

**Filozofická fakulta
Univerzita Karlova**

Bakalářská práce



Praha, 2019

Dominika Trčková

**Filozofická fakulta
Univerzita Karlova**

Ústav anglického jazyka a didaktiky

Bakalářská práce

Dominika Trčková

**Porovnání vlivu segmentálních a prozodických manipulací
na sílu cizineckého přízvuku a pocit srozumitelnosti**

*Comparing the effect of segmental and prosodic
manipulations on speaker's accentedness and
comprehensibility*

Praha, 2019

doc. Mgr. Radek Skarnitzl, Ph.D

Acknowledgements

I would like to express my sincere thanks to doc. Mrg. Radek Skarnitzl, Ph.D., who has guided me through the whole process of preparing the experiment and writing this thesis. His kind advice and insightful instruction have been most helpful. In return, I hope to be able to professionally contribute to his work and to the Department of Phonetics at the Charles University in the future. I would also like to thank Monika Vlčková, a fellow student, for her cooperation during the experiment and for allowing me to share my moments of doubts with her during our work.

Prohlášení

Prohlašuji, že jsem bakalářskou práci vypracovala samostatně, že jsem řádně citovala všechny použité prameny a literaturu a že práce nebyla využita v rámci jiného vysokoškolského studia či k získání jiného nebo stejného titulu.

V Praze dne **24. 5. 2019**

Handwritten signature of Dominika Tyčková in cursive script.

Abstrakt a klíčová slova

Cílem této bakalářské práce je porovnat a prozkoumat vliv segmentálních a prozodických manipulací na sílu cizineckého přízvuku a pocit srozumitelnosti, a to prostřednictvím experimentálního výzkumu. Zvukové manipulace spočívají v odstranění chyb typických pro české mluvčí angličtiny, a tedy v teoretickém vylepšení jejich projevu. V teoretické části jsou nejprve představeny dimenze, pomocí nichž lze popsat charakterizovat cizineckého přízvuku. Následující sekce se věnuje sociálním aspektům komunikace v řeči s cizineckým přízvukem a diskriminaci, jíž nerodilí mluvčí mohou čelit. Další sekce teoretické části se zabývá popisem segmentálních a suprasegmentálních rysů, které jsou typické pro angličtinu, a jejich rolí v osvojování cizího jazyka. Mimoto jsou popsány také základní rysy tzv. české angličtiny, tedy rysy anglického projevu českého mluvčího. V empirické části této bakalářské práce se představuje metodologické pozadí celého výzkumu, tedy výběr materiálu, jeho segmentální a prozodické manipulace, příprava percepčního testu, jeho zadání a popis statistické analýzy dat. Výsledky jsou poté představeny v následující diskuzní části. Data získaná z odpovědí šedesáti osmi respondentů ukazují, že manipulace percepci řečového projevu nerodilých mluvčích ovlivňují, ale v různé míře. Síla cizineckého přízvuku je pozitivně ovlivněna segmentálními manipulacemi a je nevýznamně ovlivněna manipulacemi prozodickými, což je předvídatelné vzhledem k tomu, že přízvuk koreluje zejména právě se segmenty. Hodnocení pocitu srozumitelnosti se ovšem chová méně předvídatelně, a to tak, že kvůli segmentálním i prozodickým manipulacím dochází ke zhoršení.

Klíčová slova: angličtina, český mluvčí, cizinecký přízvuk, pocit srozumitelnosti, výslovnost, prozodie, hodnocení

Abstract and key words

This thesis aims at comparing and investigating the effect of segmental and prosodic manipulations on speaker's accentedness and comprehensibility; experimental research is the chosen mode of investigation. The manipulations consist of the deletion of mistakes typically found in the speech of Czech speakers of English and of the theoretical improvement of their utterance. The theoretical part of this thesis introduces the three dimensions which are used for describing foreign-accented speech. The following section focuses on the social aspects of communicating in a non-native language with a foreign accent and on the discriminatory behavior that foreign speakers often face. The next section of the theoretical part presents a description of the most salient segmental and suprasegmental features of English, along with the role these features play in the acquisition of a foreign language. Characterization of Czech English is also introduced in this section. The empirical part of this thesis describes the methodology employed in this research; explanation of selection of material, its segmental and prosodic manipulations, the preparation of the perception test and the testing itself, and the description of the statistical analysis of the responses are all included in this section. The following chapter then presents the results and their interpretation. The data, obtained from 68 respondents, show that manipulations indeed mostly affect the evaluation of accentedness and comprehensibility, but to different degrees. Accentedness is significantly influenced by segmental manipulations and only marginally influenced by prosodic manipulations; this is to be expected, since accent and segmentals are correlates. Comprehensibility scores, however, behave less predictably: the evaluations of this parameter become worse as a result of both segmental and prosodic manipulations.

Key words: English, Czech speaker, accentedness, comprehensibility, pronunciation, prosody, evaluation

Table of Contents

1. Introduction	9
2. Theoretical background	11
2.1. Dimensions of foreign-acented speech	11
2.2. Social aspects of communication in a foreign language	13
2.3. Segmental and suprasegmental features of English	14
2.3.1. Selected segmental features of English	15
2.3.2. Selected suprasegmental features of English	20
2.3.3. The role of segmentals and suprasegmentals in language acquisition	23
2.4. Selected features of Czech English	26
3. Material and methodology	29
3.1. Compilation of material	29
3.2. Manipulations	29
3.3. Perception test	34
3.4. Experiment	36
3.5. Analysis	37
4. Results and discussion	38
4.1. Comprehensibility scores	38
4.2. Accentedness scores	39
4.3. Item analysis	40
5. General discussion	44
6. Conclusion	46
7. References	47
8. Resumé	49

Table of Figures

Figure 1: English monophthongs chart (Roach, 2004: 242).....	2
Figure 2: Centering and closing diphthongs (Roach, 2009: 18).....	2
Figure 3: English consonants chart (voiced sounds are on the left, voiceless sounds on the right)	2
Figure 4: Articulation of /θ, ð/ (Roach, 2004: 41).....	2
Figure 5: Articulation of /ŋ/ (Lawson, E., et al.).....	2
Figure 6: An example of replacing /d/ with /ð/	2
Figure 7: An example of [k] deletion	2
Figure 8: An example of a pitch manipulation, which aimed for a fall-tone at the end of an utterance.	2
Figure 9: An example of a pitch manipulation, which aimed for putting a stress on a chosen syllable	2
Figure 10: An example of a duration manipulation, which aimed at either prolonging a sound (red arrows) or shortening them (blue arrows).....	2
Figure 11: The final design of the test-screen	2
Figure 12: Item sequences	2
Figure 13: Interaction between the level and the manipulation factors within comprehensibility	2
Figure 14: Interactions between the level and the manipulation factors within accentedness..	2
Figure 15: An effects plot showing mean fitted values and the respective confidence intervals for accentedness	2
Figure 16: An effects plot showing mean fitted values and the respective confidence intervals for comprehensibility	2

1. Introduction

Traces of cultural shift of the urbane contemporary world can be seen everywhere: language is not an exception. The study of foreign-accented speech is becoming more available every day as people, with the options that travelling or the Internet offer, broaden their education by learning various languages. With such progress, data for the study of foreign-accented speech, its dimensions, and the effects such language has on interpersonal communication, are readily at hand. That is the case especially with English, which has become a global language and today is often used more as an acquired language, not as a native language (Meierkord, 2013: 1). Acquired languages are often distinguished from the codified standard by specific features of pronunciation, lexicon, and syntax; the case is not different with English and its spoken varieties across the world.

Speaking with an L2 accent has “strong social, psychological, and communicative consequences” (Derwing and Munro, 2009: 476). Studies into this area therefore lead to important conclusions which inform and update the instruction of English as a foreign language. This thesis focuses on native Czech speakers who speak English as their L2. As a study of the connection between accentedness and comprehensibility of English with distinguishable Czech accent has not been conducted yet to this extent, the aim of this paper is to research the variables in the evaluation of these two criteria and, possibly, to consequently improve the quality of teaching pronunciation of English to Czech native speakers.

The experiment consists of playing several recordings of Czech speakers of English to respondents, who will have to rate how accented and how understandable the recordings were. The key part of the experiment is that one half of these recordings will be segmentally or prosodically manipulated to achieve either a better, or a worse impression on the respondent; this is to determine the effect of these manipulations on listeners.

Apart from the introductory word, this thesis contains the following sections: Theoretical Background, Method, Results and Discussion, General Discussion, and Conclusion. Theoretical Background discusses three main areas: dimensions of foreign-accented speech (accentedness, comprehensibility, intelligibility), social aspects of communication language acquisition in a foreign language, selected segmental and suprasegmental features of English, and selected segmental and suprasegmental features of Czech English. In the Method section, the process of material selection, material manipulation, the creation of the perception test, and the actual experiment are described. Results and Discussion section presents data obtained from the experiment. A short summary of this discussion and its implications are presented in General Discussion. The experiment’s links to

both past and future research are presented in Conclusion. A list of sources used and consulted to construct this paper is at the very end of this thesis, along with a Czech resumé.

The practical part of this thesis is a work of collaboration. The manipulations, the design and the carrying out of the experiment, the compilation of data, as well as the administration of a perceptual test were prepared and conducted together with a student of the Metropolitan University Prague as part of her own bachelor's thesis. The reason behind this is purely practical: the scope of the investigation was better managed in a team of two as it allowed for more manipulations to be carried out and more participants being tested.

2. Theoretical background

2.1. Dimensions of foreign-accented speech

Successful communication rests on many pillars, and the ability to be understood by the listener is one of them. As Varonis and Gass state in their research, “differences in the interlocutors’ cultural and linguistic background can upset what might otherwise be a relatively straightforward exchange of information” (1984: 85). A speaker’s accent is one such linguistic marker which affects comprehension of speech. As “either dialectical differences attributable to region or class, or phonological variations resulting from L1 influence on the L2” (Derwing and Munro, 2009: 476) it is a perceptual phenomenon, which helps listeners to distinguish members of different speech communities (Derwing and Munro, 2015: Glossary). Foreign accent is most commonly characterized by the speaker’s native language (Atagi and Bent, 2011: 260) and has a considerable influence on their ability to be understood. Consequently, attempts have been made to classify dimensions of foreign-accented speech and to quantify the comprehension of such speech.

While there are several ways to approach this question, the most salient and consistent terminology was first introduced by Varonis and Gass (1982, 1984) and later elaborated by Derwing and Munro (1995). Through experiments in which listeners were exposed to non-native accented speech and tested on their ability to comprehend the content, these researchers developed a set of descriptors. Foreign-accented speech is thus defined using three different dimensions: accentedness, comprehensibility and intelligibility. In their *Pronunciation Fundamentals*, Derwing and Munro (2015) define these terms as follows:

- accentedness: “The extent of difference perceived by speakers of one linguistic variety when listening to speakers of other varieties” (175). In other words, this parameter explains the extent to which an individual’s speech is marked by a foreign accent. In the case of this particular research, this will be Czech accent in English speech.
- comprehensibility: “The ease or difficulty a listener experiences in understanding utterance” (5). Put simply, this parameter marks how easy or difficult it is for listeners to understand the content of someone else’s speech; in other words, comprehensibility is a subjective construct.
- intelligibility: “The degree of match between a speaker’s intended message and the listener’s comprehension” (5). This parameter is the connection between the speaker’s

intention and the reality: the extent to which the planned function of a discourse is fulfilled on the side of the listener. In other words, intelligibility is an objective construct.

These three parameters are interrelated, but partially independent. Derwing and Munro's data show a correlation "between intelligibility and perceived comprehensibility and between comprehensibility and accent" (1995: 90). Atagi and Bent confirm another finding from the previous study: that high accentedness influences perception but does not necessarily interfere with intelligibility (Atagi and Bent, 2011: 262). They are therefore overlapping but independent parameter, "as illustrated by the fact that even some heavily accented L2 speech can be highly comprehensible" (Saito et al., 2016: 218-219).

Saito and his colleagues further investigated concrete aspects of speech which affected the rating of accentedness and comprehensibility of L2 speakers. When it comes to accent, both segmental (individual sounds) and suprasegmental (word stress, intonation, speech rate) dimensions of speech are "equally important at all levels" of language fluency. Lexicon and grammar are important at the initial stage, while "grammatical complexity (along with pronunciation variables)" determine the rating of native-like L2 speech at the highest skill level. With comprehensibility, word stress and intonation are equally important at all levels; "attaining minimum level of segmental accuracy, fluency, lexical appropriateness, and grammatical accuracy" matter in the initial stage of learning, while "segmental precision and grammatical accuracy" mark the highest stage of learning (2016: 232-233).

Given that the aim of this thesis is to study the effects of segmental and suprasegmental aspects on accentedness and comprehensibility, there is a clear need to list what exactly is meant by these two terms. Segmentals are to be understood in quite simple terms: as phonemes – the smallest units in language that can distinguish meaning – of a given language. In this particular case, those would be consonants and vowels of the English language. Suprasegmental aspects pertaining to comprehensibility and accentedness are speech rate, pausing, stress, and pitch patterns (intonation) (Kang et al., 2010: 555). In other words, prosody (a synonymous term used to denote suprasegmentals) is *how* we say what we say. Kang et al. (2010: 564) came to the following conclusion about prosody and comprehensibility in their study: that "suprasegmental fluency, the use of mid-rising tone choices, and control over high vocal pitch" are "especially potent in determining perceived proficiency and comprehensibility;" the "use of features that mark boundaries between idea units" also came out as important.

In this thesis, accentedness and comprehensibility will be the foci of the experiment.

2.2. Social aspects of communication in a foreign language

Every single speaker – be it a native or a non-native – has an accent. It is thus an inextricable part of speech and, consequently, a part of interpersonal communication. Most researchers argue that “there is nothing inherent to accents that makes some more aesthetically pleasing than others; rather, accents serve as cues to social identities, activating either negative or positive connotations” (Dovidio and Gluzsek, 2010: 217). Deviation from what is generally perceived as the language standard by the majority population is, unfortunately, linked to stigmatization and often blamed for all kinds of problems connected with miscommunication. What is worse, it is also used as a cover-up for racism and other forms of discrimination (Derwing and Munro, 2009: 476). It is thus a key fact that the one facet of speech to which listeners are generally very sensitive is the presence of an accent: “accent features are exceptionally salient, and as a result we’re very good at detecting perceived outsiders on the basis of their speech patterns” (Scovel, 1988; cited in Derwing and Munro, 2009: 477). It is therefore vital to look at the social aspects of communication in a foreign language.

First, let us start with the most salient social benefits of speaking with a foreign accent. An L2 accent is an important signal to the interlocutors that instructs them to possibly modify their speech to match the degree of the linguistic skill of the L2 speaker in that particular communicative situation. This phenomenon is called foreigner talk (Derwing and Munro, 2009: 484). Another, rarer, social benefit of an accented L2 speech is the supposed sophistication of some foreign accents: Derwing and Munro name the conductor Leopold Stokowski as an example of a London-born speaker, who chose to publicly present himself with an eastern accent (Derwing and Munro, 2009: 484).

However, the social disadvantages of an accented L2 speech seem to outweigh its benefits. Dovidio and Gluzsek (2010) published an extensive paper, in which they attempt to map the social stigma arising from speaking with an accent which is perceived as foreign: the paper defines stigma as a deeply discrediting attribute, which reduces the speaker “from a whole and usual person to a tainted, discounted one” (Goffman, 1963; in Dovidio and Gluzsek, 2010: 216) and which “conveys a social identity that is devaluated in a particular social context” (Crocker et al., 1998; in Dovidio and Gluzsek, 2010: 216). With this definition in mind, a theory about learning and identity ought to be mentioned: Brian Morgan states that learning is a process of “acquiring an identity, of becoming someone or something” (2010; in Derwing and Munro 2015: 140). However, as Derwing and Munro

remind us, it is not possible to always acquire the identity the speaker would like, especially with L2 speakers and their accent. Non-native speakers whose speech is marked by a foreign accent may face many forms of social prejudice and discrimination, which in many cases lead to the previously mentioned social stigmatization of the L2 identity acquired through learning.

From the perspective of the listener, the following are amongst the most common examples of stigmatizing attitudes toward accented speech. In as early as 1960, research has demonstrated that speakers of both the stigmatized and the stigmatizing languages “held negative attitudes towards the stigmatized language.” Since then, similar results were obtained and a clear conclusion reached: non-natively accented speech is perceived more negatively than a speech with a native accent (Dovidio and Gluzsek, 2010: 217). A real example of such behaviour in the presence of someone who is linguistically perceived as an outsider is the case of Ian Hussey. Hussey was born and grew up in England and worked in an India-stationed telecommunications service; upon answering a question about his location, the customer reacted with “F*** off, you job-stealing Paki.” Similar “racial epithets” were a part of the service employees’ daily experience (Derwing and Munro, 2015). Numerous other negative stereotypes are associated with foreign accented speech: people with L2 accents are perceived as less intelligent, less loyal, less competent, less credible, as speaking the language poorly, or as having a lower economic and social status than native speakers. There are also many stereotypes, both positive and negative, connected with specific accents; an example of this is Korean accent and the attribute of being hard-working and intelligent (Dovidio and Gluzsek, 2010: 217). Discrimination is another dangerous facet of communicating with an accent. The following are only a few examples of the possible scenarios in which speech serves as a basis for arbitrary evaluations: discrimination in housing, employment, education, and the courts (Dovidio and Gluzsek, 2010: 218).

2.3. Segmental and suprasegmental features of English

In the following sections (2.3.1., 2.3.2.), selected segmental and suprasegmental features of English will be explained and discussed. Those presented were selected as features which are the key milestones in learning and often pose as problematic to learners of English as a foreign language, especially Czech learners of English, and which are representative of English in general. It is by no means an attempt to map English in its total;

for this purpose, further reading is recommended. Subsection 2.3.3. outlines the role of the selected features in the process of foreign language acquisition.

2.3.1. Selected segmental features of English

As any other language, English works with two major groups of sounds: vowels and consonants. Roach uses the broad definition of vowels as “sounds in which there is no obstruction to the flow of air as it passes from the larynx to the lips” (2009: 10). Vowels are generally described using two dimensions: “the vertical distance between the upper surface of the tongue and the palate” and “the part of the tongue, between front and back, which is raised highest” (Roach, 2009: 10). Using these dimensions, vowels are placed in a quadrilateral diagram, whose dimensions roughly answer to the shape of the oral cavity. *Figure 1* and *Figure 2* below illustrate these diagrams in practice as they show English vowels – both monophthongs and diphthongs. English monophthongs can be either short or long. Short monophthongs include the vowels /ɪ, e, æ, ə, ʌ, ɔ, ʊ/; long monophthongs include the vowels /i:, u:, ɔ:, ɜ:, ɑ:/.

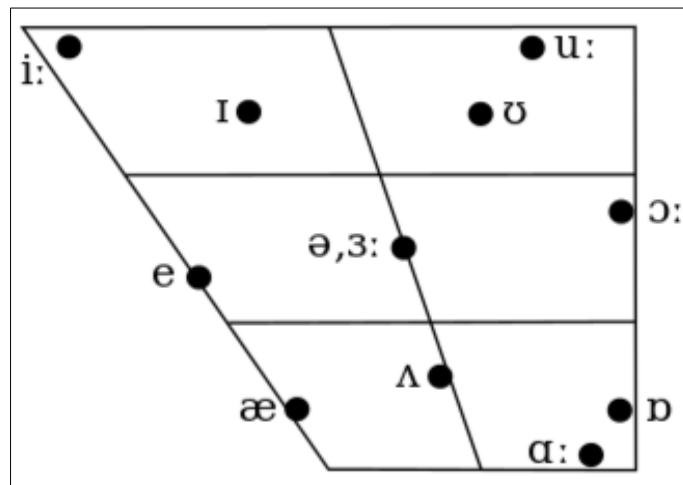


Figure 1: English monophthongs chart (Roach, 2004: 242)

Sounds, which “consist of a movement of a glide from one vowel to another” (Roach, 2009: 17) are called diphthongs. In English, the first part is, as a rule, longer and stronger than the second part. English has two kinds of diphthongs: centering and closing. Centering diphthongs end in the vowel /ə/, which will be discussed later in the chapter; closing diphthongs of English end either in /ɪ/ or in /ʊ/, which are articulated with the tongue close to the palate, hence the name closing diphthongs. The diagram below illustrates the nature of English diphthongs and the direction of the sound

movement: the image on the left shows centering diphthongs and the image on the right shows closing diphthongs.

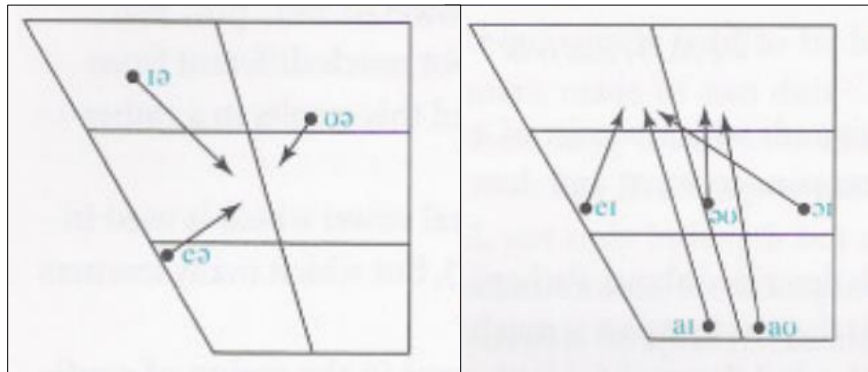


Figure 2: Centering and closing diphthongs (Roach, 2009: 18)

A sound which is worth more than a mere mention is the so-called *schwa*, marked in IPA (the International Phonetic Association's graphic representation of sounds) as /ə/. It is the most frequently occurring vowel in English (Roach, 2009: 65). As a mid-central short vowel, it originates in the centre of the oral cavity with the tongue in neutral position. An example of this phoneme are the initial sounds in *about* or *attend*.

It can be described as lax, since its production does not require much energy (Roach, 2009: 65); this feature of schwa plays a key role in the formation of English rhythm as it allows for the process called reduction to take place. More will be said about reduction in the following section.

As to consonants, it is logical to start in defining them against vowels. Where vowels form an opening in the oral cavity, consonants form an obstruction using various articulatory organs in the vocal tract. There are three key characteristics which can be used to define consonants: voicing, manner of articulation, and place of articulation. Voicing, also referred to as phonation, is a process of obstructing the air flow by bringing vocal folds together in the larynx, causing them to vibrate. Phonation can vary: vocal folds can "be made longer or shorter, more tense or more relaxed or be more or less strongly pressed together" (Roach, 2009: 25); the pressure of the stream of air can also be manipulated. English consonants divide into two large groups of sound: obstruents and sonorants. Obstruents are those consonants which have no (significant) voicing; sonorants are those consonants which are always voiced and some of them can even act as vowels as they can be the peak of a syllable. These two groups then divide into subcategories based on the manner and place of articulation. For clarity, *Figure 3* can be consulted as it presents all this information in an organized table.

		bilabial	labio-dental	dental	alveolar	post-alveolar	palatal	velar	glottal
obstruents	plosives	p b			t d			k g	ʔ
	fricatives		f v	θ ð	s z	ʃ ʒ			h
	affricates					tʃ dʒ			
sonorants	nasals	m			n		ŋ		
	approximants					r	j	w	
	lateral approximants				l				

Figure 3: English consonants chart (voiced sounds are on the left, voiceless sounds on the right)

Place of articulation is a good category to start with as it is rather straightforward; as the name suggests, this feature refers to the place where the consonant sound originates in the vocal tract. Bilabial sounds /p, b, m, w/ are created using the lips, while labiodental consonants /f, v/ combine both the lips and the teeth. Dental phonemes /θ, ð/ are made by putting the tongue against the two rows of teeth. Alveolar sounds /t, d, s, z, n, l/ are created by putting the tip of the tongue to the so-called alveolar ridge, which can be described in simple terms as the little step behind our teeth. To make the post-alveolar sounds /ʃ, ʒ, tʃ, dʒ, r/, the tongue is in contact with the area slightly further back than in the position for an alveolar sound. The palatal /j/ originates at the hard palate of the oral cavity. Velar phonemes /k, g, ŋ/ use the soft part of the palate called the velum. Lastly, the glottal (laryngeal) consonant /h/ is made at the back of the cavity by a complete closure of the vocal folds.

English consonants can also be categorized by describing their manner of articulation, meaning the way in which these sounds are created. These can be divided into obstruents and sonorants, which were already mentioned. The categories called plosives, fricatives, and affricates are obstruents; sonorants comprise of nasals, approximants, and lateral approximants. Plosives are sounds which are formed by the complete closure of the oral cavity during which the air exceeds the elasticity of the tissue and is suddenly released. /p, t, k/ are voiceless and /b, d, g/ are voiced. When fricatives are formed, the closure is not complete: there is a critical narrowing between two organs through which air escapes turbulently. /f, θ, s, ʃ, h/ are voiceless and /v, ð, z, ʒ/ are voiced. Affricates are a combination of the previous two processes: they start with plosion and end as fricatives. /tʃ/ is voiceless while /dʒ/ is voiced. Nasals are formed through a complete closure of the oral cavity and the air escaping through the nasal cavity via lowered velum; /m, n, ŋ/ are nasal sounds. When there is a narrowing, which is not critical, between an active and a passive organ, approximants are formed. They are divided into glides, which have no stable phase in articulation and thus provide a

sense of gliding; /j, w/ are glides. The second group of approximants are liquids: these are also unstable, and they seep into the neighboring sounds; /l, r/ are liquids. /l/ can also be classified as a lateral approximant, because the sides of the tongue lower during the formation of this sound and the air escapes laterally.

As this research focuses on three specific consonants, more shall be said about them in order to highlight their problematic nature and illustrate their correct realization. These are the sounds in question: /θ, ð, ɱ/. The dental fricatives /θ, ð/ are one of the most, if not the most, problematic aspect of Czech pronunciation of English, since these sounds are unfamiliar to the Czech language. /θ/ is a voiceless sound and appears in words like *thing* or *theft*. /ð/ is a voiced sound and it appears in words like *they* or *thus*. However, it is more accurate to refer to them as fortis (strong) and lenis (weak) because even voiced obstruents include hardly any significant voicing in terms of perception when compared to sonorants; this terminology applies to all obstruents. To repeat for clarity: to create a dental fricative, the tongue is put behind the teeth, “with the tip of touching the inner side of the lower front teeth and the blade touching the inner side of the upper teeth. The air escapes through the gaps between the tongue and the teeth” (Roach, 2009: 41). The fricative sound is rather weak. Unfortunately, some teachers tend to teach this consonant by telling their pupils to place their tongue between their teeth; positioning the tongue against the teeth has in reality no effect and so is unnecessary. The image on the left in *Figure 4* shows a diagram of the correct articulation of English dental fricatives.

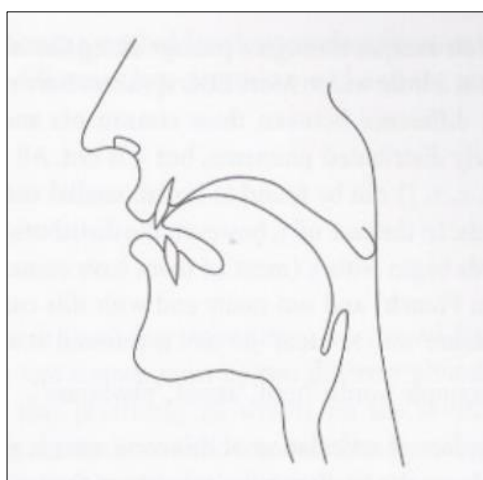


Figure 4: Articulation of /θ, ð/ (Roach, 2004: 41)

The sound of the velar nasal /ŋ/ is made when the air escapes the oral cavity through the nose, rather than the mouth as in the case of other consonants; for this to happen, the soft palate (velum) must be lowered and the back of the tongue put against it. The correct pronunciation is demonstrated in *Figure 5*.



Figure 5: Articulation of /ŋ/ (Lawson, E., et al.)

Again, this sound is quite problematic for foreign learners; the reason behind this, however, is not the difficulty of its articulation (it is phonetically simple), but its phonological complexity. In other words, the rules about the contexts in which this sound appears are hard to follow. Let us therefore have a look at its distribution. In the first place, /ŋ/ never appears initially (at the beginning of a word); along with /ʒ/ it is the only consonant in English with such distribution. This straightforward rule is generally applied without a problem by foreign speakers. Secondly, the phoneme does appear medially but is pronounced only in some cases. When it is followed by either of the plosives /k, g/, these rules apply: when the letters <nk> are found in the middle of a word, the /k/ sound is always pronounced, as in the word *tanker*; this is easily comprehended by Czech speakers, since their own language uses the same paradigm. When the letters <ng> appear in the middle of a word, the pronunciation is based on the word's morphology (the word's construction). The major difference with which foreign learners have to grapple with, especially Czech learners, is the difference between /ŋ/ as a phoneme and an allophone. In English, /ŋ/ is a separate phoneme; this means that the presence of this sound distinguishes meaning – the words *sing* with /ŋ/ means something else than *sin* with mere /n/. In Czech, however, /ŋ/ is an allophone, meaning it is only a positional variant which does not alter the meaning of a word if it appears instead of /n/. What is problematic for non-native speakers is the final position. /ŋ/ is pronounced

without the /g/ sound if it stands at the end of a morpheme, as in *singer*; if it stands within a morpheme, /ŋg/ is pronounced fully, as in *anger*. /ŋ/ is always pronounced without /g/ in the final position (at the end of a word), for example in *long*. Unfortunately, there are exceptions to these rules and those are comparative and superlative forms, whose correct pronunciation includes /g/, as in words *longer* and *longest* (Roach, 2009: 47-48).

2.3.2. Selected suprasegmental features of English

Now, before any in-depth description of English prosody – including stress placement, rhythm, intonation and pitch-range – weak and strong syllables need to be discussed first. It is now that the previous description of the vowel schwa becomes useful: English syllables can be classified either as strong or weak and the presence of this vowel is often a determining factor. Weak syllables tend to be “shorter, of lower intensity (loudness) and different in quality” (Roach, 2009: 64). Schwa is the vowel most associated with weak syllables; besides that, vowels found in the general area of /ɪ, i:/ and of /ʊ, u:/ which are symbolised as [ɪ, u] can also be the peaks of weak syllables. Consonants /l n/ can also form the peak of a syllable – these are called syllabic consonants. These weak forms most often appear in grammatical words; these words generally carry a function rather than content, e.g. pronouns (*your*), prepositions (*from*), or auxiliary modal verbs (*can*). The strong pronunciation of /jɔː, frɒm, kæn/, which appears when these words are isolated, can become /jə, fm, kən/ in connected speech. Connected speech is economical and so prevents the speaker from putting excessive effort into speaking; in other words, the speaker reduces the strong (full) forms of words as much as necessary to save their energy. It is important to note that connected speech is a natural phenomenon that occurs with ordinary conversation speech. It is thus crucial for foreign learners of English to master connected speech in order to achieve a native-like accent. Three large groups of processes are the main modes of connected speech: assimilation, elision, and linking. Assimilation is a process in which “we find a phoneme realised differently as a result of being near some other phoneme belonging to a neighbouring word” (Roach, 2009: 110). Assimilation mostly affects consonants. It is of three kinds: place, manner, and voicing. The phrase *in bed* can therefore become something along the lines of /ɪm bed/, due to the effect of /b/ on /n/. “Under certain circumstances, sounds disappear” is how Roach (2009: 113) defines elision. An elided phoneme has thus a zero realisation. Both consonants and vowels are affected: the

phrase *next please* can become /neks pli:z/ and the word *police* can be realized as /pli:s/; in the first example, the consonant /t/ disappears and in the second the schwa is deleted. Linking, as the name suggests, is a process during which words are linked across their boundaries; words are therefore not pronounced as separate units.

With this introduction into weak forms, neutralisation, and connected speech in mind, it is now time to turn to more general prosodic phenomena. Let us start with word stress. From a production point of view, a stressed syllable is produced with more muscular effort than an unstressed one; perceptually, a stressed syllable carries more prominence. This means that they are perceived as louder, longer, pronounced on some pitch (it sounds lower or higher), and has a different quality (it is not a weak syllable) (Roach, 2009: 73). Unstressed syllables therefore assume weak forms and are more likely to undergo the phenomena linked to connected speech. The problem with English stress placement is that it is highly unpredictable: unlike in languages where stress is fixed to a certain position (like Czech), English stress moves in what seems a rather random fashion. The regularities are so complex that it is better to “treat stress placement as a property of the individual word, to be learned when the word itself is learned” (Roach, 2009: 76). The regularities can be mapped onto the following areas: the morphological complexity of the word, its grammatical category, the number of syllables, and the phonological structure.

Rhythm involves a noticeable action taking place at regular intervals of time. The rhythm of English speech is classified by theory as a stress-timed rhythm: this implies that “stressed syllables will tend to occur at relatively regular intervals” and that “the times from each stressed syllable to the next will tend to be the same, irrespective of intervening unstressed syllables” (Roach, 2009: 107). These units of equal duration are referred to as feet. It is at this point that the distinction between weak and strong syllables becomes key to understanding English rhythm: weak syllables have the capacity to become reduced as much as is required by the length of a rhythmical foot. If there are only a few syllables in a foot, the need for reduction is not as strong as in those feet with numerous syllables; in those, weak syllables can become extremely reduced to keep the perceived regularity of the speech. Of course, it is important to keep in mind that this is a theoretical approach; in reality, speech is not as regular as we might wish and not all types of speech will follow the scheme of a stress-timed rhythm.

There is not one right definition of what intonation is, but roughly speaking, it involves a melody movement (a change in the vibration of vocal folds) and it provides

a system of contrasts which change the meaning of an utterance. We can thus attempt to define intonation as a linguistically significant variation in the pitch within a prosodic phrase. Pitch is an “auditory sensation experienced by the hearer” (Roach, 2009: 120) which helps them to brand it as high or low. Each speaker has their own pitch range: “a top level which is the highest pitch normally used by the speaker, and the bottom level that speaker’s pitch normally does not go below” (Roach, 2009: 122). A prosodic phrase is a tone unit – a part of an utterance which bears a concrete melody. In English, these units can have five different tones: the simple tones fall and rise, the complex tones rise-fall and fall-rise, and the static level tone. These tones generally begin on the last stressed syllable of a tone unit and it is called the tonic syllable. Each of these tones carries with them some typical functions. These can be demonstrated at a yes or no answer to a question. A fall tone provides a sense of finality. Its opposite, the rise tone, on the other hand gives the impression that something more is to follow. A fall-rise answer suggests a limited agreement and rise-fall is often used to convey strong feelings, like disapproval or surprise. Lastly, we will have a look at some of the functions of intonation. Mennen and de Leeuw (2014: 185-186) provide a comprehensible list of these and their illustrations on English. It can have a grammatical, or discourse, function, which indicates what type of a sentence is the speaker dealing with: the phrase “Mile End is in London” (185) can function as a statement if said with a falling tone, or as a question if said with a rising tone. Intonation can also be a tool of turn taking. A falling pitch combined with the lengthening of the final syllable indicates that the speaker has finished their turn, whereas a rising pitch or a high-level pitch shows that they want to continue their utterance. Besides that, prosody can make the important information stand out: such focusing of attention to a certain word is cued “primarily by acoustic patterns of fundamental frequency (F0), duration, and amplitude” (185). These are perceived as “pitch, length, and loudness, such that the stressed syllable of the emphasized word is perceived as higher, longer, and louder in comparison to the words and syllables that are not emphasized” (185). A second function of prosody is to convey lexical meaning. Languages such as English employ lexical stress, which can serve as the primary distinguishing tool in minimal pairs of words – words with no segmental differences, which differ by their stress placement. Mennen and de Leuw use *FOREbear* (a noun meaning *an ancestor*) and *forBEAR* (a verb meaning *to abstain*) to exemplify this, using capital letters to indicate the stress placement. A third function of prosody that plays a key role in English is that of “grouping constituents that belong together”

(186) – this is referred to as prosodic phrasing. With the help of this function, words can be grouped into larger blocks of speech to signal “major syntactical boundaries or paragraph boundaries” (186) to prevent ambiguities of utterances with unclear syntax. The sentence “When you learn gradually you worry more” can be divided twofold: either into the blocks “when you learn gradually” and “you worry more” or into the blocks “when you learn” and “gradually you worry more” (186). It is obvious that the meaning of the utterance changes significantly based on the choice of phrasing. These differences are again signalled by cues such as “pausing, lengthening of the syllable at the end of the phrase, a change in pitch direction, or a combination of these” (186).

2.3.3. The role of segmentals and suprasegmentals in language acquisition

In the third chapter of their highly illuminative work *Pronunciation Fundamentals: Evidence-based Perspectives for L2 Teaching and Research* (2015: 29-53) Derwing and Munro discuss various variables which influence the process of learning a foreign language. The following sequence lists the most prominent factors: language experience, motivational influences, aptitude, instruction, or lesser factors such as learning styles or personality. What is key for this research is a discussion of the type of instruction; this follows below.

Broselow and Kang’s *L2 Phonology and Phonetics* (2013) introduces a whole range of issues in second language acquisition. First, they talk about transfers from the learner’s L1, that is applying structures from one’s native tongue into the target language. Instead of accepting the Contrastive Analysis Hypothesis, which states that those aspects in L2 that are similar to L1 will be acquired easily, they use the Markedness Differential Hypothesis. This theory claims that marked (typologically uncommon and phonetically harder) L2 aspects are “more difficult to acquire than equally novel but less marked structures” (Broselow and Kang, 2013: 530). A phenomenon which is termed “equivalence classification” in the Speech Learning Model (Flege, 1995; in Broselow and Kang, 2013: 530) describes the process of assigning “L2 sounds to similar but not necessarily identical L1 phoneme categories” (Broselow and Kang, 2013: 530). Such an introduction into the matter of what constitutes issues in second language acquisition clearly maps out the necessary focus on what is especially foreign – or marked – to the learner. The following studies aim to provide a solution to this with the help of explicit instruction.

Gordon et al. (2013) offer an insightful view into the classroom and the role of explicit phonetic instruction in acquisition of English as a second language. Their classroom-based study confirms that the explicit instruction of segmentals and suprasegmentals visibly improve the comprehensibility of L2 English speakers. The experiment consisted of three intermediate speaking classes: one was taught with non-explicit instructions, while the other two were taught with “explicit phonetic instruction within a communicative methodology” (Gordon et al., 2013: 196) – one of these entailed such instruction of segmentals (vowels /i, ɪ, ε, æ/), the second focused on suprasegmentals (stress, rhythm, linking, reduction). Each lesson followed a presentation-practice-production plan, during which the experimental groups took part in activities such as “minimal-pair recognition and discrimination, or analysis of stress and rhythm in short passages and sentences” to practice and “pair and group discussions, role-plays, or information-gap activities” (Gordon et al., 2013: 197) to produce the newly learnt material. Pretest and posttest speech samples were collected as data for a later comprehensibility-rating task. The experiment confirmed that “adding only a relatively time-limited explicit pronunciation component in a primarily communicative classroom can lead to beneficial results in production for learners” (Gordon et al., 2013: 201) and that “the explicit phonetic instruction regarding possible miscommunication issues that come up as a result of suprasegmental problems and directed feedback on production resulted in improved comprehensibility rating” (Gordon et al., 2013: 200). A crucial conclusion is therefore this: it is key to make learners aware of the possible communication issues which can possibly arise from mispronunciation and faulty prosody.

Yenkimaleki and van Heuven (2016) posed a more specific question about explicit phonetic instruction. Their investigation focused on which one of two areas of explicit instruction teaches better listening comprehension skills for Farsi-English interpreter trainees, whether segmentals or suprasegmentals. The methodology applied in this experiment followed the same path of three English-teaching classes, out of which one group was controlled and two were experimental; all groups underwent tests before and after the experiment to assess their improvement. Similarly to the previous studies, Yenkimaleki’s and van Heuven’s results prove that explicit phonetic instruction of both segmental and suprasegmental features of English significantly improves the learners’ listening comprehension skills; what their research added to the discussion was the finding that where limited time is available for teaching, “a well-motivated choice

can be made to lend priority to the explicit teaching and practice of segmental properties of the target language, rather than to the teaching of suprasegmental (prosodic) properties” (Yenkimaleki and van Heuven, 2016: 12).

The following discussion should illuminate some concrete aspects of the acquisition of suprasegmental features of a second language. The acquisition of the suprasegmental features of stress and rhythm includes, as can be expected, a high level of L1 transfers: “typological similarity confers an advantage in acquiring an L2 stress system” (Broselow and Kang, 2013: 544). A language with a stress system – a stress-timed rhythm – has a rhythm which is based on a relatively regular occurrence of stress syllables whether they are separated by unstressed syllables or not. Learners whose L1 is a stress language tend to acquire the stress system of their L2 better, even if the rules of stress placement are new to them. Following that, the type of errors which appear in their L2 can also be inferred from their L1, just like their choice of “acoustic cues used to convey stress” (Broselow and Kang, 2013: 545). To take English as an example, the process of reduction, explained in the previous chapter of this thesis, tends to be more problematic for those learners whose L1 is not stress-timed (e.g. Spanish). However, there are tendencies which appear across the whole spectrum of learners, no matter their L1: there is a preference to placing stress on heavy syllables (CVV or CV:) rather than on light syllables (CV), especially when the vowel is long. A second universal is that vowels of higher sonority (those which are lower in height) tend to attract stress more; the correlation to their longer duration is a possible explanation for this. The acquisition of the remaining suprasegmental features – tone, pitch, accent and intonation – are harder to map, according to Broselow and Kang (2013). However, L1 transfers are evident again: one tendency is for the learners to struggle with distinguishing lexical contrasts “when the meaning of intonation contours differs in the two languages” (Broselow and Kang, 2013: 551). A second problematic area arises “from differences in the phonetic realization of pitch contours.” This can be exemplified on Greek and Dutch; although their declarative intonation included a non-final rise, they differ in where exactly they put the highest point of the rise. When producing utterances in their respective L2 languages, most speakers in the experiment placed the rise point somewhere between the points dictated by either Dutch or Greek – an interlanguage system was formed (Broselow and Kang, 2013: 551-552).

2.4. Selected features of Czech English

The Czech language is descended from the Slavic language family (along with languages like Slovak or Russian), while English is historically a Germanic language. Even though there is a significant historic interplay with German due to the geographical position of the Czech speaking territory, there is still a significant leap between phonology, morphology, syntax, and vocabulary of Czech and English. The easily noticed Czech accent in the L2 English of Czech speakers therefore does not come across as surprising, as there are numerous phenomena on both segmental and suprasegmental levels which are unique to either one or the other language. As the point of this research paper is not to map the Czech accent in English speech as a whole, but only the most salient features pertaining to it, the following aspects of Czech English will be discussed: dental fricatives /ð/ and /θ/, and velar nasal /ŋ/ for segmental level; stress placement, rhythm, intonation and pitch range for suprasegmental level.

Dental fricatives /ð/ and /θ/, orthographically marked as <th> in all possible positions, belong among those consonants which are characteristic of English and therefore do not appear in Czech. Even though Roy Major claims that dissimilar sounds are easier to learn than sounds which appear like the learner's L1 (Major, 2001; in Rumlová, 2018: 14), /ð/ and /θ/ keep proving themselves as rather problematic for non-native speakers of English. This is the case of Czech speakers, as the incorrect pronunciation of these consonants is one of the most telling characteristics of their foreign accent. What these learners tend to do is substitute English dental fricatives with Czech consonants that appear closest in pronunciation, those being the voiceless fricative [s] or even [f] and the voiced plosive [d]. This was confirmed in an extensive investigation of strong Czech accent in English by Jana Rumlová (2018). She found that the voiced dental fricative /ð/ was pronounced correctly in only 30% of the examined words and that in 50% of those the phoneme was substituted by the voiced plosive [d]. Interestingly, her data showed that in grammatical words (such as articles or pronouns) this kind of substitution amounted to 80%; lexical words were mispronounced in more than 60%, but [d] was used as a substitute only in 30% of the examined words. The remaining substitutes were [t], [th], and [s]. The success of the pronunciation of the voiceless dental fricative /θ/ was slightly better, but similar to its voiced counterpart: /θ/ was pronounced correctly in about 40% of the examined words. However, it has a tendency to be substituted by more Czech consonants than /ð/ has; the most common substitute was the voiceless plosive [t], followed by the voiceless fricatives

[s] and, in a few instances, by the voiceless fricatives [f] and [tʃ], the dental fricative [ð] and the consonant cluster [th]. The similarity between the data of the pronunciation success rate of both the voiced and the voiceless dental fricatives indicates that successful learning of one these is linked to mastering the other (Rumlová, 2018: 36-39).

The velar nasal consonant /ŋ/ forms an interesting example of a sound that is present in both Czech and English, but which has a different function in each. /ŋ/ appears and works as a phoneme in English: a word's meaning changes, depending on whether /ŋ/ or /n/ is present. This can be exemplified on the minimal pairs *sun/sung* or *kin/king*. In Czech, however, [ŋ] functions only as an allophone, meaning that it is only a positional variant which does not alter the meaning of a word. It is generally pronounced in the place of the alveolar nasal /n/ when it stands before the velar stops /k/ and /g/. This English phoneme was also studied by Rumlová. When analyzing this particular sound, she divided the words into three groups: <-ing> forms where /ŋ/ is supposed to be pronounced, words with /ŋ/ within a morpheme, and words which are pronounced with /ŋk/ (such as *bank*). Her expectations were confirmed in that 65% of <-ing> forms were pronounced incorrectly, and that this success rate is bigger than that of words with /ŋ/ within a morpheme, in which case up to 85% were mispronounced. The following tendencies were discovered: addition of /k/ or /g/ to /ŋ/ where it is not proper (words like *king* or *running*); some speakers also tend to pronounce [n] instead of /ŋ/ (Rumlová, 2018: 39-40).

Let us now proceed to suprasegmental aspects of Czech English. In the first place, stress plays a different role in Czech and English and its wrong realization can therefore have a strong impact on the perception of a foreign-accent. Stress is mostly fixed in the Czech language, generally being placed on the first syllable of the prosodic word; its function is thus only delimitative. The stressed Czech syllable has no positive prominence; some studies even suggests that the second, unstressed syllable has a higher fundamental frequency (Palková and Volín, 2003, and Volín, 2008; in Skarnitzl and Rumlová, 2019: 5). English, on the other hand, assigns stress placement a rather crucial, contrastive role and its mastery includes the following: “not only the placement and adequate acoustic realization of the stressed syllable but also, and perhaps more importantly, mastering the quality of the unstressed syllables” (Skarnitzl and Rumlová, 2019: 5). As studies have shown, Czech speakers struggle with this particular aspect of English prosody: their schwa were too prominent, although more advanced speakers approximated native durational and spectral patterns that those less advanced. Temporal and qualitative reduction is what gives English prosody its traditional rhythm and connected-speech processes, like assimilation,

coalescence, or linking, are tools through which it is achieved. As indicated in the previous discussion, Czech speakers do not always put this to practice. Numerous studies, as cited by Skarnitzl and Rumlová (2019: 5-6) along with their own research (2019: 15), have presented data showing that the Czech use of linking is very limited and, in its place, glottalization appears. Elision of initial [h] in weak forms is also problematic. Due to the absence of these reductive processes, the typical rhythmical patterning of native English speech is thus corrupted and may be perceived as a sign of accentedness.

Intonation and pitch range are much more varied in English than in Czech. This can possibly be explained by looking at the nature of the word order of these two languages. Word order in English is mostly fixed, hence its reliance on “melodic cues when expressing prominence” (Skarnitzl and Rumlová, 2019: 6) and the wide pitch range this function employs. Czech, on the other hand, deals with the expressive function differently: free word order or grammatical inflection may co-exist with, or even replace, intonation as a tool of expressing prominence (Rogerson-Revell, 2011; in Volín et al., 2015: 108). Volín et al. (2019: 115) provide an extensive research when it comes to the cross-linguistic comparison between Czech and English intonation; they found that the pitch range of L1 British newsreaders was narrower by 2 semitones than that of Czech broadcasters who were L2 speakers of English. In other words, the speech of Czech English speakers is flatter than that of native speakers.

3. Material and methodology

3.1. Compilation of material

Recordings of speech of six females and four males were chosen from the Prague Phonetic Corpus which comprises interviews with university students of English philology at the Faculty of Arts, Charles University, and the Metropolitan University of Prague. These were first year students, their age therefore varied from late teens to early twenties. The students were recorded as a required part of their course; they were asked to describe their most memorable holidays for approximately one minute. The material was chosen to satisfy the needs of the experiment. First, it was necessary to find such portions of recordings (of between 5 and 8 seconds in duration) where the speakers spoke relatively fluently. The students chosen also had to have noticeable Czech accents, especially in the target segmental and prosodic phenomena analysed here (the dental fricatives /θ, ð/ and the velar nasal /ŋ/, intonation, and temporal patterning). In other words, they could not have sounded too native-like for a foreign speaker and there had to be a good amount of the aforementioned phenomena. Once the interviews were chosen, they had to be cut into shorter passages and, in some cases, altered to create recordings in which the studied segmentals would appear in a usefully concentrated number. Both of these manipulations were done in a programme called Adobe Audition. Each full recording was cut into eight shorter passages; these had between five to eight seconds in length each. When the sounds /θ, ð, ŋ/ were scarce, different parts of an interview were “glued” together, with an attempt to make the recording sound as natural as possible. In the end, the ten initial interviews yielded 40 shorter recordings. 20 of these were then manipulated segmentally and 20 prosodically.

3.2. Manipulations

In order to change the pronunciation of dental fricatives to their correct forms, which would later on be compared with their original counterparts by the subjects, the following steps were taken. First, it was necessary to find properly sounding phonemes: to do that, an adult male and female recorded the words with dental fricatives in them which needed to be improved. The recording took place in a recording studio at the Department of Phonetics of the Faculty of Arts, Charles University. Their pronunciation aimed at the best possible form; the intention was to cut out their dental fricatives and put them in the place of the mispronounced dental fricatives produced by the speakers whose recordings were used in the experiment. Logically, the male recorded those words uttered by male speakers, and the

female recorded those words uttered by female speakers. The replacement process was performed in Adobe Audition, observing zero crossings, so that the manipulation did not result in audible artefacts. The whole process is shown in *Figure 6*, in which both the waveform and the spectrogram serve as a proof of the sound change resulting from the manipulation of the initial sound in the word *there*. The red arrow in the upper image points to a clear explosion, a type of release indicative of the sound [d]. The image below showcases a dental fricative put in the place of the original [d]; this time, the red arrow points to the part of the wave indicating an exemplary friction of /ð/.

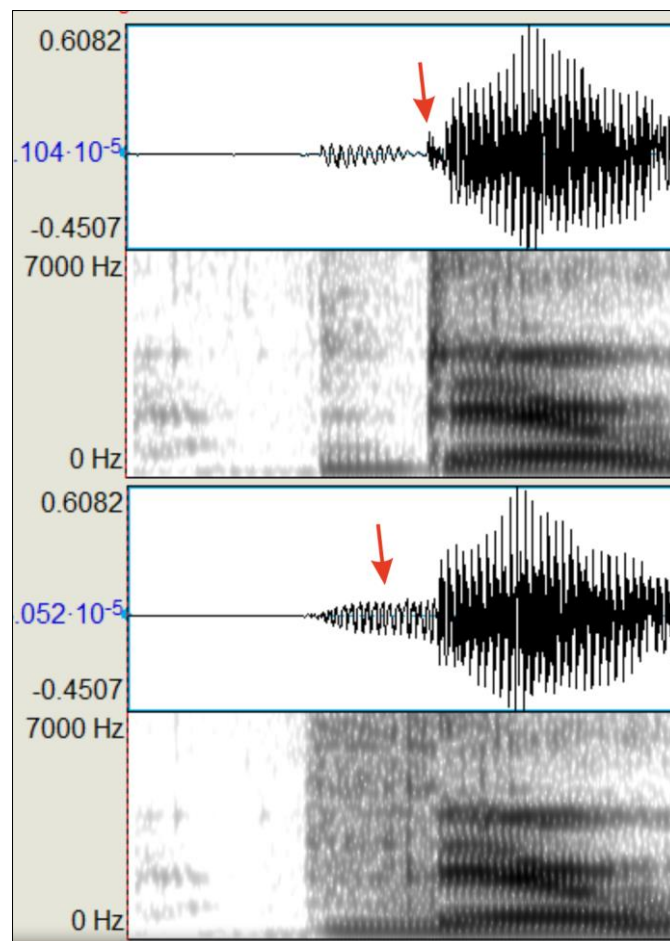


Figure 6: An example of replacing /d/ with /ð/

This, of course, had to be managed in the most natural sounding manner possible; in a case of an easily perceived manipulation, a better result was generally achieved by lowering the volume of the inserted sound, by changing its fundamental frequency closer to the original, or by lowering the amplitude in high frequencies. A problem arose with manipulations of recordings marked as *M1* and *M4* (male speakers): another male speaker had to be asked to come in and record the correct pronunciation of their dental fricatives again, because the voice in the first set of improved recordings was too low and its insertion

could not be masked. This solution proved as appropriate and the final manipulations of this type appeared without a problem.

Improving the mispronounced velar nasal /ŋ/ was a more straightforward process. As we were looking specifically at the pronunciation of this phoneme at the end of a morpheme, which is mostly realized at the end of a word, the task was simply to find the suffix <-ing> pronounced incorrectly, that is with both the nasal and the plosive sound. This time, it was the plosive which got deleted. *Figure 7* shows an example of such deletion of [k]: the red arrows in the upper image point to that part of the wavelength which is indicative of the sound of this plosive and the explosion which goes with it can be seen in the spectrum below (the two thin black lines). The image below shows the recording after a manipulation: the red arrows point to those spots which lack the prominent [k] wavelength. This faulty segment was again deleted after careful selection, noting the zero crossings of amplitude

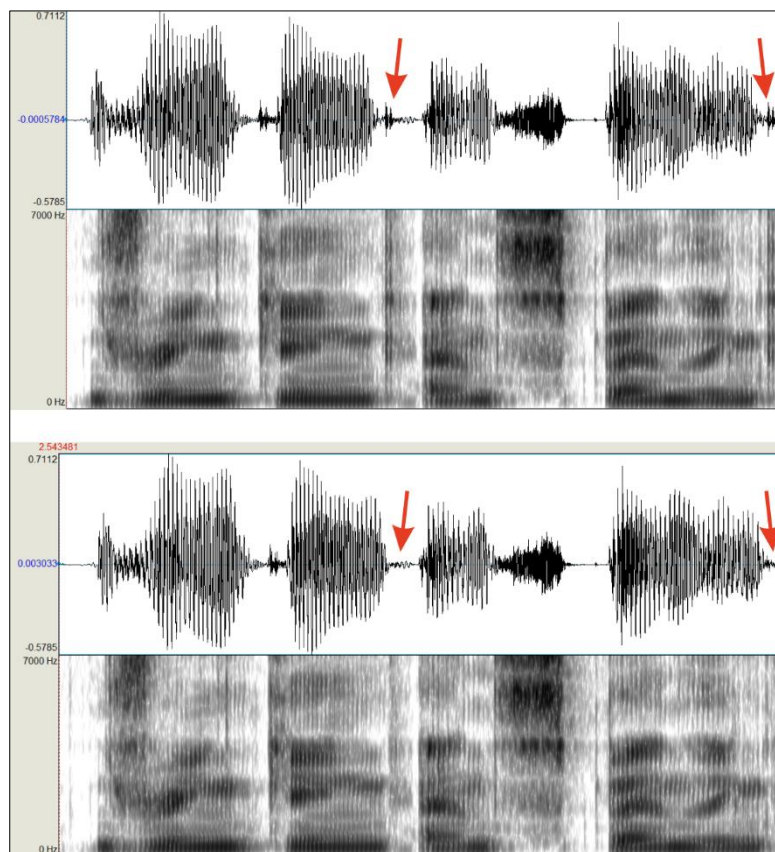


Figure 7: An example of [k] deletion

Suprasegmental manipulations took place in a programme called Praat. A sound file was opened as a Praat object and then converted into a manipulation project, using the function “To Manipulation,” with the set-up left as offered by the programme (time step: 0.01 s; minimum pitch: 75 Hz; maximum pitch: 600 Hz). The new object called *Manipulation F1-pros1-man* (an example) was then opened to be viewed and edited. The

first step in this new window was to remove all the original pitch points in order to create an environment where pitch of our own choice could be placed instead. All this happened in the field marked as “Pitch manip,” which can be seen in *Figure 8*. The point of these manipulations was to create a livelier pitch range, as Czech speakers of English tend to speak rather flatly; pitch points were thus added to the contours of the original melody (marked in grey in *Figure 8*), changing it where necessary. To prevent unnatural sound, it was important to place the new pitch points in the middle of the vocalic centre of the syllables. The final pitch point, marked red at 5.75 s, was placed as it was in order to simulate a substantial fall in intonation, which typically appears at the end of an utterance to mark its termination. To observe the traditional intonation pattern, this pitch point was placed just before the start of the last stressed syllable of the unit, the word “do” in this case. The blue arrows point to fall-rise tones, which were added to the words “back” and “unfortunately” – by placing these tones in this way, semantic emphasis was placed there as well, guiding the listener to what is important in the utterance and preventing a flat manner of speaking.

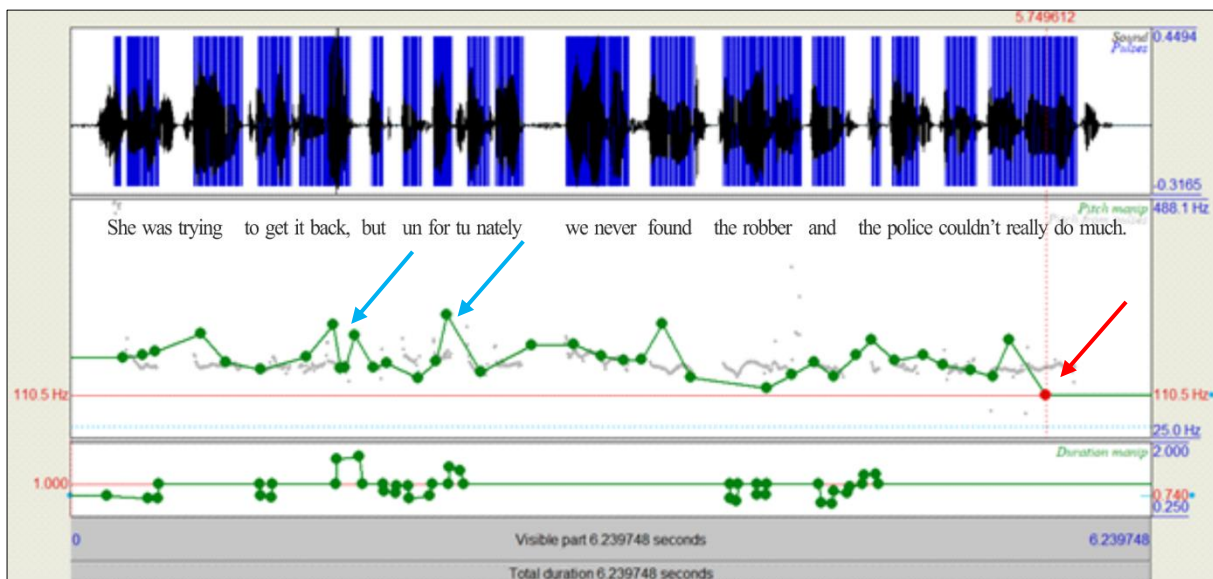


Figure 8: An example of a pitch manipulation, which aimed for a fall-tone at the end of an utterance.

Figure 9 below exemplifies another source of motivation to manipulate the speaker’s prosody in a specific manner: our aim was to put stress on the word “all.” The pitch point created for this word is placed noticeably high in comparison with the surrounding points. This produces a significant fall-rise tone in the speech; given that

intonation is one of the traits typical for stressed syllables, the word “all” gained the prominence we intended.

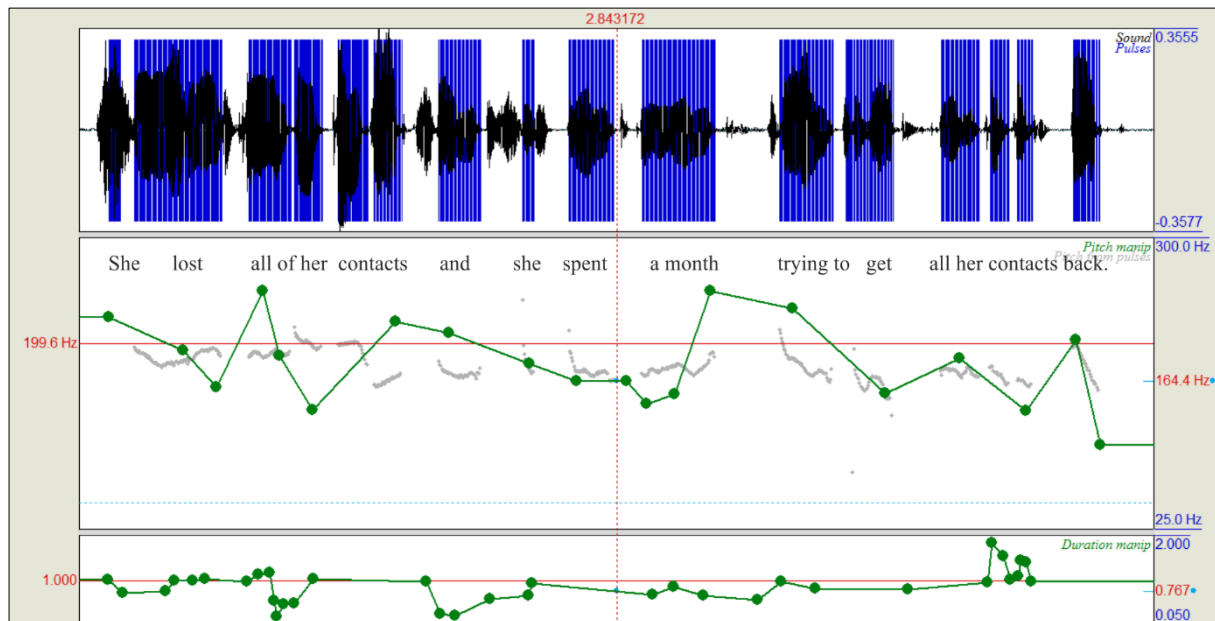


Figure 9: An example of a pitch manipulation, which aimed for putting a stress on a chosen syllable

Suprasegmental temporal manipulations took place in the identical manipulation file in Praat as prosodic manipulations did. Temporal manipulation for *F1-pros1-man* were therefore created also in the file *Manipulation F1-pros1-man*. This time, however, the manipulations were prepared in the field marked as “Duration manip,” located below the field mentioned previously and visible in *Figure 10*. Temporal manipulations consisted of either prolonging or shortening a sound – we changed the pace of the speaker’s recording. A part of the speech (a segment, a syllable, or even a whole word) needed to be lengthened in order to either make a stress placed on that syllable more noticeable, or to put the stress there in the first place, since length is one of the markers of stress. Another reason behind these manipulations was to improve the overall flow of the speech to achieve better fluency. In order to do that, two duration points were added and shifted above the line marked with number 1 and indicating the original temporal set-up. The higher this temporal “dent” was placed above this line, the longer that unit of sound became. Similarly, when a unit of sound needed to be shorter – in order to remove a wrongly placed stress, or to reduce the chosen sound to achieve a more natural, connected speech – the same process was repeated, only the duration points were placed below the original temporal line. Again, this process is demonstrated at a concrete example in *Figure 10*. The words “didn’t” and “any” at the beginning of the utterance needed more prominence in order to satisfy their stressed

character. Adding to the pitch points created previously in pitch manipulations, the first syllables in “didn’t” and “any” were also made longer; this is exemplified by the upward-pointing dents in the temporal outline of the speech. The following word “energy,” on the other hand, needed to be shortened. Originally, the speaker pronounced the second syllable with a full /e/ instead of a schwa and, as it is actually a weak syllable, the sound had to be reduced. A visible downward-pointing dent was thus created in order to make the duration of the syllable “-ner-” shorter.

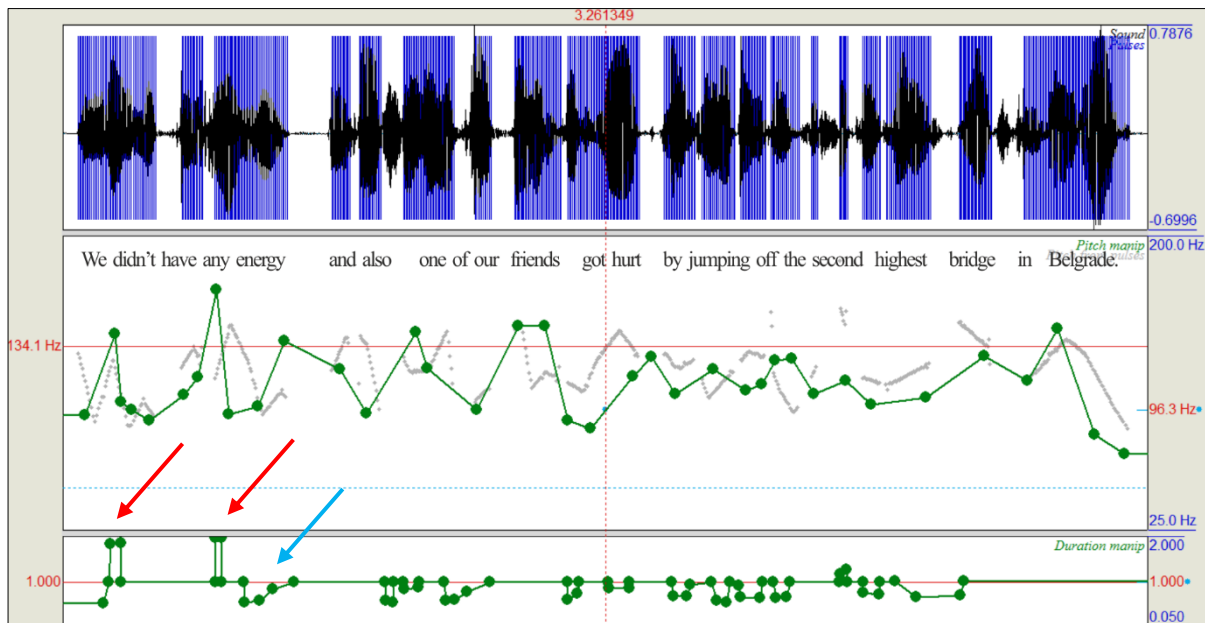


Figure 10: An example of a duration manipulation, which aimed at either prolonging a sound (red arrows) or shortening them (blue arrows).

Once both pitch and temporal manipulations were completed, the manipulation project was extracted through the function “Publish resynthesis.” This new file, called *Sound from Manipulation Editor*, was then saved as a WAV file.

3.3. Perception test

The test itself was written in Praat’s MFC setting. The final design of the test screen is pictured in *Figure 11*. A key step was to create an environment which would enable the subjects to evaluate these two aspects of each of the phrases they were going to hear: how much accented the English of the speaker was, and how easy it was for them to understand that speaker. Two 5-interval scales were chosen for this purpose. The first 5-interval scale asked the subjects to evaluate the extent of accentedness in recordings; interval 1 stood for an “almost native-like” accent, interval 3 for a “moderate accent,” and interval 5 for a “very strong accent”. The second scale asked the subjects to evaluate the extent of comprehensibility of the phrase; similarly, interval 1 stood for “no problem understanding,”

interval 3 for “moderately difficult,” and interval 5 for “very difficult.” In both scales, intervals 2 and 4 were placed as unlabeled choices in case the subject’s answer required less clarity. These two scales were placed in the middle of the screen and highlighted in colour. First, the question about accentedness needed to be answered, upon which the selected choice became marked with a different colour. Following that, the question about comprehensibility had to be answered and the process was repeated. To make sure that each subject could listen to the phrase properly, an option of three repeats was offered with a separate button in the lower-left corner. Once both questions were answered and the subject desired to continue, a button marked as “OK” was placed in the lower-right corner for them to click-on.

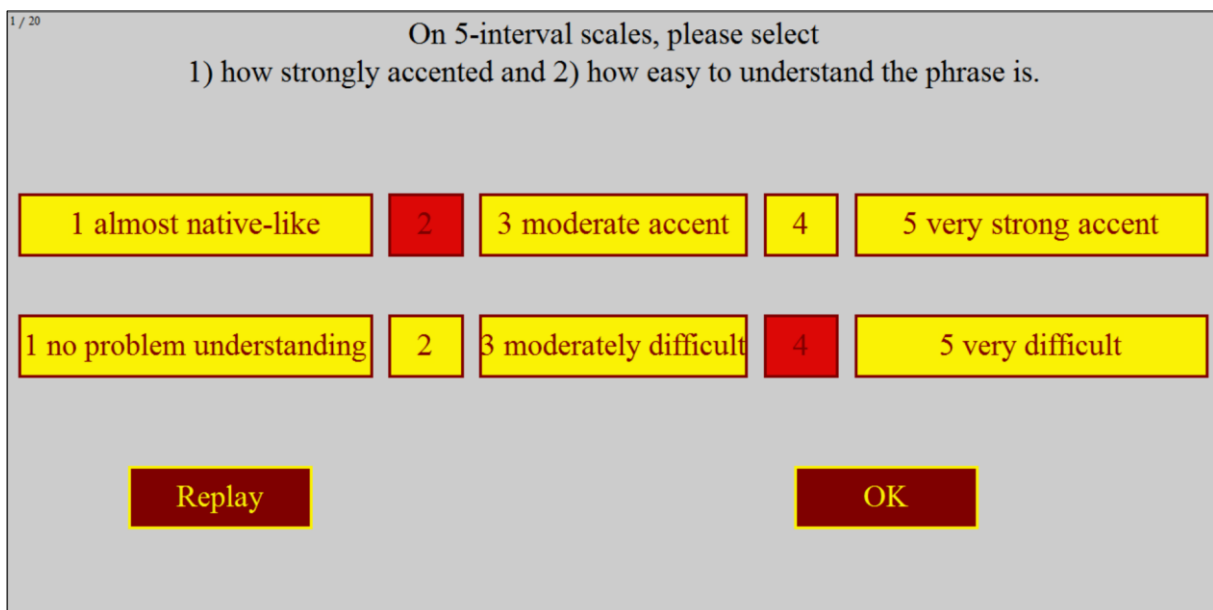


Figure 11: The final design of the test-screen

The second important choice concerned the assignment and the sequence of the phrases as they would play during the test. It was decided that two versions of the test would be run and the total number of 40 phrases was therefore divided into two sets of 20 phrases, each set assigned to either *Test 1* or *Test 2*. The first ten items of the first set were selected randomly; the remaining thirty phrases were sequence with regards to the following logic. The two tests, practically speaking, mirrored each other. Where in *Test 1* a phrase called *F1-pros-man* was placed, *Test 2* used *F1-pros-orig*. In the same spot of the sequence, the two tests used phrases focused on the same level of speech (segmental or prosodic), but which differed in whether or not they were manipulated. Secondly, within each set of phrases, the first half of the batch was opposite to the second half. *F1-pros-man* was the first item in the first set of phrases; *F1-seg-orig* was therefore put as the eleventh item,

because it differed from the the first item in both its focus and whether or not it was manipulated. The final set-up of the phrase sequences for the two tests is showed in *Figure 12* below:

Item n.	Test 1	Test 2
1	<i>F1-pros-man</i>	<i>F1-pros-orig</i>
2	<i>M1-seg-orig</i>	<i>M1-seg-man</i>
3	<i>F2-seg-orig</i>	<i>F2-seg-man</i>
4	<i>F3-pros-man</i>	<i>F3-pros-orig</i>
5	<i>M2-seg-orig</i>	<i>M2-seg-man</i>
6	<i>F4-pros-man</i>	<i>F4-pros-orig</i>
7	<i>F5-seg-man</i>	<i>F5-seg-orig</i>
8	<i>M3-pros-man</i>	<i>M3-pros-orig</i>
9	<i>F6-seg-orig</i>	<i>F6-seg-man</i>
10	<i>M4-pros-orig</i>	<i>M4-pros-man</i>
11	<i>F1-seg-orig</i>	<i>F1-seg-manip</i>
12	<i>M1-pros-man</i>	<i>M1-pros-orig</i>
13	<i>F2-pros-man</i>	<i>F2-pros-orig</i>
14	<i>F3-seg-orig</i>	<i>F3-seg-man</i>
15	<i>M2-pros-man</i>	<i>M2-pros-orig</i>
16	<i>F4-seg-orig</i>	<i>F4-seg-man</i>
17	<i>F5-pros-orig</i>	<i>F5-pros-man</i>
18	<i>M3-seg-orig</i>	<i>M3-seg-man</i>
19	<i>F6-pros-man</i>	<i>F6-pros-orig</i>
20	<i>M4-seg-man</i>	<i>M4-seg-orig</i>

Figure 12: Item sequences

Lastly, provisions had to be made in order to avoid confusion or a lack of understanding of the task on the part of the subjects. An initial slide with clear instructions was therefore provided. It also included an explanation of the relationship between accentedness and comprehensibility. The test itself was preceded by two trial items to allow the subjects to familiarise themselves with the design of the test. Once the trial items were completed, the instructional slide appeared again, after which the real test began.

3.4. Experiment

The experiment took place in a quiet environment in the offices of the Department of Phonetics at the Faculty of Arts, Charles University in Prague. The recordings were played to the subjects via identical closed headphones (Sennheiser HD-201) and the subjects were asked to label these recordings using two sets of scales presented to them in the program specifically created for this task. Each subject was provided with as much time as they deemed necessary to make their choices regarding the evaluation of the recordings.

After each subject finished their test, the results were generated and saved as a .ResultsMFC file. Finally, all these files were pooled to a single table.

3.5. Analysis

First, comprehensibility and accentedness scores were tested and analysed. The analyses we ran were conducted in R (R Core Team, 2017), and we applied linear mixed-effects (LME) modelling to assess the influence of various factors on comprehensibility and accentedness scores, using the *lme4* package (Bates, Maechler, Bolker & Walker, 2015). Comprehensibility and accentedness were the dependent variables which changed with regards to the independent variables; these fixed factors (independent variables) included whether or not the items used in the experiment were manipulated or original (manipulation factor), and whether the manipulation level was segmental or suprasegmental (level factor). Two random effects were included in the analysis – speaker and listener – to control for the fact that individual respondents were likely to differ in their sensitivity to different statements. The significance of individual effects or interactions was tested by comparing a full model, which included the factor/interaction in question, to a reduced model, in which the given factor/interaction was excluded; these came out as significant, if the difference between the full and the reduced models was significant. Standard likelihood ratio tests were used for the evaluation.

Tukey posthoc comparisons were also run; they were conducted using the R package *multcomp* (Hothorn, Bretz & Westfall, 2008), again with the inclusion of the speaker and listener random effects. These tests aimed at deducing the extent of the effect of sound manipulations within both the segmental and the prosodic level. This time, therefore, we did not analyse the difference between these two levels, but instead compared four concrete groups in a pairwise manner: segmental original items, segmental manipulated items, prosodic original items, and prosodic manipulated items. The comparisons were run for both comprehensibility and accentedness scores.

Finally, analysis of individual items was conducted to find concrete patterns or exceptional behaviour of comprehensibility and accentedness scores.

4. Results and discussion

4.3. Comprehensibility scores

First, comprehensibility scores were analysed using LME analysis. The following results were obtained to find out to what extent can the results be generalized, as opposed to the difference being a result of random effects. The results are interpreted bearing in mind that $p < 0.05$ is the limit of significance. The effect of level on comprehensibility scores is highly significant: $\chi^2(1) = 57.25, p < 0.0001$. The effect of manipulation on comprehensibility scores is significant: $\chi^2(1) = 55.39, p < 0.0001$. The interaction between the level and the manipulation factors is significant: $\chi^2(1) = 9.54, p < 0.01$. Clearly, manipulations affected comprehensibility scores of the listeners more significantly if they took place on the level of prosody (the suprasegmental level). Segmental manipulation affected the listeners' evaluation of the speakers' comprehensibility much less. The most important conclusion that this analysis offers, however, is that comprehensibility scores are made worse by manipulations. This is indicated by the higher scores of the manipulated items, since the evaluation scale considered number 1 as "no problem understanding" and number 5 as "very difficult" to understand.

Tukey posthoc comparisons were also run and aimed at finding out the extent to which manipulations affected the scores. Four groups were compared in a pairwise manner: segmental original items, segmental manipulated items, prosodic original items, and prosodic manipulated items. *Figure 13* depicts the interactions between the individual factors within the comprehensibility level. The comparison test yielded the following result to express the significance of manipulations within the segmental level: $p = 0.0542$; this result indicates only a marginal significance. Comprehensibility of non-native speakers of English was therefore minimally affected by segmental manipulations. To map the effect of prosodic manipulation on comprehensibility, the analysis produced the following result: $p < 0.001$; this, on the other hand, implies a high significance. Comprehensibility was therefore highly affected by prosodic manipulations of intonation and the temporal aspect of speech.

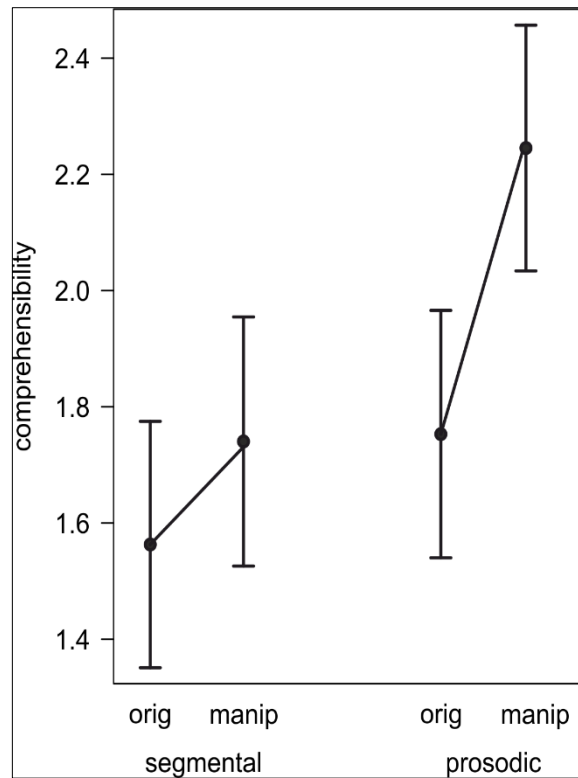


Figure 13: Interaction between the level and the manipulation factors within comprehensibility

4.4. Accentedness scores

In the second place, accentedness scores were also tested and analysed using LME analysis. The same analyses were used as with comprehensibility scores, but the dependent variable was changed to accentedness. The following results were obtained. The effect of level on accentedness scores is significant: $\chi^2(1) = 6.19, p < 0.05$; this is much less compared to the effect of level on comprehensibility, which is highly significant ($p < 0.0001$). The effect of manipulations on accentedness scores is highly significant: $\chi^2(1) = 14.51, p < 0.001$; this result is broadly comparable to the same information about comprehensibility, even though the numbers are again smaller, which shows a smaller impact of manipulations on the evaluation of accentedness than on the evaluation of comprehensibility. The interaction between level and manipulation is not significant: $\chi^2(1) = 0.25, p > 0.5$; this is a major difference from the data obtained from the analysis of comprehensibility, in which case the interaction is significant. Segmental manipulations seem to have a slightly bigger effect on accentedness than prosodic manipulations, as the average accentedness score for segmentally manipulated items is higher than that of prosodically manipulated items. However, the biggest difference from the

comprehensibility level is the opposite trend in the direction of the effect of the manipulations: accentedness scores seem to have an improved tendency after the manipulations, as the scores of manipulated items is lower (hence better) than the scores of originally sounding items.

Tukey posthoc comparison was run for accentedness as well; the results are graphically summarized in *Figure 14*. The comparison between original and segmentally manipulated items and the overall effect of segmental manipulations on the evaluation of accentedness is expressed by the equation by the following significance: $p < 0.05$; this value implies that the effect of segmental manipulations on accentedness was significant. Segmental manipulations therefore influence accentedness more than they do comprehensibility. The extent to which prosodic manipulations affected accentedness is summarized in the following result: $0.1106 > 0.1$; there is no significance.

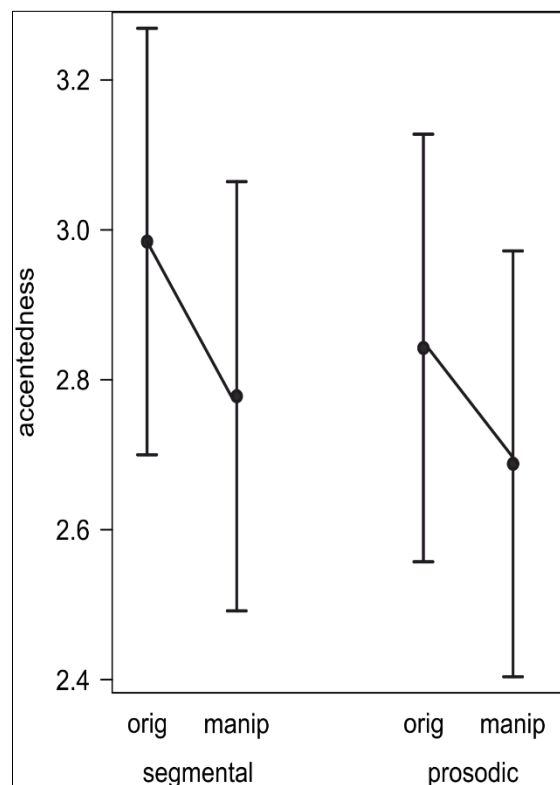


Figure 14: Interactions between the level and the manipulation factors within accentedness

4.5. Item analysis

Individual items were also analysed; this enabled us to look for interesting patterns or for exceptions to the general trends within both comprehensibility and accentedness scores. To represent average scores for comprehensibility and accentedness of individual

items in the test, effect plots showing mean fitted values and the respective confidence intervals were created; they were constructed using the *effects* package (Fox, 2003). *Figure 15* summarizes the data for average scores for comprehensibility, while *Figure 16* does so for accentedness. Originally-sounding items are marked in black; manipulated items are marked in yellow. Looking at the rough outline of the data, we can observe that items concerning the differences between the evaluation of the original and the manipulated items was more concise within the category of the segmental level; the differences between the two types of items are not as significant as those in the category of prosodic level, in which big leaps in evaluation can be seen. Accentedness, with a few exceptions, generally tended to improve with segmental manipulations and to mostly improve with prosodic manipulations. In general, comprehensibility was more affected by both types of manipulations. Again, this data offers the following interpretation of aspects of non-native English: the ability of being understood comes out as more important than accent alone.

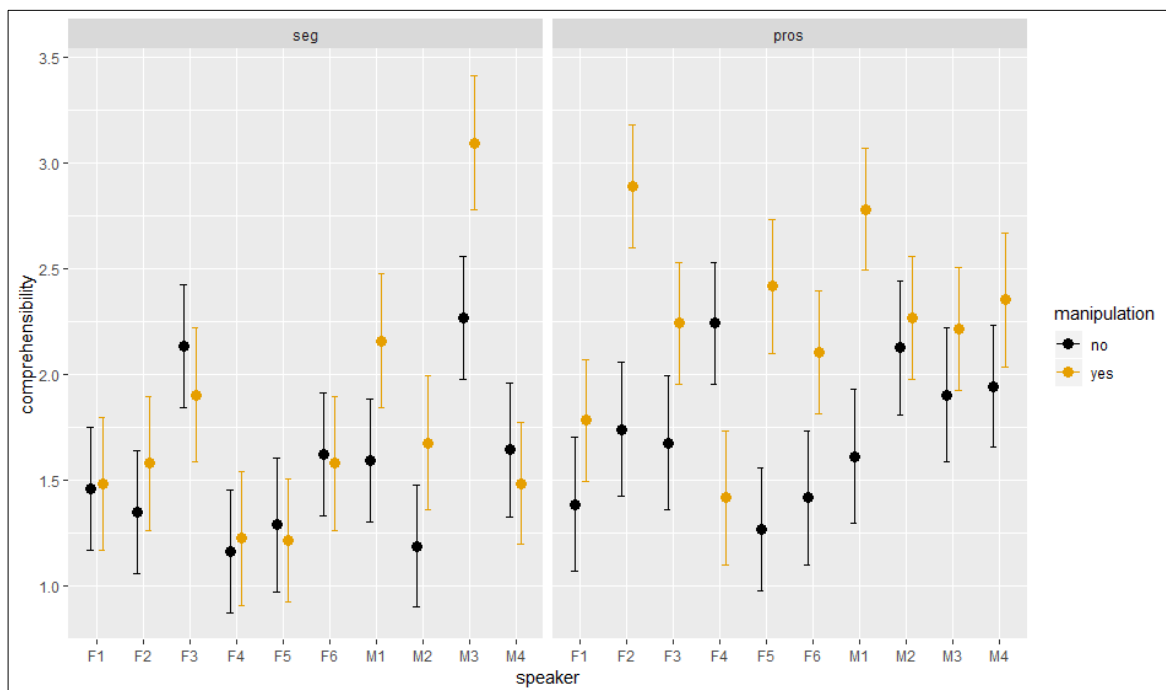


Figure 15: An effects plot showing mean fitted values and the respective confidence intervals for accentedness

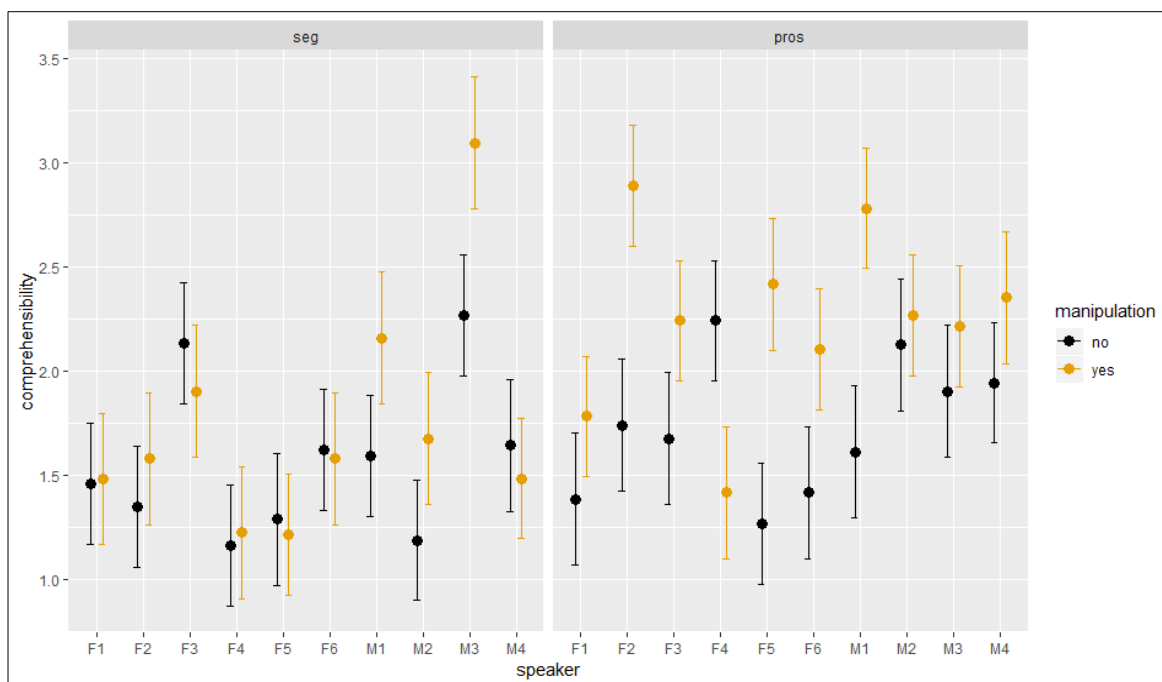


Figure 16: An effects plot showing mean fitted values and the respective confidence intervals for comprehensibility

The following discussion focuses on singularly interesting phenomena which stand out in the results.

Speaker *F2* shows opposite scores within the different categories. In the group of segmental manipulations, her comprehensibility score worsened with the manipulated item (she was assessed as more difficult to understand), but her accentedness improved (she was assessed as more native-like). In the group of prosodic manipulations, both comprehensibility and accentedness scored worse when manipulated. This is especially the case with comprehensibility, in which category her score became worse by roughly 1.2 points on the scale; this is the biggest difference out of all the categories and all the items. Having listened to the prosodically manipulated item, it has to be added that this item is one of those which do not sound entirely natural and are more likely to sound conspicuous; this nature of this item can therefore be a possible explanation of the much worse score of comprehensibility. It can be assumed that the same problem stands behind the whole tendency of worse comprehensibility scores of prosodically manipulated items, as all of them follow the same trend.

Item *F4* is a special case, as it was the only item which was manipulated to sound worse; the data obtained from the experiment were reversed so that the results would not confuse the whole set of results. The black dots therefore represent the manipulated item

(made worse) and the yellow dots represent the original item (segmentally and prosodically better). In terms of comprehensibility, the segmentally manipulated item scored better than the original item; this came out as rather unexpected, since the general trend with segmentals and accentedness is that of improvement. Accentedness was assessed as less native-like with items manipulated both segmentally and prosodically. Still, even though the item was manipulated to sound worse, it still scored better than most of the other manipulated items; this is probably due to the fact that the speaker spoke a very good English overall.

Data describing items recorded by male speakers (*M1-M4*) look much more uniform, with one exception: item *M1*, which tends to score worse than the rest within the male group. In terms of accentedness and the segmental view, *M1* was assessed as the worst item in the batch, with the manipulated item scoring better than the original. However, prosodically it did much better, scoring along the same lines as the remaining male speakers; still, manipulation again improved its score by the biggest difference in the male set. When evaluating comprehensibility within the segmental level, listeners assessed *M1* better when they heard the original item; this does not come as surprising, because the division between improved and worsened score is roughly the same within this category. With prosodic manipulations and comprehensibility, the manipulated item scored much worse than the remaining speakers, whose manipulated items also score worse, but by a smaller difference. Having listened to the prosodically manipulated item *M1*, a possible explanation arises: the speech in this item is much faster and parts of it sound rather robotic. The original item was hard to manipulate due to many hesitation phenomena, a very flat intonation, and the occurrence of creaky voice; due to these, many different parts of the original interview were glued together. This probably resulted in the unnatural sound of this item, which became more prominent with the subsequent manipulations. Compared to *M1*, items *M2*, *M3*, and *M4* consisted of a single utterance and the original bases for manipulations were more fluent. This is one explanation for the similarity between these three items, and for the divergence of *M1*.

5. General discussion

The data obtained from LME analysis, Tukey posthoc test, and item analysis showcase several trends. Let us first discuss comprehensibility. The overall effect of manipulations on comprehensibility and the interaction between the level of manipulation (segmental or prosodic) and manipulations themselves (whether they took place or not) is significant. In other words, the evaluations of the speakers by the participants of the test were influenced by our manipulations, i.e. their scores changed depending on whether or not the item they were listening to was manipulated. This result continues the trend set up by previous studies introduced in the Theoretical Background section of this thesis: the difference between various skill levels of pronunciation and between various prosodies are noticeable to listeners, who consequently evaluate them differently. Manipulations of both segmentals and suprasegmentals led to a worsening of comprehensibility evaluations of the items, meaning they were evaluated as more difficult to understand than the original items. This effect was stronger with prosodic manipulations. This occurrence has two possible explanations. First, the speakers could have sounded more native-like; they therefore lost the familiarity with the mother tongue of the listeners, who consequently comprehended the foreign speech with more difficulty. However, the results have not established a causative relationship, so this theory ought to be abandoned. Instead, the negative impact of manipulations, especially prosodic manipulations, is probably due to the insufficient quality of our sound manipulations. While this was not a problem with segmental manipulations, which actually led to some improvement in evaluations, prosodic manipulations proved problematic. We have struggled with a rather robotic tinge of the altered recordings and, despite attempts to prevent it, the subjects probably subconsciously noticed this and, in consequence, felt difficulty understanding the items. The negative influence of our prosodic manipulations seems to be confirmed by the Tukey posthoc test, which shows that comprehensibility is only a little affected by segmental manipulations, but more so by the less successfully manipulated prosodic ones.

With accentedness, we can observe a different kind of trend. The listeners' evaluations were less affected by manipulations. Unlike comprehensibility, the small effect the manipulations did have was to subjectively improve the sound of the speakers' accents. In other words, those items in which segmentals were corrected to their proper pronunciation, were assessed as more native-like than their original counterparts with incorrect pronunciation. However, it is important to remember that accentedness was not as affected as comprehensibility was and the success is therefore limited. With regards to the influence of the type of manipulation on accentedness, the posthoc test shows that accentedness is significantly

affected by segmental manipulations, but not affected at all by prosodic manipulations; the trend is therefore opposite to that of comprehensibility. This shows just how much accent is tied to the segmental level and less so to prosody. Again, however, the lack of influence on the side of prosody may be connected to the insufficient quality of this type of manipulations, as discussed previously.

Our hypothesis – that segmental and prosodic manipulations would improve the evaluations of comprehensibility and accentedness – was therefore only partially confirmed by this experiment. Accentedness improved in connection with the segmental level; comprehensibility generally worsened. Still, the experiment has some practical implications. The important fact to take away from this discussion is that comprehensibility is the dimension of foreign-accented speech that seems to matter more to the listeners of such non-native speech and that the ability to be understood without major difficulties is therefore the key to a good evaluation by others; based on the data, the ability to be understood matters more than mere accent. In other words, the style of the individual sounds (segments) is less important in understanding a non-native speech than the overall presentation of the speech, including intonation, stress and duration. Although our manipulations had a negative effect, they still show that prosody matters more when talking about the dimension of comprehensibility. The teaching of a second language should therefore focus on both segmental and suprasegmental aspects of the language in question and it ought to stress the importance of successful communication, whether it is packaged in a native-like accent or not. Explicit phonetic instruction introduced in Theoretical Background seems like a feasible option for this.

In the specific case of Czech English, it would be advisable to investigate the matter of accentedness and comprehensibility further. Given our struggle with the quality of prosodic manipulations, the improvement of the technique and possibly finding better material should especially be stressed. Including the dimension of intelligibility, the objective parameter of understanding a foreign utterance, would be an interesting addition.

6. Conclusion

We have begun this research with the aim of examining the effects that segmental and prosodic manipulations had on the evaluation of the speaker's accentedness and comprehensibility. The study took the form of an experiment. In this experiment, we played both original and manipulated items produced by the identical speaker to see whether or not, and to what extent would the evaluations of their accentedness and comprehensibility change. We expected an improvement in both these areas.

The theoretical part of this thesis introduced the three dimensions used for description of non-native speech: accentedness, comprehensibility, and intelligibility; this subsection also presented the previous research of these markers. The other topics covered in Theoretical Background included a discussion of social aspects of communication in a foreign language and a description of selected key features of native English and English produced by Czech speakers. The methodological section then presented a description of the process of material selection, sound manipulation, preparation of the perception test, the experiment itself, and of analysis. The results have shown that segmental and prosodic manipulation generally affect evaluation of the speakers, but our hypothesis was confirmed only partially. There seems to be a tendency of improvement of accentedness, especially by segmental manipulations, while the effect of prosodic manipulations was close to none. Comprehensibility scores were both quite significantly affected by both types of manipulations, but more so by prosodic manipulations, which tended to worsen the scores. It may be concluded that segmental manipulations were generally more successful in improving scores, while prosodic manipulations need to be improved in the future.

7. References

- Atagi, E. and Bent, T. (2011). Perceptual Dimensions of Nonnative Speech. In Proceedings of 17th ICPHS, pp. 260-263.
- Bates, D., Maechler, M., Bolker, B. & Walker, S. (2015). *lme4: Linear Mixed-effects Models Using Eigen and S* [R package version 1.1-8]. Retrieved from <http://CRAN.R-project.org/package=lme4>.
- Boersma, P., Weenink, D. (2016). *Praat: Doing Phonetics by Computer* (version 6.0.50). Retrieved from www.praat.org.
- Broselow and Kang. Phonology and Speech. In: Herschensohn, J. & Young-Scholten, M. (Eds.), *The Cambridge Handbook of Second Language Acquisition*, pp. 529-553. Cambridge: Cambridge University Press.
- Derwing, T.M. and Munro, M.J. (1995). Foreign Accent, Comprehensibility, and Intelligibility in the Speech of Second Language Learners. *Language Learning*, 45(1), pp. 73-97.
- Derwing, T.M. and Munro, M.J. (2009). Putting accent in its place: Rethinking obstacles to communication. *Language Teaching*, 42(4), pp. 476-790.
- Derwing, T.M. and Munro, M.J. (2015). *Pronunciation Fundamentals: Evidence-based Perspectives for L2 Teaching and Research*. Amsterdam, Philadelphia: John Benjamins Publishing Company.
- Dovidio, J. F. and Gluzsek, A. (2010). The Way They Speak: A Social Psychological Perspective on the Stigma of Nonnative Accents in Communication. *Personality and Social Psychology Review*, 14, pp. 214-237.
- Fox, J. (2003). Effect displays in R for generalised linear models. *Journal of Statistical Software*, 8, 1-27.
- Gordon, J., Darcy, I., Ewert, D. (2013). Pronunciation teaching and learning: effects of explicit phonetic instruction in the L2 classroom. In Proceedings of the 4th Pronunciation in Second Language Learning and Teaching Conference, pp. 194-206.
- Hothorn, T., Bretz, F. & Westfall, P. (2008). Simultaneous inference in general parametric models. *Biometrical Journal*, 50, pp. 346-363.
- Kang, O., Rubin, D., Pickering, L. (2010). Suprasegmental Measures of Accentedness and Judgments of Language Learner Proficiency in Oral English. *The Modern Language Journal*, 94(4), pp. 554-566.
- Lawson, E., Stuart-Smith, J., Scobbie, J. M., Nakai, S. (2018). *Seeing Speech: an articulatory web resource for the study of Phonetics*. University of Glasgow. 25th April 2019. <https://www.seeingspeech.ac.uk/>
- Meierkord, Ch. (2013). English as a Lingua Franca. In: *The Encyclopedia of Applied Linguistics*, pp.1-7. Oxford, UK: Blackwell Publishing Ltd.
- Mennen, I., and de Leeuw, E. (2014). Beyond segments: Prosody in SLA. *Studies in Second Language Acquisition*, 36, pp. 183-194.
- R Core Team (2017): *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing. Retrieved from <https://www.r-project.org/>.
- Roach, P. (2009). *English Phonetics and Phonology: A Practical Course*. Cambridge: Cambridge University Press.
- Rumlová, J. (2018). Phonetic features of strong Czech accent in English. Prague: Faculty of Arts. Unpublished BA thesis.
- Saito, K., Trofimovich, P., Isaacs, T. (2016). Second language speech production: Investigating linguistic correlates of comprehensibility and accentedness for learners at different ability levels. *Applied Psycholinguistics*, 37, pp. 217-240.

- Skarnitzl, R., Rumlová, J. (2019, in print). Phonetic aspects of strongly-accented Czech speakers of English. *AUC Philologica*.
- Varonis, F. and Gass, S. (1982). The Comprehensibility of Non-native Speech. *Studies in Second Language Acquisition*, 4, pp. 114-136.
- Varonis, F. and Gass, S. (1984). The Effect of Familiarity on the Comprehensibility of Nonnative Speech. *Language Learning*, 34(1), pp. 65-89.
- Volín, J., Poesová, K. & Weingartová, L. (2015). Speech melody properties in English, Czech and Czech English: Reference and interference. *Research in Language*, 13, pp. 107–123.
- Wickham, H. (2009): *ggplot2: Elegant Graphics for Data Analysis*. New York: Springer.
- Yenkimaleki, M., and van Heuven, V. J. (2016). Explicit teaching of segmentals versus suprasegmentals: Which would yield better listening comprehension skills for interpreter trainees? An experimental study. *British Journal of English Linguistics*, 4(6), pp. 11-22.

8. Resumé

Tato bakalářská práce se zaměřuje na výzkum vlivu segmentálních a prozodických manipulací na sílu cizineckého přízvuku a pocitu srozumitelnosti. Konkrétně jsme se zaměřili na přízvuk českých mluvčích angličtiny a manipulací jejich mluveného projevu jsme se pokusili odstranit chyby, které se v české angličtině běžně vyskytují. Mezi tyto jevy se zpravidla řadí chybná výslovnost dentálních frikativ, přítomnost exploziv v koncové pozici za velární nazálou, špatné umístění přízvuku a plochá intonace. V rámci výzkumu byl měřen rozdíl manipulované a nemanipulované položky jednoho mluvčího.

V teoretické části (kapitola 2) této práce jsou nejprve představeny obecné parametry, pomocí nichž lze popisovat charakter cizineckého přízvuku. Tento systém se užívá jednotně ve studiích tohoto typu a je tedy vhodným kandidátem i k popisu českého přízvuku v angličtině. Jedná se o tři konkrétní dimenze popisu: síla cizineckého přízvuku, pocit srozumitelnosti, a porozumění. Síla cizineckého přízvuku popisuje míru přítomnosti cizineckého přízvuku v dané promluvě. Pocit srozumitelnosti se používá ke zhodnocení obtíže, se kterou posluchač projevu porozuměl; jedná se o čistě subjektivní ohodnocení, které se skutečným poznatkem obsahu řeči nesovisí. K tomu slouží třetí parametr, porozumění, které hodnotí, do jaké míry posluchač opravdu pochopil to, o čem mluvčí rozpravuje. Tyto tři faktory jsou propojené, jeden na druhém ale nezávislé. Studie z minulosti ukazují, že pocit srozumitelnosti a porozumění spolu souvisí, stejně jako pocit srozumitelnosti a síla cizineckého přízvuku. Vysoká míra cizineckého přízvuku nemusí nutně ovlivnit pocit srozumitelnosti, a posluchači mohou tedy porozumět i projevu se silným přízvukem. Podkapitola 2.2. na tuto diskuzi navazuje popisem sociálních aspektů komunikace s cizineckým přízvukem, tedy popisem toho, jakým způsobem přízvuk ovlivňuje roli mluvčího ve společnosti. Mezi malým množstvím výhod (v porovnání s nevýhodami) řeči s přízvukem je funkce signalizace, která rodilým účastníkům komunikační situace říká, že je třeba přizpůsobit úroveň svého projevu úrovni nerodilého mluvčího. Druhou, ojedinělou výhodou je stereotypizovaná sofistikovanost konkrétních cizineckých přízvuků (například východní přízvuk ve světě hudby). Bohužel, společenské nevýhody zatím převažují. Cizinecký přízvuk hraje roli při utváření identity v osvojovaném jazyce, na jeho základě jsou ovšem cizinci v očích rodilých mluvčích často odsuzováni k negativním národnostním stereotypům. Typicky jsou tak považováni za chudší, méně vzdělané, méně důvěryhodné, a jiným podobným nálepkám. Taková stigmatizace je příznakem diskriminace na základě cizineckého přízvuku, potažmo na základě rasy.

Následující tři podkapitoly (2.3., 2.4., 2.5.) se zabývají vybranými segmentálními a suprasegmentálními charakteristikami rodilé angličtiny a české angličtiny a rolí těchto aspektů

v osvojování cizího jazyka. Podkapitola 2.3.1. představuje anglický segmentální systém, v rámci kterého nejprve zmiňuje krátké samohlásky /ɪ, e, æ, ə, ʌ, ɔ, ʊ/, dlouhé samohlásky /i:, u:, ɔ:, ɜ:, a:/ a dvojhlásky. Důraz je mimoto kladen na středovou hlásku *šva*, která hraje důležitou roli v suprasegmentálních procesech. Anglické souhlásky jsou popsány z hlediska místa sonority, z hlediska místa tvoření (obouretné, retozubné, dásňové, zadodásňové, předopatrové, zadopatrové, hlasivkové) a z hlediska způsobu tvoření (okluzivy, nazály, frikativy, semiokluzivy, vibranty, aproximanty). Detailně popsané jsou poté dentální frikativy /θ, ð/ a velární nazála /ŋ/, které jsou obzvlášť problematické pro české mluvčí angličtiny. Ti mají tendenci nahrazovat /θ, ð/ jinými souhláskami, zatímco /ŋ/ vyslovují chybně v kombinaci s okluzivou /k/. Podkapitola 2.3.2. věnuje pozornost suprasegmentální charakteristice angličtiny. Konkrétně se zabývá tzv. silnými a slabými slabikami, které jsou rozlišeny plnou či redukovanou výslovností. Právě v redukovaných (slabých) slabikách se objevuje zmíněná šva. To vše navazuje na diskuzi o tzv. vázané řeči, která je utvářena redukčními procesy, jako je asimilace, elize, či spojování slov. Nadále se tato podkapitola věnuje anglickému rytmu, který je založen na principu pravidelně se opakujících přízvuků, v čemž se tento jazyk podstatně liší od češtiny. Intonace, intonační rozpětí a anglické tóny (stoupavý, klesavý, stoupavo-klesavý, klesavo-stoupavý, statický); na závěr jsou stručně rozebrány funkce intonace (diskurzivní, lexikální, funkce dělení řeči). Podkapitola 2.3.3. rozebírá roli segmentálních a suprasegmentálních složek při osvojování cizího jazyka: věnuje se tématu jazykových transferů a důležitosti explicitní fonetické instruktáže, která vede ke znatelně lepším schopnostem porozumění a řečového projevu. Poslední podkapitola (2.4.) této sekce se zaměřuje na typické rysy českého přízvuku v angličtině. Na segmentální rovině se k těmto rysům řadí chybná výslovnost dentálních frikativy, které jsou nejčastěji nahrazovány hláskami [s, f, d]. Ve spojení s velárními nazálami se diskutuje o jejich typickém výskytu a chybné české realizaci, v rámci které se tato nazála objevuje i se správně nevyslovovanou okluzivou. Problematická je také rozdílná funkce tohoto segmentu: zatímco v češtině /ŋ/ funguje pouze jako alofon, v angličtině je tento segment plnohodnotný foném, a rozlišuje tedy významy slov. Co se prozodie týče, čeští mluvčí mají tendenci mluvit s velice plochým intonačním rozpětím a zápasí s redukcí, která se vyskytuje ve vázané řeči. Jejich projev tedy nesdílí plynulost anglického projevu.

Materiál a metodologie jsou ústředním tématem třetí kapitoly, která je rozdělena na několik podkapitol. Sekce 3.1. popisuje výběr deseti mluvčích z Pražského fonetického korpusu, jejichž řečového projevu jsme rozkouskovali do čtyřiceti kratších nahrávek, se kterými jsme dále pracovali. Důležitým faktorem při výběru byla přítomnost českého přízvuku a

typických chyb, které jsme si vybrali pro tento výzkum. Toto množství nahrávek jsme nadále upravovali v programech Praat a Adobe Audition, a to tak, že jsme manipulovali segmentální a suprasegmentální úrovně mluvčích. Tyto manipulace jsou popsány v sekci 3.2. V rámci segmentálních úprav jsme odstraňovali okluzivy z chybné kombinace s koncovými velárními nazálami. Chybné dentální frikativy jsme nahrazovali správně vyslovenými verzemi těchto hlásek, které jsme sami nahráli. Prozodické úpravy spočívaly v odstranění původní intonace a vytvoření nové intonace tak, aby obsahovala větší intonační rozpětí a v rámci které je přízvuk vložen na ty slabiky, na které patří. Tímto způsobem jsme tedy opravili nejen plochost řeči, ale i špatné rozmístění přízvuku a chyby ve vázanosti projevu. Takto upravené nahrávky jsme nadále spojili do menšího množství delších položek v rámci zachování realistické délky percepčního testu. Vytvoření tohoto testu se popisuje v podkapitole 3.3., ve které se vysvětluje postup při tvorbě grafické podoby testu. Důležité je vysvětlení systému hodnocení nahrávek, na kterém celý výzkum spočívá. Každá z dvaceti položek testu se musí zhodnotit ze dvou hledisek: na kolik se přízvuk zdá posluchači rodilý či cizinecký, a zdali posluchač položce rozumí bez obtíží či s obtížemi. Toto hodnocení se odehrálo na škále od 1 do 5, kde 1 představovala téměř rodilý přízvuk a bezproblémové porozumění a 5 představovala silný přízvuk a obtížné porozumění. Poslední dvě podkapitoly hovoří o zadání tohoto testu a o postupu vyhodnocení výsledků.

Čtvrtá kapitola se zabývá výsledky a diskuzí, a prezentuje tak jednotlivá východiska výzkumu pomocí LME analýzy a Tukeyho posthoc testu. S pomocí pěti grafů ukazuje, že manipulace vyústily v rozdílné hodnocení manipulované a nemanipulované nahrávky stejného mluvčího. Manipulace měly větší vliv na hodnocení pocitu srozumitelnosti než na hodnocení síly přízvuku. Pocit srozumitelnosti byl více ovlivněn prozodickými manipulacemi, a to zhoršením hodnocení. To připisujeme pravděpodobné nedostačující kvalitě tohoto druhu manipulací. Síla přízvuku byla pozitivně ovlivněna segmentálními manipulacemi a nijak zvlášť ovlivněna manipulacemi prozodickými. Vybraní mluvčí jsou mimoto rozebráni individuálně, a to z důvodu zvláštnosti jejich hodnocení či viditelných trendů v rámci skupiny. Důležitým východiskem studie je význam porozumění v hodnocení řečového projevu, který jasně převyšuje význam přízvukovosti. Potvrzuje se, že nejen segmentální, ale zejména suprasegmentální složky jazyka jsou v tomto směru důležité a jejich studium by se nemělo podceňovat. Pátá kapitola, závěr, práci shrnuje, komentuje její omezení a navrhuje směr budoucího výzkumu. Šestá kapitola obsahuje seznam použité literatury.