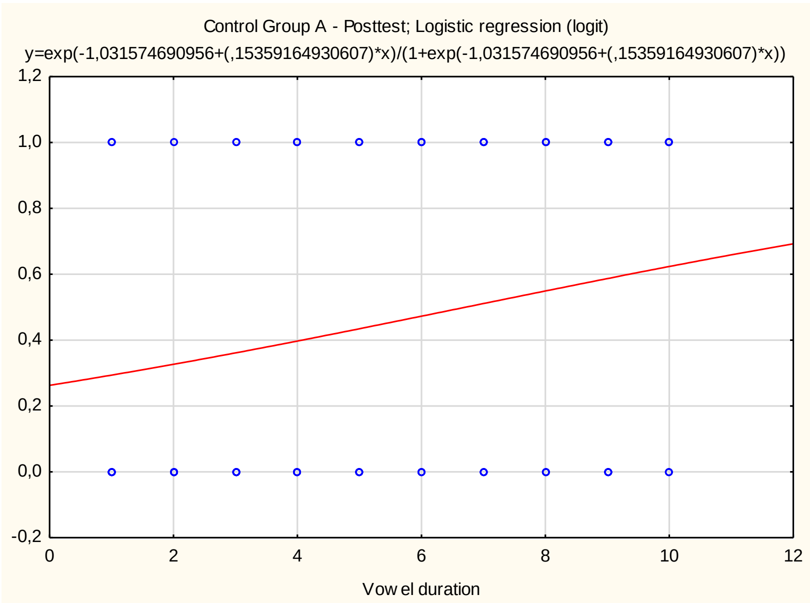


Figure 12 Control Group A – Posttest Logistic Regression: The Main Effect of Vowel Duration

Similar to the above, a non-linear logistic regressions for control group B of phonologically aware speakers to document the progress from pretest to posttest were calculated. The curve in Figure 13 resembles the the curve in Figure 12 in its steepness which means the responses of control group B in its pretest phase were similar to the responses of control group A in its posttest phase. The main effect of vowel duration is in this case significant (Chi2(1)=25,226 p=,00000).

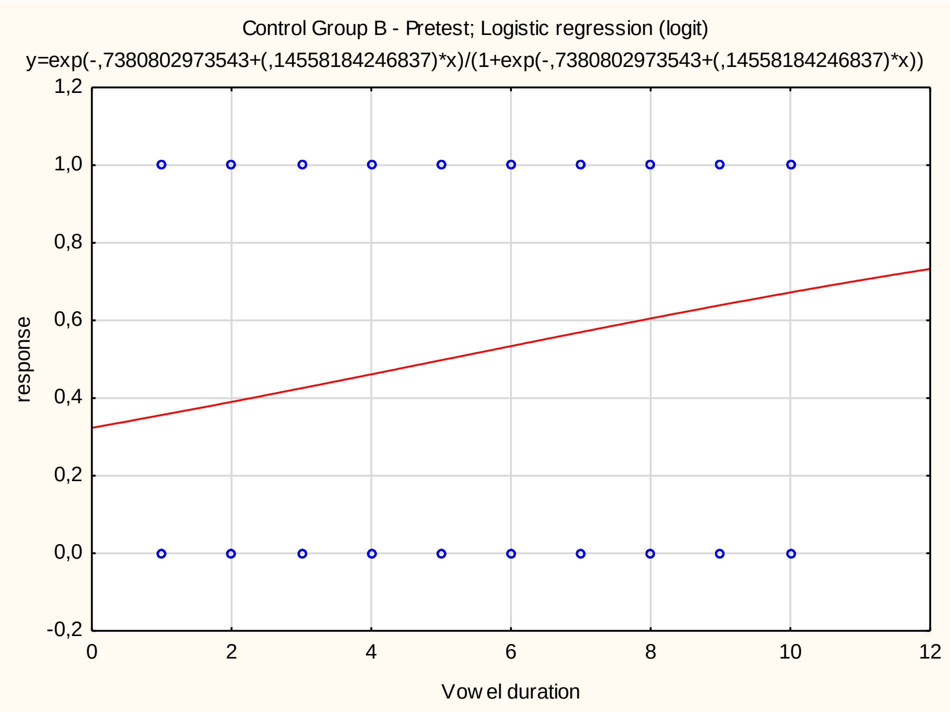


Figure 13 Control Group B – Pretest Logistic Regression: The Main Effect of Vowel Duration

The graph for posttest phase of control group B (Figure 14) is almost identical to the graph for pretest phase of control group B (Figure 13) except a slight difference during shorter vowel durations. The main effect of vowel duration iss significant (Chi2(1)=25,226 p=,00000).

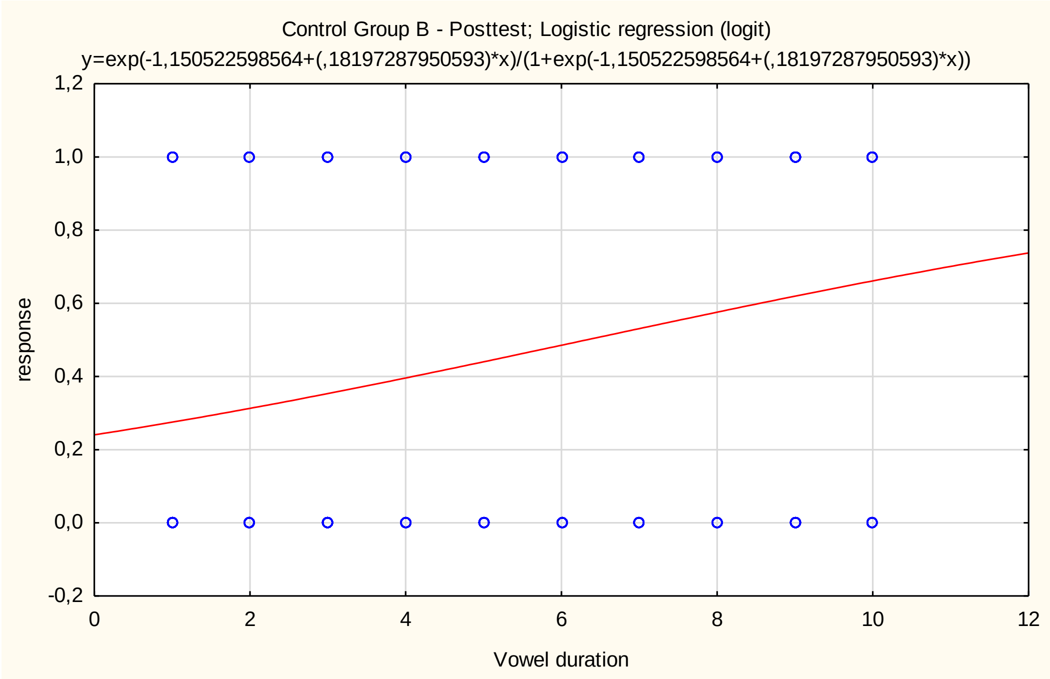


Figure 14 Control Group B – Pretest Logistic Regression: The Main Effect of Vowel Duration

The same logistic regression models were calculated for experimental group. Pretest phase (Figure 15) indicates a higher likelihood of the response being [yig] during shorter vowel duration than in other groups' test phases. Main effect is significant (Chi2(1)=25,226 p=,00000).

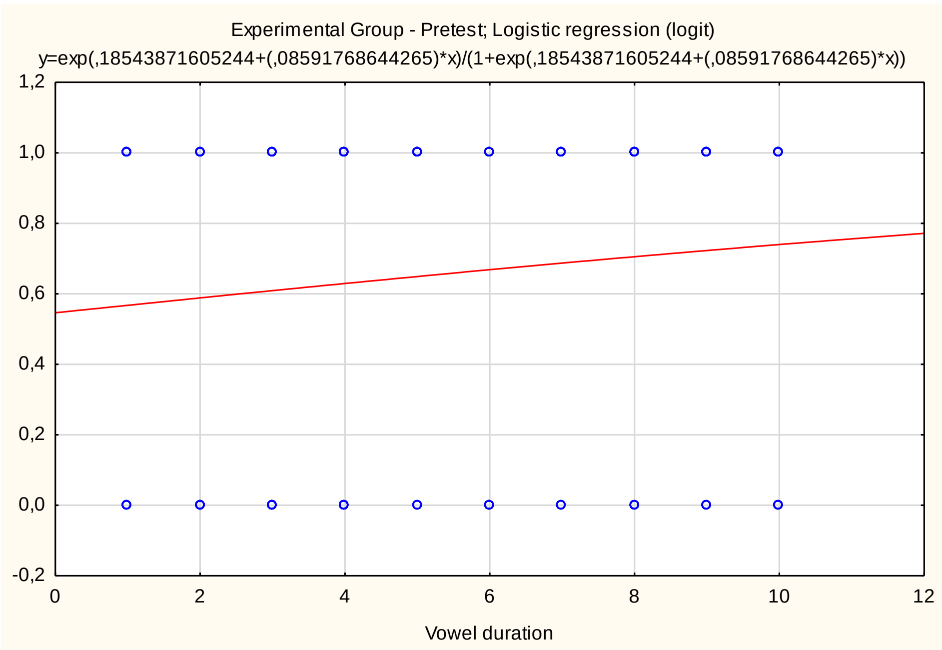
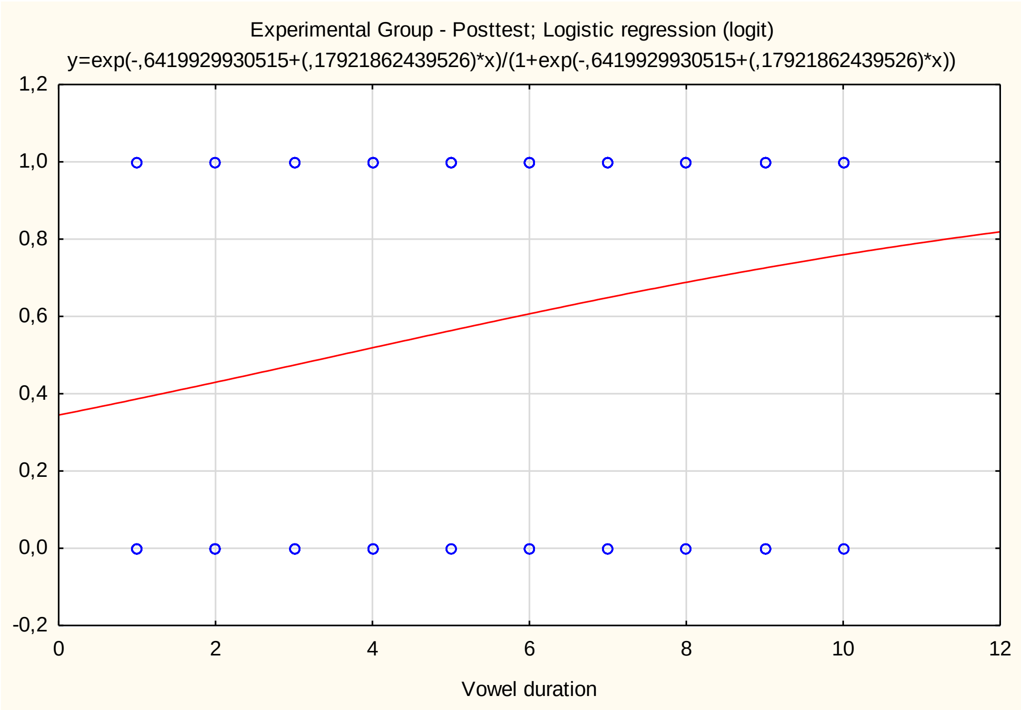


Figure 15 Experimental Group – Pretest Logistic Regression: The Main effect of Vowel Duration

Last graph shows how the likelihood of responses developed in the posttest (Figure 16) for the experimental group. There was a significant decrease in [yig] responses for the first half of vowel duration increase. This likelihood remained relatively unchanged for the second half of increased vowel durations. The main effect of vowel duration was also found significant (Chi2(1)=38,448 p=,00000)

Figure 16 Experimental Group – Pretest Logistic Regression: The Main effect of Vowel Duration



4. Discussion

In the present chapter, results that were obtained in the experiment will be commented upon in here. They will be interpreted with regard to the hypotheses that were formulated at the end of 1. Introduction. Additionally, possible influences on the results will be acknowledged. Recognizing the unwanted factors and flaws in my experiment could prove helpful in further research in which the methodology can be adjusted to avoid these flaws and yield more reliable results.

The first hypothesis (H_1) states that listeners that underwent perceptual training would improve their ratio of correct responses in posttest when compared to pretest. Even though there has been a significant main effect of test found, we cannot draw any reliable conclusion to either confirm or deny the hypothesis. Figure 3 hints at improvement in responses in posttest for all groups. All groups show a possible slight improvement. But even then, the degree of improvement seems to be quite similar across the groups and experimental group does not seem to show a steeper improvement as would have been expected due to the hypothesized impact of imitation training to which the experimental grop was exposed.

I hypothesized in H_2 that the reaction times needed to decide on coda voicing would be shorter in posttest. As can be seen in Figure 10, this expectation was fulfilled only for control groups. Experimental group's reaction times in fact increased.

This suggests that perceptual training is likely the root cause. Seeing that this increase was present for experimental group, there is a chance that one of the factors could be the newly acquired phonological knowledge which respondents had to consciously apply in evaluation of coda voicing. Overriding their default response directed by the cues which they naturally use and trying to apply the new cues based on phonological rules which were taught to them could have resulted in the increase of response time. On the other hand, this could also be a positive sign if we assume that attention to aspects of L2 sound patterns is a prerequisite for their successful acquisition.

Another aspect which should be considered is the psychological occurence of respondent's feeling of duty to perform better. The perceptual training of experimental group and its progression was fully traceable (for obvious methodological reasons) and participants were aware of this fact. It could have emphasized the feeling of responsibility and caused the participant to overcommit and carefully asses their responses during the posttest to show the researcher what they perceived as the desirable result.

Another hypothesis (H_3) was concerned with the vowel duration of the nonsense stimuli. I hypothesized that listeners would take vowel duration into account more when deciding on coda voicing in nonsense words as these words are unknown to them which avoids possible bias. The two-dimensional non-linear logistic regression (Figure 8 in 3.3. Logistic Regression) shows that listeners were likely to judge the non-sense stimulus as word-finally voiced depending on how long was the preceding vowel. There was an almost direct proportion between vowel duration and the /yig/ responses from respondents which provides support to the hypothesis that preceding vowel duration is used as a cue to coda voicing.

As was also hypothesized in the last hypothesis H_4, the listeners with previous phonological knowledge outperformed the other two groups in correct responses overall as well as having shown significantly lower response times in both tests. Even though the main effect of interaction between test and group was not significant and no reliable conclusion can be drawn in this regard, there was an inclination to higher ratio of correct responses in both pretest and posttest when compared to other groups in each respective test.

One of the influences that possibly distorted the results obtained from respondents could have been the disregard of stimuli which were not known by the respondent. For example, after hearing the word "cud", respondent was presented with two choices – "cut" and "cud." Since respondent does not have the word "cud" in his mental vocabulary, he is likely to choose a word which he is more familiar with, in this case the word "cut."

Notable lack of sample in control group of naive speakers – only three persons – failed to provide reliable results. Thus, increasing the sample of this control group would generalize the findings as well as reduce any discrepancies formed by outliers in the sample.

As was explained in 2.2 Stimuli, the set of stimuli used in posttest was identical to the set of stimuli used in pretest. This could have affected the results by introducing an unwanted factor of memory. The period of one week between pretest and posttest could have enabled the participants to memorize the stimuli, discarding the novelty of the sounds. By memorizing certain stimuli in the pretest, the respondents would be able to contrast the memorized stimulus to its minimal-pair counterpart before hearing the actual representation of the latter in the posttest. All groups could have been affected and inclined to higher ratio of correct responses (as can be seen in Figure 2 in 3.1 Proportion Correct).

Another threat to this study's validity which is also closely related to the feeling of duty to perform better that I mentioned earlier in this chapter could be Hawthorne's effect when participants perform differently due to their immediate knowledge of being under supervision and being subjects of a study. Mellow, branded it as "the single most serious threat to studies of spontaneous language use." (Mellow et al. 1996, 332). Only this effect would not be limited to experimental group only but could have a significant effect on all participants.

From methodology standpoint, purely quantitative research was used to gain and interpret data. As Dörneyi (2005) points out, while quantitative research is systematic, involves precise measurements and produces reliable and replicable data, it has also severe disadvantages – it does not provide researcher with a reason for particular observation which results in limited exploratory capacity. (Dörneyi 2005, 32-35)

While quantitative research is particularly effective in natural sciences, it cannot cover social sciences to the same extent. Fundamental differences between the two sciences exist – researched subject, humans, show variation over time and display within-individual variation. (Dörneyi 2005, 44)

This has lead researchers to invest in mixed methods research. Mentioned method utilizes both QUAN and QUAL approaches in order to neutralize weaknesses of these methods and improve on the data analysis. (Dörneyi 2005, 44-45) Therefore, further research which would incorporate mixed methods could uncover more regarding this phenomenon and even lead to new findings.

5. Conclusion

Word-final obstruent voicing is governed by various factors in human speech and the weigh of these factors differs across languages (see 1.3.). Therefore, it is widely researched topic and my thesis contributes to this research by investigating the details of perception of this obstruent voicing and also investigating ability of native Czech learners of English to improve their perception of word-final voicing by explicit learning.

Unlike English, Czech neutralizes word-final obstruent voicing which is typical for majority of Slavic languages (see 1.3.3.). English also uses duration of preceding vowel as a significant cue to coda voicing (see 1.3.)

To explore the possibility of training native Czech speaker to utilize the vowel duration as a cue to voicing, an experiment which included one training group and two control groups (naive learners, phonologically aware). All groups participated in two perceptual tests with training group receiving an additional learning session in between the tests (see 2.3.).

Results confirmed two hypotheses. The control group with previous phonological knowledge was able to significantly outperform other two groups in regard to overall correct responses and their lower reaction times. Results of the experiment conducted for perception of coda voicing in nonsense continua (yik-yig) provide substantial support to the claim that native Czech speakers are able to use preceding vowel duration as a cue to word-final obstruent voicing (see 3.).

However, the improvement in perception of final-obstruent voicing for the experimental group did not confirm the hypothesis that experimental group would achieve a higher increase in correct response. The degree of increase which is reported is similar for all groups involved, thus not showing a visible impact of the imitation training itself.

Topic of perception of coda voicing in English by native Czech speakers should be subject to further research as my thesis used a single methodological approach and as mentioned in 4. Discussion, refining the methods or using another approach could yield different results.

Resumé

Cílem této práce je rozšířit poznatky v problematice znělosti koncových souhlásek, respektive zkoumat odlišnosti jejich fonologických pravidel mezi českým a anglickým jazykem a dále pak upřesnit jak čeští rodilí mluvčí vnímají realizaci této znělosti v anglickém jazyce. Výsledky experimentů lze také uplatnit v praktické rovině, a to zejména posouzení efektivnosti použitých metod a testů a jejich další využití.

V úvodní části je popsána obecná teorie a pravidla znělosti koncových souhlásek jak pro český, tak i anglický jazyk. Dochází rovněž ke srovnání obou jazyků a k poukázání na jejich rozdíly týkajících se této problematiky. Uvedeny jsou také předešlé výzkumy v této oblasti, jejich výsledky a použité metody.

V dalších kapitolách jsou pak detailně zdokumentovány metody i materiál použit v experimentu společně s průběhem jednotlivých cvičení. Výsledky jsou znázorněny graficky a popsány.

V koncové části jsou dosažené výsledky interpretovány a posouzeny s hypotézami, které byly stanoveny v úvodní kapitole. Zmíněny jsou zde i faktory, které mohly mít možný vliv na průběh experimentu a měření.

Ačkoliv byl experiment nácviku percepce znělosti koncových souhlásek podle anglických fonologických pravidel v určitých ohledech neúspěšný, ukázalo se, že čeští rodilí mluvčí, kteří byli vystaveni znalosti fonologických pravidel angličtiny jsou schopni dosáhnout lepších výsledků. Výsledky poukazují také na fakt, že čeští rodilí mluvčí jsou do určité míry schopni určit znělost koncové souhlásky na základě délky předcházející samohlásky. Toto pravidlo je pro český jazyk neobvyklé, pro anglický jazyk však esenciální.

References

Alanen, R. (1995). Input enhancement and rule presentation in second language acquisition. *Attention and awareness in foreign language learning*, 259-302.¨

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

Baart, J. L. (2011). *A field manual of acoustic phonetics*. Dallas, TX: SIL International.

Baleghizadeh, S., & Derakhshesh, A. (2017). Measuring the Effectiveness of Explicit and Implicit Instruction through Explicit and Implicit Measures. *Iranian Journal of Applied Linguistics*, 20(1), 81-111.

Barriuso, T. A., & Hayes-Harb, R. (2018). High Variability Phonetic Training as a Bridge from Research to Practice. *CATESOL Journal*, 30(1), 177-194.

Berendsen, E. (1983). *Final devoicing, assimilation, and subject clitics in Dutch* (pp. 21-29). Instituut De Vooys.

Berg, T. (1994). The sensitivity of phonological rimes to phonetic length. *AAA: Arbeiten aus Anglistik und Amerikanistik*, 63-81.

Bičan, A. (2012). *Phonotactics of Czech* (Doctoral dissertation, Masarykova univerzita, Filozofická fakulta).

Bradlow, A. R., Pisoni, D. B., Akahane-Yamada, R., & Tohkura, Y. I. (1997). Training Japanese listeners to identify English/r/and/l: IV. Some effects of perceptual learning on speech production. *The Journal of the Acoustical Society of America*, *101*(4), 2299-2310.

Broersma, M. (2005). Perception of familiar contrasts in unfamiliar positions. *The Journal of the Acoustical Society of America*, *117*(6), 3890-3901.

Broersma, M. (2008). Flexible cue use in nonnative phonetic categorization. *The Journal of the Acoustical Society of America*, *124*(2), 712-715.

DeKeyser, R. (2008). 11 Implicit and Explicit Learning. *The handbook of second language acquisition*, 27, 313.

Denes, P. (1955). Effect of duration on the perception of voicing. *The Journal of the Acoustical Society of America*, *27*(4), 761-764.

Dornyei, Z. (2005). Research methods in applied linguistics.

Ellis, N.C. (2002). Frequency effects in language processing: A review with implications for theories of implicit and explicit language acquisition. *Studies in Second Language Acquisition*, 24(2), 143–188.

Embarki, M. (2016). Voicing Effects an Absolute Universal or Language Specific: New Evidence from Modern Arabic.

Charles-Luce, J. (1985). Word-final devoicing in German and the effects of phonetic and sentential contexts. *The Journal of the Acoustical Society of America*, *77*(S1), S85-S85.

Chen, M. (1970). Vowel length variation as a function of the voicing of the consonant environment. *Phonetica*, 22(3), 129-159.

Kessler, B., & Treiman, R. (1997). Syllable structure and the distribution of phonemes in English syllables. *Journal of Memory and language*, 37(3), 295-311.

Kulikov, V., & McMurray, B. (2017). Incomplete neutralization in voice assimilation and final devoicing in Russian: Toward a model of gradient “cumulative” voicing. *Unpublished ms*.

Ladefoged, P., & Johnson, K. Jan. 2014. *A Course in Phonetics, 7th Edition.* Cengage Learning.

Machač, P., & Skarnitzl, R. (2007). Temporal compensation in Czech. *Proceedings of 16 th ICPhS,* 537-540.

Mellow, J., Reeder, K., & Forster, E. (1996). Using Time-Series Research Designs to Investigate the Effects of Instruction on SLA. *Studies in Second Language Acquisition*, 18(3), 325-350.

Mitleb, F. (1982). Voicing effect on vowel duration is not an absolute universal. *The Journal of the Acoustical Society of America*, *71*(S1), S23-S23.

Napoli, D. J. (1996). *Linguistics: an introduction*. New York: Oxford University Press.

Podlipsky, V. J. (2008). Vowel duration as a cue for coda voicing and the perception of second-language vowel quantity. *Proceedings of Acoustics’08 Paris*, 6155-6159.

Podlipský, V. J., & Chládková, K. (2007). Vowel duration variation induced by coda consonant voicing and perceptual short/long vowel categorization in Czech. In *Speech Processing: 17th Czech–German Workshop. Prague: AS CR* (pp. 68-74).

Rebuschat, P. (Ed.). (2015). *Implicit and explicit learning of languages* (Vol. 48). John Benjamins Publishing Company.

Reichl, P., Egger, S., Schatz, R., & D'Alconzo, A. (2010, May). The logarithmic nature of QoE and the role of the Weber-Fechner law in QoE assessment. In *2010 IEEE International Conference on Communications* (pp. 1-5). IEEE.

Scheer, T. (2017). Voice-induced vowel lengthening. *Papers in Historical Phonology*, *2*, 116-151.

Stevens, K. N. (2000). Diverse acoustic cues at consonantal landmarks. *Phonetica, 57*(2-4), 139-151.

Simonchyk, Ala. (2017). Acquisition of the feature [+voice] in word-final position by native speakers of Russian. *IULC Working Papers*, 17. Indiana University Linguistics Club Working Papers. 17.

Slowiaczek, L. M., & Dinnsen, D. A. (1985). On the neutralizing status of Polish word-final devoicing. *Journal of phonetics*.

Slowiaczek, L. M., & Szymanska, H. J. (1989).Perception of word-final devoicing in Polish. *Journal of Phonetics*, 13, 325-341.

Smith, B. L., Hayes-Harb, R., Bruss, M., & Harker, A. (2009). Production and perception of voicing and devoicing in similar German and English word pairs by native speakers of German. *Journal of Phonetics, 37*(3), 257-275.

Waniek-Klimczak, E. (2005). *Temporal parameters in second language speech: An applied linguistic phonetics approach*. Wydawnictwo Uniwersytetu Łódzkiego.

Wong, J. (2012). Training the Perception and Production of English/ε/and/Θ/of Cantonese ESL Learners: A Comparison of Low vs. High Variability Phonetic Training. In *Proceedings of the 14th Australasian International Conference on Speech Science and Techniology*. Sydney, Australia (pp. 3-6).

Wong, J. W. (2014). The effects of high and low variability phonetic training on the perception and production of English vowels/e/-/æ/by Cantonese ESL learners with high and low L2 proficiency levels.

Zonneveld, W. (1983). Lexical and phonological properties of Dutch voicing assimilation. *Sound structures: Studies for Antonie Cohen, ed. by Marcel van den Broecke, Vincent van Heuven, and Wim Zonneveld*, 101-35.

List of Figures

Appendix 1: List of stimuli used for the pretest and the posttest

Voiceless	Voiced (natural)
buck	bug
cap	cab
clock	clog
cop	cob
cup	cub
dock	dog
duck	dug
duff	dove
fat	fad
-	cud

Voiceless	Voiced closure, voiceless release (natural)
back	bag
bat	bad
bop	bob
coat	code
cot	cod
dose	doze
-	bud
-	cog
-	fade
-	greed

Voiceless	Devoiced (natural)
base	baize
bet	bed
bit	bid
calf	calve
feet	feed
fuss	fuzz
grace	graze
grate	grade
-	bead

Voiceless	Voiced coda swap to voiceless (manipulated)
hack	hag
hit	hid
chuck	chug
kit	kid
mat	mad
mate	maid
muck	mug

nip	nib
pat	pad
pick	pig
pot	pod
pup	pub
sack	sag
seat	seed
site	side
sop	sob
strife	strive
tap	tab
tuck	tug
-	hob
-	paid
-	peg
-	phase
-	ridge
-	save
-	shag
-	stag
-	swede
-	tide
	Scaling from voiced to voiceless (artificial)
	yig

Appendix 2: List of stimuli used for the imitation training

Voiceless	Voiced
/f/	*/v/*
life	live
proof	prove
/s/	*/z/*
race	raise
rice	rise
/t/	*/d/*
site	side
wait	wade
/p/	*/b/*
nap	nab
rope	robe
/k/	*/g/*
leek	league

/ʧ/	/ʤ/
rich	ridge
batch	badge
/θ/	/ð/
teeth	teethe

Appendix 3: PRAAT script of the pretest and the posttest

All PRAAT scripts below were written and provided by Mgr. Václav Jonáš Podlipský, PhD.

```
initialPauseDur = 0.35
pauseAfterEvry = 50
;minDistanceFromFirst = 6

pairs = Read from file: "stimuli/pairs.txt"
neutralizedVoicingList = Read from file: "stimuli/neutralized-voicing-list.txt"
allStimList = Create Strings as file list: "stimList", "stimuli\*.wav"
Randomize
nAllStim = Get number of strings
playListTab = Create Table with column names: "playListTab", nAllStim, "subject task
file neutralized-coda-swap underlying-voicing leftButton rightButton response rw rt"
for j to nAllStim
	select allStimList
	stim$ = Get string: j
	select playListTab
	Set string value: j, "file", stim$
endfor

trainList = Create Strings as file list: "stimList", "stimuli\train\*.wav"
Randomize
nTrain = Get number of strings

beginPause: "A short questionnaire"
	comment: "Please fill out this short questionnaire."
	word: "Initials", "AB"
	word: "Age", "22"
	choice: "Sex", 1
		option: "female"
```

```
            option: "male"
        choice: "Test", 1
            option: "pre-test"
            option: "post-test"
            option: "post-test2"
        natural: "Participant number", "1"
clicked = endPause: "Continue", 1

subjCode$ = "'participant_number'_'initials$'_'age$'_" + left$ (sex$, 1)

select playListTab
Save as tab-separated file: "results\'subjCode$'_'test$'.txt"

label INTRO0
demoWindowTitle: "Listening task"
demo Black
demo Times
demo Font size: 24
demo Select inner viewport: 0, 100, 0, 100
demo Axes: 0, 100, 0, 100
demo Paint rectangle: "silver", 0, 100, 0, 100
demo Text: 50, "centre", 55, "half", "##Maximize this window now.#"
demo Text: 50, "centre", 45, "half", "Click to show instructions."
while demoWaitForInput ( )
        goto INTRO1 demoInput ("•→")
endwhile

label INTRO1
demo Erase all
demo Select inner viewport: 0, 100, 0, 100
demo Paint rectangle: "silver", 0, 100, 0, 100
demo Text: 50, "centre", 85, "half", "You will hear words recorded by different people."
demo Text: 50, "centre", 75, "half", "(They are either existing English words or not, that's
not important)."
demo Text: 50, "centre", 65, "half", "Decide which of two words shown on the screen you
heard"
demo Text: 50, "centre", 55, "half", "and click on the corresponding button."
```

```
demo Text: 50, "centre", 35, "half", "Click to hear four examples."
while demoWaitForInput ( )
	goto TRAINING demoInput ("•→")
	goto INTRO0 demoInput ("←")
endwhile

label TRAINING
for i to nTrain
	select trainList
	file$ = Get string: i
	soundRaw = Read from file: "stimuli\train\'file$'"
	sampleRate = Get sampling frequency
	soundMono = Convert to mono
	pause = Create Sound from formula: "pause", 1, 0, initialPauseDur, sampleRate,
"0"
	select soundMono
	soundCopy = Copy: "soundCopy"
	select pause
	plus soundCopy
	sound = Concatenate
	soundDur = Get total duration
	@screen
endfor

label INTRO2
demo Erase all
demo Paint rectangle: "silver", 0, 100, 0, 100
demo Black
demo Font size: 24
demo Select inner viewport: 0, 100, 0, 100
demo Text: 50, "centre", 65, "half", "Is the volume fine? (To play the examples again
press the spacebar.)"
demo Text: 50, "centre", 45, "half", "The task shouldn't last more than 10 minutes."
demo Text: 50, "centre", 35, "half", "You'll be given the chance to take a break."
demo Text: 50, "centre", 25, "half", "Click to run the task."
while demoWaitForInput ( )
	goto THETEST demoInput ("•→")
```

```
		goto INTRO1 demoInput ("←")
		goto TRAINING demoInput (" ")
endwhile

label THETEST
for i to nAllStim
		select playListTab
		file$ = Get value: i, "file"
		shortName$ = file$ - ".wav"
		soundRaw = Read from file: "stimuli\'file$'"
		sampleRate = Get sampling frequency
		soundMono = Convert to mono
		pause = Create Sound from formula: "pause", 1, 0, initialPauseDur, sampleRate,
"0"
		select soundMono
		soundCopy = Copy: "soundCopy"
		select pause
		plus soundCopy
		sound = Concatenate
		soundDur = Get total duration
		@screen
		if left$ (file$, 4) <> "stim"
			select neutralizedVoicingList
			findRow = Search column: "file", shortName$
			swap$ = Get value: findRow, "neutralized-coda-swap"
			voicing$ = Get value: findRow, "underlying-voicing"
		endif
		select playListTab
		Set string value: i, "subject", subjCode$
		if left$ (file$, 4) <> "stim"
			Set string value: i, "neutralized-coda-swap", swap$
			Set string value: i, "underlying-voicing", voicing$
		endif
		Set string value: i, "task", test$
		Set string value: i, "response", response$
		Set string value: i, "rw", rw$
		Set string value: i, "leftButton", opt1$
```

```
		Set string value: i, "rightButton", opt2$
		Set numeric value: i, "rt", rt
		Save as tab-separated file: "results\'subjCode$'_'test$'.txt"
		if i mod pauseAfterEvry = 0 and i < nAllStim
			@pause
		endif
endfor

label END1
demo Erase all
demo Black
demo Font size: 24
demo Select inner viewport: 0, 100, 0, 100
demo Paint rectangle: "silver", 0, 100, 0, 100
demo Text: 50, "centre", 60, "half", "You've finished. Your answers have been recorded."
demo Text: 50, "centre", 40, "half", "Thanks very much! You can close this window
now."
select all
Remove

procedure screen
	demo Erase all
	demo Font size: 34
	demo Select inner viewport: 0, 100, 0, 100
	demo Paint rectangle: "silver", 0, 100, 0, 100
	demo Black
	demo Text: 50, "centre", 80, "half", "Which word did you hear?"
	demo Grey
	demo Line width: 3
	demo Draw rounded rectangle: 32, 46, 42, 58, 3
	demo Draw rounded rectangle: 54, 68, 42, 58, 3
	corOpt = randomInteger(1,2)
	if corOpt = 1
		other = 2
	elif corOpt = 2
		other = 1
	endif
```

```
if left$ (file$, 4) = "stim"
	opt'corOpt'$ = "yik"
	opt'other'$ = "yig"
else
	opt'corOpt'$ = file$ - ".wav"
	select pairs
	pairRow = Search column: "vl", opt'corOpt'$
	if pairRow = 0
		pairRow = Search column: "vd", opt'corOpt'$
		opt'other'$ = Get value: pairRow, "vl"
	else
		opt'other'$ = Get value: pairRow, "vd"
	endif
endif
	select sound
	stopwatch
	Play
	demo Paint rounded rectangle: "cyan", 32, 46, 42, 58, 3
	demo Paint rounded rectangle: "cyan", 54, 68, 42, 58, 3
	demo Black
	demo Text: 39, "centre", 50, "half", opt1$
	demo Text: 61, "centre", 50, "half", opt2$
	while demoWaitForInput ( )
		if demoClickedIn (32, 46, 42, 58)
			rawRt = stopwatch
			rt = rawRt - soundDur
			response$ = opt1$
			if response$ = opt'corOpt'$
				rw$ = "r"
			else
				rw$ = "w"
			endif
			goto NEXTTRIAL
		elsif demoClickedIn (54, 68, 42, 58)
			rawRt = stopwatch
			rt = rawRt - soundDur
			response$ = opt2$
```

```
                if response$ = opt'corOpt'$
                    rw$ = "r"
                else
                    rw$ = "w"
                endif
                goto NEXTTRIAL
            endif
        endwhile
        label NEXTTRIAL
endproc

procedure pause
        demo Erase all
        demo Black
        demo Font size: 24
        demo Select inner viewport: 0, 100, 0, 100
        demo Paint rectangle: "silver", 0, 100, 0, 100
        demo Text: 50, "centre", 60, "half", "Now you can take a short break if you like."
        demo Text: 50, "centre", 40, "half", "Click to resume the task."
        while demoWaitForInput ( )
            goto RESUMED demoInput ("•")
        endwhile
        label RESUMED
endproc
```

Appendix 4: PRAAT script of the imitation training

```
Imiation trainer. Version 2.5.
#    This is a script for the program Praat, <http://www.praat.org/>.
#
#    If you use this script for your research, please cite us.
#
#    Copyright (C) 2016 Václav Jonáš Podlipský
#
#    vaclav.j.podlipsky@upol.cz
#
#    This program is free software: you can redistribute it and/or modify
#    it under the terms of the GNU General Public License as published by
```

```
#	the Free Software Foundation, either version 3 of the License, or
#	(at your option) any later version.
#
#	This program is distributed in the hope that it will be useful,
#	but WITHOUT ANY WARRANTY; without even the implied warranty of
#	MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE.  See the
#	GNU General Public License for more details.
#
#	A copy of the GNU General Public License can be found at
#	<http://www.gnu.org/licenses/>.
#
######
# SET UP
######
# Set up the microphone gain (0 - 1).
gain = 0.7
######
# Put the model files (.wav and .TextGrid) in the folder 'sounds'. They should contain
about 0.3s of silence
# before and after the actual recording.

######

# Play model files in random order? 1 is yes, 0 is no.

randOrder = 0

######

# Show original annotation over your recording? 1 = yes, 0 = no.

showYourGrid = 1

######

# Set up recording duration: if adaptRecDur = 1, then the recording duration will be
# the same as the duration of the model file + extraRecDur. if adaptRecDur = 0, a fixed
```

```
# recording time will be used, which you define as fixedRecordDur (in seconds).

adaptRecDur = 1
extraRecDur = 0
fixedRecordDur = 1.5

######

# Set up whether a log file gets created: if makeLog = 1, a table with user activity
# will be saved automatically. If you don't need this, set makeLog to 0.

makeLog = 0

# Layout set-up

fromx = 5
tox = 100 - fromx
gap = 1.5
leftgap = fromx
if showYourGrid = 1
	waveHeight = 11
	specHeight = 22
	gridHeight = 8.2
	y1 = 100 - gap
	y2 = y1 - waveHeight
	y3 = y2 - specHeight
	y4 = y3 - gridHeight
	y5 = y4 - gap
	y6 = y5 - waveHeight
	y7 = y6 - specHeight
	ybot = y7 - gridHeight
else
	waveHeight = 11.5
	specHeight = 23
	gridHeight = 8.7
	y1 = 100 - gap
	y2 = y1 - waveHeight
```

```
	y3 = y2 - specHeight
	y4 = y3 - gridHeight
	y5 = y4 - gap
	y6 = y5 - waveHeight
	y7 = y6 - specHeight
	ybot = y7
endif
ybot2 = ybot - gap/2

# Sounds longer than squeezeLimit (in s) get shrunk temporally to make them fit on the screen.
# For shorter sounds, left-to-right screen distance correspond to time in absolute terms.
squeezeLimit = 1.3

# The script attempts to automatically align the users recording to the model so that the first stressed vowel
# are shown to begin at the same time in the graphs. These values can be tweaked to improve the performance.
# The lower the noise threshold (in dB) the more immunity to background noise but less sensitivity to speech.
noiseThreshold = -10
noiseThresholdorig = -14
voiceThreshold = -8


######
# END OF SET UP
######

if makeLog = 1
	beginPause: "Identification"
		comment: "Please fill in your data."
		word: "First name", ""
		word: "Last name", ""
		choice: "Course", 1
			option: "AFO2"
			option: "AFO4"
```

```
                    option: "AF11"
                    option: "other"
               choice: "Your teacher", 1
                    option: "Sarka"
                    option: "Jonas"
     clicked = endPause: "Continue", 1
     subjCode$ = "'first_name$'_'last_name$'_'course$'_'your_teacher$'"
     logTab = Create Table with column names: "logTab", 0, "student word played
recorded yours"
endif

list = Create Strings as file list: "list", "sounds/*.wav"
if randOrder = 1
     Randomize
endif
nSounds = Get number of strings

demoWindowTitle: "Imitation training"
demo Black
demo Times
demo Font size: 20
demo Select inner viewport: 0, 100, 0, 100
demo Axes: 0, 100, 0, 100
demo Erase all
demo Paint rectangle: "silver", 0, 100, 0, 100
demo Text: 50, "centre", 70, "half", "You'll hear a word. Click on ""play again"" to hear it
once more."
demo Text: 50, "centre", 60, "half", "Then click ""record"" and immediately repeat the
word."
demo Text: 50, "centre", 50, "half", "Then click on ""show graphs""."
demo Text: 50, "centre", 30, "half", "Click to start."
while demoWaitForInput ( )
     goto NEXT demoInput ("•→")
endwhile

label NEXT
for i to nSounds
```

```
label LOAD_WORD
rec = 0
pl = 0
yrs = 0
enSpec = 0
fromx = leftgap
select list
file$ = Get string: i
gridOrig = Read from file: "sounds/" + file$ - ".wav" + ".TextGrid"
nInt = Get number of intervals: 1
for j to nInt
	intStart'j' = Get starting point: 1, j
	intEnd'j' = Get end point: 1, j
	plPiecesound'j' = 0
	plPiecetimeShifted'j' = 0
endfor
sound = Read from file: "sounds/" + file$
origDur = Get total duration
if origDur < squeezeLimit
	fromx = 50 - (origDur/2*(50-fromx))/(squeezeLimit/2)
	tox = 100 - fromx
endif
if adaptRecDur = 1
	recordDur = origDur + extraRecDur
elsif adaptRecDur = 0
	recordDur = fixedRecordDur
endif
if index (file$, "_") = 0
	word$ = file$ - ".wav"
else
	word$ = right$ (file$, length (file$) - index (file$, "_")) - ".wav"
endif
grid0 = To TextGrid (silences): 90, 0, noiseThresholdorig, 0.01, 0.01, "silent",
"sounding"
soundStart0 = Get end point: 1, 1

select sound
```

```
	sounding = Extract part: soundStart0, origDur, "rectangular", 1, "yes"
	pitch = noprogress To Pitch: 0, 85, 300
	sine = To Sound (sine): 44100, "at nearest zero crossings"
	grid = To TextGrid (silences): 90, 0, voiceThreshold, 0.01, 0.01, "silent",
"sounding"
	label$ = Get label of interval: 1, 1
	if label$ = "silent"
		soundStart = Get end point: 1, 1
	else
		soundStart = soundStart0
	endif

	label SCR1
	@trialScr: word$
	select sound
	Play
	while demoWaitForInput ()
		if demoClickedIn (30, 70, 54, 70) or demoInput ("p")
			pl = pl + 1
			demo Insert picture from file: "sounds/play-icon.png", 26, 30, 59,
65
			select sound
			Play
			demo Paint rectangle: "silver", 26, 29.5, 59, 65
		elsif demoClickedIn (30, 49, 35, 45) or demoInput ("r")
			rec = rec + 1
			demo Paint rectangle: "silver", 0, 100, 72, 90
			demo Paint circle: "red", 21, 40, 0.9
			input'rec' = Record Sound (fixed time): "Microphone", gain, 0.5,
"44100", recordDur
			;Scale peak: 0.99
			demo Paint circle: "silver", 21, 40, 0.91
			demo Teal
			demo Text special: 50, "centre", 80, "half", "Times", 50, "0",
word$
			demo Black
		elsif demoInput (" g") or demoClickedIn (51, 70, 35, 45)
```

```
				if rec = 0
					demo Colour: "red"
					demo Text special: 50, "centre", 80, "half", "Times", 30,
"0", "Record yourself first."
					demo Colour: "black"
				else
					goto MOVE
				endif
			elsif demoInput ("→")
				goto FASTMOVE
			elsif demoInput ("←") and i > 1
				i -= 1
				goto LOAD_WORD
			endif
		endwhile
		label MOVE
		select input'rec'
		gridCz0 = nowarn To TextGrid (silences): 90, 0, noiseThreshold, 0.01, 0.01,
"silent", "sounding"
		label$ = Get label of interval: 1, 1
		if label$ = "silent"
			soundStartCz0 = Get end point: 1, 1
		else
			soundStartCz0 = 0
		endif
		select input'rec'
		soundingCz = Extract part: soundStartCz0, origDur, "rectangular", 1, "yes"
		pitchCz = noprogress To Pitch: 0, 85, 300
		sineCz = To Sound (sine): 44100, "at nearest zero crossings"
		gridCz = nowarn To TextGrid (silences): 90, 0, voiceThreshold, 0.01, 0.01,
"silent", "sounding"
		label$ = Get label of interval: 1, 1
		if label$ = "silent"
			soundStartCz = Get end point: 1, 1
		else
			soundStartCz = soundStartCz0
		endif
```

```
	if soundStartCz > soundStart
		select input'rec'
		part1 = Extract part: soundStartCz - soundStart, origDur, "rectangular", 1,
"no"
		part2 = Create Sound from formula: "silence", 1, 0, soundStartCz -
soundStart, 44100, "0"
		select part1
		plus part2
		timeShifted = Concatenate
	elsif soundStartCz < soundStart
		part1 = Create Sound from formula: "silence", 1, 0, soundStart -
soundStartCz, 44100, "0"
		select input'rec'
		part2 = Extract part: 0, origDur - (soundStart - soundStartCz),
"rectangular", 1, "no"
		select part1
		plus part2
		timeShifted = Concatenate
	else
		timeShifted = input'rec'
	endif

	@trialScr2
	select sound
	plus timeShifted
	Play
	while demoWaitForInput ()
		if demoClickedIn (fromx, tox, y3, y1) or demoInput ("o")
			select sound
			Play
			pl = pl + 1
		elsif demoClickedIn (fromx, tox, y7, y5) or demoInput ("y")
			select timeShifted
			Play
			yrs += 1
		elsif demoInput ("n →") or demoClickedIn (50+(leftgap/2), 100-leftgap,
(ybot2/2)-5, (ybot2/2)+5)
```

```
				goto MOVE2
			elsif demoInput ("a←") or demoClickedIn (leftgap, 50-(leftgap/2),
(ybot2/2)-5, (ybot2/2)+5)
				select timeShifted
				if soundStartCz <> soundStart
					plus part1
					plus part2
				endif
				for j to nInt
					if plPiecetimeShifted'j' = 1
						plus piecetimeShifted'j'
						plPiecetimeShifted'j' = 0
					endif
				endfor
				plus pitchCz
				plus sineCz
				plus gridCz
				plus spectrogCz
				plus gridCz0
				plus soundingCz
				Remove
				goto SCR1
			endif
			for j to nInt
				if j = 1
					if demoClickedIn (fromx, fromx+(tox-
fromx)/(origDur/intEnd'j'), y4, y3)
						@playpiece: "sound"
					endif
				else
					if demoClickedIn (fromx+(tox-
fromx)/(origDur/intStart'j'), fromx+(tox-fromx)/(origDur/intEnd'j'), y4, y3)
						@playpiece: "sound"
					endif
				endif
			endfor
			if showYourGrid = 1
```

```
				for j to nInt
					if j = 1
						if demoClickedIn (fromx, fromx+(tox-
fromx)/(origDur/intEnd'j'), ybot, y7)
							@playpiece: "timeShifted"
						endif
					else
						if demoClickedIn (fromx+(tox-
fromx)/(origDur/intStart'j'), fromx+(tox-fromx)/(origDur/intEnd'j'), ybot, y7)
							@playpiece: "timeShifted"
						endif
					endif
				endfor
			endif
		endwhile
		label MOVE2

;		select sound
;		for ii to rec
;			plus input'ii'
;		endfor
;		plus timeShifted
;		if soundStartCz <> soundStart
;			plus part2
;			plus part1
;		endif
;		plus gridOrig
;		plus pitch
;		plus sine
;		plus grid
;		plus spectrog
;		plus sounding
;		plus grid0
;		plus pitchCz
;		plus sineCz
;		plus gridCz
;		plus spectrogCz
```

```
;	plus gridCz0
;	plus soundingCz
;	for j to nInt
;		if plPiecesound'j' = 1
;			plus piecesound'j'
;		endif
;		if plPiecetimeShifted'j' = 1
;			plus piecetimeShifted'j'
;		endif
;	endfor

select all
minus list
	Remove
	if makeLog = 1
		select logTab
		Append row
		Set string value: i, "student", subjCode$
		Set string value: i, "word", word$
		Set numeric value: i, "played", pl
		Set numeric value: i, "recorded", rec
		Set numeric value: i, "yours", yrs
		Save as tab-separated file: subjCode$ + "_imitationLog.Table"
	endif
	label FASTMOVE
	if i = nSounds
		demo Erase all
		demo Paint rectangle: "silver", 0, 100, 0, 100
		demo Text: 50, "centre", 50, "half", "No more words left. Press left arrow
to go back or click to finish."
		while demoWaitForInput ()
			if demoInput ("•")
				goto THEEND
			elsif demoInput ("←")
				i -= 1
				goto LOAD_WORD
			endif
```

```
				endwhile
			endif
endfor

label THEEND
demo Erase all
demo Paint rectangle: "silver", 0, 100, 0, 100
demo Text: 50, "centre", 50, "half", "You've reached the end. Thanks for using this trainer."

select list
if makeLog = 1
	plus logTab
endif
Remove

procedure trialScr: .word$
	demo Erase all
	demo Colour: "black"
	demo Line width: 2
	demo Select inner viewport: 0, 100, 0, 100
	demo Axes: 0, 100, 0, 100
	demo Paint rectangle: "silver", 0, 100, 0, 100
	demo Draw rounded rectangle: 30, 70, 54, 70, 2
	demo Text special: 50, "centre", 62, "half", "Times", 20, "0", "#play again"
	demo Draw rounded rectangle: 30, 49, 35, 45, 2
	demo Text special: 39.5, "centre", 40, "half", "Times", 20, "0", "#record"
	demo Draw rounded rectangle: 51, 70, 35, 45, 2
	demo Text special: 60.5, "centre", 40, "half", "Times", 20, "0", "show #graphs"
	demo Text special: 0, "left", 0, "bottom", "Times", 12, "0", "'i'/'nSounds'"
	demo Text special: 50, "centre", 0, "bottom", "Times", 12, "0", "Use ← arrows → to move between examples."
endproc

procedure trialScr2
	demo Erase all
	demo Colour: "black"
```

```
demo Line width: 0.5
demo Select inner viewport: 0, 100, 0, 100
demo Axes: 0, 100, 0, 100
demo Paint rectangle: "silver", 0, 100, 0, 100
demo Paint rectangle: "white", fromx, tox, y3, y1
demo Select inner viewport: fromx, tox, y2, y1
select sound
demo Draw: 0, 0, 0, 0, "no", "Curve"
demo Draw inner box
if enSpec = 0
	spectrog = noprogress To Spectrogram: 0.005, 7500, 0.002, 20,
"Gaussian"
	enSpec = 1
else
	select spectrog
endif
demo Font size: 12
demo Select inner viewport: fromx, tox, y3, y2
demo Paint: 0, 0, 0, 0, 100, "yes", 50, 6, 0, "no"
demo Draw inner box
demo Font size: 25
demo Line width: 1
demo Select inner viewport: fromx, tox, y4, y1
select gridOrig
demo Draw: 0, 0, "yes", "yes", "no"
demo Font size: 25
demo Select inner viewport: fromx, tox, y4, y1
demo Text left: "no", "Original"

demo Select inner viewport: 0, 100, 0, 100
demo Axes: 0, 100, 0, 100
demo Line width: 0.5
demo Paint rectangle: "white", fromx, tox, y7, y5
demo Select inner viewport: fromx, tox, y6, y5
select timeShifted
demo Draw: 0, 0, 0, 0, "no", "Curve"
demo Draw inner box
```

```
	demo Select inner viewport: fromx, tox, y7, y6
	spectrogCz = noprogress To Spectrogram: 0.005, 7500, 0.002, 20, "Gaussian"
	demo Paint: 0, 0, 0, 0, 100, "yes", 50, 6, 0, "no"
	demo Draw inner box
	if showYourGrid = 1
		demo Grey
		demo Font size: 25
		demo Line width: 1
		demo Select inner viewport: fromx, tox, ybot, y5
		select gridOrig
		demo Draw: 0, 0, "yes", "yes", "no"
		demo Black
	endif

	demo Font size: 12
	demo Select inner viewport: fromx, tox, ybot, y6
	demo One mark bottom: 0, "yes", "yes", "no", ""
	demo One mark bottom: 'origDur:3', "yes", "yes", "no", ""
	;demo One mark bottom: 'soundStart:3', "yes", "yes", "yes", ""
	demo Text bottom: "no", "Time (s)"
	demo Font size: 25
	demo Select inner viewport: fromx, tox, ybot, y5
	demo Text left: "no", "Yours"
	demo Select inner viewport: 0, 100, 0, 100
	demo Axes: 0, 100, 0, 100
	demo Text special: 0, "left", 0, "bottom", "Times", 12, "0", "'i'/'nSounds'"
	demo Line width: 2
	demo Draw rounded rectangle: leftgap, 50-(leftgap/2), (ybot2/2)-3.5,
(ybot2/2)+3.5, 2
	demo Text: leftgap+((50-(leftgap/2)-leftgap)/2), "Centre", ybot2/2, "Half", "try
#again"
	demo Draw rounded rectangle: 50+(leftgap/2), 100-leftgap, (ybot2/2)-3.5,
(ybot2/2)+3.5, 2
	demo Text: 50+(leftgap/2)+((50-(leftgap/2)-leftgap)/2), "Centre", ybot2/2, "Half",
"go to #next"
endproc
```

```
procedure playpiece: .sound$
		if plPiece'.sound$"j' = 0
			select '.sound$'
			beg = Get nearest zero crossing: 1, intStart'j'
			end = Get nearest zero crossing: 1, intEnd'j'
			piece'.sound$"j' = Extract part: beg, end, "rectangular", 1, "no"
			Play
			plPiece'.sound$"j' = 1
		else
			select piece'.sound$"j'
			Play
		endif
endproc
```

Appendix 5: PRAAT script used for manipulation of stimuli

```
longVdur = 0.18

shortVdur = 0.1

longCdur = 0.12

shortCdur = 0.06

vSteps = 9

cSteps = 2


source = Read from file: "source.adjusted.F0.wav"

man = To Manipulation: 0.01, 75, 600

durTier = Extract duration tier

grid = Read from file: "source.TextGrid"

gridTab = Down to Table: "no", 10, "yes", "no"

vBeg = Get value: 1, "tmin"

vEnd = Get value: 1, "tmax"

cBeg = Get value: 2, "tmin"

cEnd = Get value: 2, "tmax"
```

```
for i to cSteps+1
	select durTier
	durC'i' = Copy: "durC'i'"
	appendInfoLine: "c'i' ", e^(ln(shortCdur)+((ln(longCdur)-
ln(shortCdur))/cSteps*(i-1)))
	Add point: cBeg, 1
	Add point: cBeg+0.001, (e^(ln(shortCdur)+((ln(longCdur)-
ln(shortCdur))/cSteps*(i-1))))/(cEnd-cBeg)
	Add point: cEnd-0.001, (e^(ln(shortCdur)+((ln(longCdur)-
ln(shortCdur))/cSteps*(i-1))))/(cEnd-cBeg)
	Add point: cEnd, 1
	for ii to vSteps+1
		select durC'i'
		durC'i'V'ii' = Copy: "durC'i'V'ii'"
		appendInfoLine: "v'ii' ", e^(ln(shortVdur)+((ln(longVdur)-
ln(shortVdur))/vSteps*(ii-1)))
		Add point: vBeg, 1
		Add point: vBeg+0.001, (e^(ln(shortVdur)+((ln(longVdur)-
ln(shortVdur))/vSteps*(ii-1))))/(vEnd-vBeg)
		Add point: vEnd-0.001, (e^(ln(shortVdur)+((ln(longVdur)-
ln(shortVdur))/vSteps*(ii-1))))/(vEnd-vBeg)
		Add point: vEnd, 1
		plus man
		Replace duration tier
		select man
		stimC'i'V'ii' = Get resynthesis (overlap-add)
		if ii < 10
			Save as WAV file: "stimuliLog/stimC'i'V0'ii'.wav"
		else
			Save as WAV file: "stimuliLog/stimC'i'V'ii'.wav"
```

```
        endif
    endfor
endfor
```

Appendix 6: Instruments used for participant recording in imitation training

Headphones: Sennheiser HD 202 II Wired headphone HiFi Stereo - Black

Microphone: Trust Mico USB Microphone

Software: PRAAT.exe version 5.4.19 developed by Boersma and Weenink (2014)

Device: MSI laptop GE60 0ND

Appendix 7: List of candidate nonsense words considered for the pretest and the posttest

Voiceless	**Voiced**
/t/	*/d/*
ploot	plood
shpoot	shpood
/p/	*/b/*
wope	wobe
/k/	*/g/*
moke	moge
shlick	shlig
wuk	wug
yik	yig