Filozofická fakulta
Univerzita Karlova

Bakalářská práce

Praha 2019 Nina Laketić

Filozofická fakulta

Univerzita Karlova

Ústav anglického jazyka a didaktiky

Bakalářská práce

Nina Laketić

Vliv rytmické struktury na subjektivní srozumitelnost české angličtiny

Effect of Rhythmic Structure on the Comprehensibility of Czech English

Praha 2019

vedoucí práce: doc. Mgr. Radek Skarnitzl, Ph.D.

Acknowledgements

I would like to express my sincere gratitude to doc. Mgr. Radek Skarnitzl, Ph.D. whose guidance and kind encouragement were of an utmost importance during the process of conducting the experiment and writing the thesis itself. His patience and immense knowledge made this research possible and I hope to be able to contribute to the field of phonetics and to the Institute of Phonetics at the Charles University in the future.

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 **5. 8. 2019**

……

Nina Laketić

Abstrakt a klíčová slova

Cílem této práce je zkoumat vliv rytmické struktury řeči na subjektivní srozumitelnost české angličtiny. Centrálním bodem práce je percepční test formou vynuceného výběru, který je zadán čtyřiceti českým posluchačům rozděleným do dvou skupin podle jejich znalostí angličtiny (*anglofonní* a *ostatní*). Respondenti poslouchají dvojice frází české angličtiny, jejichž rytmická struktura byla zmanipulována tak, aby se rytmus jedné verze blížil anglickému rytmu a rytmus druhé verze českému rytmu. Posluchači pak určují, která verze se jeví jako více srozumitelná nebo u které z verzí cítí silnější cizinecký přízvuk. Teoretická část práce vymezuje termín prozodie řeči a poukazuje na její význam při osvojování jazyků. Tato kapitola dále popisuje konkrétní prozodický jev, kterým je rytmus. Ten se řadí mezi nejvýznamnější prozodické jevy v angličtině, a to díky své klíčové roli v rámci efektivní komunikace, jelikož umožňuje, aby nejdůležitější prvky v informační struktuře byly zvýrazněny. Praktická část práce obsahuje podrobný popis uplatněné metodologie: výběr materiálu, temporální manipulace, tvorbu percepčního testu, výběr respondentů a popis analýzy dat. V závěru praktické části je zpracován detailní popis výsledků testu a jejich zakomponování do širší perspektivy spolu se zhodnocením dílčích závěrů. Z výsledků vyplývá, že se srozumitelnost české angličtiny zvyšuje, pokud se její rytmická struktura přibližuje rytmu angličtiny, a zároveň se zvyšuje dojem přítomnosti cizineckého přízvuku, pokud se její rytmická struktura blíží rytmu češtiny. Signifikantní výsledky byly zaznamenány pouze u skupiny *anglofonních*, a proto jsou hypotézy potvrzeny pouze částečně. Poznatky této práce poukazují na potřebu posílit učební metody zaměřující se na temporální členění osvojovaného jazyka.

Klíčová slova: rytmus řeči, prozodické jevy, subjektivní srozumitelnost, cizinecký přízvuk, osvojování jazyků, česká angličtina

Abstract and key words

The aim of this thesis is to examine the effects of rhythmic structure on the comprehensibility of Czech English. The focal point of this study is a *pairwise comparison* perception test administered to 40 Czech listeners divided into two groups according to their familiarity with native English (*Anglophones* and *others*). They are presented with pairs of Czech English phrases from which one is manipulated to approximate English rhythmic structure, the other to approximate Czech rhythmic structure, and are asked firstly to select which of the versions is more comprehensible, and secondly which is more foreign-accented. The theoretical background provides brief description of prosody of languages focusing on its significance in Second Language Acquisition; this section then moves to one specific prosodic feature, the rhythm of speech. Rhythm classifies amongst the most significant prosodic features of English due to its key role in an efficient communication; it enables the most important elements in an information structure to be highlighted. The practical part of the thesis contains a thorough description of the methodology employed during the experiment: the selection of material, temporal manipulations, creation of the perception test, respondents, and the analysis of the gathered data. These are followed by a detailed report of the test results and a general discussion. The results show that rhythmic structure which emulates native English rhythm increases the comprehensibility of speech; and simultaneously, that rhythm which emulates Czech rhythm is perceived as more foreign-accented. However, these results were significant only for the *Anglophone* group. The hypotheses have therefore been partially confirmed. This outcome advocates the importance of training focused on the temporal organization of the L2 speech.

Key words: rhythm, prosody, comprehensibility, foreign accent, second language acquisition, Czech English

Table of Contents

Table of Figures

1. Introduction

In this day and age, the individuals learning foreign languages are opening themselves to new cultures and perspectives on a social level, and to the prospects of a successful future on a personal level. This era tracks an increasing demand for peoples' linguistic proficiency and consequently, researchers are continuously attempting to enhance the methods associated with the acquisition of second languages. One of the main aspects discussed in this respect is the temporal organization of speech – specifically rhythm – and its significance in human communication. Accordingly, the questions are being raised what are the consequences of the "negative transfer" (Tajima, Port & Dalby, 2007: 2) of temporal features which emerge from the learner's native language (L1) and how the research focused on the rhythmic structure of languages might enrich the existing teaching practices. The present study intends to contribute toward this advancement.

The main aim of this thesis is to examine how the rhythmic structure of speech affects the comprehensibility and perceived accentedness of Czech English. This will be analysed from the results of a *pairwise comparison* perception test. Prior to the test itself, short English phrases recorded by Czech speakers shall be temporally manipulated using the programme Praat (Boersma & Weening, 2016). Thereby two versions of each phrase will be acquired; the temporal organization of the first version will emulate the rhythm of native English, whereas the temporal organization of the second version will be manipulated to appear *even more* Czech. Compiled pairs of phrases will be used for the perception test: two groups of Czech listeners (the students of Anglophone studies and the subjects of different background/non-Anglophones) will be presented with two parts of the test – one focused on comprehensibility, the other on accentedness (each containing 40 items) – in which they will be asked to decide which of the versions is a) *more comprehensible/understood with more ease*, and b) *more foreign-accented/Czech-like*, respectively.

As the premise of this thesis has been introduced, what follows is the outline of its structure. This introductory chapter will be followed by chapters: *Theoretical background* (2), *Research questions and hypotheses* (3), *Material and methodology* (4), *Results and discussion* (5), *General discussion* (6), and *Conclusion* (7). The theoretical background consist of two main sections: 2.1 briefly defines the term *prosody*, describes its role during the L2 acquisition, and introduces the concept of prosodic typology of languages. Section 2.2 then moves to the main subject matter of the thesis, the *rhythm*, by describing its role in life in general and in speech in particular. This section introduces the principle of isochrony and presents

several alternative principles of English rhythm. Hypotheses and related questions of this study are stated in chapter 3. The practical part of the thesis begins with chapter 4; this chapter describes the methodology of this experiment, including the compilation of material, the individual temporal manipulations, the process of creating the perception test, the selection of subjects, and the analysis of the gathered data. The results of the experiment are illustrated in detail in chapter 5 and put into larger perspective in chapter 6. These sections are followed by a conclusion in chapter 7. Here the outcome of this study will be linked to the past research, and suggestions for further research will be made. The thesis closes with the list of cited sources and a resumé in Czech.

2. Theoretical background

2.1 Prosody and its role in Second Language Acquisition

Before proceeding with the description of English rhythm itself, it is necessary to focus on the prosody of speech (under which rhythm is subsumed) and its role in Second Language Acquisition. Some scholars choose to define *prosody* from an abstract, phonological standpoint, while others prefer to examine it from the point of view of its actual role in speech. Specifically, Shattuck-Hufnagel and Turk (1996; in Mennen & de Leeuw, 2014: 184) write that in the first case the term is defined as "the phonological organization of segments into higher-level constituents and to the pattern of relative prominences within these constituents". This definition disregards the effect of extralinguistic factors, such as the speaker's emotional state, demeanour, and social and regional group, as they are not considered to be "channeled through prosody" (Mennen & de Leeuw, 2014: 184). On the opposite side of the spectrum is the definition by Cutler, Dahan, and van Donselaar (1997; in Mennen & de Leeuw, 2014: 184) who claim that from the point of view of its role in speech *prosody* stands for "the realization itself" implying that it is used synonymously with the term *suprasegmental features,* such as pitch, tempo, length, duration, etc. This view, for instance, does not regard syllable structure as relevant to the study of prosody (Ibid.). Mennen and de Leeuw claim that the most desirable definition unites both of the statements above: *prosody* therefore represents "the linguistic structure which determines suprasegmental properties of utterances" (Cutler et al., 1997; in Mennen & de Leeuw, 2014: 185). It is a feature which occurs at all times in every speech of any language.

Due to some of its functions, *prosody* is of an utmost importance for an efficient human communication. Firstly, it has a *grammatical* function (e.g., it may determine a sentence type; for instance, whether a certain sentence represents a declaration or a question). Secondly, it has a *discourse* function which facilitates governing a dialogue (e.g., during *turn-taking*; the falling pitch signifies that the speaker had finished talking and he or she expects the listener to react). It also has a *focusing* function meaning that it emphasizes the most important elements in a particular utterance; this is mostly accomplished "by acoustic patterns of fundamental frequency (F0), duration, and amplitude" (Ibid.) which are heard as pitch, length, and loudness by the listener. Another function of prosody is to signify *lexical meaning*; for instance, if two words have an identical phonemic structure, the placement of *stress* will make the difference in their meaning evident. It should be noted, however, that in the majority of words stress is not the only indicator of meaning. Stressed syllables mostly contain vowels with different quality

which likewise enables the listener to understand the meaning (Ibid., 186). Lastly, prosody has the function of *prosodic phrasing* which signifies that it contributes towards an increase in the comprehensibility of utterances. In other words, by the insertion of pauses, the emphasis of lengthening, and by altering the pitch, the speaker disambiguates the syntactic structures which could be interpreted incorrectly if different prosodic structure was employed (Ibid.).

It has been acknowledged that achieving a native-like prosody is one of the most difficult tasks for the individuals attempting to learn a foreign language and that in reality, the majority of these learners will never succeed to do so (Atoye, 2005; Banjo, 1979; Cruz-Ferreira, 1989; in Mennen & de Leeuw, 2014: 186-187). The principal cause of this struggle is a large amount of differences between the prosodic systems of languages and the vast number of ways in which prosodic properties may be carried out by pitch, tempo, loudness, and duration (Cutler et al., 1997; in Mennen & de Leeuw, 2014). Accordingly, it is believed that the learner's first language (L1) strongly influences his or her perception of the prosodic system of the second language (L2), therefore the speaker modifies L2 prosody according to the prosodic regularities of L1 (Mennen & de Leeuw, 2014: 185). Learners are therefore faced with a task of learning what are the individual features of the prosodic system of L2 and secondly, understanding how are these features used in an actual speech in order to transmit a certain message in the most efficient way (Mennen & de Leeuw, 2014). They should understand which features serve to highlight significant or new information; for instance, whether the target language emphasizes the key information by the word order, by a specific pitch accent, or by assigning stresses to significant components while de-stressing the components of lesser importance (Ibid., 185).

2.1.1 Prosodic typology

Broselow and Kang (2013) claim that languages may be divided into three groups of *prosodic typology* and that this categorization may help clarify why some L2 learners of English of a particular L1 succeed to produce a native-like prosody in a shorter amount of time and with more ease than L2 learners of a different L1. According to their prosodic regularities, languages are categorised into *tone*, *pitch accent*, and *stress* languages. In *tone* languages the meaning of morphemes is distinguished by pitch (e.g., Mandarin Chinese), whereas in *pitch accent* languages "the inventory of pitch patterns on words is generally restricted, with specific syllables within a word (accented syllables) associated with invariant tonal contours" (Ibid., 543), for instance, Tokyo Japanese. In *stress* languages (e.g., English) every lexical word carries

one or multiple stresses; although, it is important to note that stressed syllables may become unstressed if the context requires it (Ibid.).

Considering the fact that English classifies as a *stress* language, it is suitable to define the concept of *stress* and to present some of its characteristics. Roach (2009: 73) explains that *stress* may be defined firstly from the standpoint of production which indicates what are the speaker's actions when creating stressed syllables: during the production of stressed syllables, there is a significant increase in generated muscular energy which is tied to a greater effort of the speaker's vocal tract. Secondly, from the point of view of perception we examine what traits make the listener perceive the particular syllable as stressed: Roach explains that one attribute shared by all stressed syllables is *prominence* which is generated by loudness, length, pitch, and quality (Ibid.). Usually all of these factors are employed simultaneously, but occasionally, stress may be produced using only one or two of them; thus, it is evident that individual factors differ in their importance: pitch is the most significant factor, it is then followed by length, and the least important factors are loudness and quality (Ibid., 73-74). If we narrow down our focus to English, *stress* represents a major factor when resolving whether a syllable has a strong or weak character (Ibid., 76). While stressed syllables are always strong, unstressed syllables can be either strong or weak (Ibid., 77). A syllable consists of an *onset* and a *rhyme* and the latter can be further divided into a *peak* and a *coda* (Ibid., 60); accordingly, an unstressed syllable is considered *strong* if its rhyme consists of:

- **a.** a peak realized by a long vowel or a diphthong with an optional coda (*whose* /hu:z/, *both* / bəʊθ /);
- **b.** a peak realized by one of the short vowels [ɪ, e, æ, ʌ, ɒ, ʊ] followed by a coda of the minimum of one consonant (*but* /bʌt/) (Ibid., 76).

On the other hand we consider the syllable *weak* if its rhyme consists of:

- **a.** a peak realized by vowels [i] or [u] not followed by a coda, or by [ə] with an optional coda (*lazy* /ˈleɪzi/, *influence* /ˈɪnfluəns/, *sofa* /ˈsəʊfə/);
- **b.** a peak that contains a syllabic consonant (*sudden* /ˈsʌdn̩/);
- **c.** a peak which consists of the vowel [ɪ] followed by a consonant that represents the onset of the subsequent syllable (*event* /ɪˈvent/); elsewhere, [ɪ] is the peak of stressed, i.e., *strong*, syllables (Ibid., 77).

Now that the concept of stress has been defined, it also important to examine what are the basic conventions of stress placement in *stress* languages. In some languages, stress is assigned to the syllables according to their position in a word; namely, Czech places stress on the initial syllable, Polish stresses the penultimate syllable, and French assigns stress to the

ultimate syllable (Ibid., 76). However, deciding where stress occurs in English words is far more complicated. Scholars have agreed that perhaps the most favourable approach towards learning English stress placement is to learn it simultaneously with the individual words (Ibid.). Roach believes that since native English speakers are able to correctly assign stress without encountering the word previously, there ought to be some systematic rules that they unconsciously apply which might be recorded and put into practice by non-native speakers, and indeed, he attempts to introduce a few of these rules. He nonetheless admits that they do not capture the rich variation of stresses in English, stating that there are a lot of exceptions to them, and for their complexity he acknowledges that it might be more viable to learn stresses while extending one's vocabulary (Ibid.).

As the concept of stress has been defined, we may now return to the prosodic typology of languages. Broselow and Kang (2013) claim that when acquiring the prosody of L2, some learners may have an advantage due to the similarities in the prosodic structure of L1 and L2. They list several articles which demonstrate how the learners whose L1 belonged to the same prosodic category as L2 were significantly better at producing L2 stress than the learners whose L1 belonged to a different category. One of them is the study by Altmann (2006; in Broselow & Kang, 2013) who showed that the learners of English whose L1 was Arabic (a *stress* language) were more successful at the production of stresses than the learners whose L1 was Mandarin (a *tone* language). Additionally, the overlapping stress systems of L1 and L2 have also been reported to represent a certain advantage for L2 learners. Kijak (2009; in Broselow & Kang, 2013), for instance, observed that students learning Polish (a language that places stress on the penultimate syllable) had a greater success at assigning stress if their L1 had the option of the penultimate stress. He showed that the students whose L1 was English, German, Italian, Russian, and Spanish had an advantage over the students whose L1 was French (stress on the final syllable), Czech (the initial stress), and Mandarin (a *tone* language).

Furthermore, it is to be acknowledged that the difficulties with the acquisition of the prosody of L2 are not always caused by the prosodic system of L1. Archibald (1997; in Mennen & de Leeuw, 2014) claims that, in fact, some learners are able to produce prosody which does not occur in L1 nor in L2. Altmann (2006; in Mennen & de Leeuw, 2014) wrote that there have also been several instances of the learners being able to *perceive* the prosody of L2 well, yet failing to *produce* the correct prosody in L2 speech, and vice versa, some students might not succeed at the perception, but they are able to produce the correct L2 prosody. Rare situations like these may be associated with the person's limits of motor production or with his or her

difficulties with keeping the prosodic information in the long-term memory (Mennen & de Leeuw, 2014: 188).

2.1.2 The effect of prosody on the perception of L2 speech

Researchers have been inspecting how a non-native prosody influences the *perception* of L2 speech by native speakers of the target language. Studies have shown that prosody significantly contributes towards noticing a *foreign accent* and even helps recognize the speaker's specific L1 background (Boula de Mareüil & Vieru-Dumulescu, 2006; Mennen & de Leeuw, 2014). Most experiments have been concerned with the effect of intonation on the perception of non-native accent, but various studies have proven that the perception of accentedness arises from other prosodic properties as well; namely, Kang (2010; in Mennen & de Leeuw, 2014) studied pitch range and stress, speaking rate has been examined by Munro and Derwing (2001; in Mennen & de Leeuw, 2014), the effect of timing has been described by Tajima et al. (1997; see also the section 2.2), and several studies have focused on phonotactics and rhythm (Carter, 2005; Grenon & White, 2008; Gut, 2003; White & Mattys, 2007; all in Mennen & de Leeuw, 2014).

Prosody may also affect the intelligibility and comprehensibility of speech. Atagi and Bent (2011: 260) define *intelligibility* as "listener's ability to accurately report the words that a talker has produced" while *comprehensibility* is "the subjective, perceived ease with which listeners understand speech". The following studies showcase the correlation of prosody with comprehensibility and intelligibility; most of them were performed by employing digital signal processing and manipulation techniques. Derwing and Munro (1995; 1997; in Atagi & Bent, 2011 and in Mennen & de Leeuw, 2014) have shown that speech with non-native segmental and suprasegmental features – which resulted in a high degree of accentedness – could still be perceived as very intelligible. They disclosed that although certain features affect comprehensibility and perceived accentedness of speech, they do not reduce its intelligibility. Mennen & de Leeuw mention a few similar studies, such as the one by Braun, Lemhöfer, and Mani (2011; in Mennen & de Leeuw, 2014) who had demonstrated that abnormalities which occurred in "the language-specific implementation of stress (i.e., by placing stress on the correct syllable but using the wrong acoustic cues)" (Mennen & de Leeuw, 2014: 189) may negatively affect the comprehensibility of speech. Comparable results have been presented in an experiment by Braun, Dainora, and Ernestus (2011; in Mennen & de Leeuw, 2014) in which the occurrence of foreign intonational contrasts obscured and decelerated lexical access, thus disturbing the comprehensibility of speech. A recent experiment by Trčková (2019) showed that comprehensibility is facilitated more by the correct prosody (intonation, stress, and

duration) than by the correct segments. Mennen and de Leeuw (2014) conclude that further research is needed to help describe which prosodic properties improve the intelligibility and comprehensibility of speech. This would essentially enrich the practices applied during the acquisition of second languages. This thesis further narrows down its focus to a specific prosodic feature, the rhythm.

2.2 Rhythm

The definition provided by Roach (2009: 107) is that *rhythm* "involves some noticeable event happening at regular intervals of time." Similarly, Wade (2004; in Ravignani & Madison, 2017: 2) defines it as a "pattern of events in time." Volín (2010: 290) likewise defines *rhythm* as "a temporal organization of recursive patterns" and adds that these patterns are consisted of contrasting elements. In speech these elements are "the alternate stressed and unstressed syllables and the alternate high and low pitches" (Fries, 1943 [emphasis original]; in Dickerson, 2016: 40). For the purposes of this thesis, the alternating high and low pitches shall not be considered a part of the rhythmic structure of speech; *rhythm* is therefore to be understood purely as a temporal organization of speech established by the alternation of elements of higher and lower levels of prominence which is signified by their longer and shorter duration, respectively.

Before examining more closely the role of rhythm in speech it is important to recognize its pivotal role in life in general. Various studies have pointed out that inclination towards rhythmicality is perhaps something people are born with (Ravignani & Madison, 2017; Volín, 2010). This assumption is most apparent in babies' and small children's preference for something rhythmical over that which is arrhythmical; e.g., nursery rhymes (Volín, 2010: 291). The reasons for this may be biological; that is, physiological processes such as heartbeat, breathing, and locomotion are more or less rhythmical. Even pulses of the nervous system work in rhythmical patterns, and their arrhythmicality can be one of the symptoms of neurological disorders such as Parkinson's disease (Grahn, 2012; in Ravignani & Madison, 2017; Volín, 2010). Another reasoning behind general human predilection for the rhythmical rather than the arrhythmical may be evolutionary. Ravignani and Madison (2017) claim that there is a possibility of our ancestors having to produce loud, synchronized, rhythmical noises so that their signal reached farther distances, possibly to attract the females, i.e., for procreational reasons. Nowadays, rhythmical chanting is used for expressing political or religious ideas at

protests, cheering during sport's events, or other situations where a crowd intends to convey a common objective (Volín, 2010: 291).

Volín discusses two universal principles in regard to the role of rhythm in speech; firstly that "the regular is easier than irregular" and secondly that "regular allows for the coordination of actions" (Ibid., 290). Instances of the latter have been briefly displayed above, therefore are not to be discussed here. All living organisms are driven by a *principle of least effort* (Ibid.). This rule applies even when talking about spontaneous speech; that is to say, speakers of all languages incline towards regular patterns – i.e., *rhythmicality* – during speech, since arrhythmical speech production is significantly more laborious for the vocal tract. Moreover, speech becomes easier if the syllables that were made acoustically prominent are followed by unstressed syllables, which represent brief resting phases. However, speech rhythm cannot be maintained in a perfectly regular state, because it ought to satisfy the conditions imposed by the content, that is, it is necessary for the principal idea of the sentence to be properly transmitted (Ibid.).

Not only does the rhythm of speech help its production, but it also enables effortless speech perception. Listeners will perceive regular speech signals more efficiently than irregular ones because their brains have to perform so-called *act of resonance*, which is defined as:

> a series of synchronous activations of neurons which occurs when the expectational neural representations (based on memory traces activated by the concurrent context analysis) meet with the input neural representations (based on the properties of the incoming signal) (Grossberg, 2003; in Volín, 2010: 294).

That is, if the neural activity invoked by the perception of speech and the neural activity invoked by our expectation of the given speech meet at the same time, the two streams reach neural synchrony, i.e., *neural resonance*, resulting in a smooth identification of individual segments of speech, such as words, syllables, morphemes, etc. (Ibid.).

Volín lists several studies which showcase the dependence of perception on regular rhythm; Huggins (1979; in Volín, 2010) learns that natural rhythm configurations increase the intelligibility of speech, while unnatural ones cause the number of word errors to rise. Buxton (1983; in Volín, 2010) shows that in various tasks, the reaction time of listeners to a target word is longer if the regular temporal structure of speech is manipulated and thereby disrupted. Similarly, Quéne and Port's (2005; in Volín, 2010) study demonstrates that reaction time decreases with regular, and thus predictable, temporal structure. The experiment by Ghitza and Greenberg (2009; in Volín, 2010) showed that the intelligibility of recordings – which were previously sped up three times – increased when silent intervals were added in between every

two neighbouring 40-millisecond fragments, with the duration of the silent intervals periodically fluctuating between 20 and 120 milliseconds.

The research by Tajima et al. (1997) demonstrated how the manipulation of rhythm influences its intelligibility. The authors performed two modifications of temporal patterning of speech. Firstly, they intended to improve the intelligibility of English phrases recorded by a native Chinese speaker. They did so by modifying their temporal structure using the identical set of phrases recorded by a native English speaker as a template. Secondly, they wanted to worsen the intelligibility of English phrases recorded by a native English speaker, so through manipulations they emulated the original temporal patterning of the Chinese speaker. The test had been administered to a group of native English listeners and the results have shown significant changes in intelligibility ratings of both speakers; the intelligibility of the Chinese speaker improved from 39% to 58% correctness, while the correctness of the English speaker's phrases decreased from 94% to 83%. Through this study Tajima et al. (Ibid.) promoted the idea that the intelligibility of learners' speech could increase significantly if they are provided with training specifically focused on the temporal features of speech.

The regularity of rhythm and temporal structuring of speech has also proved to be significant in the field of psychophonetics. Volín, Poesová, and Skarnitzl (2014; in Berkovcová, Černikovská, and Skarnitzl, 2016) have carried out an experiment that examined the effect of speech rhythm on the perception of speaker's *neuroticism*. The recordings of 14 professional speakers of English were manipulated, so that the duration of stressed vowels was reduced by half, and unstressed vowels were modified to be twice as long (the idea of this method was based on the fact that in English, stressed vowels tend to be longer and unstressed vowels shorter). Recordings with a distorted rhythmical structure have been rated more negatively on a scale of perceived neuroticism by the listeners (Ibid.). Furthermore, Berkovcová et al. (2016) studied how the temporal organization of speech affects the listeners' perception of the speakers' *competence*, but instead of focusing on speech rhythm they focused on its tempo. The authors have completed temporal manipulations of the recordings of four native Spanish speakers so that the tempo of their speech fluctuated; this has been accomplished by speeding up or slowing down individual words (or phrases of two words) in the recordings. 40 Spanish-speaking Czechs rated the speakers in regard to their perceived competence, i.e., the speaker's ability and readiness to solve assignments effectively (Hřebíčková, 2011; in Berkovcová et al., 2016: 9). The results have shown that a distorted temporal structure of speech had a negative effect on the perception of speakers' competence; the speakers whose recordings involved a fluctuating speech tempo were rated as less competent than the speakers with original,

unmanipulated speech tempo. These two studies showed that temporal regularity and rhythmicality of speech represent critical factors in human communication. The fact that irregular speech may stigmatize the speakers supports Volín's (2010) notion of people's inherent preference for regular patterns.

2.2.1 Isochrony and other principles of English rhythm

In order to define *isochrony* it is necessary to distinguish two subordinate terms first; these are the *idealized isochrony* and *empirical isochrony*. Ravignani and Madison (2017: 2, [emphases original]) explain that *idealized isochrony* denominates "a rhythmic pattern where all intervals have *equal duration*", but as the term implies, is rather non-existent in regular speech. For this reason, they claim that what is generally understood under the term *isochrony* is, in fact, *empirical isochrony* which is defined as "a rhythmic pattern where all intervals have *roughly* equal duration" or "a rhythmic pattern obtained by jittering events in an idealized isochronous pattern" (Ibid.).

Traditionally people thought of a two-way distinction in rhythmic properties of languages, according to the level, at which isochrony (i.e., rhythmic pattern) occurs; languages were labelled as either *stress-timed* or *syllable-timed* (later, the third group, *mora-timed*, was added, but it shall not be discussed in the thesis) (Ibid.). Units of rhythm are called *feet* and Roach (2009: 108) explains that a *foot* "begins with a stressed syllable and includes all following unstressed syllables up to (but not including) the following stressed syllable". In *stress-timed* languages (not to be interpreted as the above-mentioned prosodic type of *stress* languages), *feet* are thought to be of a roughly similar duration. Accordingly, stressed syllables, which mark their beginning, appear at intervals of a relatively regular duration, i.e., *isochronously*. It was thought that in order to satisfy the requirements of an isochronous pattern of stressed syllables, a varying number of unstressed syllables was being "compressed" into the time intervals between stressed syllables. For a long time, English, amongst other languages such as Russian or Arabic, had been considered a *stress-timed* language (Roach, 2009: 107-108; Ravignani & Madison, 2017: 2). In languages that have a *syllable-timed* rhythm, all syllables whether stressed or unstressed tend to occur isochronously (i.e., at regular intervals), meaning that with the increasing number of unstressed syllables increases also the duration of the time interval between stressed syllables and vice versa (Ibid.).

However, it has been acknowledged that division into these groups is too trivial to encompass the richness of different languages (Volín, 2010: 300). Countless experiments have

failed to provide enough evidence for the exact distinction between *stress-timed* and *syllable-timed* languages. Specifically, regarding English, a number of techniques for temporal measurement of speech have shown that the duration of *feet* and the intervals between stressed syllables are not as regular as the theory of isochrony suggested (Roach, 2009: 110). Despite the criticism of isochronous principle of rhythm, many teachers attempt to improve the pronunciation of L2 learners of English by performing exercises which emphasize the regularity of rhythm. These methods involve clapping or beating on the stressed syllables. Thought not perfect, these techniques help the students of English develop a sense of differentiation between strong and weak syllables, i.e., the aforementioned tendency to the compression of unstressed syllables which is one of the most important aspects of a nativelike pronunciation of English. Such exercises are especially useful for students whose first language does not involve contrast of syllables as significant as that in English (e.g., Spanish, Hungarian, or Czech). However, teachers ought to be careful when using this method so that their students are not apt to applying the rhythm of recitation to a connected speech (Ibid.).

Despite there being a prevalence of scientists who had rejected the categorization of languages into *stress-timed* and *syllable-timed* (and *mora-timed*), the 1990's tracked the emergence of a new approach towards the study of speech rhythm. These are the computational techniques termed *rhythm metrics* (Volín, 2017). Their attractiveness spread among phoneticians due to their apparent ability to describe temporal properties of individual languages by quantifying them (Nolan & Jeon, 2014: 3). Volín (2017: 81) explains that their potential to provide numerical material seemed precise and sophisticated enough to help scholars dispute claims that linguistics belongs amongst the group of "easier" sciences. Another reason for the popularity of *metrics* is the belief that they would finally validate the categorization of languages according to their isochronous speech patterns (Ibid.).

Nolan and Jeon (2014: 3) list a few of these techniques. Firstly, Ramus (1999, 2003; in Nolan & Jeon, 2014) introduced the proportion of vocalic intervals in an utterance (%*V*), the standard deviation of duration of vocalic intervals (Δ*V*), and the standard deviation of duration of consonantal intervals (Δ*C*). Secondly, Dellwo and Wagner (2003; Dellwo, 2006; in Nolan & Jeon, 2014) modified Δ*V* and Δ*C* into VarcoV and VarcoC, respectively. Thirdly, a *pairwise variability index* or PVI was by Low (1998; Low, Grabe & Nolan, 2000; in Nolan & Jeon, 2014) and it examined "the degree of variability [e.g., the variability of duration] between successive acoustic segments or phonological units" (Nolan & Jeon, 2014: 3) by which Low usually referred to syllables and vocalic intervals. Low *average pairwise difference* signalled the regularity and high *average pairwise difference* signalled the irregularity of measured

properties (Ibid.). Later, *normalized* PVI (nPVI) was introduced and it proved useful in situations where the basic PVI was insufficient, such as in the cases of gradually increasing or decreasing tempo of the measured speech (Ibid.). This is because it "divides each v-interval difference by the sum of the respective two intervals" (Dankovičová & Dellwo, 2007: 1243). In Dankovičová and Dellwo's (2007) study nPVI results showed that Czech is *syllable-timed*; however, the research showed mixed overall results therefore the classification of Czech into one particular category of isochrony was not possible. Furthermore, the experiment by Volín (2017) showed that *vocalic* PVI (nPVI-V) of English (specifically, Southern British Standard) is nearly doubled in comparison with Czech (i.e., 37.2 for English and 19.5 for Czech) implying that vocalic durations vary more in English than in Czech. Initially, this seemed rather surprising considering the fact that vowel length is a phonological feature in Czech, but Volín suggested that the reason behind this is the tendency for English vowels to become reduced in unstressed syllables and exaggerated in stressed syllables, while in Czech neither of this occurs (Ibid.).

However, there is much criticism towards the notion of *rhythm metrics*. Volín (2017: 81) states that these techniques should not be named *rhythm metrics* (nor *rhythm-class metrics* as earlier research suggested) in the first place as they perform quantification of exclusively durational measures. They entirely disregard the perceptual phenomena of prominence which results from the variation of pitch, loudness, and timbre (Ibid.) whose modifications further cause changes of rhythmic structure (Barry, Andreeva & Koreman, 2009; Brugos & Barnes, 2014; Cumming, 2011; in Volín, 2017: 81). Volín (2017: 81) states that these metrics ought to be termed *durational variation metrics* (DVM).

Many researchers who disagree with a clear-cut concept of *stress-timed* and *syllable-timed* languages claim that isochrony of speech is a perceptual phenomenon which arises from the ability of humans to *perceptually regularize* anisochronous speech signal, and therefore does not originate in its actual physical properties (Ravignani & Madison, 2017; Roach, 2009; Volín, 2010). However, Cauldwell (2002) opposes both the theory that languages may be divided into categories of *stress-timed* and *syllable-timed* languages, and the theory that such distinction is a matter of the listener's perception. He believes that both concepts should be abandoned altogether, in spite of them being the most "comfortable" ones. In his view, in order for the speakers of a particular language to be able to perceive certain speech as *stress-timed* or *syllable-timed*, they need to have some predisposition to notice one or the other, yet that is impossible since there is no such thing as *stress-timed* or *syllable-timed* language in the first place. He adds that rhythmicality may be spotted only in tone units which have three or more

points of prominence, i.e., the third tonic prominence confirms the regularity set by the first and the second tonic prominence, but the occurrence of such utterances in spontaneous speech is scarce. Figure 1 from Cauldwell's (2002: 17) article shows a rare instance of an utterance with four points of tonic prominence (*he's currently thinking of moving again*) where rhythmical pattern may be spotted (stressed syllables are signified by the upper-case letters and the symbol "X", unstressed syllables by the lower-case letters and the symbol "x"). However, if certain words are replaced by different words without changing the meaning of the utterance within its context, the rhythm is likely to become disrupted, such as in the phrase *he's now thinking of doing it all over again*. The second phrase could be produced in a way that it would carry regular rhythm as well, but only if the speaker intended to speak in this manner, i.e., rhythmical speech would be the speaker's conscious choice (Ibid.).

word accents	he's	CUR rent ly	THIN king of	MO ving a	GAIN
Rhythm	x	X x x	X x x	X x x	X
word accents	he′s	NOW	THINking of	DOing it all over a	GAIN
Rhythm	x	X	X x x	X x x x x	X

Figure 1: Rhythmical pattern of clauses realized by different words, but carrying identical meaning within its context (Cauldwell, 2002: 17).

Cauldwell recognizes only two instances of rhythm; *elected rhythmicality* which the speaker uses during reciting, e.g., a verse or a book title, and *coincidental rhythmicality* which is specific for a rhetoric that seeks some social purposes (Ibid.). Most importantly Cauldwell hypothesises that speech is essentially *functionally arrhythmic*. This arrhythmicality allows for the most important elements to be highlighted and therefore enables the most effective transmission of meaning to take place (Ibid.).

Furthermore, Dickerson (2016: 40-43) explains that for decades the descriptions and teachings of English rhythm have been based on three pillars, two of which he renounces in his article (in Figure 2, the breakage shows which of the pillars have been rejected). The first pillar – *stress alternation* – states that rhythm rests in the altering prominence of syllables (Fries, 1943; in Dickerson, 2016) and it has been recognized as the only stable pillar. The second pillar, named *time intervals* or *stress-timed rhythm*, presents us with a theory which has been mentioned as well; that all interstress intervals are of a similar duration (Pike, 1945b; in Dickerson, 2016). However, as was mentioned earlier, no actual regularity has been identified; contrariwise, the duration of intervals has been shown to correlate with the number of unstressed syllables in a stress group. These findings unveil the unsustainability of the second pillar. The third pillar proposes the idea that accent should appear with every content word

within a phrase (Prator, 1951; in Dickerson, 2016), but Dickerson argues against this, interestingly, using Prator's own claims that such stress-placement might sound unnatural to a native (Prator, 1957; in Dickerson, 2016). The only reason why this theory survived so long is that it is the simplest way of deciding where stress should be placed thus being the most accessible method for the students of English (Prator, 1972; in Dickerson, 2016). Furthermore, Dickerson continues with Pike's statement that "a conversation style is characterized by *few centers of special attention* [stresses] and by *many repressed lexical stresses*" (Pike, 1945a, [emphasis original]; in Dickerson, 2016).

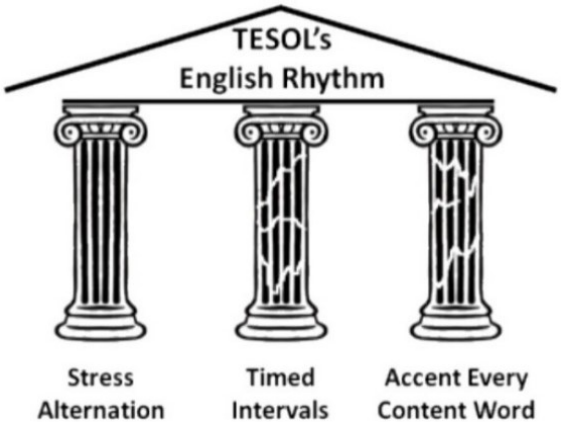

Figure 2: Three pillars of English rhythm (Dickerson, 2016: 41).

The concept of *stress-timing* has been disproved, but the question of how one should talk about the rhythm of English remains to be answered. Dickerson's concept of the *two-peak profile* may represent a satisfactory solution. The two-peak profile further develops the idea that there are only one or two points of prominence in most prosodic phrases of spontaneous speech which has already been discussed in relation to Cauldwell's *functional arrhythmicality* (Pike, 1945a; in Dickerson, 2016; Cauldwell, 2002). Originally these points of prominence have been labelled *onset* and *nucleus*, where nucleus is obligatory (Dickerson, 2016: 43-44). The terms Dickerson uses for the two-peak profile are *anchor peak* and *primary peak* (Figure 3) and they are the two words that on their own summarize the essential meaning of a phrase. A native listener is able to infer this key message under two conditions; a) that the peaks are in such proximity that they can be perceived as one thought; and b) that the prominence is given to no other element which might disrupt the chance of the listener's successful cognitive thinking, i.e., all elements besides anchor and primary peak shall be unstressed, contributing to the expected speech rhythm (Ibid., 44-47). Figure 4 shows several examples of the two-peak profile.

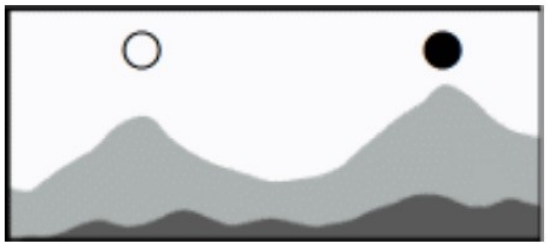

Figure 3: The two-peak profile of English rhythm; the hollow bullet is the primary peak, filled bullet, the anchor peak (Dickerson, 2016: 45).

			○				●
Mrs.	White's	little	boy	went	to	the	house.
		○			●		
Could	you	tell	me	the	time?		
		●					
Do	you	remember	him?				
○					●		
Buy	me	some	big	brown	potatoes.		

Figure 4: Examples of the two-peak profile (Fries, 1943; Pike 1942; in Dickerson, 2016: 44).

3. Research questions and hypotheses

The theoretical background delineated the importance of prosody in Second Language Acquisition and further focused on the rhythm of English. On the grounds of what was provided in the previous chapter, this section shall introduce what are the main aims and arguments of the thesis. Firstly, it should be noted that rhythm (and its structure) remains to be a feature which lacks unified definition among the academic sphere and thus is open to a further discussion. However, for the purposes of this study we shall return to the concept of *stress-timed* and *syllable-timed languages* (see 2.2.1). In general terms it is apparent that duration has a different role in Czech and English; in Czech, duration represents a phonological feature (distinguishing long and short vowels), while in English duration (yielded by stress) signalizes information structure (the most important elements gain prominence by an increased duration whereas the least significant elements become reduced in duration). For this reason we may accept the notion that Czech *approximates* syllable-timed rhythm while English *approximates* stress-timed rhythm. This has been implied by Palková (1994: 159), Roach (2009: 107), and partially by the results (nPVI) of Dankovičová and Dellwo (2007: 5). Based on this presumption our aim is to examine how Czech listeners perceive Czech English with native English rhythmic structure and Czech rhythmic structure. Therefore, the hypotheses of the research are: 1) the speech which approximates English rhythm will be perceived as more comprehensible than the speech which approximates Czech rhythm; 2) the speech with Czech-like rhythm will be perceived as more foreign-accented than the speech with English-like rhythm; and 3) the listeners who are more familiar with native English will be more perceptive of the differences between English-like and Czech-like rhythmic structure than the listeners who are less familiar with native English. In order to verify these arguments we shall perform temporal manipulations to the recordings of Czech English speech which will be described in detail in the following chapter.

4. Material and methodology

4.1 Compilation of material

Speech recordings of ten Czech female students (aged 19-25 years) without any speech impediments and with prominent Czech accents were chosen from the archive of the Institute of Phonetics at the Faculty of Arts, Charles University (though see section 5.4 where we mention that it was later found that one speaker did have a mild speech impediment). Recordings were acquired in the sound-treated recording studio of the Institute of Phonetics, using the microphone AKG C 4500 B-BC. The speakers were asked to read excerpts from the BBC newsletter and they were allowed to read and prepare the text before the recording itself. Having chosen older recordings we have avoided the familiarity of the perception-test listeners with the voice and diction of these speakers, which could potentially affect the results of the test comprehensibility-wise. From each recording of these ten females, four shorter recordings were chosen for temporal manipulations. The duration of these sections was approximately 3.00 to 3.50 seconds, as it has been agreed on this being the optimal length for the listeners to still be able to focus on the differences between the manipulated recordings. However, a faster speech tempo of some of the speakers, which caused a denser concentration of words in a recording, allowed us to choose even shorter sections (the shortest being 2.48 seconds) and a slower tempo with fewer words allowed us to choose some longer sections (the longest was 4.31 seconds long). The individual recordings were being chosen in regard to their suitability for the temporal manipulations aiming at the rhythm of speech, i.e., besides the contrast between stressed and unstressed syllables, the preferred material contained syllable peaks realized by diphthongs (the different duration ratio of diphthong constituents in Czech – 1:1 – and English – 2:1 or 3:1 – allowed us to create the impression of native and foreign rhythm by shortening or lengthening these constituents) or the chains of multiple grammatical words following each other (due to a different tendency towards assigning prominence to grammatical words in Czech and English speech: Czechs tend to pronounce grammatical words with equal prominence as lexical words, whereas the English tend to reduce the entire chain of grammatical words). On the other hand, we have avoided selecting the recordings which would limit the options of manipulations, such as those with many instances of pre-fortis shortening (in stressed syllables with pre-fortis shortening, the vowel mostly cannot be lengthened since this would result in a rhythm that does not appear neither English nor Czech) or cases of creaky phonation (as its manipulation would result in an unnaturally sounding recording).

To obtain suitable material, different segments of one speaker's recording frequently had to be "stuck" together; these edits had been performed in programme Adobe Audition. That way we have attempted to create recordings with the potential of showing the most promising results in the perception test, i.e., where English rhythm could contribute to the comprehensibility of speech the most. As a result of cutting and connecting different pieces of recordings together, some unnatural sounds and intonational "jumps" appeared at times. To eradicate their possible negative effect on the results, they were edited out either in Praat, where the jumps in the F0 contour were levelled out, or in Adobe Audition in which the two connected words had been "blended" so as to make the interval between them sound less disruptive and more natural. Furthermore, as the speakers read the same set of BBC news we have also been choosing the recordings with the intention of avoiding frequent repetitions of identical words across the recordings of various speakers. Finally, forty recordings (four for each speaker) were compiled for subsequent temporal manipulations. Each recording was to be manipulated twice, firstly to approximate English rhythm, secondly to approximate Czech rhythm.

4.2 Temporal manipulations

The selected sounds were manipulated in Praat. The original WAV file was opened in the programme and converted into a Manipulation file using the programme's pre-set time step of 0.01, minimum pitch 75 Hz, and maximum pitch 600 Hz. The newly created file was then manipulated using the "view and edit" function. Praat enables users to perform manipulations of the fundamental frequency and duration of speech; however, this experiment focuses only on the effects of temporal manipulations, therefore the F0 contour (*Pitch manip*) remained intact.

For the sake of convenience, firstly all duration points which marked the segments that could potentially be accelerated or decelerated were added to the duration axis (*Duration manip*) using the function "add duration point at…/at cursor". This way all duration points appeared at coefficient 1.0 of the axis (which marks the original relative duration of the recording). In this step, the manipulation file had been opened simultaneously with the WAV file so as to be able to see not only the waveform, but also the spectrogram of the sound, facilitating a more precise placement of duration points. If some duration points later proved unnecessary, they were removed. Manipulation file prepared in this way (seen in Figure 5) served as a starting point for the manipulations approximating English-like and Czech-like

rhythm; these will be henceforward referred to as the "English" and "Czech" version, respectively.

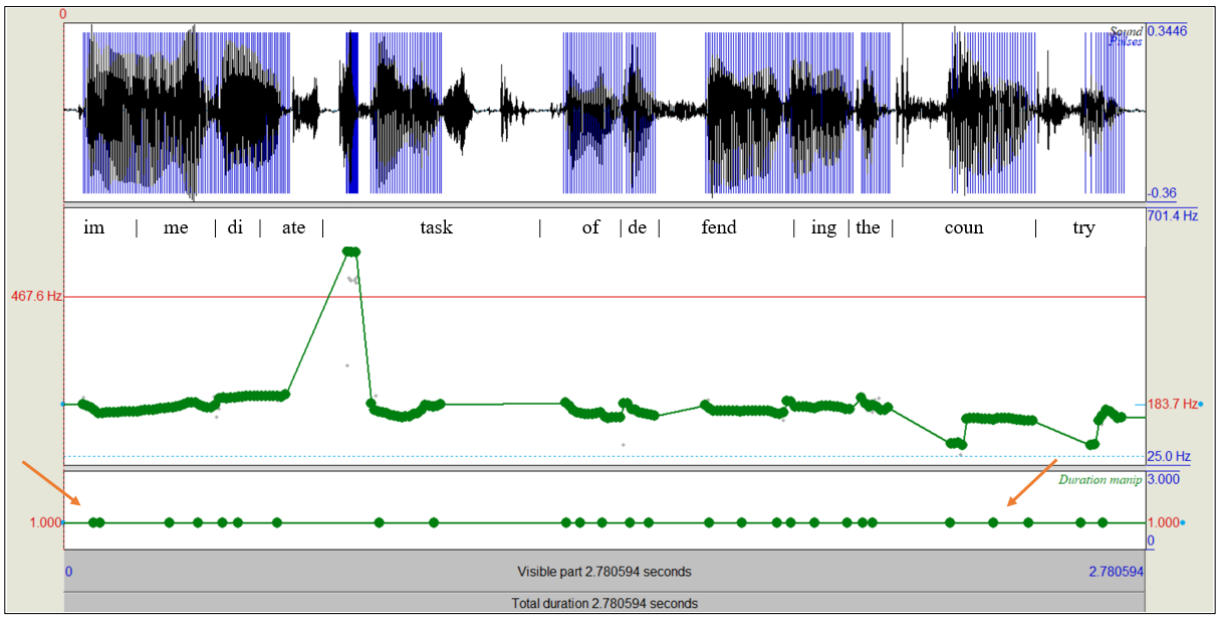

Figure 5: Duration points added to the *Duration manip* axis (orange arrows)

The primary principle according to which temporal manipulations were performed is that the rhythmic structure of English consists of the variation of points of greater and of lesser or no prominence. With this thought in mind, the most basic approach was to create a greater contrast between the duration of stressed and unstressed syllables in the version manipulated into sounding more English and to reduce this contrast in the Czech version. In the English version, stressed syllables with the peak realized by a monophthong were lengthened, i.e., their duration was increased by moving the *Duration manip* axis up (above the relative duration of 1.0), such as in Figure 6, where blue arrows point to the peak vowels of stressed syllables that have been lengthened, namely, /u:/ in *new* and /i:/ in *agreement*. However, as implied earlier, this manipulation could not be done in the case of pre-fortis shortening where in a single syllable the peak vowel is followed by a fortis consonant; in these cases the vowels could not be made too long otherwise the speech would sound foreign-accented (the pronunciation would sound almost Italian). Occasionally, syllables with pre-fortis shortening did allow a minor increase in duration, but such manipulations had to be done very carefully so as to not disrupt the impression of English rhythm.

The monophthong peak vowels of unstressed syllables were then shortened in order to reduce their prominence by moving the *Duration manip* axis down (below 1.0) which may again be seen in Figure 6 where the orange arrows point to all instances of shortening (it is apparent that in all syllables with peak vowel /ə/, this vowel has been shortened, reducing the

syllable's prominence). The shortening of vowel duration also has certain limits, such as in the case of final lengthening where the final syllable of a phrase tends to be longer even if it is unstressed; its shortening would therefore seem unnatural. Besides occurring at the ends of phrases, instances of phrase-final lengthening appeared in the middle of some of the recordings and they either had to remain unmanipulated or be manipulated with caution so as to avoid the end result sounding too "rushed".

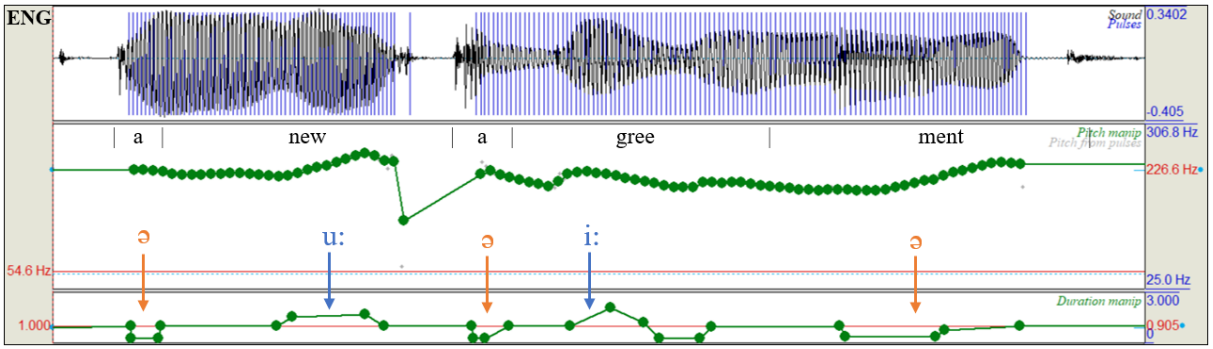

Figure 6: Temporal manipulations imitating English rhythm, orange arrows point to shortened (unstressed) segments, blue arrows point to lengthened (stressed) segments.

On the other hand, in the manipulations which resulted in recordings with a more Czech-sounding rhythm the goal was to reduce the contrast between stressed and unstressed syllables to create a rhythmic structure reminiscent of a *staccato* (Rumlová, 2018: 9). This was accomplished by shortening the peak monophthongs in stressed syllables and their lengthening in the unstressed syllables, if necessary (as all of the speakers already had a very Czech-accented pronunciation, these manipulations were not always required). The examples of these manipulations are shown in Figure 7 (which contains the same phrase as Figure 6); it may be noticed that all manipulations from Figure 6 and Figure 7 (except for the manipulation of the syllable *-ment* in *agreement*, whose creaky phonation did not allow us to lengthen the vowel) represent near mirror images of one another.

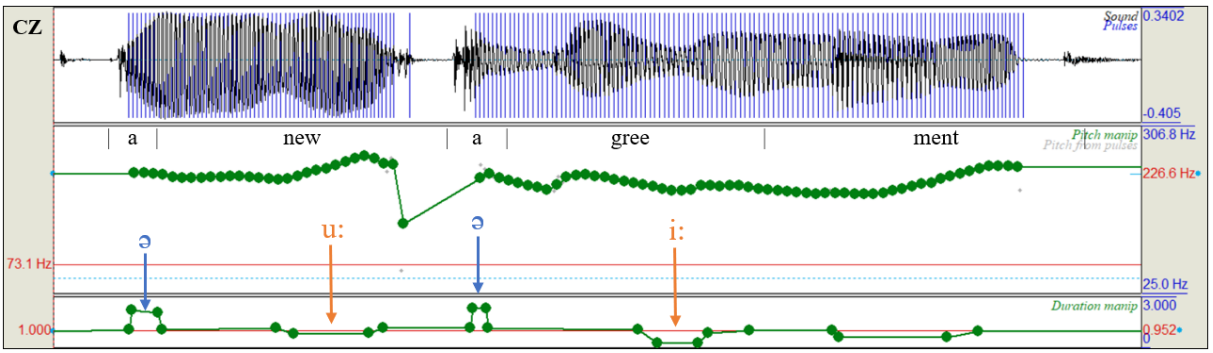

Figure 7: Temporal manipulations imitating Czech rhythm, orange arrows point to shortened (stressed) segments, blue arrows point to lengthened (unstressed) segments.

It is also important to note that length represents a phonological feature in the Czech vowel system (Dankovičová & Dellwo, 2007: 1242), therefore not all long vowels which occurred in stressed syllables could be radically shortened in the Czech versions. Actually, it is rather unlikely that Czech speakers would pronounce these vowels as significantly shorter than native English speakers. These manipulations have been completed intuitively, depending on individual speakers and their vowel realizations, which explains why the long vowel /u:/ in *new* (Figure 7) has been slightly shortened, while the duration of the vowel /i:/ in *free* (in the Czech version of Figure 8) remained the same. In Figure 8, the contrast between the duration of vowel /i:/ in the two versions was therefore created only by prolonging the peak in the English version, which is more probable to occur in speech.

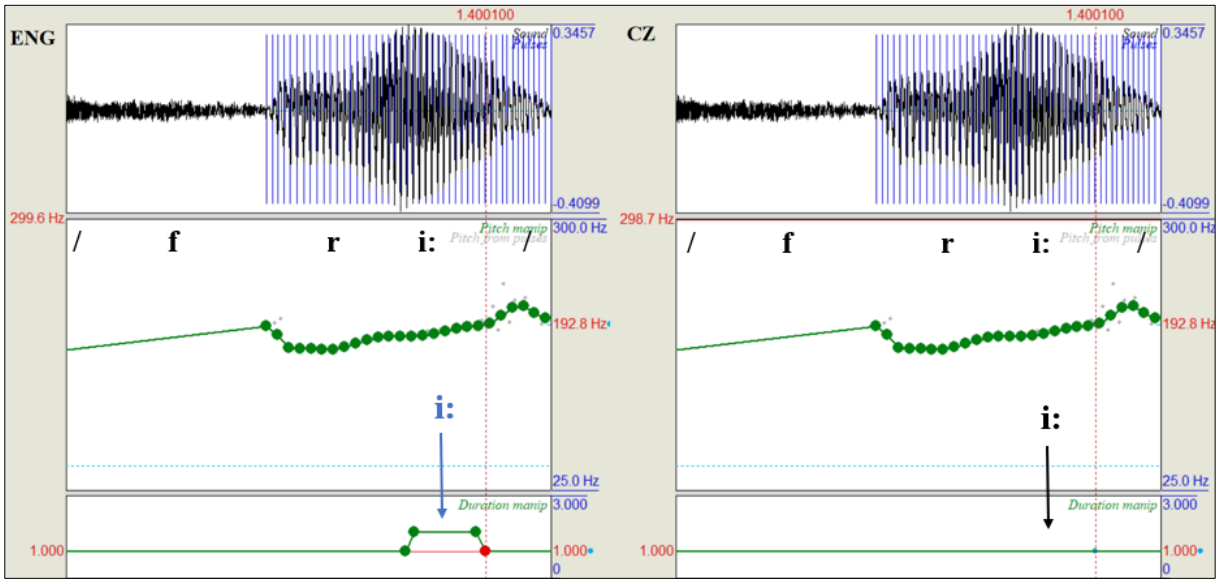

Figure 8: Example of the temporal manipulation of a long vowel; since length represents a phonological feature in Czech, the vowel /i:/ in *free* was not shortened in the Czech version (right).

Diphthongs were treated differently from monophthongs. Here, the main principle according to which the manipulations were completed was a different duration ratio between vowel constituents of the diphthong in English, where the ratio tends to be 2:1 or even 3:1, and in Czech, where the ratio of the two constituents is around 1:1. Therefore, in the English version the initial component was lengthened and the second component shortened, while in the Czech version the constituents were manipulated to sound equal in duration, usually by shortening the initial component and prolonging the second component. One instance of such manipulation may be seen in Figure 9, where the temporal manipulation of diphthong /eɪ/ in *break* is shown; in the English version (left) /e/ is lengthened (*Duration manip* axis moved up) and /ɪ/ is shortened (axis moved down) creating the impression of a 3:1 ratio; in the Czech version (right) /e/ is shortened (axis moved down) and /ɪ/ lengthened (axis moved up) to produce a balanced

1:1 ratio. If, however, the diphthong appeared before a fortis consonant, the attention was paid not to create a diphthong too long in the English version.

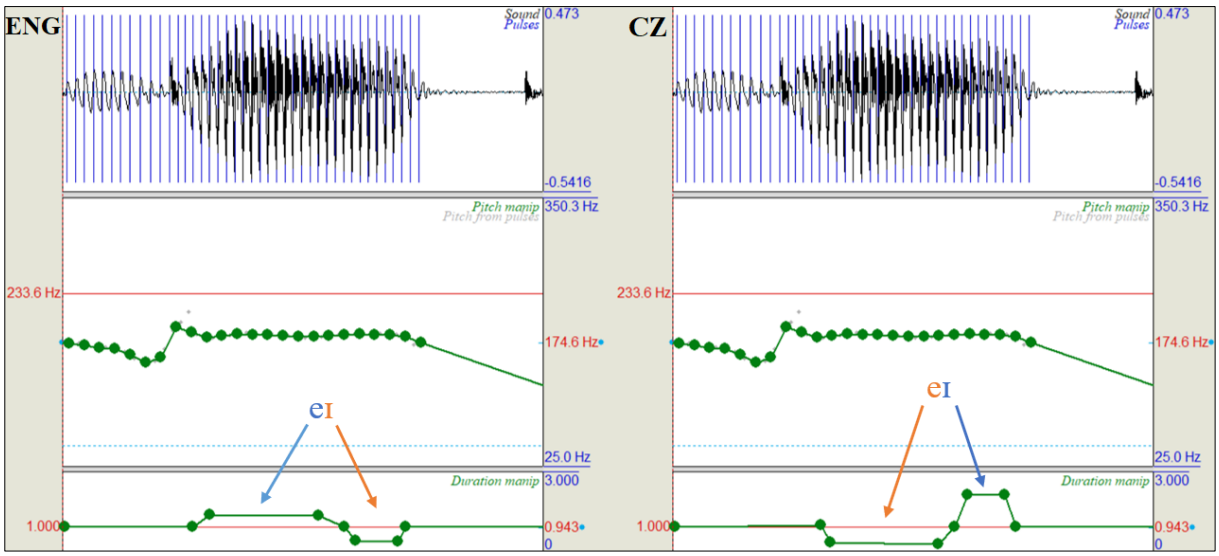

Figure 9: The manipulation of diphthong /eɪ/ in the English (left) and Czech (right) version. Blue arrows point to the lengthened segments, orange arrows point to the shortened segments.

As stated earlier, temporal manipulation has been completed in case multiple grammatical words occurred in a chain. Presuming that all of these syllables are unstressed, the idea was that in the English version, all words could be taken as one segment which may be sped up, reducing the prominence of all of its syllables while simultaneously compressing the words in between stressed syllables. In the Czech version, taking these words as one segment and lengthening it (the exact opposite of the manipulation in the English version) was not possible as the lengthening of consonantal obstructions may create various distractive noises resulting in an unnaturally-sounding recording. Grammatical words therefore had to be treated individually, lengthening only their peak vowels. The difference between the manipulations of groups of grammatical words in the English and Czech version is shown in Figure 10.

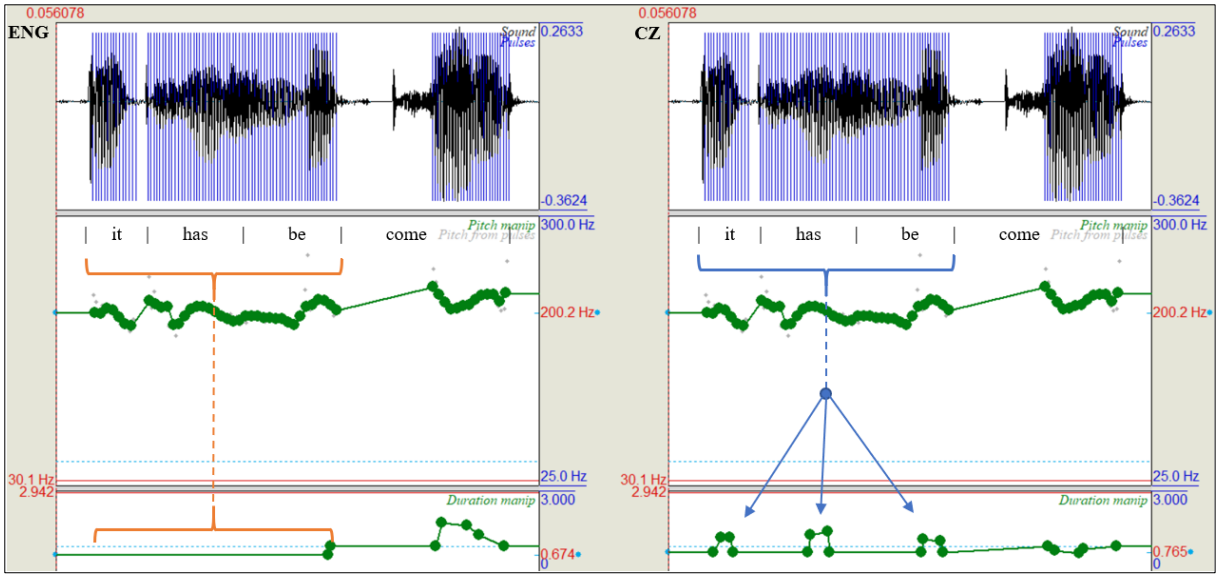

Figure 10: The temporal manipulation of a chain of grammatical words in the English (left) and Czech (right) version.

After both English and Czech manipulations have been created, their overall comparison and correction was necessary. This is due to the fact that various temporal manipulations often caused the English and Czech version of one phrase to differ significantly in the total duration. Instead of focusing on the rhythm of language, the listeners could be influenced by the discrepancy in the total duration of the two recordings; this could affect the results of the perception test, since it is known that faster speech is evaluated as more proficient. To prevent this from happening, it has been decided that the difference between total duration of the English and Czech version ideally should not be more than 20 ms. This was completed by taking all duration points of the recording (or just the duration points of its part) and simultaneously moving them up or down the axis to fasten or slow down the entire phrase. Figure 11 shows a phrase manipulated to imitate the Czech rhythm, but which was considered too slow after the final comparison with the English version; for this reason all duration points have been dragged down, so that the points initially occurring at the coefficient 1.0 of the relative duration were now at the coefficient 0.876.

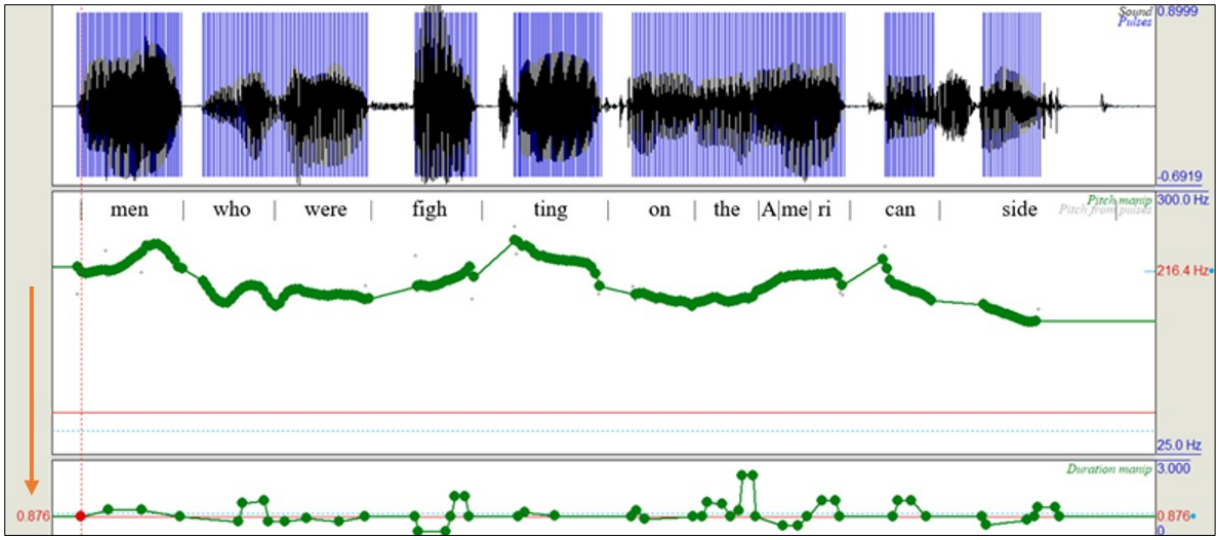

Figure 11: The entire phrase sped up by simultaneously moving all duration points lower in *Duration manip* field.

80 manipulation files prepared this way were converted into sounds using the function "get PSOLA resynthesis" and saved as WAV files. The individual sounds were then carefully examined with the purpose of finding and removing the rest of unwanted background noises which might distract the listeners. The set of sounds was then listened to at once to unify the volume of the recordings. All of these edits were again completed in Adobe Audition.

4.3 Perception test

The perception test was created using Praat's ExperimentMFC, which is suitable for the type of test chosen in this experiment, the *pairwise comparison* (two-alternative forced choice, 2AFC). Listeners were to be presented with two recordings and were to be asked which of the recordings satisfies a certain condition more. The decision that the perception test shall be a pairwise comparison was based on a recent experiment by Trčková (2019) who studied the effect of segmental and suprasegmental manipulations on accentedness and comprehensibility, wherein the *evaluative* type of test proved to be rather inadequate. In addition, a recent summary (McAleer & Belin, 2019) shows that evaluation of the same phenomenon on a Likert scale may yield different results from a pairwise comparison. Furthermore, the primary focus of the experiment was to examine how rhythmic structure influences the comprehensibility of Czech English; however, for the sake of better comparison with Trčková's (Ibid.) experiment, it has later been decided that the effect of rhythm on accentedness will be examined too.

The goal was to create a perception test that consists of two parts, where one part focuses on the comprehensibility and the other on the accentedness of speech. Each part contained 40 items and in each item the subjects listened to the Czech and English version of previously

manipulated recordings (this meant that the original, unmanipulated recordings were not used in the actual test). In both the part that focused on comprehensibility and the part focusing on accentedness, the subjects were initially being informed that one phrase will be said in two slightly different versions. Then, depending on the focus of the given part, the listeners were introduced to their task, i.e., either to decide which version sounds *more foreign-accented/more Czech-like* (accentedness) or to decide which of the two versions is *easier to understand/understandable with less effort* (comprehensibility). To prevent listeners from forgetting the task of the perception test, each item contained a briefly formulated instruction which served as the task reminder. In the item itself, the first recording was preceded by the initial silence of 1 second and it was followed by the medial silence of 1 second which represented the optimal pause between the two recordings. Listeners were allowed to repeat the pair of phrases up to three times (i.e., they could listen to the recordings up to four times); after the third repetition, the "*Replay*" button disappeared, and the subjects had to choose one of the versions. After doing so, the "*OK*" button appeared in the bottom right corner and upon its clicking the test proceeded with the next pair of phrases. Listeners were provided with the option of taking a short break after every 10 items. There was no maximum time set for this break, therefore the subjects decided individually how long it was going to be. The design of the perception test along with the instructions is shown in Figure 12.

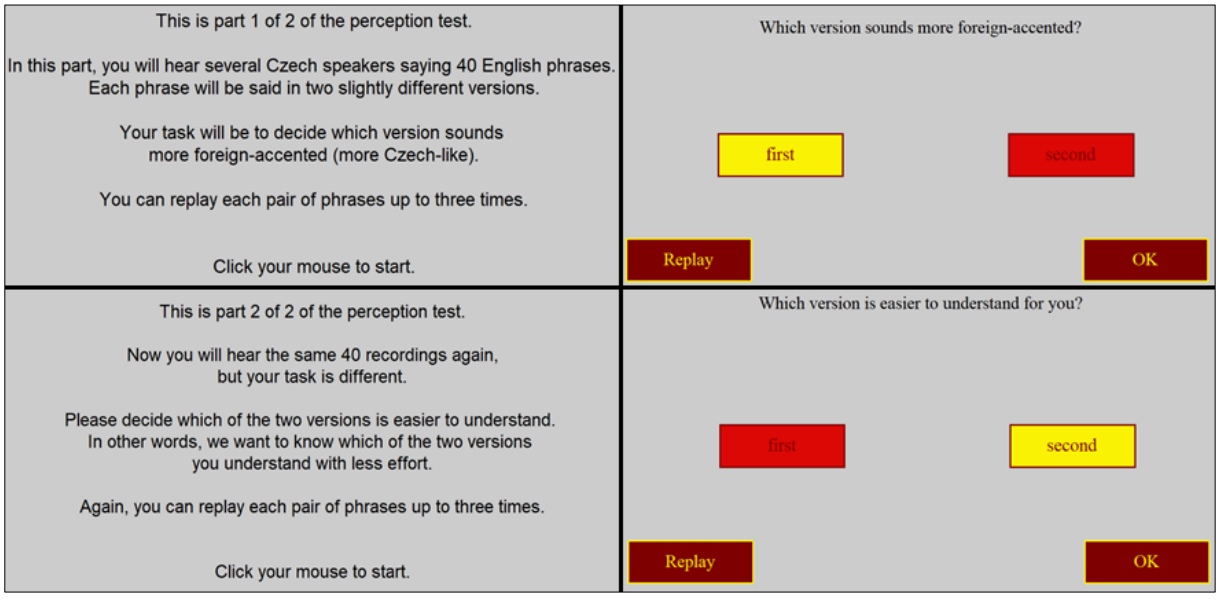

Figure 12: The design of the perception test

As the subjects were to be presented with the same set of pairs of 40 phrases in both the comprehensibility-focused and accentedness-focused part, we were challenged with a task of creating the perception test in such way so as to eradicate, as much as possible, the effects of

the following factors on the results of the test: the order of the two main parts (accentedness and comprehensibility), the order of the 40 stimuli in these two parts, and the order of Czech and English version within each item. In order to counterbalance all these effects, four versions of the perception test have been created.

Firstly, it was programmed in Praat script of the perception test that the order of 40 stimuli (i.e. 40 items) was to be random. The four versions of the test were named *"1A"*, *"1B"*, *"2A"*, and "*2B"*. The number signifies the order of the two main parts of the perception test: in the versions marked with number *"1"*, the subjects were first presented with the part focused on accentedness and second, with the part focused on comprehensibility. Opposed to that, in the versions marked with number *"2"*, the comprehensibility part preceded the accentedness part. The letter *"A"* or *"B"* then signifies the order in which the English and Czech version of a phrase would be presented. In the test version *"A"*, the first part contained items with a randomized order in which the English and Czech version of the recordings were presented; in the second part this order represented the exact opposite (e.g. if the first part presented the recordings in *CZ-EN* order, the second part presented them in *EN-CZ* order). In the test version *"B"*, the first part presented the English and Czech version of a phrase in order opposite to that which occurred in the first part of *"A"* version, while the second part of *"B"* presented Czech and English version in order opposite to that which occurred in the second part of *"A"* test version (e.g. if the order of recordings in the first part of *"A"* version was *CZ-EN*, the order in the first part of *"B"* was *EN-CZ*; and accordingly, if the second part of *"A"* presented recordings in *EN-CZ* order, the second part of *"B"* presented them in *CZ-EN* order). That way we secured counterbalance of factors which could influence the outcome of the experiment: the order of parts focused on accentedness and comprehensibility, the order of the 40 stimuli, and the fact that the subjects will hear the same recordings twice.

Furthermore, a preparatory part preceded the actual perception test; it contained two trial items and it served mainly for the subjects to become acquainted with the design of the test. There were two versions of the preparatory part, one focused on the accentedness and the other on the comprehensibility of phrases. The test version that the listener was to receive determined the version of the preparatory part, i.e., subjects whose test version was *"1A"* or *"1B"*, meaning that the first part of their test focused on accentedness, were given the accentedness-focused preparatory part; subjects with test version *"2A"* or *"2B"* with the comprehensibility-focused part being the first one, had been given the comprehensibility-focused preparation.

4.4 Subjects and experiment

The perception test was administered to 40 respondents, made up of two groups of 20 subjects, because we wanted to see how the subjects who may be expected to be more and less familiar with native English are sensitive to the temporal manipulations described above. First group, *Anglophones*, were the students of daily bachelor programme of Anglophone studies at the Faculty of Arts, Charles University. All of these students were native speakers of Czech whose ages varied from late teens to twenties and two subjects were in their forties. Due to their major, these subjects were expected to have a higher level of English language skills (minimum of C1 according to *Common European Framework of Reference for Languages*, or *CEFR*) and overall more frequent contact with the language in a daily life. It was presumed that their extensive practical and theoretical knowledge of English would make it easier for them to spot nuances between the Czech and English version. The second group, *others*, which consisted of the other 20 people, were the subjects of various backgrounds, but a vast majority of them were the students of other philological or non-philological majors at the Faculty of Arts (such as French and Hispanic studies, History, Psychology, etc.). The subjects of this group were native speakers of Czech too, and their level of English was not lower than B2. Their ages again varied from late teens to thirties. Each group was divided into smaller groups of five persons; depending on their group, the subjects were given one of the four versions of the test. Subjects were assigned to these group according to the order in which they had arrived at the testing location.

The bigger portion of testing took place in a quiet lecture room of the Institute of Phonetics at the Faculty of Arts. The schedule of subjects' arrivals was completed in advance using an online sign-up sheet, which enabled testing to progress smoothly and in a steady pace. 34 subjects have been given the test in this way (20 *Anglophones* and 14 *others*) and the remaining 6 persons from the group *others* have been tested in a calm and quiet environment in their homes. All forty subjects completed the test using one type of headphones, Sennheiser HD 201, and they were instructed to adjust the volume of the sound themselves so as to secure as much comfort as possible. The thesis author or supervisor were present at all times during the perception test in case the subjects (especially from the group *others*) needed assistance with understanding the instructions. In the end, each of the versions of the test (*"1A", "1B", "2A",* and *"2B"*) was completed by 5 *Anglophones* and 5 *others*.

4.5 Analysis

The results were extracted from Praat into an Excel table. From the total of 3,200 responses – 40 listeners × 40 stimuli × 2 focus groups (*accentedness* and *comprehensibility*) – it was first necessary to eliminate responses which were considered invalid. For this, we used the reaction time provided by Praat. It has been decided not to include answers that were submitted in under 5 seconds in the analysis: the shortest recordings were approximately 2.40 seconds long and each stimulus contained two recordings; if we consider that initial and medial silence make up 2 seconds, the shortest time in which the two recordings could be fully listened to is approximately 6.80 seconds. If we accept the possibility that the listener could have decided which of the recordings satisfies a particular condition (comprehensibility or accentedness) more without listening to the second recording till the end, it would have been possible to complete the task in 5 seconds, but not quicker. All results with the reaction time (which is the time from the moment a new stimulus appeared, till the moment the listener pressed the *"OK"* button) being less than 5 seconds (115 results) were therefore removed. The final table used for the analysis thus consisted of the remaining 3,085 results (examples of results are shown in Table 1).

subject	listener	focus	testorder	stimulus	first	response orig	response	speaker	reaction time	group
A01_1A-accent	A01	accentedness	acc-compr	KLIA-FMA-01EN,KLIA-FMA-01CZ	EN	first	-1	KLIA	9,023	Anglophone
A01_1A-accent	A01	accentedness	acc-compr	PAUA-EMA-03EN,PAUA-EMA-03CZ	EN	second	1	PAUA	10,095	Anglophone
NA11_2A-compr	NA11	comprehensibility	compr-acc	BMA-DL-01CZ,BMA-DL-01EN	CZ	second	1	BMA	9,467	other
NA11_2A-compr	NA11	comprehensibility	compr-acc	PLDA-AMA-01EN,PLDA-AMA-01CZ	EN	second	-1	PLDA	5,776	other

Table 1: Examples of the results used for the analysis: the listeners marked with "A" belong to the *Anglophone* group; the listeners marked with "NA" (Non-Anglophone) belong to the group *others*. Highlighted column shows whether the answer agrees (1) or disagrees (-1) with the hypotheses.

The column *response* in Table 1 marks whether the listeners' responses agree (1) or disagree (-1) with the hypotheses, which are: a) that English rhythmic structure improves the comprehensibility of Czech English, and b) that Czech rhythmic structure makes the speech seem more accented than in the English version. Specifically, in the part focusing on the comprehensibility (where the task was to choose the recording which is *easier to understand*), if the listener selected the English version the response was marked as "1" implying this result is in line with the hypothesis. If the listener selected the Czech version, the answer is in line with the hypothesis and was therefore marked as "-1". In the part focusing on the accentedness (where the task was to select the recording which sounds more *foreign-accented/Czech-like*), if the listener chose the Czech version, the response was marked as "1" (in line with the hypothesis) and if the listener's choice was the English version, the response was "-1" (not in

line with the hypothesis). For each analysed group the mean value and estimated confidence intervals were calculated using the bootstrap method with a significance level of 0.05 (Bonferroni-corrected for multiple testing). This means that a null hypothesis cannot be rejected if the confidence interval in the charts presented in the following section includes the value of 0.

5. Results and discussion

5.1 Accentedness

The results were processed using the programme RStudio. The same programme was used for creating individual graphs. Figure 13 shows the comparison of how the listeners from the *Anglophone* and *other* group selected the versions of phrases according to their accentedness. As stated earlier, value "1" corresponds to the answers that support the hypothesis (that speech approximating Czech rhythm is perceived as more foreign-accented than the speech approximating English rhythm), while the value "-1" does not support the hypothesis. In other words, the higher the confidence interval is positioned in the chart, the greater was the number of answers that agreed with the hypothesis. The mean value (further referred to as $\bar{x}$) of the *Anglophone* group is 0.20, and since the confidence interval does not include the value 0, this result may be considered as significantly in line with the hypothesis. Although the mean value of *Anglophones'* answers was expected to be higher, the results in general do show that the *Anglophone* listeners tended to select the version imitating Czech rhythm as more accented. On the other hand, the confidence interval of the group *other* includes the value 0 ($\bar{x}$ = 0.054), and the result therefore does not confirm nor reject the hypothesis. We may only say that there was a slight tendency for *others* to select the Czech version as more accented.

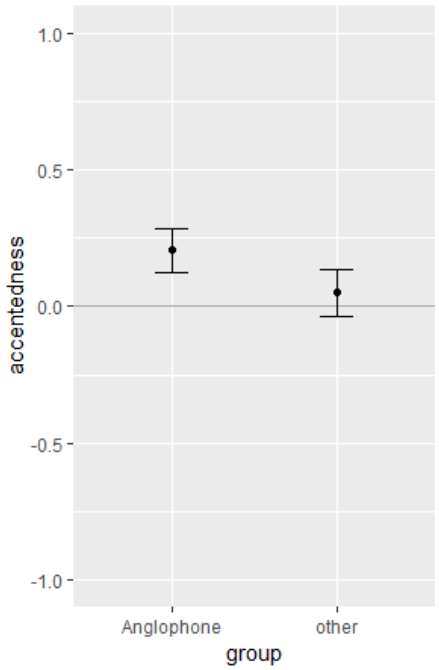

Figure 13: The mean values and their confidence intervals for the results of accentedness ratings.

5.2 Comprehensibility

The second of the two main test parts focused on the perceived comprehensibility of speech. Figure 14 shows the comparison of the results of the *Anglophone* and *other* group in this regard. The results of the *Anglophone* group show that these listeners tended to select the recordings manipulated into sounding more English (i.e., to approximate English rhythm) as more comprehensible ($\bar{x}$ = 0.13), which supports the hypothesis stating that the improvement of the rhythm of Czech English improves its comprehensibility. In regard to the group *others*, the confidence interval again includes the value 0 ($\bar{x}$ = -0.05), therefore this result is insignificant. However, it is interesting that there was a slight tendency for non-Anglophone subjects to actually choose the Czech version as more comprehensible, contrary to the hypothesis. This might be caused either by the respondents' lower level of English skills, or by a greater exposure to the Czech English with a strong Czech-like temporal structure. Due to this familiarity with English spoken with Czech rhythm, the listeners from the *other* group might find this form more comprehensible.

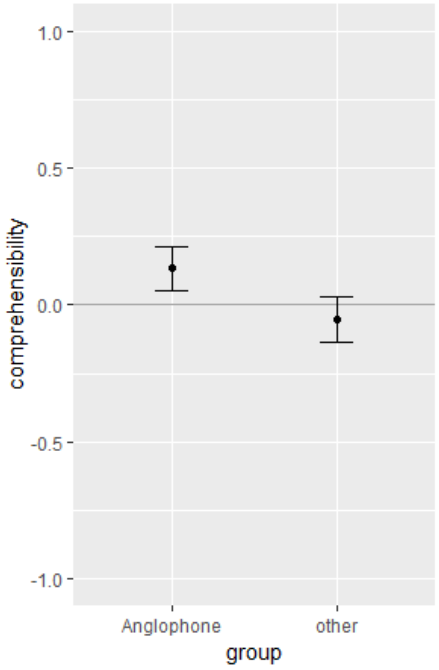

Figure 14: The mean values and their confidence intervals for the results of comprehensibility ratings.

5.3 Results based on the order of the accentedness/comprehensibility-focused part

We may consider the order in which the two main parts of the perception test were administered (one focusing on accentedness, the other on comprehensibility) as a factor which could affect the outcome of the experiment. As shown in Figure 15, this is indeed the case. Let us first consider the *Anglophone* group. Strangely, the respondents who had received the test versions *"1A"* or *"1B"* (in which the accentedness-focused part came first and the comprehensibility-focused part second, i.e., we are examining the left-most confidence intervals in both Figure 15A and B, shown in red) performed at both tasks more in line with the hypothesis than the respondents who were given test versions *"2A"* or *"2B"* (where comprehensibility preceded accentedness; shown in blue, second confidence intervals from the left). In other words, only one half of the *Anglophone* respondents are responsible for the statistically significant result of this group observed in Figure 13 and Figure ***14***.

At first sight, the reason for this stark difference in pairwise preferences is not apparent. However, we believe that these results were the outcome of the effect of various external factors, such as the way in which respondents were being assigned to the groups for the versions of the test. Test versions were being administered to the respondents in the order in which they came to complete the perception test. As the listeners given the versions *"1A"* and *"1B"* completed the test earlier in the day, whereas the *"2A"* and *"2B"* tests were completed by the listeners in the afternoon, it is possible that the tiredness of the respondents in the afternoon hours (though no respondent started the test after 5 p.m.) may have caused difficulties with concentration. A more important factor might be that several of the respondents with *"2A"* and *"2B"* version came to the testing location shortly after finishing an exam from another subject; a certain amount of stress that the respondents felt might have also affected their ability to concentrate. This would explain why the second half of *Anglophones* had altogether less favourable results (with respect to our hypothesis) than the first half of *Anglophones*. It is also conceivable that the respondents' levels of English differed between those who took test version *"1A"* or *"1B" vs. "2A"* or *"2B"* (although this is unlikely).

Regarding the group *others*, the results in Figure 15A show that there is a subtle tendency for these respondents to select the Czech versions of recordings as more accented if this task was presented in the second part of the test (although this tendency is insignificant). This applies also to the comprehensibility-focused test (Figure 15B), i.e., non-Anglophones to whom the task to choose a more easily understandable version was presented in the second part

of the perception test, were more likely to select the English-like versions. The aforementioned suggests that all subjects from the group *other* were able to perform the tasks that were administered in the second part more in line with our hypotheses than the tasks presented in the first part. This may be caused by the fact that in the second part, the respondents heard the identical set of recordings again, and even though the order of the stimuli was different, the subjects were already subconsciously trained to be more perceptive of the differences between phrase versions. They were therefore able to perform the task itself more favourably (with respect to our hypotheses). Furthermore, the listeners from the group *other* presented with the comprehensibility task in the first part tended to choose the Czech version significantly more often than the predicted English version ($\bar{x}$ = -0.16, the confidence interval does not include the value 0).

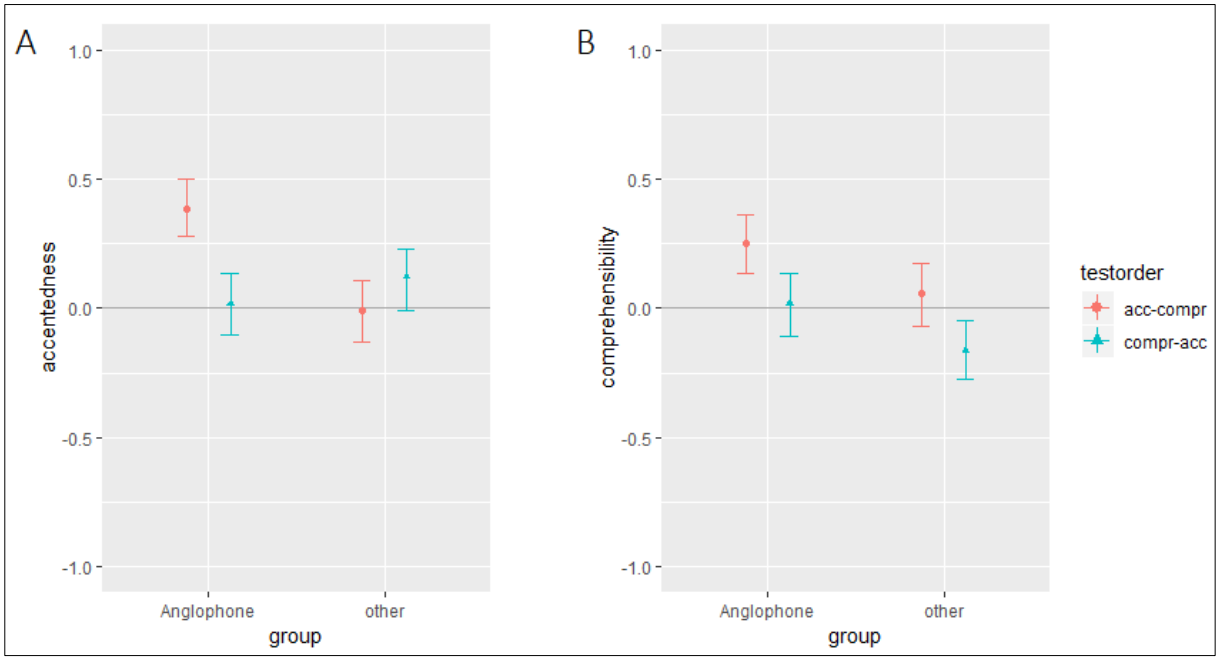

Figure 15: The mean values and their confidence intervals for the results of accentedness (A) and comprehensibility (B) ratings depending on the order in which the two test parts were administered. The order in versions *"1A"* and *"1B"* was accentedness-comprehensibility, and in version *"2A"* and *"2B"* comprehensibility-accentedness.

5.4 Speaker-dependent results

It is easily imaginable that phrases provided more scope for effective manipulations for some speakers and less so for others, and that this would be reflected in the respondents' assessments. That is why in this section we are looking at results for individual speakers. Figure 16(A) and (B) show accentedness and comprehensibility results for each of the 10 speakers. From the Figure 16(A) it is apparent that the accentedness results of the *Anglophone* group

regarding four speakers (JABA, MUPA, SSA, and VLHA) significantly confirm our hypothesis; however, all the remaining results of *Anglophones* and *others* include the value 0, thus are insignificant. From the four listed speakers with significant results the one whose ratings were most in line with our hypothesis was the speaker SSA ($\bar{x}$ = 0.39), i.e., SSA's Czech versions were most often selected as more foreign-accented by *Anglophones*. Other significant results were obtained from *Anglophones*' ratings of MUPA ($\bar{x}$ = 0.33), JABA ($\bar{x}$ = 0.32), and VLHA ($\bar{x}$ = 0.28). The group *other* had the most positive (but insignificant) results accentedness-wise with the speakers BMA ($\bar{x}$ = 0.24) and PLDA ($\bar{x}$ = 0.24). On the contrary, the speaker with the biggest number of English versions selected as more foreign-accented (across both groups of subjects) was MPA ($\bar{x}_{Anglophones}$ = -0.06; $\bar{x}_{others}$ = -0.23), although, as already stated, this result is insignificant. This did not come as a surprise because even during the primary selection of the recordings for manipulations, it was fairly difficult to find good excerpts from the set of recordings of this particular speaker, who, as implied in section 4.1, we later realized had a mild speech impediment which affected her /r/ sounds.

The Figure 16(B) shows speaker-dependent results in regard to comprehensibility. Again, the chart shows that the majority of results is insignificant, meaning they do not confirm nor reject the hypothesis. We may see that only the results of two speakers (VLHA and SSA) rated by the *Anglophone* group are significantly in line with our hypothesis ($\bar{x}_{VLHA}$ = 0.41; $\bar{x}_{SSA}$ = 0.32). This figure also shows that the results of one speaker, JABA ($\bar{x}$ = -0.29), which were obtained from the responses of *others* significantly oppose our hypothesis. This speaker had the largest number of Czech versions chosen as more comprehensible instead of the predicted English versions by *others*. We examined JABA's manipulated recordings in attempt to explain what the cause of this trend might be, but no atypical features or other abnormalities were found.

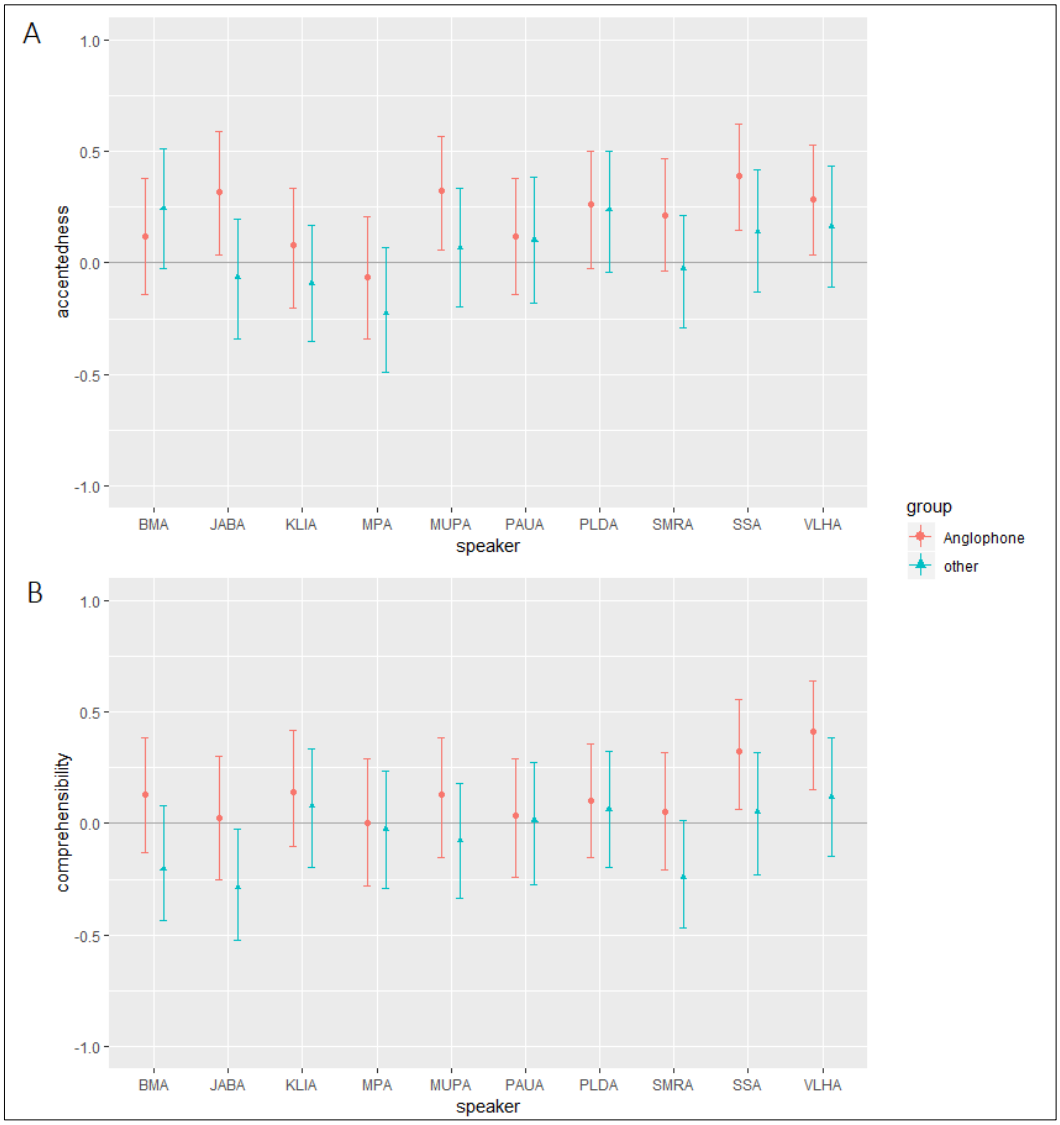

Figure 16: The mean values and their confidence intervals for the results of accentedness (A) and comprehensibility (B) ratings depending on the speaker.

5.5 Listener-dependent results

Similarly, just as the results may vary across individual speakers, so can they depend on the listeners; hence, this section present results in regard to individual subjects to whom the perception test was administered. Figure 17(A) and (B) show the results of the accentedness-focused part. Let us first examine the Figure 17(A) which focuses on the *Anglophone* group. It may be seen that the confidence intervals of twelve *Anglophones* indicate insignificant results. On the other hand, the results of six *Anglophones* are significantly in line with our hypothesis that Czech-like rhythmic structure increases the perceived accentedness of speech; these are: A01 ($\overline{x}$ = 0.65), A05 ($\overline{x}$ = 0.65), A09 ($\overline{x}$ = 0.65), A06 ($\overline{x}$ = 0.60), A10 ($\overline{x}$ = 0.50), and A03 ($\overline{x}$ = 0.45). On the opposite side of the accentedness chart (i.e. below the value of 0) in Figure 17(A)

are the listeners whose responses significantly are not in line with the accentedness hypothesis; there are two: A19 ($\bar{x}$ = -0.58) and A17 ($\bar{x}$ = -0.35).

Furthermore, Figure 17(B) shows the results of subjects from the group *others* ("NA" in the graph stands for "Non-Anglophone"). We may once again notice that the majority of *others* show insignificant results (seventeen subjects). This being said, the only two listeners whose results significantly agree with the accentedness hypothesis are NA17 ($\bar{x}$ = 0.50) and NA07 ($\bar{x}$ = 0.40). On the contrary, the results of one listener, NA05 ($\bar{x}$ = -0.70), significantly oppose the hypothesis.

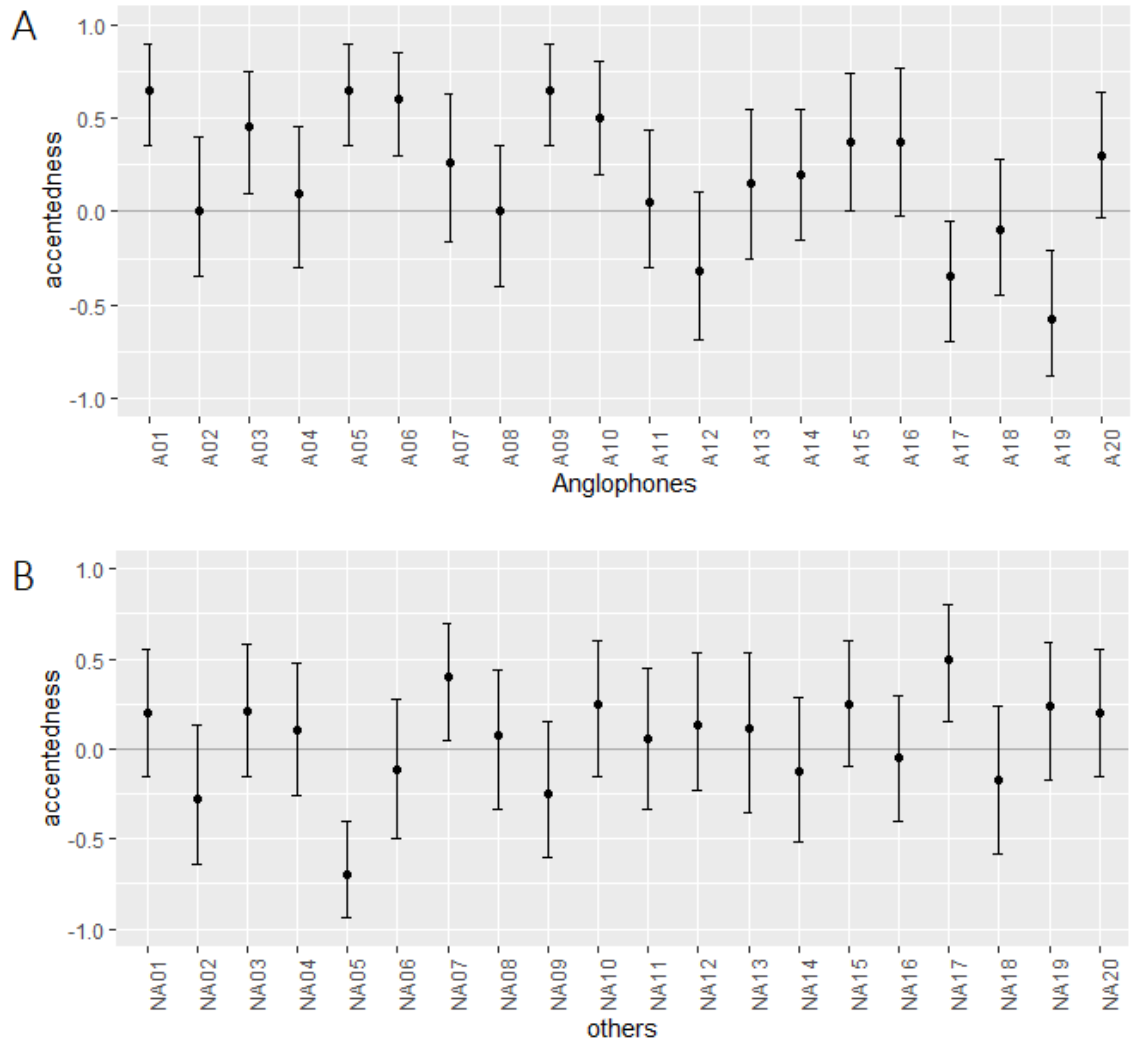

Figure 17: The mean values and their confidence intervals for the results of accentedness ratings depending on the listeners from the *Anglophone* (A) and *others* (B) group.

We shall now move to Figure 18(A) and (B) which show mean values and their confidence intervals of individual listeners on the comprehensibility chart. We may see that a greater portion of the results of both groups is insignificant. Specifically, in Figure 18(A), which shows the results of *Anglophones*, there are seventeen insignificant results. The three remaining

results are significantly in line with the comprehensibility hypothesis (that the speech approximating English rhythmic structure is understood with more ease than the speech which approximates Czech rhythm); these listeners are A05 ($\overline{x}$ = 0.59), A09 ($\overline{x}$ = 0.55), and A15 ($\overline{x}$ = 0.35). Thus, no *Anglophone*'s result significantly disagreed with the hypothesis.

Lastly, the comprehensibility chart of *others* may be seen in Figure 18(B). From a total number of twenty results of non-Anglophones, eighteen results prove to be insignificant (they include the value 0). From the two remaining significant results only that of the speaker NA06 shows significant tendency towards selecting English versions as more comprehensible ($\overline{x}$ = 0.38). Opposite to that, the only non-Anglophone respondent whose results significantly oppose the comprehensibility hypothesis is NA20 ($\overline{x}$ = -0.40).

A few rather interesting observation have been made; firstly, NA17 whose result on the accentedness chart was significantly in line with the hypothesis ($\overline{x}$ = 0.50) had a mean value of -0.20 on the comprehensibility chart which is one of the lower (though insignificant) scores. Thus, there is a slight discrepancy between the respondent's mean values in both charts. Secondly, as evident from the values stated above, the respondent with the most favourable results (with respect to our hypothesis) in electing which version of the phrase is more foreign-accented or comprehensible based on its rhythmic structure is A05. Besides being a student of Anglophone studies, this respondent has a lot of experience with music, being able to sing and play various instruments. Since the ability to perceive rhythm is something all musicians possess, musical skills might represent a certain advantage in tasks based on differences in the rhythmic structure of speech.

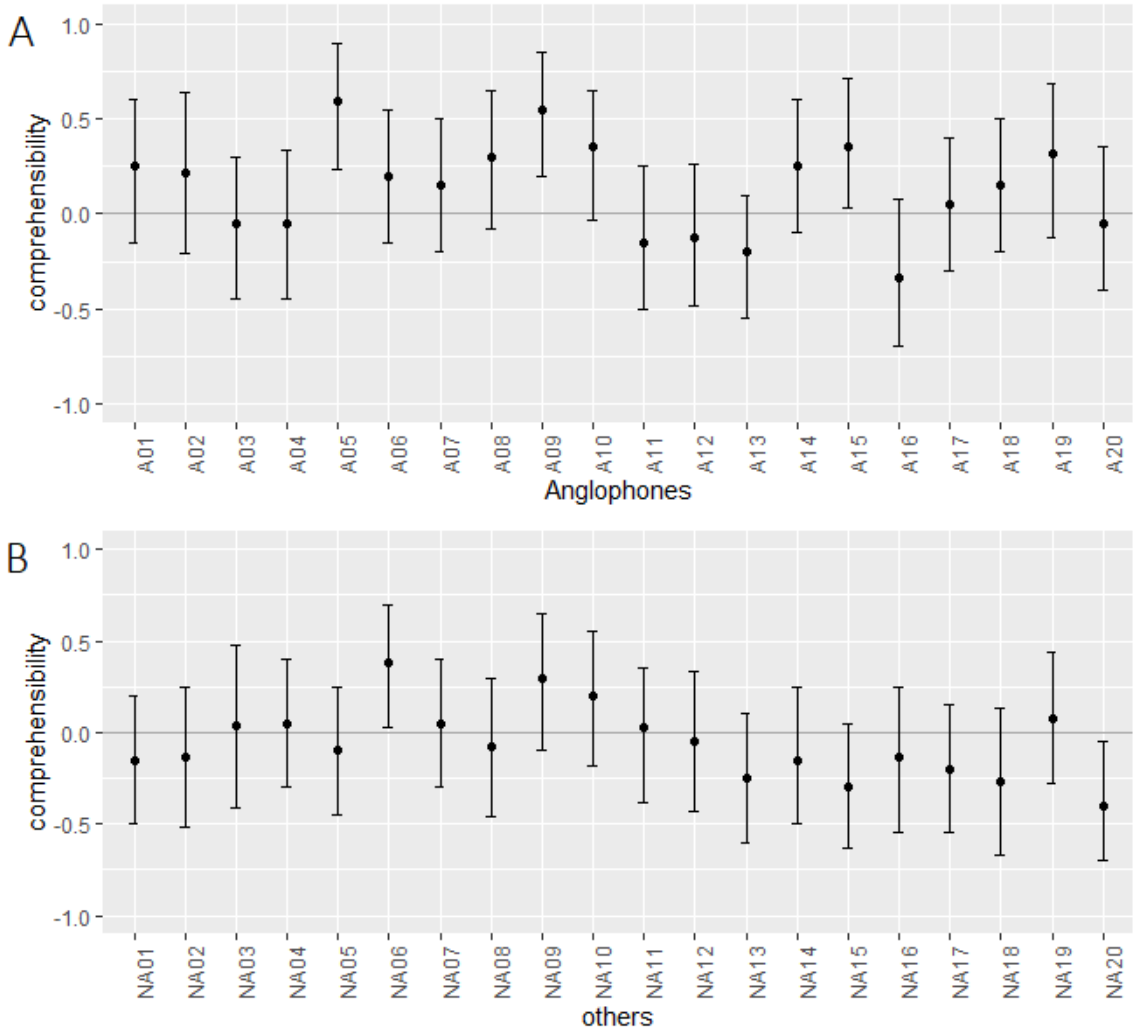

Figure 18: The mean values and their confidence intervals for the results of comprehensibility ratings depending on the listeners from the *Anglophone* (A) and *others* (B) group.

6. General discussion

The main aim of this thesis was to examine whether the rhythmic structure of speech affects the perceived comprehensibility (and accentedness) of Czech English. The previous section presented the specific results of the perception test, thereby in this section, these results shall be discussed within a larger context. A few notable observations have been made; firstly, only the group of *Anglophones* showed results which were statistically significant in regard to both comprehensibility and accentedness; contrary to that, the results of *others* reveal mere tendencies (see Figure 13 and Figure ***14***). This affirms our assumption that the listeners' familiarity with English correlates with their sensitivity towards perceiving slight differences in temporal organization of speech manipulated in Praat; the results imply that only the speakers whose English is approximately at C1 or higher, and who are in direct contact with English on a daily basis, are sensitive to differences in the duration of segments.

Essentially it may be said that both hypotheses 1) and 2) have been partially confirmed (*cf.* chapter 3); Czech English whose rhythmic structure was improved by imitating the rhythm of native English speech was indeed more comprehensible (for the *Anglophone* group) than Czech English whose rhythm was changed by emulating even more the Czech-like rhythm. Concurrently, the versions approximating Czech rhythm were mostly perceived as more foreign-accented. To a certain extent, the findings of this experiment support the idea proposed by Tajima et al. (1997) who claimed that with the training focused on the temporal features of English, the intelligibility of learners' speech might increase. The outcome of this experiment suggests that a training of this sort may improve the perceived accentedness and comprehensibility of the L2 speech as well.

The experiment has overall reached a rather satisfactory result, although initially we expected the differences between Czech and English phrases to be more apparent, especially among the *Anglophone* group. It is therefore fitting to reassess what might be the causes of the results' lower significance. First of all, as has been stated in section 4.1, some recordings chosen for manipulations proved to be less suitable than others. We had to work with the recordings which were available in the database, and identifying four phrases per speaker where the speech would be fluent and contain words which would fulfil the criteria described in section 4.2 was sometimes difficult. The desired differences between the English and Czech version may therefore have been less apparent to the listeners. One might speculate that the English versions of recordings would perhaps end up having a slightly different temporal structure if they were

revised also by a native English speaker; however, due to the meticulous procedure described earlier we consider this scenario unlikely.

Moreover, it is important to note that the comprehensibility of speech and sensitivity towards noticing subtle differences in speech rhythm are affected by external factors associated with the current mental state of listeners and by the conditions in which a perception test like the one presented in this study is carried out. Indeed, this has been implied in several studies listed by Tajima et al. (1997: 4). Specifically, the results of the *Anglophone* group hinted that students who completed the perception test probably while being either tired or nervous were less perceptive of the differences in the temporal structure of phrases (see section 5.5).

Let us next turn to the difference between the results of the *Anglophone* and *other* group. As has been implied, the results in Figure 13 and Figure ***14*** show that *Anglophones* were more perceptive of the nuances in the temporal structure of the recordings, thus confirming the hypothesis 3): the familiarity with native English indeed facilitates the perception of speech rhythm. Our assumption is that it is possible that only after acquiring high skills in English, foreign learners are able to perceive subtle differences in prosodic features which they were unable to attend to earlier. This may be explained by the differences between the L1 and L2. Specifically for this thesis, Czech and English differ within the scope of prosodic typology, which was discussed in section 2.1.1. Despite the fact that both languages are classified as *stress* languages, they differ significantly in their treatment of prosodic features, such as stress (and therefore, rhythm). The rhythmic structure of English is conveyed by stress which highlights the most important elements, thus facilitating comprehensibility. However, stress does not take on this role in Czech (due to its fixed initial position in a word), therefore, perhaps native Czech listeners are not inclined towards reaching for a correct rhythmic structure when trying to understand English speech. Broselow and Kang (2013: 547) use the term *stress deafness* to describe the situations in which "learners fail to attend to stress in the L2 input because L1 stress is fully predictable". This indicates that Czech learners would certainly profit from a training that would cultivate their ability to "hear" the stresses; they would put this ability to use when recognizing the most important elements of the information structure, not only in English, but also in other target languages.

7. Conclusion

The experiment presented in this thesis was conducted with the intention to examine what are the effects of rhythmical structure on the comprehensibility and perceived accentedness of Czech English. In this regard, we have designed a perception test in the form of a *pairwise comparison*, in which Czech listeners divided into two groups according to their familiarity with English (*Anglophones* and *others*) were presented with two versions of English phrases recorded by Czech speakers. These two version were created using "manipulate" function in Praat; one version was temporally manipulated to emulate the rhythmic structure of native English, whereas the second was manipulated to imitate even more the rhythmic structure of Czech. The perception test was administered in two parts, one focusing on the comprehensibility and the other on the perceived accentedness of speech; the listeners were asked to decide which of the two versions is *more comprehensible/understood with more ease* or which is *more foreign-accented/Czech-like*, respectively.

Chapter 2 of the thesis, the *Theoretical background*, was divided into two main parts; the section 2.1 defined *prosody* and characterized its role during Second Language Acquisition; the section 2.2 discussed more in detail the prosodic feature of *rhythm*. Here we have introduced several principles of English rhythm, such as Dickerson's *two-peak profile* and Cauldwell's *functional arrhythmicality* (see 2.2.1), but most importantly we described the notion of *isochrony*, which we had employed in the practical part of the experiment. The practical part contained chapters 4 and 5; here we described the methodology of the experiment and discussed the outcome of the perception test. The overall results of the experiment show that the *Anglophone* group significantly selected the English-like versions as more comprehensible and Czech-like versions as more foreign-accented. The results of the group *others* proved insignificant. We may therefore say that the hypotheses have been partially confirmed: the rhythmic structure manipulated to approximate native English rhythm indeed enhanced the comprehensibility of Czech English; and the rhythmic structure manipulated to approximate he Czech rhythm did increase the impression of a foreign-accent; however the ability to attend to the subtle differences in the temporal structure of speech occurred only with advanced speakers of English (*Anglophones*). This also confirmed the research question 3 (see chapter 3). It should also be noted that the perception of rhythmic structure may be negatively affected by various external factors, such as emotional stress or tiredness (see 5.5).

Overall, the findings of this experiment are in line with the previous research which showed that the improvement of the temporal structure of speech significantly improves either

its comprehensibility or intelligibility (see 2.1.2 and 2.2). Similar to what Trčková (2019) implied in her research, a future experiment focused on the effect of rhythmic structure on the *intelligibility* of speech could prove to be highly informative. Lastly, future research oriented on the rhythm of speech, or other prosodic features, might additionally consider the aforementioned external factors as aspects that might influence the ability to perceive slight nuances in speech.

References

Atagi, E. & Bent, T. (2011). Perceptual dimension of nonnative speech. In: *Proceedings of the 17th International Congress of Phonetic Sciences (ICPhS XVII): August 17-21, 2011*, pp. 260-263. Hong Kong: City University of Hong Kong.

Berkovcová, Z., Černikovská, Š. & Skarnitzl, R. (2016). Vliv temporálních manipulací na vnímání kompetence mluvčího [Effect of temporal manipulations on the perception of speaker competence]. *Studie z aplikované lingvistiky, 1,* pp. 7-19.

Boersma, P., Weenink, D. (2016). *Praat: Doing Phonetics by Computer* (version 6.0.50). Retrieved 25 September 2016 from www.praat.org.

Boula de Mareüil, P. & Vieru-Dumulescu B. (2006). The contribution of prosody to the perception of foreign accent. *Phonetica*, *63*, pp. 247-267.

Broselow, E. & Kang, Y. (2013). Phonology and Speech. In: Herschensohn, J. & Young-Scholten, M. (Eds.), *The Cambridge Handbook of Second Language Acquisition*, pp. 529-553. New York: Cambridge University Press.

Cauldwell, R. (2002). The functional irrhythmicality of spontaneous speech: A discourse view of speech rhythms. *Apples – Journal of Applied Language Studies, 2.1,* pp. 1-23.

Dankovičová, J. & Dellwo, V. (2007). Czech speech rhythm and the rhythm class hypothesis. In: *Proceedings of the 16th International Congress of Phonetic Sciences ICPhS XVI, 6-10 August 2007,* pp. 1241-1244.

Dickerson, W. B. (2016). A practitioner's guide to English rhythm: A return to confidence. In: Levis, J., Le, H., Lucic, I., Simpson, E., & Vo, S. (Eds), *Proceedings of the 7th Pronunciation in Second Language Learning and Teaching Conference*, pp. 39-50. Ames, IA: Iowa State University.

McAleer, P. & Belin, P. (2019). The perception of personality traits from voices. In: Frühholz, S. & Belin, P. (Eds.), The Oxford Handbook of Voice Perception, pp. 585–605. Oxford: Oxford University Press.

Mennen, I. & de Leeuw, E. (2014). Beyond segments: Prosody in SLA. *Studies in Second Language Acquisition, 36*, pp. 183-194.

Nolan, F. & Jeon H.-S. (2014). Speech rhythm: a metaphor? *Philosophical Transactions of the Royal Society B, 369,*pp. 1-11.

Palková, Z. (1994). *Fonetika a fonologie češtiny*. Prague: Karolinum.

Ravignani, A. & Madison, G. (2017). The paradox of isochrony in the evolution of human rhythm. *Frontiers in Psychology, 8.1820*, pp. 1-13.

Roach, P. (2009). *English Phonetics and Phonology: A Practical Course* (4th ed.). Cambridge: Cambridge University Press.

Rumlová, J. (2018). Phonetic features of strong Czech accent in English. Prague: Faculty of Arts, Charles University. Unpublished BA thesis.

Tajima, K., Port, R. & Dalby, J. (1997). Effects of temporal correction on intelligibility of foreign-accented English. *Journal of Phonetics, 25,* pp. 1-24.

Trčková, D. (2019). Comparing the effect of segmental and prosodic manipulations on speaker's accentedness and comprehensibility. Prague: Faculty of Arts, Charles University. Unpublished BA thesis.

Volín, J. (2010). On the significance of the temporal structuring of speech. In: Malá, M & Šaldová, P. (Eds.), *...for thy speech bewrayeth thee (A Festschrift for Libuše Dušková).* Prague: Faculty of Arts, Charles University, pp. 289-305.
Volín, J. (2017). Appeal and disrepute of the so-called global rhythm metrics. *AUC Philologica 3*, pp. 79-94.

Resumé

Tato bakalářská práce se zabývá vlivem rytmické struktury na srozumitelnost řeči. Konkrétně je zkoumán vliv temporálních manipulací na subjektivně vnímanou srozumitelnost české angličtiny. Temporální organizace nahrávek deseti českých mluvčích angličtiny byla upravena tak, aby výsledkem manipulací každé nahrávky byly dvě verze s odlišnou rytmickou strukturou: v první verzi byla rytmická struktura upravena tak, aby se co nejvíce podobala rytmu rodné angličtiny, zatímco ve verzi druhé byl rytmus zmanipulován tak, aby se jeho výsledné znění co nejvíce podobalo rytmu češtiny. Tyto nahrávky byly poté použity v percepčním testu, z jehož výsledků byly vyvozeny patřičné závěry.

Bakalářská práce je rozdělena na část teoretickou a praktickou. Teoretická část je uvozena definicí a funkcemi prozodie (tj. suprasegmentálních jevů), stručným popisem prozodických prostředků jazyka a prozodické typologie se zaměřením na jejich roli v procesu osvojování jazyků (2.1). Produkce správné prozodie při studiu cizího jazyka je velmi náročným avšak nutným úkonem při snaze dosáhnout mluveného projevu podobného projevu rodilých mluvčích. Odhaduje se však, že pouze minimum studentů je ve skutečnosti schopno takové prozodie dosáhnout. Dle předešlých výzkumů rodný jazyk značně ovlivňuje percepci i produkci prozodie jazyka osvojovaného. Broselow a Kang se toto snaží objasnit řazením jazyků do několika prozodických kategorií, v rámci kterých jsou čeština a angličtina řazeny k jazykům kladoucím důraz na přízvuk (*stress languages*), funkce přízvuku je však v obou jazycích; přízvuk v češtině značí začátek slova, zatímco angličtina jím klade důraz na prominentní prvky v rámci informační struktury. V této sekci uvádím několik předešlých výzkumů, které prokázaly, že podobnosti v prozodickém systému rodného a osvojovaného jazyka (konkrétně v distribuci přízvuků) představovaly pro studenty cizích jazyků značnou výhodu.

V podkapitole 2.2 je pak diskuze vztažena ke konkrétnímu suprasegmentálnímu jevu, kterým je rytmus. Pojednává se zde nejen o rytmu řeči, ale i o obecné preferenci lidí (a organismů obecně) k rytmičnosti, což je zapříčiněno tzv. principem nejnižšího úsilí; tento princip odráží obecný fakt, že pravidelnost je oproti nepravidelnosti vnímána jako „snazší“. Sekce 2.2.1pojednává o izochronii, která je definována jako rytmická struktura, v níž všechny intervaly mají podobné trvání. Dle tohoto principu se angličtina řadí mezi jazyky, které počítají přízvuky neboli *stress-timed languages*: všechny přízvuky v promluvě se objevují v časově podobných intervalech a aby délka těchto intervalů byla dodržena, všechny nepřízvučné slabiky jsou zkomprimovány, a tudíž zkráceny. Čeština pak patří mezi jazyky, které počítají slabiky neboli *syllable-timed languages*: všechny slabiky (přízvučné i nepřízvučné) se objevují zhruba

ve stejných intervalech. Toto obecné rozdělení je však rozporováno řadou vědců, kteří zároveň navrhují další způsoby, jak rozumět rytmu jazyků. V případě rytmu angličtiny byl navržen model *two-peak profile* – tento princip udává, že anglický mluvený projev v rámci jedné fráze zdůrazňuje pouze jeden či dva informačně nejdůležitější prvky a ostatní elementy redukuje. Ještě před touto teorií byla Cauldwellem navržena teorie angličtiny jako *funkčně arytmické*; nejdůležitější prvky promluvy, které se neobjevují se v pravidelných intervalech, jsou v promluvě zvýrazněny. Pro účely této práce přesto operujeme s izochronním pojetím rytmu: označení *stress-timed* a *syllable-timed* jsou striktní, avšak angličtinu lze považovat za více *stress-timed* a češtinu za více *syllable-timed*. Na základě tohoto předpokladu jsme v kapitole 3 stanovili hypotézy výzkumu: 1) česká angličtina, jejíž rytmická struktura napodobuje rytmus rodilých mluvčích angličtiny, bude vnímána jako více srozumitelná než česká angličtina, jejíž rytmus se blíží rytmu českému; 2) česká angličtina s rytmickou strukturou češtiny bude vnímána jako ta, která nese více patrný cizinecký přízvuk; 3) posluchači s vyšší znalostí angličtiny budou více schopni zaznamenat rozdíly v temporální struktuře dvou forem české angličtiny než posluchači s nižší znalostí angličtiny.

Po uvedení hypotéz přechází práce k části praktické, která je uvozena kapitolou 4, která stanovuje metodologii výzkumu. V sekci 4.1 je popisována kompilace materiálu: z archivu fonetického ústavu Univerzity Karlovy byly vybrány krátké anglické fráze deseti českých žen (čtyři fráze od každé mluvčí). Fráze byly vybírány s ohledem na jejich potenciál vykázat po dokončení temporálních manipulací co nejpříznivější výsledky. Najít vhodný materiál bylo často obtížné, proto musely být nahrávky v mnoha případech nejprve upraveny v programu Adobe Audition. Sekce 4.2 popisuje konkrétní manipulace v programu Praat, jejichž cílem bylo vytvořit dvě verze každé fráze: jednu, která se svým rytmem blíží rytmu češtiny (dále jako „česká verze“) a druhou, která se blíží rytmu angličtiny (dále jako „anglická verze“). Cílem bylo, aby v českých verzích byly slabiky přibližně stejně dlouhé, zatímco aby v anglických verzích byly přízvučné slabiky prominentnější (tedy delší) a nepřízvučně slabiky redukované (tedy kratší). Toho bylo docíleno nejprve zadáním bodů na osu relativního trvání a následným prodlužováním (posunem osy nad koeficient 1.0) či zkracováním (posunem pod 1.0) určitých prvků řeči. Mezi těmito prvky bylo již zmíněné trvání přízvučných a nepřízvučných slabik, a dále poměr trvání konstituentů diftongů (v češtině 1:1, v angličtině přibližně 2:1 či 3:1) či krácení/prodlužování celých skupin gramatických slov. V sekci 4.3 popisuji sestavení percepčního testu v programu Praat. Test byl rozdělen do dvou částí (první byla zaměřená na vnímanou srozumitelnost řeči, druhá na vnímaní cizineckého přízvuku). V každé části se nacházelo 40 stimulů, z nichž každý obsahoval verzi českou a verzi anglickou. Posluchači se

pak v závislosti na konkrétní části testu rozhodovali, která verze je více srozumitelná nebo která nese zřetelnější cizinecký přízvuk. Sekce také obsahuje popis sestavení testu tak, abychom zamezili ovlivnění výsledků externími faktory, kterými jsou pořadí stimulů či pořadí, ve kterém jsou předloženy dvě hlavní části testu. Tyto faktory byly vyváženy vytvořením čtyř různých verzí testu. Sekce 4.4 obsahuje popis respondentů percepčního testu a průběh testování. Test byl zadán dvěma skupinám (20 studentů anglofonních studií a 20 subjektů nestudujících anglofonní studia neboli „ostatních“) v tiché učebně Fonetického ústavu Univerzity Karlovy či v domovech respondentů. Při testu byla použita sluchátka Sennheiser HD 201. Sekce 4.5 pak obsahuje popis analýzy výsledků. Z celkového počtu 3 200 výsledků (40 posluchačů × 40 stimulů × 2 části testu zaměřené na srozumitelnost a cizinecký přízvuk) zbylo po odečtení výsledků odeslaných za dobu kratší než 5 sekund (jelikož splnění úkolu bylo v kratší době nemožné) celkem 3 085 výsledků, které byly následně zpracovány v programu RStudio.

Výsledky percepčního testu jsou uvedeny v kapitole 5. Jednotlivé grafy zachycují průměrné hodnoty a konfidenční intervaly všech analyzovaných skupin. Pokud konfidenční interval zahrnuje koeficient 0, výsledek hypotézu nepotvrzuje ani nevyvrací. Figure 13 zachycuje signifikantní tendenci skupiny *anglofonních* volit české verze nahrávek jako verze s více patrným cizineckým přízvukem, výsledky na Figure 14 pak zachycují signifikantní tendenci této skupiny volit anglické verze jako ty srozumitelnější. Oba výsledky jsou v souladu s hypotézami. Na druhou stranu výsledky skupiny *ostatních* jsou nesignifikantní, nelze z nich tedy vyvodit patřičné závěry. Zbylé sekce v této kapitole poukazují na další možné parametry výzkumu; zachycují výsledky v závislosti na a) pořadí dvou částí testu, b) mluvčích a c) posluchačích. Kapitola 6 obsahuje obecnou diskuzi, ve které jsou výsledky propojeny s poznatky z teoretické části; zároveň jsou zde uvedeny možné limity práce a další externí faktory, které mohly ovlivnit průběh či výsledky práce. Z výsledků vyplývá, že hypotézy byly potvrzeny částečně: anglická rytmická struktura skutečně napomáhá srozumitelnosti české angličtiny a zároveň česká rytmická struktura zvyšuje vnímání cizineckého přízvuku, avšak drobné rozdíly mezi temporálními strukturami dvou forem české angličtiny jsou podvědomě vnímány zejména posluchači ze skupiny *anglofonních*, kteří ovládají anglický jazyk na úrovni C1 a vyšší a jsou v kontaktu s angličtinou v každodenním životě. V budoucích výzkumech by bylo zajímavé zaměřit se na externí faktory ovlivňující vnímání rozdílů v temporální struktuře jazyka, jelikož výsledky této práce naznačují, že tato schopnost klesá, pokud jsou posluchači unavení nebo pokud pociťují stres.