UNIVERZITA KARLOVA
Filozofická fakulta
Ústav anglického jazyka a didaktiky

Bakalářská práce

Hana Hledíková

Prosodic Phrasing in Good Speakers in English and Czech

Prozodické členění u dobrých mluvčích v angličtině a češtině

Praha, 2019

doc. Mgr. Radek Skarnitzl, Ph.D.

Acknowledgements

I would like to thank my supervisor doc. Mgr. Radek Skarnitzl, Ph.D. for his guidance, advice and support, his endless patience and readiness to help, and for the incredible amount of time and energy he has dedicated to me throughout the development of this BA thesis.

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 .....

I declare that the following BA thesis is my own work for which I used only the secondary literature that is listed in the resources. This thesis was not used as a part of any other university study, nor was it used to gain a different university degree.

Prague, ...

Abstract

The aim of this BA thesis was to compare prosodic phrasing in good public speakers of English and Czech. Naive observations of English and Czech spoken in everyday communication suggest that Czech intonation is more flat and that Czech speakers divide the flow of speech into longer prosodic phrases than English speakers. We focused on the speech of good public speakers to see whether there are differences in the temporal and melodic characteristics between the two languages in this stylistic domain. We analysed segments from speeches by 10 TEDTalk speakers in Czech and American English and measured the length of prosodic phrases, speaking rate, standard deviation of the fundamental frequency in each prosodic phrase and in the nuclear part of the phrase (measure of pitch span), and Cumulative Slope Index in each prosodic phrase (measure of melodic variability). The number of syllables per prosodic phrase was found to be higher in Czech than in English, although phrases were generally very short in both languages. Speaking rate was found to be faster in Czech than in English. Pitch span in both the whole prosodic phrase and the nuclear part of the phrase was found to be wider in English than in Czech. Melodic variability was found to be higher in English than Czech. These results show that there are differences between Czech and English prosodic phrasing in good public speakers and that prosodic phrasing is influenced by speaking style.

Key words: Czech, English, prosodic phrasing, pitch span, melodic variability, speaking rate

Abstrakt

Cílem této bakalářské práce bylo porovnat prozodické členění u dobrých mluvčích angličtiny a češtiny. Na základě neformálního pozorování mluvené angličtiny a češtiny v každodenní komunikaci se zdá, že česká intonace je plošší a mluvčí češtiny rozdělují proud řeči na delší prozodické fráze než mluvčí angličtiny. Zaměřili jsme se na projev dobrých řečníků, abychom zjistili, zda jsou v této stylistické oblasti mezi těmito dvěma jazyky rozdíly v temporálních a melodických charakteristikách. Analyzovali jsme úseky z projevů 10 řečníků na konferenci TEDTalk v češtině a americké angličtině a změřili délku prosodických frází, mluvní tempo, směrodatnou odchylku základní frekvence v rámci prozodické fráze a v melodémové části (míra intonačního rozpětí) a Cumulative Slope Index v rámci prozodické fráze (míra melodické variability). Počet slabik za frázi byl naměřen vyšší v češtině než v angličtině, ačkoli fráze byly obecně velmi krátké v obou jazycích. Mluvní tempo bylo naměřeno rychlejší v češtině než v angličtině. Intonační rozpětí jak v rámci celé fráze, tak v melodémové části bylo naměřeno větší v angličtině než v češtině. Melodická variabilita byla naměřena vyšší v angličtině než v češtině. Tyto výsledky ukazují, že existují rozdíly v prozodickém členění mezi češtinou a angličtinou u dobrých řečníků a že prozodické členění je ovlivněno mluvním stylem.

Key words: čeština, angličtina, prozodické členění, intonační rozpětí, melodická variabilita, mluvní tempo

Table of Contents

1. Introduction 8
2. Theoretical Background 9
2.1 Prosody 9
2.1.1 Melody of speech 10
2.1.2 Temporal characteristics of speech 12
2.2 Prosodic phrasing 15
2.2.1 The Prosodic phrase 15
2.2.2 Prosodic Boundaries 16
2.2.3 Relationship between Prosodic Structure and Syntactic Structure 18
2.2.4 Other Factors Influencing Prosodic Phrasing 21
2.3 Prosody in Czech and English 24
3. Method 26
3.1 Material 26
3.2 Analysis 26
4. Results 29
4.1 Temporal characteristics 29
4.1.1 Number of syllables per prosodic phrase 29
4.1.2 Number of words per prosodic phrase 31
4.1.3 Speaking rate 33
4.2 *f0* characteristics 34
4.2.1 Standard deviation of *f0* in each prosodic phrase 34
4.2.2 Standard deviation of *f0* in the nuclear part of the phrase 36
4.2.3 Cumulative Slope Index in each prosodic phrase 37
5. Discussion 39
5.1 Temporal characteristics 39
5.2 *f0* characteristics 42
References 43
Resumé 47

List of Figures

Figure 1. Number of syllables per prosodic phrase depending on LANGUAGE (Czech × English) and TYPE OF PROSODIC BREAK (BI3 × BI4). 29
Figure 2. Number of syllables per prosodic phrase in BI4 without hesitation depending on LANGUAGE (Czech × English). 31
Figure 3. Number of syllables per prosodic phrase depending on LANGUAGE (Czech × English) and TYPE OF PROSODIC BREAK (BI3 × BI4). 31
Figure 4. Number of words per prosodic phrase in BI4 without hesitation depending on LANGUAGE (Czech × English). 32
Figure 5. Speaking rate depending on LANGUAGE (Czech × English) and TYPE OF PROSODIC BREAK (BI3 × BI4). 33
Figure 6. Speaking rate in BI4 without hesitation depending on LANGUAGE (Czech × English). 34
Figure 7. Standard deviation of *f0* in each phrase depending on LANGUAGE (Czech × English) and TYPE OF PROSODIC BREAK (BI3 × BI4). 34
Figure 8. Standard deviation of *f0* in each phrase ending with BI4 without hesitation depending on LANGUAGE (Czech × English). 35
Figure 9. Standard deviation of *f0* in the nuclear part of the phrase depending on LANGUAGE (Czech × English) and TYPE OF PROSODIC BREAK (BI3 × BI4). 36
Figure 10. Standard deviation of *f0* in the nuclear part of the phrase in BI4 without hesitation depending on LANGUAGE (Czech × English)..... 37
Figure 11.CSI in each prosodic phrase depending on LANGUAGE (Czech × English) and TYPE OF PROSODIC BREAK (BI3 × BI4). 37
Figure 12. CSI in each phrase ending with BI4 without hesitation depending on LANGUAGE (Czech × English). 38
Figure 13. Histogram of syllable counts per prosodic phrase for Czech and English. 39
Figure 14. Histogram of syllable counts per prosodic phrase without hesitations for Czech and English. 40

List of Tables

Table 1. Number of prosodic phrases with and without hesitation (“inter”)..... 30

1. Introduction

Prosodic features of speech, such as its melodic and temporal organization, have been shown to be crucial in the process of communication, playing an important role in the listener's comprehension of what is being said (e.g. Frazier, Carlson & Clifton, 2006) and the impression the speaker makes on the listener (e.g. Rosenberg & Hirschberg, 2009). This BA thesis focuses on prosody in Czech and English, specifically on prosodic phrasing and its melodic and temporal characteristics in the speech of good public speakers.

Our principal aim was to describe the use of prosodic phrasing by good speakers in the stylistic domain of public speaking and to find possible differences between the two languages, addressing a common opinion based on naive observation of Czech and English used in everyday conversation: that Czech prosody sounds flatter and more monotonous than English prosody. It may be supposed that there are language specific differences in Czech and English use of prosodic phrasing which may also be found in other speaking styles, such as public speeches given by skilled speakers.

As material for our analysis, we used segments of 20 TEDTalks (speeches on a wide range of topics given to a general audience, usually popularizations of scientific research, motivational speeches or personal stories), 10 in Czech and 10 in American English. We measured several acoustic features of the speech signal to find objective evidence that may account for the aforementioned impressions of differences between Czech and English prosody. We focused on the prosodic phrase, a basic unit into which the flow of speech is divided, and measured its length (in words and syllables), speaking rate (in syllables per second), standard deviation of fundamental frequency (in semitones) and Cumulative Slope Index (in semitones per second), all in the domain of the prosodic phrase. These measures, which capture the temporal and melodic characteristics of prosodic phrasing, are explained in detail in the Method section.

The following theoretical part will introduce basic terms and concepts relevant for our research and give some information about previous findings concerning prosodic phrasing. In the research part, we will describe the material and how it was chosen and analysed, the statistical methods used, and finally the results and the conclusions which can be drawn from them.

2. Theoretical Background

2.1 Prosody

Prosody is a term which describes the characteristics of speech above the level of individual phonemes. These include features such as melody, temporal characteristics (e.g. tempo, rhythm or length of prosodic constituents), loudness and voice quality, and their acoustic correlates – fundamental frequency (*f0*), duration, intensity, spectral characteristics.

These features characterize all spoken utterances and are not random or merely decorative, but carry important meanings and functions. For example, they organize the flow of speech and give it structure by dividing it into smaller units and grouping certain words together or separating them. The role of rhythm, which can be understood as a regular pattern of prominences, is also important in organizing the flow of speech so that it can be easily processed. Prosodic features can also help to indicate the syntactic or information structure of the sentence, carry information about the speaker's intention, give the listener cues to identify the speaker's attitude and stance to what is being said (e.g. help him to discern irony from a genuine statement or a joke from a serious statement) or indicate the speaker's emotional state.

Skarnitzl, Šturm & Volín (2016: 128–133) identify these general functions of prosody: lexical, grammatical, discourse, accentual, affective and indexical. The lexical function is only relevant for certain languages (such as Chinese or Vietnamese) where a tone is an integral part of a lexical unit, i.e. the use of different tones can distinguish between different lexical meanings. The grammatical function serves to indicate the syntactic and semantic structure of the sentence. This will be discussed in more detail in section 2.3. It also includes the indication of sentence type. For example, a typical English declarative sentence ends with a falling intonation, while a yes-no question ends with a rising intonation. In the case of declarative questions (e.g. *The restaurant is closed?*), rising intonation is the only feature that distinguishes them from a declarative sentence (*The restaurant is closed.*), as the word order is the same in both sentence types (in writing, the distinction is indicated by punctuation). The discourse function relates to regulating the course of communication between the communication partners. For example, it can be used to manage turn-taking (i.e. indicate whether we want to continue speaking or whether we want our partner to continue), to indicate which reaction is expected from the communication partner (e.g. whether we are asking a genuine question and expecting an answer, or merely looking for some indication of agreement with what has been said) or to imply that we want to end the conversation. Intonation can also be used to emphasise certain words with high information value and draw attention to them. This function is called accentual and may be subsumed under discourse function. We will discuss it in more detail in section 2.4. The affective function means that prosody carries information about the speaker's affective state (their emotions, moods, interpersonal stances,

attitudes and personality traits). To give an example how this function can be tested experimentally, Ladd, Silverman & Tolkmitt et al. (1985) investigated how manipulations of *f0* range, intonation contour type and voice quality affected listeners' judgements of the speaker's affect. They used 8-point scales to rate "arousal-related states" (relaxed/aroused, open/deceitful, annoyed/content, insecure/arrogant, indifferent/involved) and "cognitive attitudes" (emphasis, cooperativeness, contradiction, surprise, reproach) (Ladd et al., 1985: 437). Harsh voice quality and wide *f0* range led to higher ratings of arousal, annoyance and involvement in the arousal-related set of attitudes, and higher ratings of emphasis, contradiction and reproach in the cognitive set of attitudes. Harsh voice quality was also correlated to judgements of the speaker as less cooperative, more deceitful and more arrogant. *f0* range and contours were manipulated using resynthesis, voice quality manipulation was achieved by instructing the speaker to speak a certain way ("normal, relaxed, friendly", "annoyed, irritated, angry"), making it hard to control exactly across speakers as they may not produce comparable changes of voice quality. But the effects of *f0* range were clearly replicated across different speakers. Lastly, the indexical function of prosody describes the fact that prosodic features act as markers of the speaker's identity or a social group that he or she belongs to, e.g. his or her education, age, gender, etc.

2.1.1 Melody of speech

In this thesis, we will focus on some melodic and temporal characteristics of speech. The melody of speech, or intonation, refers to "patterned variation in voiced source pitch" (Beckman & Venditti, 2010: 603), that is to sequences of tones of different pitch realized during the course of an utterance. The acoustic correlate of perceived pitch is the fundamental frequency *f0*, which reflects the speed of vocal fold vibration (the higher the speed, the higher the fundamental frequency and the higher the perceived pitch). Although this means that *f0* is only present in voiced segments of speech, our perception of melody is continuous – as listeners, we are able to fill in the empty spaces in the melodic contour and do not perceive any interruptions (Skarnitzl, Šturm & Volín, 2016: 128). While perceiving speech melody, we also do not take into account microintonation, that is phenomena such as fluctuations in *f0* in voiced obstruents (consonants that are produced by creating a constriction in the vocal track while vocal folds are vibrating at the same time) or in the onset of phonation after a voiceless obstruent, which create small movements of *f0* irrelevant for the overall melody contour of the utterance (Skarnitzl, Šturm & Volín, 2016: 127). To analyse the intonation of recorded speech, the *f0* track can be extracted automatically by autocorrelation, smoothed out to exclude these small, perceptually irrelevant movements and interpolated to create a continuous contour without interruptions in unvoiced segments.

When describing speech melody and its patterns, the pitch track is represented using some kind of elemental units into which it is segmented. Different transcription systems use different basic units – for example, they can be defined as "melodically equivalent glissandi" (e.g. a rising movement or a falling movement), or we can choose to decompose these movements into successions of "endpoint notes" or tonal targets (e.g. a high tone or a low tone) and take these as the elemental units instead (Beckman & Venditti, 2010: 609). These units are anchored to "phonologically significant events" of two basic types: stressed syllables and edges of prosodic phrases (see section 2.1 for definition of prosodic phrase) (Beckman & Venditti, 2010: 607). The "Standard British" transcription system uses these basic units: "high and low versions of level, rise, fall, fall-rise and rise-fall," and combines them with single and double bars to mark edges of prosodic phrases (Roach, 1994: 93). The American ToBI transcription system (the acronym standing for "Tones and Break Indices") uses high and low tones assigned to stressed syllables (marked by *) or ends of prosodic phrases (marked by %). In addition to this, it also has a different level for marking the nature of breaks between words using numbers 0 to 4, going from fully linked (0) to clearly separated by a pause (4) (Roach, 1994: 95).

The difference between the lowest and highest values of *f0* used by a speaker is called a pitch range. More specifically, this difference, which reflects how wide the range of frequencies covered by the speaker is, is referred to as pitch span, while pitch level is a term used to describe how high or low the speaker's voice is set overall (Ladd, 1996; cited in Patterson, 2000). Pitch span is usually measured by looking at the long term distribution of *f0* values, measuring the difference between minimum and maximum, the difference between 90^{th} and 10^{th} percentile or 75^{th} and 25^{th} percentile, or standard deviation. Patterson (2000) criticizes this approach by pointing out that *f0* is not normally distributed around the mean and that *f0* movements are "fundamentally linked to tonal targets (…) i.e. that *f0* contours are structured at a phonological level" (Patterson, 2000: 36), meaning that they are linked to those basic units of intonation which we have talked about in the previous paragraph. He suggests an alternative approach to these global measurements by suggesting to measure *f0* span using certain linguistic targets, such as sentence initial highs and sentence final lows or non-initial accent peaks and post accent valleys. He gives evidence that this approach corresponds better to listeners' judgements. Compared to measurements of *f0* span based on the long term distribution of *f0*, which can be computed automatically, these measurements are more time consuming and labor intensive.

Although the basic unit of frequency is hertz (Hz), it is better to measure pitch span in semitones (ST), because semitones reflect the way differences in pitch are actually perceived. Because we perceive the same tone but an octave higher when its frequency is doubled, the absolute

difference between two tones expressed in hertz will be larger the higher the voice level gets, while it is perceived as the same difference in pitch. For example, the difference between one-lined A (440 Hz) and two-lined A (880 Hz) is 440 Hz, but the difference between two-lined A (880 Hz) and three-lined A (1760 Hz) is 880 Hz, that is the difference in hertz is twice as large in a higher octave, while the perceived difference in pitch is still the same – one octave, that is 12 semitones.

The range of frequencies that a speaker is physically able to produce (a speaker's vocal range) is larger than that which is actually used in speech (Skarnitzl, Šturm & Volín, 2016: 123). The pitch range used in speech has been observed to differ across languages. Andreeva, Demenko & Wolska, et al. (2014) found that German and English speakers have a lower pitch level and narrower pitch span compared to Bulgarian and Polish speakers. Mennen, Schaeffler & Docherty (2012) found female speakers of English to have a higher pitch span and level than female speakers of German. They also found differences between the realization of prominent peaks of *f0* in the prosodic phrase – phrase initial prominent peaks were significantly higher in English than in German, but non-initial peaks were higher in German. That is, the realization of prominent peaks in English depends on their position in the phrase, initial peaks being higher than non-initial ones, while the realization of prominent peaks in German remains the same across the whole phrase, without any significant differences between initial and non-initial peaks. English speech contained a higher number of prominent peaks overall, while German speech contained a higher number of prominent valleys. Keating & Kuo (2012) compared pitch range in English and Mandarin in three types of utterances – isolated words, a read passage and a fairytale story performed using different voices for different characters. Mandarin was found to have a higher pitch level than English in both isolated words (mean *f0* of 229 Hz in Mandarin versus 186 Hz in English) and reading passage (Mandarin 171 Hz versus English 151 Hz), but not in the fairytale story (Mandarin 228 Hz versus English 241 Hz). While performing the fairytale story, English speakers went both higher and lower with their voice, resulting in a significantly higher pitch span than that of the Mandarin speakers. This study shows cross-language differences in pitch range as well as the significance of the type of speech material compared.

2.1.2 Temporal characteristics of speech

Speech is temporally organized not only on the level of segments (duration of individual vowels and consonants), but also on the suprasegmental level. This type of temporal organization is connected to aspects of speech such as rhythm, tempo, or length of prosodic constituents. Rhythm can be defined as regular "alternation of weak and strong elements" (Fletcher, 2010: 550). Studies in auditory perception have shown that listeners tend to perceive rhythm even when the stimulus is not actually rhythmically structured. For example, listeners perceive an unstructured succession of

sounds (i.e. all sounds being equal and having equal time intervals between them) as structured into smaller, equivalent groups of sounds (e.g. triads), the first sound of the group being perceived as more prominent (Fletcher, 2010: 550). This means that rhythm falls into the domain of perception and it cannot easily be explained by simple measurements of constituent durations (Skarnitzl, Šturm & Volín, 2016: 149–150).

Different languages were observed to have different rhythms and classified into three major groups: stress-timed, syllable-timed and mora-timed (Fletcher, 2010: 552). This classification depends on which unit tends to have a constant duration – the stress group in stress-timed languages, the syllable in syllable-timed languages, the mora in mora-timed languages. It should be noted that this classification only reflects the overall tendency of a given language (Skarnitzl, Šturm & Volín, 2016: 149).

Tempo can be defined as the "speed of speaking which is best measured by rate of syllable succession" (Abercrombie, 1967; cited in Fletcher, 2010: 569). Measures of speaking tempo include speech rate, which is the number of syllables per second measured in stretches of speech including pauses, and articulation rate, which is the number of syllables per second not including pauses (Fletcher, 2010: 570). It is more informative to measure articulation rate and the duration and frequency of pauses separately, as different speakers have different strategies in their use of pausing (Dankovičová, 1997: 287–288).

Speech tempo varies in the course of speaking. For example, a speaker may slow down when pronouncing important or difficult words or at the end of his turn in a dialogue (Fletcher, 2010: 569). Dankovičová (1997) found that in Czech, articulation rate varies in a systematic way in the domain of the prosodic phrase (see section 2.1 for definition of prosodic phrase). There is a tendency to slow down as the phrase progresses, with the last word showing a strong tendency to have the slowest articulation rate. This tendency is related to phrase-final lengthening (see section 2.2).

The speaker's overall tempo can be influenced by his emotional state; emotions such as sadness, grief and boredom are associated with a slower tempo, while happiness, but also anger, rage and fear are associated with a faster tempo (Trouvain, 2004: 15). The communicative situation and communication partner are an important factor as well – speakers tend to slow down to "improve the communication channel between hearer and listener" (Fletcher, 2010: 569), or when speaking to small children (Van de Weijer, 1997; cited in Trouvain, 2004: 15). Tauroza & Allison (1990) investigated speaking rate in British English in four different types of utterances: scripted radio monologues, spontaneous conversations, non-scripted interviews and lectures given to audiences consisting mainly of non-native speakers of English. Speech in spontaneous conversation

was found to be the highest, but the most salient difference was in the speaking rate in lectures, which was significantly slower than in the other three types of utterances. This shows the significance of the type of speech material used for measuring speaking tempo.

2.2 Prosodic phrasing

2.2.1 The Prosodic phrase

The structure of speech can be described using a hierarchy of prosodic constituents. The flow of speech is segmented into smaller units defined on different levels. Above the level of individual speech sounds, these units can be defined as the syllable, the prosodic word, the prosodic phrase and the utterance (Skarnitzl, Šturm & Volín, 2016: 119).These units somehow correspond to units of the levels of language – morphemes, words, phrases and sentences, but it is not a case of one-to-one correspondence (Skarnitzl, Šturm & Volín 2016: 120). The constituents at a higher level in the hierarchy are composed of the constituents at a lower level in the hierarchy. The exact number of units in this hierarchy differs in different descriptions, for example the intermediate phrase or phonological phrase may be introduced at the level immediately below the prosodic phrase (Shattuck-Hufnagel & Turk, 1996: 205–206). Also, other terms are used for what we call the prosodic phrase: the intonational phrase, the prosodic unit, the tone unit.

Prosodic phrases are units into which the flow of speech is divided, which helps to organize it by grouping certain words together and plays a significant role in the listener's comprehension of what is being said. (Frazier, Carlson & Clifton, 2006: 246). The prosodic phrase is a unit defined as "the domain of a perceptually coherent intonational contour" (Shattuck-Hufnagel & Turk, 1996: 210) that is delimited by prosodic boundaries. The strongest prominence of the prosodic phrase is the nuclear stress (or tonic stress, pitch accent) which is realized on the tonic syllable. It tends to occur in a word near the end of the phrase, usually on the stressed syllable of the last content word (Féry, 2017: 61). It is possible to place nuclear stress on the stressed syllable of a different word to emphasize it, give it focus. Generally, the semantically most important word in the phrase receives nuclear stress (Roach, 1991: 173). The tonic syllable together with the remaining syllables in the phrase carry the tone – an elementary melodic movement which carries a general meaning (such as finality or non-finality).

In tonal languages, tones are basic melodic movements which have a lexical function, meaning that the choice of tone can distinguish between different lexical meanings. A typical example that is usually given is from Chinese, where the word "ma" can mean "mother," "hemp" or "scold" depending on which tone is used (Roach, 1991: 136). English and Czech are not tonal languages, but intonation languages, which means that their tones do not have a contrastive function to distinguish between different lexical meanings, but have more general meanings related to indicating sentence type and attitude.

According to Roach (1991: 169–168), English has an inventory of five basic tones: rise, fall, rise-fall, fall-rise and level tone. The tones are not realized solely on the tonic syllable, but continue

over the tail – the rest of the syllables following the tonic syllable until the end of the phrase. If there is more than one syllable following the tonic syllable, the rise-fall is realized in such a way that the syllable immediately following the tonic syllable has the highest pitch and the rest are lower, and the fall-rise is realized in such a way that the tonic syllable starts in high pitch, a falling movement follows and a rise is realized on the last or last stressed syllable of the phrase. Roach (1991: 167–168) gives some basic meanings of the English tones: fall – finality, definiteness; rise – general question, listing, "more to follow", encouraging; fall-rise – uncertainty, doubt, requesting; rise-fall – surprise, being impressed. But these are to be taken only as very general outlines of the possible meanings, the specific meaning is always heavily dependent on the context.

The inventory of tones in Czech includes the falling conclusive tone, the rising conclusive tone and the non-conclusive tone. The conclusive tones appear at the end of an utterance, the non-conclusive tone is found inside an utterance delimiting phrase boundaries. The particular realizations of these tones have differing pitch contours, the tones are abstractions. For example, besides a simple gradual fall, the falling conclusive tone is often realized by a rise on the syllable following the tonic syllable and subsequent fall. The non-conclusive tone exhibits the greatest variety of realizations, the most typical being a rising melodic contour (Skarnitzl, Šturm & Volín, 2016: 134–137).

2.2.2 Prosodic Boundaries

Prosodic phrases are separated by prosodic boundaries. They are usually signalled by melodic and temporal features, sometimes by a pause. Ladd (2008: 288) points out that prosodic boundaries have been defined both as audible boundaries and as boundaries predicted by a given internal prosodic structure of a phrase. He also looks at how the relationship between syntactic/semantic structure and prosodic structure influences our perception of prosodic boundaries: "If we hear an audible break in a syntactically or semantically 'impossible' location, we may be tempted to say that it is a hesitation rather than a ProsP [prosodic phrase] boundary; conversely, if we fail to observe a clear boundary where our rules lead us to expect one, we may be tempted to conclude that one is present anyway, but that it is hard to hear" (Ladd, 2008: 289). In the ToBI system of transcription, there are five break indices to indicate the strength of separation of each two words, 0 being the strongest degree of connection (meaning that the words are connected by connected speech processes), and 4 being the strongest degree of disconnection. Break indices 3 and 4 are defined as marking the end of higher level prosodic constituents – the intermediate and intonation phrase (Beckman & Elam, 1997: 31–32).

The phonetic cues for identifying a prosodic boundary are mainly pauses, melodic movements and temporal features. The melodic movement at the very end of a prosodic phrase is

described using a set of boundary tones in the ToBI system (Beckman& Elam, 1997: 12). As for temporal features, final lengthening, that is the increase in duration of the speech sounds at the end of the phrase, has been observed in many languages (Duběda, 2005: 190), but it is not universal. "There is a tendency to lengthen the final elements in an utterance, particularly the last vowel, before a pause (for French, English, German and Spanish, Italian, Russian, Swedish). However, there seem to be languages in which there is little (if any) final lengthening, e. g. in Finnish, Estonian and Japanese" (Vaissiere, 1983: 60).

The beginning of a new phrase can also be signalled by intonation reset, that is by returning to a pitch level higher relative to the previous phrase, whose ending has been affected by declination (Duběda, 2005: 177–178). There is a global tendency of *f0* to decline over the course of the utterance (Féry, 2017: 107) and "reset refers to the return of the *f0* to a high level after a prosodic domain of the size of an ι-phrase [intonation phrase] ends and a new one begins" (Féry, 2017: 112). Roach (1991: 159) sees rhythm as one of the ways to identify prosodic boundaries, because speech inside the prosodic phrase has a regular rhythm that is interrupted by the boundary. Changes in voice quality also contribute to marking prosodic boundaries. Creaky phonation at the end of the phrase is quite frequent and can be used by listeners as a cue for identifying a prosodic boundary (Crowhurst, 2018).

Also, looking at prosodic boundaries from the perspective of speech production, articulation at the end of a prosodic phrase slows down. There is an increase in duration and decrease in peak velocity of articulatory gestures – "gestures get longer, larger and farther apart. (…) Each of these effects can be viewed as increasing the perceptual salience of (i.e., ''marking'') a phrase boundary for a listener" (Byrd & Saltzman, 2003: 159). Krivokapić & Byrd (2012) found that the strength of a boundary perceived by the listeners reflected the actual articulation of the speakers as measured by electromagnetic articulography. Features showing strong correlation with judgements of perceived boundary strength were: pre-boundary movement duration, duration between peak velocity of the opening movement of the pre-boundary consonant and peak velocity of the closing movement of the post-boundary consonant, and duration between peak velocity of the closing movement of the pre-boundary vowel and peak velocity of the retraction movement of the post-boundary vowel (Krivokapić & Byrd, 2012: 437). The last two measures can be taken as measures of boundary duration (Krivokapić & Byrd, 2012: 434).

De Pijper & Sanderman (1994) looked at the relationship between naive listeners' judgements of perceived boundary strength and different suprasegmental features: pauses, melodic discontinuities (speakers used four basic pitch patterns followed by a discontinuity manifesting itself in the pitch resuming at a different level after the boundary (de Pijper & Sanderman, 1994:

2041)), baseline reset, and final lengthening. Boundaries were generally perceived stronger when more of these phonetic cues were associated with them (de Pijper & Sanderman, 1994: 2045). Even though the experiment was carried out on Dutch, the results should be relevant across languages, because it was found that listeners really judge boundary strength based on phonetic cues, not syntactic, semantic or lexical information specific to the language, as results were similar in normal and delexicalized speech (de Pijper & Sanderman, 1994: 2045).

2.2.3 Relationship between Prosodic Structure and Syntactic Structure

As has been mentioned earlier, the prosodic phrase is somehow analogous to the sentence or syntactic phrase. To a certain degree, prosodic structure is related to the syntactic structure of an utterance. Prosodic boundaries tend to occur at major syntactic boundaries (Shattuck-Hufnagel & Turk, 1996: 196), so that the edge of a prosodic constituent is aligned with the edge of a syntactic constituent. Some syntactic structures seem to have an especially strong tendency to be realized in separate prosodic phrases, such as parentheticals, tags, non-restrictive relative clauses (and non-restrictive modifiers in general), sentential adverbs, vocatives and appositives (Nespor & Vogel, 1987, cited in Watson & Gibson, 2004: 715; Shattuck-Hufnagel & Turk, 1997: 197). For example, it is very likely that the following utterances will be produced with a prosodic boundary in the marked places. (The examples are taken from the speech material analysed in the thesis; the material will be described and the codes explained in the Method section.)

Parenthetical:

(1) *And I'm like* | ***you know*** | *knock discomfort upside the head.* (EN01)

(2) *zkusili* | ***pokud přijmete moje argument y***| *měnit* | *vzdělávací kulturu* (CZ08)

Question tag:

(3) *It's an oxymoron* | ***isn't it?*** (EN03)

Appositives:

(4) ***The famed therapist*** | ***M. Scott Peck*** (EN05)

(5) *tohle je například rozhovor* ***s jedním z nejzajímavějších lidí na světě*** | ***s Bjarkem Ingelsem*** | ***architektem*** (CZ02)

Vocative:

(6) ***Cal*** | *social media* | *is one of the fundamental technologies* (EN07)

Sometimes the placement of prosodic boundaries can help to determine the syntactic structure of the sentence. Because the aforementioned syntactic elements tend to be separated by a boundary, their status can be differentiated from other syntactic elements using prosody. For example, if the utterance *Don't worry | Anna* is produced with a prosodic boundary between *worry* and *Anna* as indicated, *Anna* will most likely be considered a vocative, but if the utterance is produced without the prosodic boundary, *Anna* may also be interpreted as the object of the verb.

Prosodic grouping of certain words together can also resolve attachment ambiguities, i.e. it can help to decide to which syntactic constituent a certain other syntactic constituent is attached, one that is lower or higher in the hierarchy of constituents in a sentence (high vs. low attachment). That is, if we represent the structure of the sentence by drawing a syntax tree, low attachment means attaching the constituent to the lowest node possible, and high attachment means attaching the constituent to a higher node. Producing words in the same prosodic phrase can signal their close syntactic relationship, so the syntactic constituent should be produced in the same prosodic phrase as the other syntactic constituent it is attached to. For example, Speer, Warren & Schaffer (2011) looked at how speakers disambiguated high and low attachment of prepositional phrases in a cooperative task on sentences such as these:

> "*I want to change the position of the square with the triangle.*" (low attachment)
> "*I want to change the position of the square with the cylinder.*" (high attachment)
>
> (Speer, Warren & Schaffer, 2011: 42)

The speaker knew from context that in the first sentence, the prepositional phrase *with the triangle* modifies the noun phrase *square* (low attachment), and in the second sentence, the prepositional phrase *with the cylinder* modifies the verb phrase *change* (high attachment). A prosodic boundary before the prepositional phrase suggests a high attachment interpretation, because it separates *square* and *with the cylinder*, whereas a missing boundary suggests a low attachment interpretation, because it groups *square* and *with the triangle* together, signalling their syntactic closeness. The results showed that there were indeed more prosodic boundaries produced by speakers before the prepositional phrase in the high attachment condition.

The listener is then able to use this information to interpret the meaning of the utterance when deciding between high and low attachment. Frazier, Carlson & Clifton (2006) found that not only the presence of a prosodic boundary before the ambiguously attached phrase, but also its relative strength in comparison with other prosodic boundaries in the utterance was used as a cue in the listeners' interpretation of high versus low attachment.

Jun (2003) also found a relationship between prosodic boundary placement and high versus low attachment interpretation of relative clauses, even across different languages (English, Greek, Spanish, French, Farsi, Japanese, Korean).

As for the assumption that non-restrictive clauses tend to require their own prosodic phrase, Watson & Gibson (2004), for example, have examined how often speakers use prosodic boundary placement to give cues for disambiguating between restrictive and non-restrictive clauses. They used target sentences which were restrictive in one context, and non-restrictive in another, as illustrated by this example:

> "*A group of film critics praised a director at a banquet and another director at a film premiere. The director who the critics praised at a banquet insulted an actor from an action movie during an interview.*" (restrictive interpretation)
>
> "*A group of film critics praised a director and a producer. The director who the critics praised at a banquet insulted an actor from an action movie during an interview.*" (non-restrictive interpretation)
>
> (Watson & Gibson, 2004: 747–748)

They found that prosodic boundaries were produced more often before non-restrictive than restrictive relative clauses. But even before non-restrictive clauses, the boundary was produced less than half of the times. This result may have been affected by the way the experiment was conducted, because the sentences which the subjects were reading out were all written without commas, even though non-restrictive relative clauses are normally separated by commas in writing.

Overall, there is good amount of evidence that prosody can be used to disambiguate syntactic ambiguities to a certain degree. Cutler, Oahan, & van Donselaar (1997), for example, cite a number of other studies that support this as well. They also suggest that speakers produce fewer prosodic cues when the context strongly supports one interpretation over the other (Cutler, Oahan, & van Donselaar, 1997: 162–163). The evidence that speakers and listeners use prosody to indicate and interpret the syntactic structure of utterances suggests a close relationship between syntax and prosody.

However, syntax and prosody are not isomorphic, and prosody is not fully determined by syntax. A sentence with a given syntactic structure can have more possible prosodic realizations, and predictions about prosodic boundary placements in an utterance can be made in terms of likelihood rather than givenness. "[Prosodic structure] allows for variation which is both qualitative

(alternative patterns which are acceptable) and quantitative (more vs. less strongly realized markers)" (Cutler, Oahan & Van Donselaar, 1997: 170).

In spontaneous speech, many prosodic breaks are realized because of hesitation. Boomer (1965; cited in Blaauw, 1994) found that there is a great number of pauses after the first word in a clause, usually a conjunction. This is probably a result of delay at the level of message generation – the speaker wants to continue talking, but the message to be articulated is not yet developed sufficiently. To indicate his or her intention to continue speaking, so that he or she is not interrupted, the speaker utters the first word of the following clause. This can be illustrated by the following example:

(7) *Ale my se musíme podívat na tu ekonomiku jako na celek* ***a*** | *ne to* | *vytrhnout* | *ze souvislosti* (CZ07)

There are other examples in our data of prosodic boundaries in a place not motivated by the syntactic structure of the utterance, such as boundaries between a noun and its modifying adjective or a noun and a demonstrative pronoun:

(8) *the three most* ***common*** | ***objections*** *I hear* (...) (EN07)
(9) *and then we remember* ***that*** | ***time*** *that we met* | *Hugh Jackman* (EN05)
(10) *úrokové sazby jsou v* ***daném*** | ***okamžiku*** *vlastně* | *velmi nízké* (CZ06)
(11) *A když už* ***ten*** | ***čtverec*** *mají* (CZ08)

2.2.4 Other Factors Influencing Prosodic Phrasing

Even if we disregard hesitations in spontaneous speech, there are factors other than syntactic structure that influence the prosodic properties of a given utterance, including the placement of prosodic boundaries. "There are often a variety of prosodic possibilities for the utterance of a sentence, and in some cases, these well-formed prosodic structures appear to violate syntactic structure" (Shattuck-Hufnagel & Turk, 1996: 197–198). Some of these factors are balance, constituent size, semantic closeness, information structure and emphasis, stance and attitude.

Shattuck-Hufnagel & Turk (1996: 203) cite symmetry and balance as a factor influencing the way an utterance is divided into prosodic phrases. The speakers have a "tendency to divide the spoken utterance into equal parts." They give examples of studies where speakers placed a boundary inside a syntactic constituent if it resulted in dividing the utterance into prosodic phrases of a similar size.

Watson & Gibson (2004) tested the effect of constituent size on prosodic phrasing. They hypothesised that "as the size of a syntactic constituent increases, the likelihood of a boundary following that constituent increases" (Watson & Gibson, 2004: 727), because there may be a need for a refractory period after producing a long constituent, and "as the size of an upcoming syntactic constituent increases, the likelihood of a boundary before that constituent increases" (Watson & Gibson, 2004: 728), because there may be a need for extra time to plan a long constituent. They found that this hypothesis is a good predictor of the boundary placements that speakers actually produced while reading sentences aloud.

However, there are other effects that interact with the effect of constituent size. Watson, Breen & Gibson (2006) found that the distinction between obligatory and non-obligatory elements can influence prosodic boundary placement. Because a head and its obligatory argument (e.g. a verb and its object) are closely semantically related, they tend to be produced in the same prosodic phrase, while non-obligatory elements (e.g. adjuncts) tend to be preceded by a prosodic boundary. This is consistent with previous findings that more prosodic boundaries are produced after nouns than verbs (Watson, Breen & Gibson, 2006: 1047), because elements modifying nouns are non-obligatory, while verbs and their arguments have a closer relationship – the valency of the verb requires certain obligatory complements.

The information structure of a sentence is also a significant factor influencing the prosodic realization of a sentence. Information structure describes the roles which different parts of the sentence have in developing and transmitting the information that is being communicated. It works with concepts such as givenness and newness, topic and focus. The part of the sentence which expresses the thing that is being talked about, usually known or somehow accessible to both communicants, already mentioned previously or inferable from the context, is called a topic (or theme). The part of the sentence which expresses what is being said about the topic is called the focus (or rheme). It contains new information that is being highlighted and is usually defined in terms of "presence of alternatives" (Féry, 2017: 139) – we are choosing one possibility from a set of possibilities of what could be true about the topic and contrasting it with all these possibilities.

There are different possibilities of marking the informational structure of a sentence, such as morphological markers, word order, focus sensitive particles and prosody. Languages use these means differently, and each to a different degree. If we look at Czech and English, one prominent difference is the use of word order. Czech has free word order, which means it is not strictly given by the syntactic structure of the sentence and can be systematically used for other purposes, such as to signal information structure. The newer and more important a word is (that is, the less given by context and the more focused), the closer it is to the end of the sentence. Focus sensitive particles

(such as *právě*, *zrovna*, *jen*, *ne* etc.) are also frequently used in Czech to signal the following word as the focus of the sentence. Because English word order is grammatical, its use as a marker of information structure is very limited. Even though English also uses focus sensitive particles and some specific syntactic constructions, such as cleft sentences, to signal the focus of a sentence, it "relies predominantly on intonational cues when expressing contrastivity and signalling major information" (Volín, Poesová & Weingartová, 2015: 107).

A focused word tends to be made prosodically prominent, so as to signal its importance. It "seeks to be prosodically initial or final" (Féry, 2017: 147), to be placed at the beginning of the phrase or at the end of it, and to carry a prominent nuclear stress, to be the relatively most prominent part of the phrase (Féry, 2017: 153). A topic tends to be phrased separately and its givenness is marked by deaccentuation (Féry, 2017: 148, 150). This means that putting emphasis on different parts of the sentence in relation to their information status can influence the shape of the tonal contour as well as the placement of prosodic boundaries.

The following examples illustrate how a focused word is marked by having its own prosodic phrase or being phrase final. In addition, the pitch movement on the emphasised word has a larger span (marked here as bold):

(12) *When you perceive* | ***choice*** | *you perceive* | ***motivation*** | *you're more* | *motivated* (EN03)

(13) *Be in* ***that*** | *moment* (EN05)

Even though prosodic marking of information structure is perhaps more typical of English, Czech also makes use of it. The following example illustrates how the contrastive topics are separated from the foci by prosodic boundaries, and how the focused verbs are emphasised by a larger pitch movement:

(14) *A to jedno dvojče* | ***mělo*** *schizofrenii* | *a to druhé dvojče* | ***nemělo*** *schizofrenii* (CZ03)

The alternative way to express the information structure of this sentence would be to use word order and put the focused verbs at the end of the clauses: *A to jedno dvojče schizofrenii mělo, a to druhé dvojče schizofrenii nemělo.* (or perhaps more naturally without repetition in the second clause: *A to jedno dvojče schizofrenii mělo, a to druhé nemělo.*). This option would not be available in English, where the order of verb and object is fixed.

Prosody is also used to expresses affective states and attitudes of the speaker towards the contents of the sentence, the communication partner or the context and communicative situation, and this is a factor which shapes the prosodic realization of an utterance significantly. Whether this function of prosody is linguistic or paralinguistic in nature may be discussed (Féry, 2017: 170), but stances and attitudes are an integral part of the pragmatics of language, and these meanings are often vital to the correct interpretation of an utterance. The interaction between the content of the utterance, prosodic factors and the context leads to perception of attitudinal meanings and interpersonal stances such as friendliness, rudeness, urgency, scorn, ridicule etc. (Wichmann, 2005). For example, Wichmann (2005) looked at the ways in which the word "please" was uttered in different situations, expressing different meanings. An unaccented or low pitch accented "please" functions as a mere pragmatic marker with an interpersonal function of expressing politeness. But "please" realized in its separate phrase with a high pitch accent conveys a sense of greater urgency, in combination with the request preceding it being unaccented it implies "that the request is already mutual knowledge (...) but has not been complied with" (Wichmann, 2005: 242–243) and expresses its reinforcement and emphasis. In some cases, an attitude of annoyance can arise from the combination of this prosody and the contents and context of the message. The type of terminal contour used on the word expresses what kind of reaction is expected from the communication partner. A fall occurs mostly in asymmetrical discourse and indicates that the listener is bound to comply, that the request is non-negotiable and is closer to a command. A rise occurs where the listener's compliance is optional and they can refuse (Wichmann, 2005: 237–238). This is consistent with the general meanings of falling and rising tones, the former indicating "closure" and "finality," and the latter indicating "openness" and "non-finality" (Wichmann, 2005: 233).

2.3 Prosody in Czech and English

Jun (2003: 220) remarks that the effects of factors influencing prosodic phrasing differ across languages. For example, in some languages, focus can block prosodic boundary placement after the focused word or deaccent the following words, while in other languages, it has no effect on the phrasing after the focused word. Moreover, the typical phrase length is different across languages, and "languages differ in the mapping between a syntactic structure and a prosodic structure" (Jun, 2003: 219). In this thesis, we are specifically interested in the difference between Czech and English.

Compared to English, Czech intonation is generally quite flat. This is largely due to an overall narrower pitch range, which has been investigated by Volín, Poesová & Weingartová (2015), addressing "the popular beliefs about the melody of Czech-accented English, which typically sounds flat and monotonous to both native and proficient non-native ears, as if signalling

boredom, disinterest or lack of involvement" (Volín, Poesová & Weingartová 2015: 109). Pitch span of English and Czech professional news-readers was measured using standard deviation, variation range, 80-percentile range and quartile range of the fundamental frequency (*f0*), and the results showed that English speakers make use of a larger range of frequencies in their speech. For example, the 80-percentile range for English speakers was 7.1 ST for females and 8.1 ST for males, and for Czech speakers, it was only 5.2 ST for females and 6.1 ST for males. Assuming that this difference influences native Czech speakers when they speak English, Czech-accented English was investigated to see if there is an effect of interference and the measures of pitch span in Czech-accented English lie between those of native Czech and English. It was found to have even narrower pitch span than Czech, which may be an effect of the speakers' uncertainty or anxiety while speaking a foreign language. These results support the claim that Czech speakers use a narrower pitch span than English speakers, but it should be noted that they come from a specific domain, the speech of professional news-readers is different from the speech of non-professionals in everyday conversations.

Based on naive observation of everyday speech, Czech also seems to be characterized by longer prosodic phrases. Dividing the utterance into a smaller number of longer prosodic phrases also contributes to the perceived flatness of intonation, because there are longer stretches of speech without stronger melodic movements.

In this thesis, we investigate the differences between Czech and English prosodic phrasing in good public speakers. We predict that the two languages differ in their prosodic phrasing and that prosodic phrasing is influenced by stylistic factors, such as the communicative situation and the speaker's competence. Analysing the speech of one professional and two non-professional speakers, de Pijper & Sanderman (1994) found that the professional speaker produced more prosodic cues than non-professional speakers and used intonation reset, which was hardly used by the non-professional speakers. More dynamic, lively intonation, such as wider pitch range and *f0* variation, was found to be used by skilled speakers and to correlate with perceptions of charisma (e.g. Rosenberg & Hirschberg, 2009; Strangert, 2005). So although the intonation of spontaneous Czech speech in everyday communication is characteristically flat, good public speakers may use different strategies in their prosody to make a good impression on the audience.

We used measures of prosodic phrase length, speaking tempo, pitch range and melodic variability in the domain of the prosodic phrase to find objective evidence that could account for the perceived differences between Czech and English, and to describe how good public speakers use prosodic phrasing in both languages.

3. Method

3.1 Material

As material for the analysis, we have chosen TEDTalks in American English and Czech. TEDTalk conferences are events where speakers present a wide range of topics from different fields of research in an attractive, entertaining way to a general audience. 15 speakers in each language were selected based on subjectively perceived speaker quality from approximately 40 talks found on YouTube. We assume that the fact that the speaker has taken part in a TEDTalk conference is a certain guarantee of his or her competence in itself, but to ensure the speakers' quality, we have conducted a perception test in which participants were asked to evaluate the 15 selected speakers in each language. They were played a 30-second segment from their speeches and asked to express their willingness to employ the speaker as their spokesperson on a 7-point scale. English and Czech speakers were evaluated separately, we have asked two different groups of 8 participants to evaluate the 15 recordings in each language. English speakers were evaluated by participants with high level of English proficiency (mostly English Studies university students), but with the exception of one participant, they were not native speakers. 10 speakers in each language who have received the best overall score were selected for the analysis. The speakers were labelled with codes including language (CZ/EN) and a number; these are the codes we use in this thesis for reference when citing examples from the data.

This choice of material enables us to compare English and Czech on fairly homogenous samples. The talks are given in similar conditions, they are comparable in length (all are between 15 and 20 minutes long), and they are also homogenous stylistically – they are public speeches which, although they have been prepared and practiced beforehand, are not written down and read, given to a large audience by a competent speaker.

3.2 Analysis

The recordings were divided into shorter segments (approximately 60 seconds long) and automatically segmented by means of Prague Labeller (in the case of Czech; Pollák, Volín & Skarnitzl, 2007) and P2FA (in the case of English; Yuan & Liberman, 2008) forced alignment, which yielded the approximate placement of phone boundaries. Segments number 3–7 (that is approximately 5 minutes of speech) from each speaker were analysed; the first two minutes were not analysed because the speaker may need some time to get started and find his speaking style, which means that the very beginning may differ slightly from the rest of the speech (e.g. in frequency of hesitations, tempo).

Prosodic phrasing was labelled manually in Praat. We used break indices 3 and 4, 4 for the strongest type of boundary which is clearly perceptible and usually has both a clear melodic

movement and final lengthening or in some cases a pause, 3 for a weaker boundary with a smaller melodic movement. Additional marking of "p" is used "to convey some sort of prosodic disfluency – for example, an abrupt cutoff after a false start or a perceptible prolongation or pause which sounds as if the speaker were hesitating while searching for the next word" (Beckman & Elam, 1997: 32). We have also marked the word carrying nuclear stress in each phrase.

We used the annotated data to measure temporal and *f0* characteristics. The variables measured were:

- number of syllables per prosodic phrase
- number of words per prosodic phrase
- speaking rate in syllables/second
- standard deviation (SD) of *f0* in each prosodic phrase in semitones (ST)
- SD of *f0* in the nuclear part of the phrase (i.e., during realization of the tone) in ST
- Cumulative Slope Index (CSI) in each prosodic phrase in ST/syllable (Hruška & Bořil, 2017)

The number of words and syllables per prosodic phrase was extracted using a script in Praat. Words were counted directly from the automatically segmented word tier, syllables were counted as the number of vowels in English, and as the number of vowels and syllabic consonants in Czech. Syllabic consonants in Czech were defined as [r] or [l] between two consonants. The cases where a word ends in a syllabic consonant and is followed by a word starting with a vowel were not counted (but the data includes only a very small number of these cases). There was no need to define syllabic consonants in English, because words where a syllabic consonant can appear, such as "intervention" or "electromagnetism," were always automatically transcribed with [ə] followed by a consonant. Speaking rate in syllables/second was measured by dividing the extracted number of syllables per prosodic phrase by phrase duration in seconds.

Fundamental frequency was extracted using autocorrelation in Praat with the default settings, except for pitch ceiling – *f0* was extracted in the frequency range of 75–320 Hz for male speakers and 75–450 Hz for female speakers. The extracted Pitch objects were smoothed using a 10-Hz filter (to exclude very small *f0* movements which do not affect intonation) and interpolated (to create a continuous *f0* contour even in unvoiced segments, which reflects the way intonation is perceived more accurately). Finally, the Pitch objects were converted into PitchTier objects which were used to measure the SD of *f0* in ST in each phrase and during realization of the tone, and CSI in each phrase. Semitones are perceptual units – they reflect the way we hear pitch differences, and their relationship to hertz is non-linear. Measuring in semitones allows us to compare intonation

ranges between speakers with different pitch levels using SD (which would not be possible in hertz, as perceptually equal ranges would have differing ranges in hertz depending on their pitch level).

CSI is a measurement of *f0* variation calculated as "the sum of absolute frequency differences between subsequent pitch points divided by the duration of the measured speech segment" (Hruška & Bořil, 2017: 37). In other words, unlike the standard deviation, CSI takes into account multiple melodic movements in a phrase. In this case, the duration of the measured speech segment (prosodic phrase) is the number of syllables per prosodic phrase.

We tested the significance of the effects of language and type of prosodic break on the measured variables. We used linear mixed-effects (LME) modelling to determine how significant the influence of language and type of prosodic break was on the temporal and *f0* characteristics while also taking into account possible individual differences between speakers. The analysis was conducted in R (R Core Team, 2017) using the *lme4* package (Bates, Maechler, Bolker & Walker, 2015). LME is used to analyse the influence of fixed effects, i.e. the independent variables under our control, and random effects, i.e. other factors that are not under our control but may have influenced the measured dependent variables. In this case, the fixed effects were LANGUAGE (Czech × English) and TYPE OF PROSODIC BREAK (BI3 × BI4), and the random effect was SPEAKER intercept (which accounts for the fact that individual speakers may significantly differ from each other in the measured characteristics) and by-SPEAKER slope for the effect of TYPE OF PROSODIC BREAK (which accounts for the fact that individual speakers may significantly differ in their realization of each type of the prosodic break). The model assumes that the residuals are normally distributed and homoscedastic, so the residual plots were visually inspected for deviations from normality and homoscedasticity. The significance of individual effects or interactions was tested by comparing the full model to a reduced model with the given factor or interaction excluded. We conducted Tukey post-hoc tests using the R package *multcomp* (Hothorn, Bretz & Westfall, 2008) to test specific pairwise comparisons (e.g. the significance of the difference in speaking rate between BI3 in English and BI3 in Czech). Plots showing mean values of the measured variables and confidence intervals were created using the *effects* package (Fox, 2003).

4. Results

4.1 Temporal characteristics

4.1.1 Number of syllables per prosodic phrase

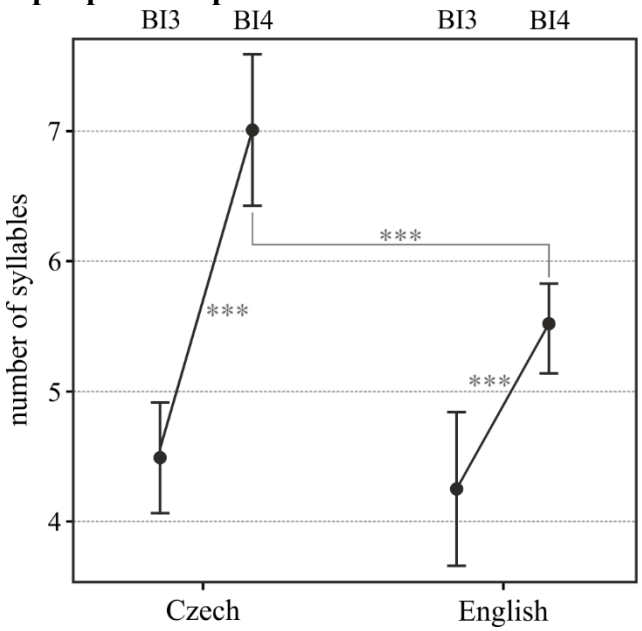

Figure 1. Number of syllables per prosodic phrase depending on LANGUAGE (Czech × English) and TYPE OF PROSODIC BREAK (BI3 × BI4). The asterisks in this figure and all subsequent figures show statistical significance: ***p < 0.001, ** p < 0.05.

Mean values and confidence intervals are shown in **Figure 1.** Likelihood ratio tests comparing the full model (syllables ~ language + BI + (1+BI|speaker)) with the model without the effect in question show that both the effect of LANGUAGE and TYPE OF PROSODIC BREAK is significant. LANGUAGE significantly affected the number of syllables per prosodic phrase ($\chi^2(1) = 7.80$, $p < 0.01$): phrases in English are generally shorter by about 0.89 (± 0.25 standard errors) syllables. TYPE OF PROSODIC BREAK significantly affected the number of syllables per prosodic phrase ($\chi^2(1) = 25.18$, $p < 0.0001$), phrases ending in a stronger prosodic break are generally longer by about 0.5–3.7 syllables (the differences in length vary between individual speakers).

The residuals show a certain degree of heteroscedasticity.

The test of interaction between LANGUAGE and TYPE OF PROSODIC BREAK is on the borderline of convergence, the interaction is significant: $\chi^2(1) = 6.98$, $p < 0.01$.

Tukey post-hoc tests show that the difference between BI3 and BI4 is significant both in Czech and in English ($p < 0.001$), the difference between BI4 in Czech and BI4 in English is significant ($p < 0.001$) and the difference between BI3 in Czech and BI3 in English is not significant ($p > 0.9$).

Because some phrases are realized with a hesitation, usually a filled pause while the speaker is searching for the next word (labelled "p" in our data), we excluded those and also looked at the length of phrases without hesitations. **Table 1** shows that hesitations are much more frequent in phrases ending in BI3 (236 out of 573, that is almost a half, is realized with a hesitation), meaning that there is only a small number of phrases ending in BI3 without a hesitation left, which poses a problem for the analysis. Therefore, only phrases ending in BI4 without hesitations are considered in some partial analyses.

language	BI	inter	n
Czech	3	no	136
Czech	4	no	1750
English	3	no	201
English	4	no	2133
Czech	3	yes	168
Czech	4	yes	110
English	3	yes	68
English	4	yes	38

Table 1. Number of prosodic phrases with and without hesitation ("inter")

The results are shown in **Figure 2**. LANGUAGE significantly affected the number of syllables per prosodic phrase ($\chi^2(1) = 15.60$, $p < 0.0001$), phrases in English are generally shorter by about 1.72 ($\pm$ 0.37 standard errors) syllables. The residuals are normally distributed and slightly heteroscedastic. Compared to the difference between English and Czech in all prosodic phrases (i.e. also in those ending with BI3 and with hesitations), which was about 0.89 syllables, the difference here is almost twice as large. It seems that many of the shorter phrases ending in BI4 in Czech can be accounted for by an interruption due to the speaker's hesitation.

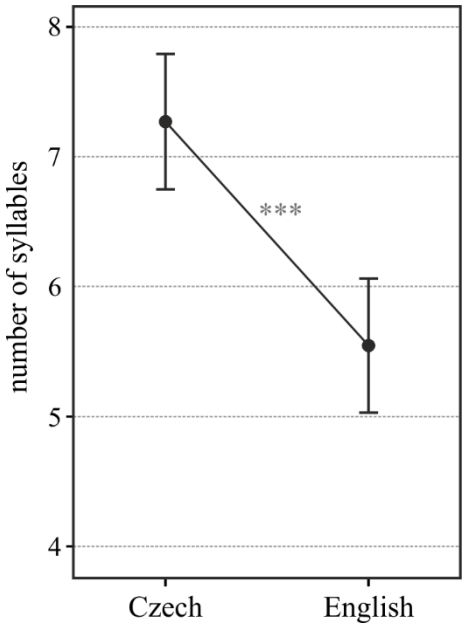

Figure 2. Number of syllables per prosodic phrase in BI4 without hesitation depending on LANGUAGE (Czech × English).

4.1.2 Number of words per prosodic phrase

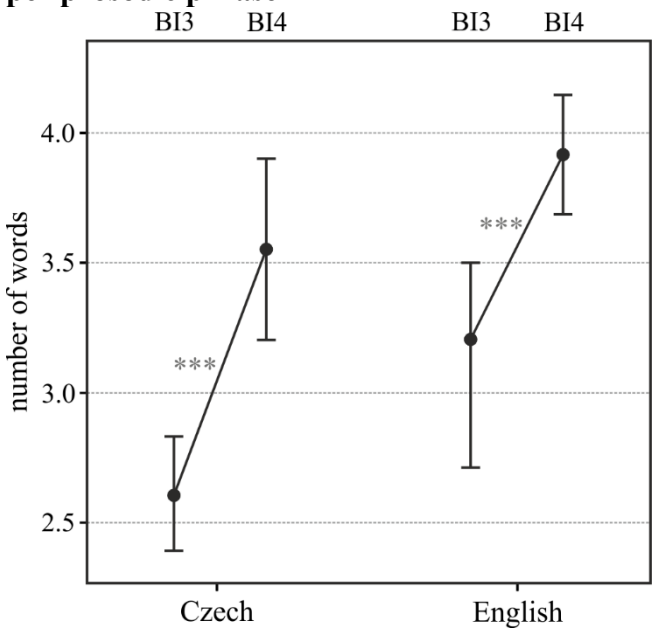

Figure 3. Number of syllables per prosodic phrase depending on LANGUAGE (Czech × English) and TYPE OF PROSODIC BREAK (BI3 × BI4).

Mean values and confidence intervals are shown in **Figure 3**. The effect of LANGUAGE on the number of words per prosodic phrase is marginally significant ($\chi^2(1) = 2.88$, $p < 0.1$), phrases in English are generally longer by about 0.43 (± 0.13 standard errors) words. This is to be expected, because even though Czech phrases are generally longer in terms of syllables, English, which is an analytical language, uses many short words with a grammatical function (such as articles), as opposed to Czech. TYPE OF PROSODIC BREAK significantly affected the number of words per

prosodic phrase ($\chi^2(1) = 23.47$, $p < 0.0001$): phrases ending in a stronger prosodic break are generally longer by about 0.5–1.6 words (the differences in length vary between individual speakers, and there is one exceptional speaker whose phrases ending with BI3 are longer than those ending with BI4 by about 0.02 words).

The whole model manifests borderline singularity, the residuals show a satisfactory level of normality and homoscedasticity.

The test of interaction between LANGUAGE and TYPE OF PROSODIC BREAK is singular and could not be carried out.

Post-hoc tests show that the difference between BI3 and BI4 is significant both in Czech and in English ($p < 0.0001$). The difference between BI3 in Czech and BI3 English and BI4 in Czech and BI4 in English is not significant ($p > 0.1$).

Figure 4 shows the results for phrases ending in BI4 and without hesitation. The effect of LANGUAGE on the number of words per prosodic phrase is not significant ($\chi^2(1) = 1.66$, $p > 0.1$). The residuals are slightly heteroscedastic.

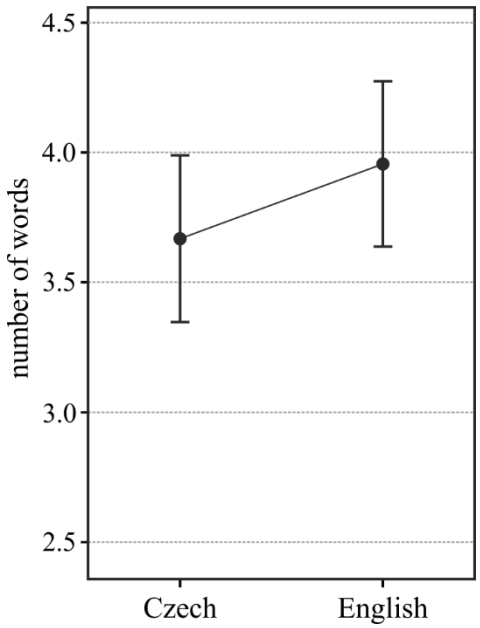

Figure 4. Number of words per prosodic phrase in BI4 without hesitation depending on LANGUAGE (Czech × English).

4.1.3 Speaking rate

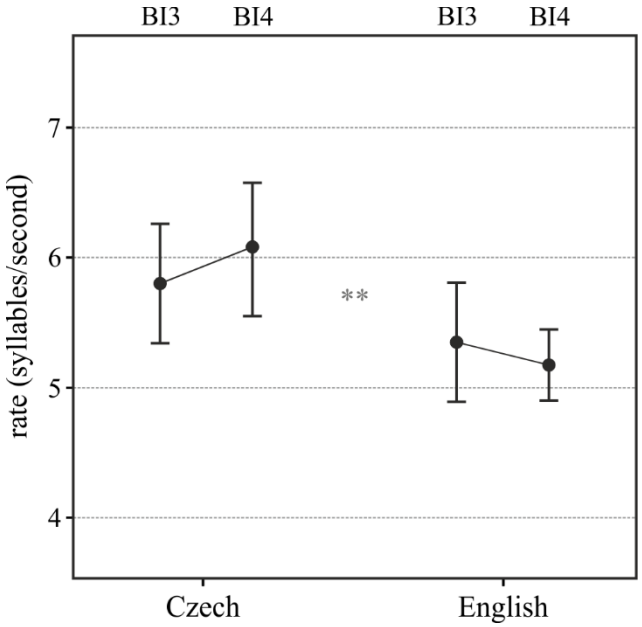

Figure 5. Speaking rate depending on LANGUAGE (Czech × English) and TYPE OF PROSODIC BREAK (BI3 × BI4).

Mean values and confidence intervals are shown in **Figure 5**. LANGUAGE significantly affected speaking rate ($\chi^2(1) = 5.38$, $p < 0.05$), speaking rate in English is generally lower by about 0.76 (± 0.29 standard errors) syllables/second. The effect of TYPE OF PROSODIC BREAK is not significant ($\chi^2(1) = 0.26$, $p > 0.6$), differences in rate in phrases ending with BI3 and BI4 are different in individual speakers, some having higher speaking rate in BI4, some lower.

The residuals are normally distributed and homoscedastic.

The test of interaction between LANGUAGE and TYPE OF PROSODIC BREAK failed to converge and could not be carried out.

Figure 6 shows the results for phrases ending in BI4 and without hesitation. LANGUAGE significantly affected speaking rate ($\chi^2(1) = 8.69$, $p < 0.001$), speaking rate in English is generally lower by about 0.96 (± 0.31 standard errors) syllables/second, which is comparable to the difference in speaking rate between English and Czech across all prosodic phrases. The residuals are normally distributed and homoscedastic.

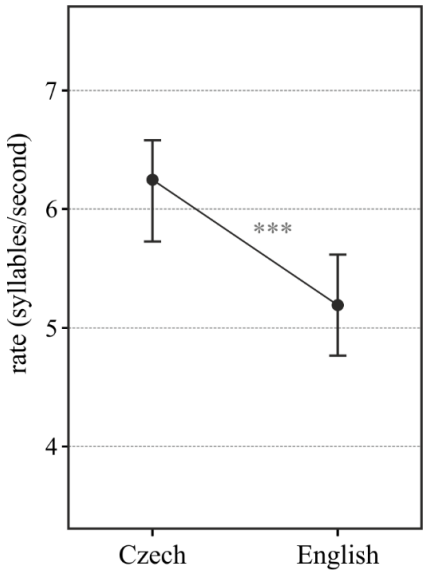

Figure 6. Speaking rate in BI4 without hesitation depending on LANGUAGE (Czech × English).

4.2 *f0* characteristics

4.2.1 Standard deviation of *f0* in each prosodic phrase

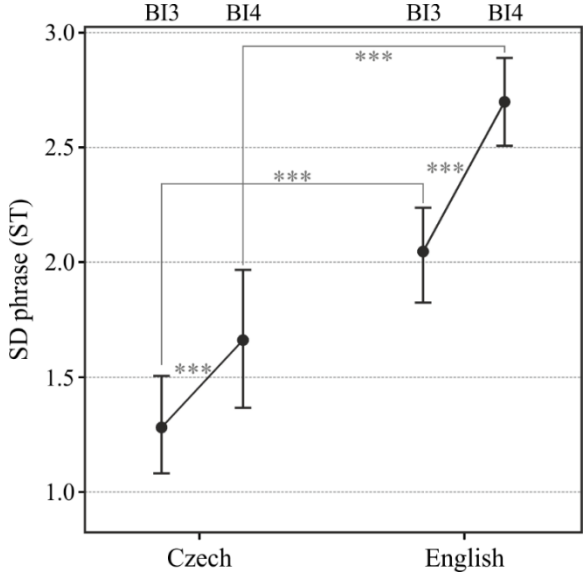

Figure 7. Standard deviation of *f0* in each phrase depending on LANGUAGE (Czech × English) and TYPE OF PROSODIC BREAK (BI3 × BI4).

Mean values and confidence intervals are shown in **Figure 7**. LANGUAGE significantly affected the SD of *f0* in each prosodic phrase ($\chi^2(1) = 15.98$, $p < 0.0001$), SD in English is generally higher by about 0.78 (± 0.15 standard errors) ST. TYPE OF PROSODIC BREAK significantly affected the SD of *f0* in each prosodic phrase ($\chi^2(1) = 26.74$, $p < 0.0001$), the differences between SD in prosodic phrases ending with BI3 and BI4 range from a slight difference in some speakers (about 0.08 ST in the speaker with the lowest difference) to a considerably large difference (about 0.94 ST in the speaker with the highest difference).

The residuals show a certain degree of heteroscedasticity.

The test of interaction between LANGUAGE and TYPE OF PROSODIC BREAK is singular and could not be carried out.

Post-hoc tests show that the difference between BI3 and BI4 is significant both in Czech and in English ($p < 0.0001$). The difference between BI4 in Czech and BI4 in English is significant ($p < 0.0001$) and the difference between BI3 in Czech and BI3 in English is also significant ($p < 0.001$).

Figure 8 shows the results for phrases ending in BI4 and without hesitation. LANGUAGE significantly affected the SD of *f0* in each prosodic phrase ($\chi^2(1) = 19.48$, $p < 0.0001$), SD in English being higher by about 1.02 ($\pm$ 0.19 standard errors) ST, which is comparable to the difference between English and Czech across all prosodic phrases, but a little higher. The residuals are slightly heteroscedastic.

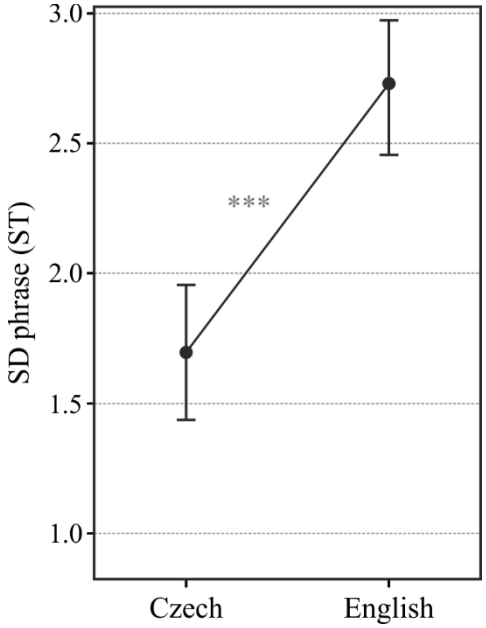

Figure 8. Standard deviation of *f0* in each phrase ending with BI4 without hesitation depending on LANGUAGE (Czech × English).

4.2.2 Standard deviation of *f0* in the nuclear part of the phrase

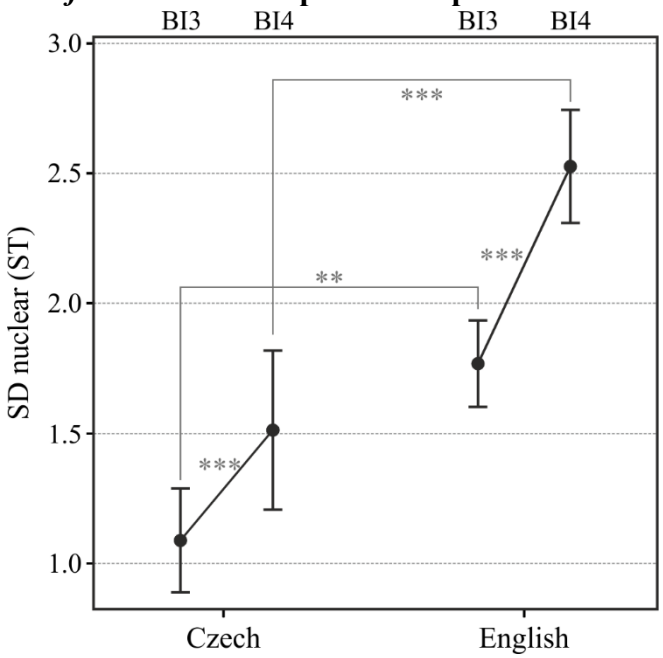

Figure 9. Standard deviation of *f0* in the nuclear part of the phrase depending on LANGUAGE (Czech × English) and TYPE OF PROSODIC BREAK (BI3 × BI4).

Mean values and confidence intervals are shown in **Figure 9**. LANGUAGE significantly affected the SD of *f0* in the nuclear part of the phrase ($\chi^2(1) = 15.97$, $p < 0.0001$): SD in English is generally higher by about 0.67 (± 0.13 standard errors) ST. The model is on the borderline of convergence. TYPE OF PROSODIC BREAK significantly affected the SD of *f0* in the nuclear part of the phrase ($\chi^2(1) = 27.99$, $p < 0.0001$), the differences between SD in prosodic phrases ending with BI3 and BI4 range from a slight difference in some speakers (about 0.001 ST in the speaker with the lowest difference) to a considerably large difference (about 1.01 ST in the speaker with the highest difference).

The residuals show a satisfactory level of normality and homoscedasticity.

The test of interaction between LANGUAGE and TYPE OF PROSODIC BREAK is singular and could not be carried out.

Post-hoc tests show that the difference between BI3 and BI4 is significant both in Czech and in English ($p < 0.0001$). The difference between BI4 in Czech and BI4 in English is significant ($p < 0.0001$) and the difference between BI3 in Czech and BI3 in English is also significant ($p < 0.01$).

Figure 10 shows the results for phrases ending in BI4 and without hesitation. LANGUAGE significantly affected the SD of *f0* in in the nuclear part of the phrase ($\chi^2(1) = 17.83$, $p < 0.0001$), SD in English being higher by about 1 (± 0.2 standard errors) ST. Compared to the difference between English and Czech across all prosodic phrases, the difference here is about 1.5 times larger. The residuals are slightly heteroscedastic.

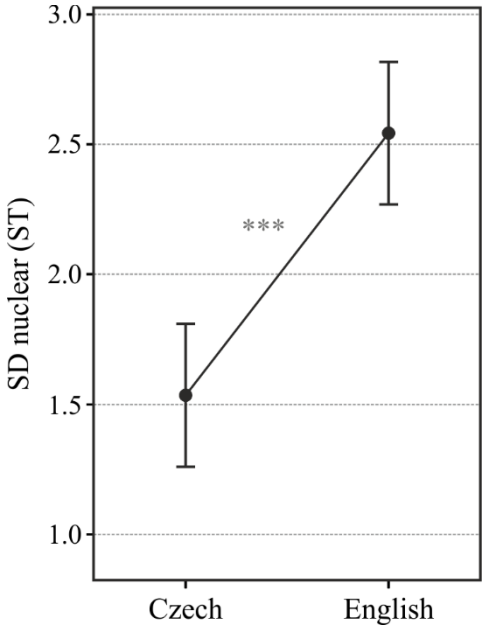

Figure 10. Standard deviation of *f0* in the nuclear part of the phrase in BI4 without hesitation depending on LANGUAGE (Czech × English).

4.2.3 Cumulative Slope Index in each prosodic phrase

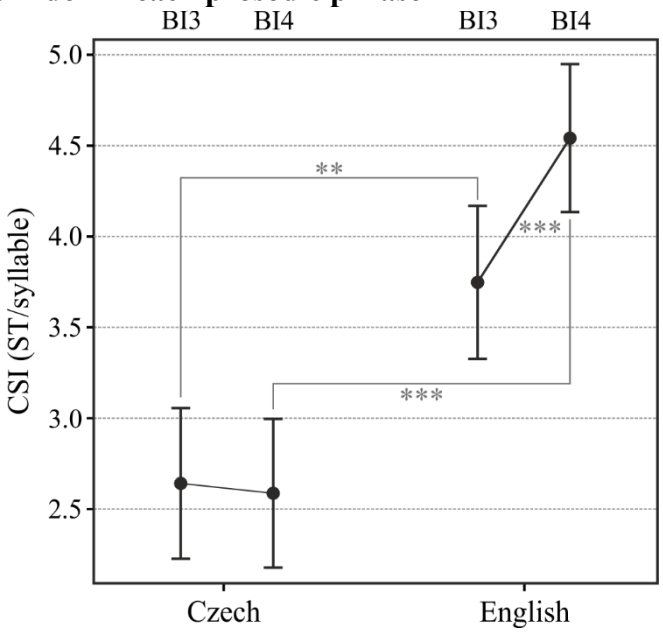

Figure 11.CSI in each prosodic phrase depending on LANGUAGE (Czech × English) and TYPE OF PROSODIC BREAK (BI3 × BI4).

Mean values and confidence intervals are shown in **Figure 11**. LANGUAGE significantly affected the SD of *f0* in the nuclear part of the phrase ($\chi^2(1) = 15.73$, $p < 0.0001$), CSI in English being higher by about 1.62 ST/syllable (± 0.27 standard errors).The test of the effect of TYPE OF PROSODIC BREAK is singular and could not be carried out. The difference between CSI in prosodic phrases ending with BI4 and BI3 is around 0 in Czech speakers, while English speakers tend to have higher CSI in prosodic phrases ending with BI4.

The residuals show a certain degree of non-normality and heteroscedasticity.

The interaction between LANGUAGE and TYPE OF PROSODIC BREAK is significant ($\chi^2(1) = 9.35$, $p < 0.01$).

Figure 12 shows the results for phrases ending in BI4 and without hesitation. LANGUAGE significantly affected the SD of *f0* in in the nuclear part of the phrase ($\chi^2(1) = 24.29$, $p < 0.0001$), SD in English being higher by about 1.99 (± 0.30 standard errors) ST/syllable. The residuals show a certain degree of heteroscedasticity.

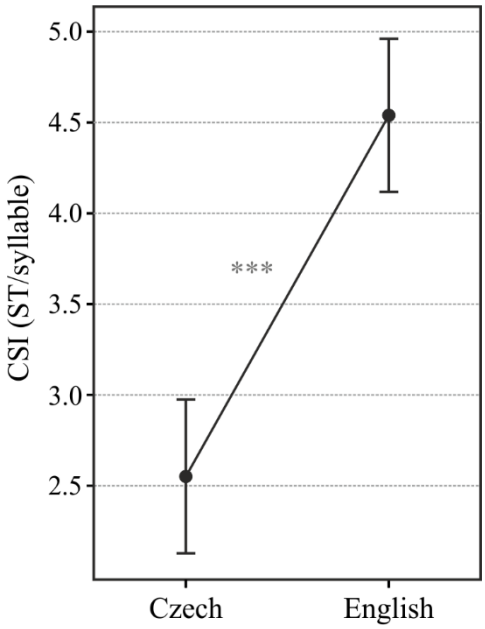

Figure 12. CSI in each phrase ending with BI4 without hesitation depending on LANGUAGE (Czech × English).

5. Discussion

Differences in prosody between Czech and English public speakers have been found both in the temporal and the *f0* characteristics. In accordance with expectations based on naive observation, Czech speakers were shown to produce significantly longer prosodic phrases, narrower pitch range and less melodic variation than English speakers. This may account for Czech intonation sounding flatter and more monotonous in comparison to English.

There were significant differences between prosodic phrases ending with a stronger (BI4) and weaker (BI3) prosodic boundary. Phrases ending with BI4 were longer in both syllables and words and exhibited higher pitch range both in the whole phrase and in the nuclear part of the phrase. Melodic variability as measured by CSI was found to be higher in phrases ending with BI4 in English, but not in Czech.

In this thesis, we examined the speech of good public speakers. Because prosody is influenced by stylistic factors related to the communicative situation and speaker characteristics, our results show characteristics of speech produced in this specific stylistic domain and cannot be easily generalized to the use of Czech and English in other contexts.

5.1 Temporal characteristics

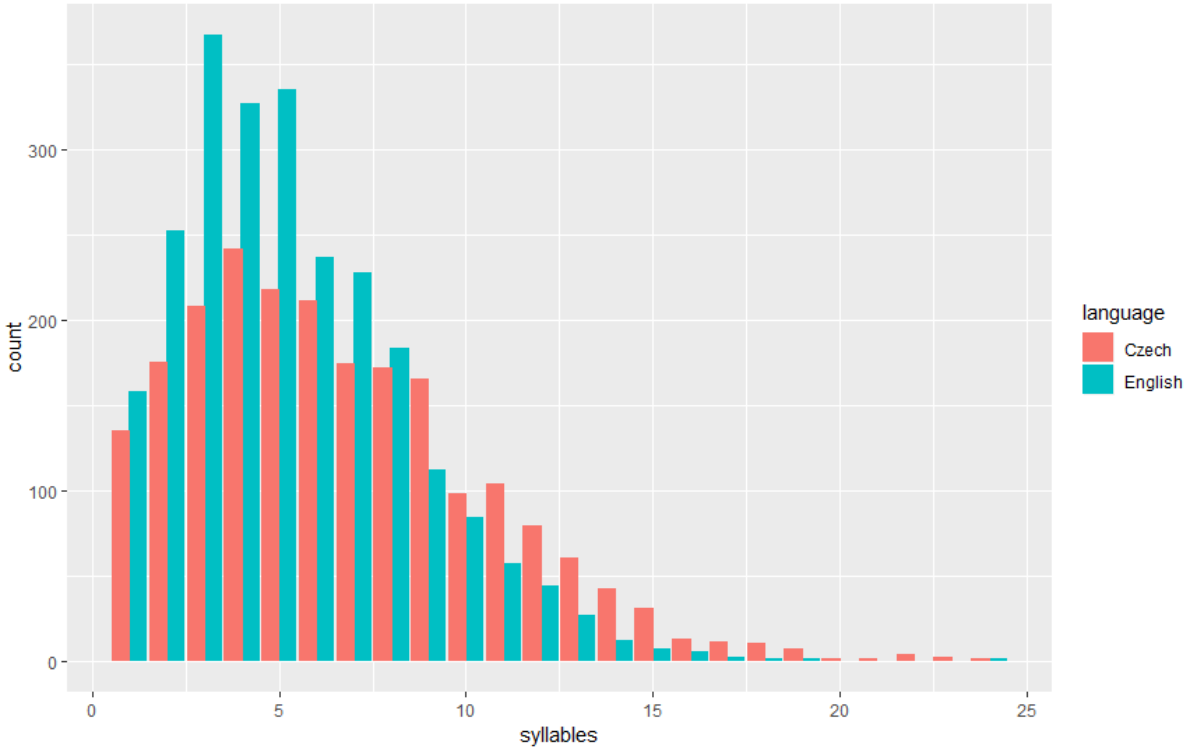

Figure 13. Histogram of syllable counts per prosodic phrase for Czech and English.

Looking at the histogram of syllable counts per phrase in **Figure 13**, we see that the most frequent phrase length for both languages is around 5 syllables (English phrases most frequently

being 3 syllables long, Czech phrases most frequently being 4 syllables long), and the amount of phrases longer than 10 syllables is considerably lower and steadily declining, the counts for Czech being consistently higher than for English after the 10-syllable mark.

This means that most phrases were very short in both languages, even though Czech had longer phrases than English overall. If we exclude phrases realized with a hesitation, the relative amount of 1- and 2-syllable long phrases decreases, as shown in **Figure 14**. This suggests that many of these are cases where the speaker starts speaking without having planned the next segment properly and interrupts and corrects him or herself after the first word or the first syllable.

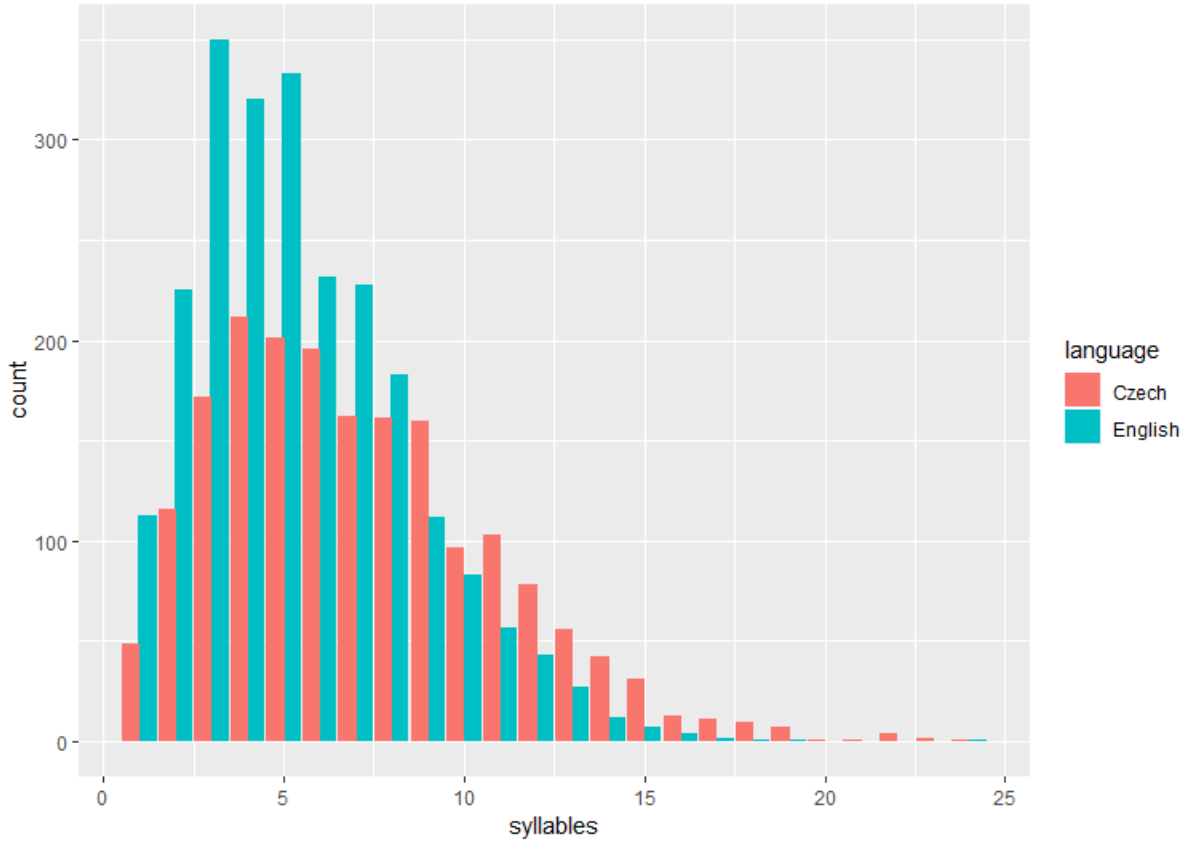

Figure 14. Histogram of syllable counts per prosodic phrase without hesitations for Czech and English.

But the overall trend remains the same, most prosodic phrases in both languages are still very short. This may be accounted for by the stylistic domain. It seems that good speakers in both Czech and English use shorter prosodic phrases, that is they divide their flow of speech more often by prosodic breaks, to achieve a better effect on the audience. A more structured speech makes the process of understanding the message easier, and shorter segments are easier to process. Public speakers will also often use emphasis, segmenting words that they want to stress into their own prosodic phrases, as shown by these examples:

(15) *it turned out to be* | *shame* (EN01)

(16) *every* | *single* | *time* (EN10)

(17) *a říká se tomu* | *exotika* (CZ09)

(18) *co řešíme* | *právě* | *my* (CZ15)

Strangert (2005) examined the speech of a Swedish professional newsreader and politician, both having a reputation of being skilled speakers, and found that their prosodic phrasing is characterized by a high frequency of breaks, often in non-syntactic positions before semantically important words. This suggests that shorter prosodic phrases may be characteristic of skilled public speakers.

However, data from other domains of communication in both Czech and English would be needed to compare the length of prosodic phrases in good public speakers with the length of prosodic phrases in other speaking styles, such as spontaneous communication in everyday situations.

We now move on from discussion of phrase length to discussion of speaking rate. Speaking rate was found to be slightly slower in English than in Czech. We found no significant effect of type of prosodic break on speaking rate in both languages.

Speaking rate has been shown to be dependent on speaking style. Jacewicz & Fox (2010) found that speech rate in English differs significantly in read speech and spontaneous conversational speech (mean value of 3.20 syllables/s in read speech compared to 4.96 syllables/s in spontaneous speech) and that speech rate can be dependent on speaker characteristics such as dialect, age and gender. Smiljanić & Bradlow (2008) found that speech rate in English is slower in clear speech (i.e. speech produced by a speaker who is trying to sound more intelligible to overcome communication barriers, for example when speaking to listeners who have hearing impairment) compared to conversational speech. Veroňková & Poukarová (2017) measured speech rate in professional Czech radio newsreaders. When they compared their results (mean value of 5.8 syllables/second) to results from studies of speech rate in other speaking styles, the speech of professional newsreaders was shown to be faster compared to read speech, speech of guests performing in the radio and direct sports reports (Veroňková & Poukarová, 2017: 104).

We have measured speaking rate by dividing the number of syllables in a prosodic phrase by the duration of the prosodic phrase in seconds, not including pauses. This means that our measures are comparable to measures of articulation rate rather than speech rate, which include pauses. In their study of Czech radio newsreaders, Veroňková & Poukarová (2017) also measured articulation rate in the domain of prosodic phrases, as we have done in this thesis, and their resulting mean

value of 6.1 syllables/s is quite close to our result (mean value of 6.05 syllables/s), suggesting that the speech tempo of Czech professional newsreaders and good public speakers is similar.

Berger, Niebuhr, & Peters (2017) examined prosodic features of charismatic speech and found out that speech manipulated to have a slower speech rate (-1 syllable/s) lead to the speaker being perceived more negatively than unaltered and higher speech rate (+1 syllable/s). Rosenberg & Hirschberg (2009: 647) also found a correlation between positive charisma judgements and faster speaking rate. Stoltzman (2006: 33) found a strong correlation between a higher voicing rate, that is "the number of voiced segments (essentially, syllables) per unit time," (Stoltzman, 2006: 21) and ratings of persuasiveness. This suggests that a higher speaking rate has a positive effect on the audience, perhaps making the speaker sound less hesitant and thus more convincing.

5.2 *f0* characteristics

We measured lower values of standard deviation of *f0* as well as CSI in Czech speakers, which is in accordance with our expectations based on naive observation of the differences between Czech and English intonation.

But, as has already been pointed out earlier, our data describes a specific style of speaking, not everyday communication. Melodic variability and pitch range are dependent on speaking style and the differences between the languages observed here may be significantly larger or smaller other speaking styles.

Dynamism in *f0* characteristics has repeatedly been shown to correlate with ratings of the speaker's charisma, as it leads to the impression of enthusiasm and expressiveness. Especially wider pitch range was found to influence judgements of charisma significantly (Rosenberg & Hirschberg, 2009; Niebuhr, Skarnitzl & Tylečková, 2018; Berger et al., 2017; Strangert & Gustafson, 2008). Using *f0* dynamics for focus and emphasis of important words and phrases also seems to be a successful strategy (Strangert, 2005: 3403). This may lead us to hypothesise that the speech of good public speakers will exhibit a wider pitch range and melodic variability than everyday speech. An interesting question, that also cannot be answered here, is whether the differences in *f0* measurements between Czech and English would be more or less pronounced in everyday speech. Good speakers of Czech may use strategies such as widening their pitch range and speaking with more melodic variation to compensate and sound less monotonous in comparison to their everyday speech, perhaps making the difference between Czech and English less pronounced in the domain of public speaking.

References

Andreeva, B., Demenko, G., Wolska, M., Möbius, B., Zimmerer, F., Jügler, J., & Trouvain, J. (2014). "Comparison of Pitch Range and Pitch Variation in Slavic and Germanic Languages", *Speech Prosody, Dublin, Ireland, 2014*, 776–780.

Bates, D., Maechler, M., Bolker, B. & Walker, S. (2015). *lme4: Linear Mixed-effects Models Using Eigen and S [R package version 1.1-8]*. Retrieved from <http://CRAN.R-project.org/package=lme4>.

Beckman, M. E., & Elam, G. A. (1997). *Guidelines for ToBI Labelling* (version 3.0). The Ohio State University Research Foundation. Retrieved from <http://www.cs.columbia.edu/~agus/tobi/labelling_guide_v3.pdf>.

Beckman, M. E. & Venditti, J. J. (2010). Tone and Intonation. In: Hardcastle, J., Laver, J. & Gibbon, F. (eds.), *Handbook of Phonetic Sciences*, 603–652. Oxford: Blackwell.

Berger, S., Niebuhr, O. & Peters, B. (2017). Winning Over an Audience – A Perception-based Analysis of Prosodic Features of Charismatic Speech. *Proceedings of the 43rd Annual Conference of The German Acoustical Society*, 1454–1457.

Blaauw, E. (1994). The Contribution of Prosodic Boundary Markers to the Perceptual Difference between Read and Spontaneous Speech. *Speech Communication,* 14, 359–375.

Byrd, D., & Saltzman, E. (2003). The Elastic Phrase: Modeling the Dynamics of Boundary-adjacent Lengthening. *Journal of Phonetics*, 31, 149–180.

Crowhurst, M. J. (2018). The Joint Influence of Vowel Duration and Creak on the Perception of Internal Phrase Boundaries. *The Journal of the Acoustical Society of America,* 143(3), 147–153.

Cutler, A., Oahan, D., & van Donselaar, W. (1997). Prosody in the Comprehension of Spoken Language: A Literature Review. *Language and Speech*, 40(2), 141–201.

Dankovičová, J. (1997). The Domain of Articulation Rate Variation in Czech. *Journal of Phonetics*, 25, 287–312.

de Pijper, J. R., & Sanderman, A. A. (1994). On the Perceptual Strength of Prosodic Boundaries and Its Relation to Suprasegmental Cues. *The Journal of the Acoustical Society of America*, 96(4), 2037–2047.

Duběda, T. (2005). *Jazyky a jejich zvuky: univerzálie a typologie ve fonetice a fonologii.* Praha: Karolinum.

Féry, C. (2017). *Intonation and Prosodic Structure*. Cambridge: Cambridge University Press.

Fletcher, J. (2010). Timing and Rhythm. In: Hardcastle, J., Laver, J. & Gibbon, F. (eds.), *Handbook of Phonetic Sciences*, 523–602. Oxford: Blackwell.

Fox, J. (2003). Effect Displays in R for Generalised Linear Models. *Journal of Statistical Software*, 8, 1–27.

Frazier, L., Carlson, K., & Clifton, C. Jr (2006). Prosodic Phrasing is Central to Language Comprehension. *Trends in Cognitive Sciences,* 10(6), 244–249.

Hothorn, T., Bretz, F. & Westfall, P. (2008). Simultaneous Inference in General Parametric Models. *Biometrical Journal*, 50, 346–363.

Hruška, R., & Bořil, T. (2017). Temporal Variability of Fundamental Frequency Contours. *Phonetica Pragensia – AUC Philologica*, 3, 35–44.

Jacewicz, E., Fox, R. A., & Wei, L. (2010). Between-speaker and Within-speaker Variation in Speech Tempo of American English. *The Journal of the Acoustical Society of America,* 128(2), 839–850.

Jun, S.-A. (2003). Prosodic Phrasing and Attachment Preferences. *Journal of Psycholinguistic Research*, 32(2), 219–249.

Keating, P., & Kuo, G. (2012). Comparison of Speaking Fundamental Frequency in English and Mandarin. *The Journal of the Acoustical Society of America*, 132(2), 1050–1060.

Krivokapić, J., & Byrd, D. (2012). Prosodic Boundary Strength: An articulatory and Perceptual Study. *Journal of Phonetics*, 40(3), 430–442.

Ladd, D. R. et al. (1985). Evidence for the Independent Function of Intonation Contour Type, Voice Quality, and F0 Range in Signaling Speaker Affect. *The Journal of the Acoustical Society of America*, 78(2), 435–444.

Ladd, D. R. (2008). *Intonational Phonology* (2nd ed.). New York: Cambridge University Press.

Mennen, I., Schaeffler, F., & Docherty, G. (2012). Cross-language Differences in Fundamental Frequency Range: A Comparison of English and German. *The Journal of the Acoustical Society of America*, 131(3), 2249–2260.

Niebuhr, O., Skarnitzl, R., & Tylečková, L. (2018). The Acoustic Fingerprint of a Charismatic Voice – Initial Evidence from Correlations Between Long-term Spectral Features and Listener Ratings. *9th International Conference On Speech Prosody 2018*, 359–363.

Patterson, D. (2000) *A Linguistic Approach to Pitch Range Modelling* (Doctoral Dissertation). University of Edinburgh.

Pollák, P., Volín, J. & Skarnitzl, R. (2007). HMM-based Phonetic Segmentation in Praat Environment. In: R. K. Potapova (ed.), *The 12th International Conference "Speech and Computer"*, SPECOM 2007, 15–18 October 2007, Moscow, Russia: Proceedings. Moscow: Moscow State Linguistic University, 537–541.

R Core Team (2017). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing. Retrieved from https://www.r-project.org/.

Roach, P. (1991). *English Phonetics and Phonology* (2nd ed.). Cambridge: Cambridge University Press.

Roach, P. (1994). Conversion between Prosodic Transcription Systems: "Standard British" and ToBI. *Speech Communication*, 15(1/2), 91–99.

Rosenberg, A., & Hirschberg, J. (2009). Charisma Perception from Text and Speech. *Speech Communication,* 51(7), 640–655.

Shattuck-Hufnagel, S., & Turk, A. E. (1996). A Prosody Tutorial for Investigators of Auditory Sentence Processing. *Journal of Psycholinguistic Research*, 25(2), 193–245.

Skarnitzl, R., Šturm, P. & Volín, J. (2016). *Zvuková báze řečové komunikace*. Praha: Karolinum.

Smiljanić, R., & Bradlow, A. R. (2008). Temporal Organization of English Clear and Conversational Speech. *The Journal of the Acoustical Society of America,* 124(5), 3171–3182.

Speer, S. R., Warren, P., & Schafer, A. J. (2011). Situationally Independent Prosodic Phrasing. *Laboratory Phonology*, 2(1), 35–98.

Stoltzman, W. T. (2006). *Toward a Social Signaling Framework: Activity and Emphasis in Speech* (Master's thesis). Massachusetts Institute of Technology.

Strangert, E. (2005). Prosody in Public Speech: Analyses of a News Announcement and a Political Interview. *Interspeech 2005*, 3401–3404.

Trouvain, J. (2004) *Tempo Variation in Speech Production: Implications for Speech Synthesis*. (Doctoral dissertation). Saarland University.

Strangert, E., & Gustafson, J. (2008). What Makes a Good Speaker? Subject Ratings, Acoustic Measurements and Perceptual Evaluations. *Interspeech 2008*, 1688–1691.

Vaissiere, J. (1983). Language-Independent Prosodic Features. In Cutler, A. & Ladd, R. (Eds.), *Prosody: Models and Measurements*, 53–66. Berlin, Heidelberg: Springer.

Veroňková, J., & Poukarová, P. (2017). The Relation between Subjective and Objective Assessment of Speaking Rate in Czech Radio Newsreaders. *Phonetica Pragensia – AUC Philologica*, 3, 95–107.

Volín, J., Poesová, K., & Weingartová, L. (2015). Speech Melody Properties in English, Czech and Czech English: Reference and Interference. *Research in Language*, 13(1), 107–123.

Watson, D., & Gibson, E. (2004). The Relationship between Intonational Phrasing and Syntactic Structure in Language Production. *Language and Cognitive Processes*, 19(6), 713–755.

Watson, D., Breen, M., & Gibson, E. (2006). The Role of Syntactic Obligatoriness in the Production of Intonational Boundaries. *Journal of Experimental Psychology: Learning, Memory, And Cognition*, 32(5), 1045–1056.

Wickham, H. (2009): *ggplot2: Elegant Graphics for Data Analysis*. New York: Springer.

Wichmann, A. (2005). Please – from Courtesy to Appeal: The Role of Intonation in the Expression of Attitudinal Meaning. *English Language and Linguistics*, 9(2), 229–253.

Yuan, J., & Liberman, M. (2008). Speaker Identification on the SCOTUS Corpus. *Proceedings of Acoustics '08*. Retrieved from http://www.ling.upenn.edu/~jiahong/publications/c09.pdf

Resumé

Hlavním cílem této bakalářské práce bylo popsat prozodické členění u dobrých mluvčích v češtině a angličtině. Zaměřili jsme se na prozodickou frázi a její temporální a melodické charakteristiky – měřili jsme její délku a mluvní tempo, intonační rozpětí a melodickou variabilitu v jejím rámci. Analyzovali jsme části veřejných projevů dobrých řečníků v češtině a americké angličtině. Výsledky ukazují, že mezi oběma jazyky jsou v prozodickém členění významné rozdíly jak v melodických, tak v temporálních charakteristikách.

Práce má teoretickou část, ve které jsou představeny základní koncepty a některé relevantní výsledky předchozích výzkumů, a praktickou část, ve které je představen samotný výzkum. V teoretické části začínáme vysvětlením obecného termínu prozodie. Prozodií se rozumí jevy, které se v řeči projevují nad úrovní jednotlivých hlásek, tedy jevy jako melodie, temporální charakteristiky (např. tempo, rytmus, délka prozodických jednotek), hlasitost a kvalita hlasu. Těm odpovídají akustické charakteristiky jako základní frekvence (*f0*), trvání, intenzita a spektrální charakteristiky. Prozodie plní v procesu komunikace řadu funkcí, které lze rozdělit na těchto šest základních: lexikální, gramatická, diskurzní, vytýkací (kterou je možné zahrnout pod diskurzní), afektivní a indexovou. Lexikální funkce slouží k rozlišování lexikálního významu (melodický průběh je součástí samotné lexikální jednotky), gramatická funkce slouží k indikaci větného typu a syntaktické struktury věty, diskurzní funkce slouží k regulaci průběhu komunikace mezi jejími účastníky, vytýkací funce slouží ke zdůraznění určitých slov, afektivní funkce nese informace o afektivním stavu mluvčího, indexová funkce nese informace o identitě mluvčího a jeho sociální příslušnosti.

V této bakalářské práci byly zkoumány melodické a temporální charakteristiky řeči, proto je jim v teoretické části věnována bližší pozornost. Melodie řeči, tj. intonace, označuje posloupnost tónů různých výšek realizovanou v rámci promluvy. Akustickým korelátem vnímané výšky je základní frekvence (*f0*), která odráží rychlost kmitání hlasivek. Melodický průběh výpovědi lze popsat pomocí elementárních jednotek, základných melodických pohybů nebo tónů a jejich kombinací. K popisu toho, jak široký je rozsah frekvencí, které mluvčí během řeči používá, se používá termínu intonační rozpětí. Běžně se měří jako rozdíl mezi minimem a maximem *f0*, rozdílem mezi 90. a 10. percentilem nebo 75. a 25. percentilem *f0* a směrodatnou odchylkou *f0*. Temporální charakteristiky řeči souvisí s rytmem a řečovým tempem. Rytmus lze definovat jako vnímanou pravidelnost ve střídání kontrastních prvků. Tempo lze definovat jako rychlost řeči a lze jej měřit ve slabikách za sekundu, buď včetně pauz, nebo bez pauz. Tempo řeči může být ovlivněno různými faktory, např. afektivním stavem mluvčího, komunikační situací, komunikačním partnerem a mluvním stylem.

Konkrétně se bakalářská práce zaměřuje na prozodické členění. Tím se rozumí členění proudu řeči na menší jednotky, které jsou definované na různých úrovních. Nad úrovní jednotlivých hlásek lze rozlišovat slabiky, mluvní takty, prozodické fráze a promluvy. Prozodická fráze je jednotka tvořená koherentním, kompletním melodickým pohybem a ohraničená prozodickými předěly. V rámci fráze je realizován melodém, jeden z inventáře základních melodických pohybů daného jazyka. Většinou začíná na přízvučné slabice posledního autosémantika ve frázi, ale může být realizován i na jiném slově, pokud je mluvčí chce zdůraznit. Hranice prozodických frází jsou tvořeny prozodickými předěly, které jsou většinou signalizovány více či méně výrazným melodickým pohybem, závěrovým zpomalením (delším trváním hlásek na konci fráze) nebo pauzou.

Prozodické členění úzce souvisí se syntaktickou strukturou věty. Prozodickou frázi lze vidět jako jednotku do určité míry analogickou větě nebo syntaktické frázi a prozodické předěly se často kryjí s důležitými syntaktickými hranicemi. Prozodické členění tak může indikovat syntaktickou strukturu věty (jak již bylo zmíněno v popisu základních funkcí prozodie). Slova, která k sobě významově patří nebo jsou na sobě syntakticky závislá, bude mluvčí mít tendenci realizovat v rámci jedné prozodické fráze, a naopak slova, která k sobě nepatří, bude mít tendenci rozdělit prozodickými předěly. Prozodické členění však není syntaktickou strukturou zcela determinováno, mnoho prozodických hranic je v řeči realizováno na místech, která nejsou syntakticky motivovaná, přestože promluva zní zcela přirozeně. Na základě syntaktické struktury věty lze tedy pouze predikovat, kde budou při její realizaci prozodické hranice umístěny více pravděpodobně, a kde míně pravděpodobně. Prozodické členění je ovlivněno i dalšími faktory, např. informační strukturou věty, tendencí produkovat stejně dlouhé fráze nebo délkou nebo postojem a afektivním stavem mluvčího.

Různé jazyky zachází s prozodickým členěním různým způsobem. Výzkum provedený v rámci této bakalářské práce zkoumal rozdíly mezi angličtinou a češtinou. Neformální pozorování naznačují, že v běžné řeči jsou prozodické fráze v češtině ve srovnání s angličtinou velmi dlouhé a intonačně ploché. Byly také zjištěny rozdíly mezi intonačním rozpětím českých a anglických profesionálních hlasatelů – v češtině bylo intonační rozpětí významně menší. Náš výzkum se zaměřil na oblast veřejných projevů dobrých řečníků.

Výzkum spočíval v analýze nahrávek mluvčích češtiny a americké angličtiny. Jednalo se o příspěvky na konferenci TEDTalk, tj. popularizační přednášky o různorodých atraktivních tématech pro širokou veřejnost. Předem jsme z cca 40 mluvčích, jejichž přednášky jsou volně k dispozici na YouTube, vybrali 15 mluvčích každého jazyka na základě vlastního hodnocení jejich kvality. Předpokládáme, že samotná účast na konferenci TEDTalk je určitou zárukou, že mluvčí je dobrý

řečník, ale abychom zajistili, že tomu tak skutečně je, provedli jsme percepční test, ve kterém jsme nechali dvě skupiny po 8 respondentech ohodnotit předem vybraných 15 mluvčích buď v češtině, nebo v angličtině (tj. mluvčí češtiny a angličtiny byli hodnoceni zvlášť). Respondenti slyšeli asi 30sekundové úseky z projevu každého mluvčího a měli na 7bodové škále vyjádřit svou ochotu zaměstnat ho jako svého tiskového mluvčího. 10 nejlépe hodnocených mluvčích jsme pak dále analyzovali v Praatu.

Nahrávky byly rozděleny na asi minutu dlouhé úseky a automaticky segmentovány na fonémy. Úseky 3–7 (tj. celkem asi 5 minut) od každého mluvčího jsme dále analyzovali. Prozodické členění bylo označeno manuálně na základě poslechu. V souladu s transkripčním systémem ToBI jsme označili hloubku prozodických předělů pomocí indexů BI3 a BI4. BI4 značí nejvýraznější prozodický předěl, realizovaný s výrazným melodickým pohybem a závěrovým zpomalením, případně pauzou. BI3 značí méně výrazný prozodický předěl, realizovaný s méně výrazným melodickým pohybem a menší mírou závěrového zpomalení. Pokud byl prozodický předěl realizován se zaváháním, přidali jsme označení „p.“ Také jsme označili, na kterém slovu ve frázi začíná realizace melodému.

Na takto anotovaném materiálu jsme měřili tyto temporální a melodické charakteristiky:

- počet slabik ve frázi
- počet slov ve frázi
- mluvní tempo ve slabikách/sekundu
- směrodatnou odchylku základní frekvence *f0* v rámci prozodické fráze v půltónech (ST)
- směrodatnou odchylku *f0* v melodémové části v ST
- Cumulative Slope Index (CSI) v rámci prozodické fráze v ST/sekundu

Počet slabik ve frázi byl extrahován pomocí skriptu jako počet vokálů v angličtině a jako počet vokálů a slabičných konsonantů v češtině. Slabičné konsonanty v češtině byly definovány jako [r] a [l] mezi dvěma konsonanty. Základní frekvence byla extrahována pomocí autokorelace v Praatu, vyhlazena pomocí 10Hz filtru (aby se odstranily malé, pro vnímanou intonaci irelevantní melodické pohyby) a interpolována (aby se vytvořila spojitá křivka nepřerušovaná v neznělých úsecích). CSI, tj. míra melodické variability, se počítá jako součet rozdílů ve frekvencích mezi po sobě následujícími body na křivce průběhu *f0* podělený trváním daného úseku.

Významnost vlivu jazyka (angličtina vs. čeština) a typu prozodického předělu (BI3 vs. BI4) na tyto charakteristiky jsme zjišťovali pomocí lineárních modelů se smíšenými efekty s faktorem mluvčí jako náhodným efektem. Náhodný efekt intercept (výchozí hodnota) u mluvčího zohledňuje to, že jednotliví mluvčí se mohou v měřených charakteristikách od sebe významně lišit, a náhodný efekt slope (sklon) u mluvčího zohledňuje to, že jednotliví mluvčí mohou oba typy předělů

realizovat různě. Významnost fixních efektů jazyka a typu prozodického předělu jsme zjišťovali srovnáním plného modelu s modelem, ve kterém byl daný efekt odebrán.

Výsledky ukázaly, že mezi češtinou a angličtinou jsou významné rozdíly v temporálních i melodických charakteristikách. V angličtině jsou prozodické fráze celkově o o 0,89 slabiky kratší než v češtině, pokud odhlédneme od frází realizovaných se zaváháním dokonce o 1,72 slabiky kratší. V angličtině je tempo v rámci fráze celkově o 0,76 slabiky/s nižší než v češtině. Směrodatná odchylka *f0* je v angličtině vyšší jak v rámci celé prozodické fráze (0,79 ST), tak v melodémové části (o 0,67 ST). CSI v rámci fráze je v angličtině o 1,62 ST/slabiku vyšší než v češtině. Zároveň se ukázaly i rozdíly mezi frázemi zakončenými výzraznějším (BI4) a méně výrazným (BI3) prozodickým předělem. Fráze zakončené předělem BI4 byly delší a směrodatná odchylka *f0* zde byla vyšší jak v rámci celé fráze, tak v melodémové části. CSI bylo vyšší ve frázích zakončených BI4 pouze v angličtině.

Potvrdil se tedy předpoklad, že v oblasti veřejných projevů produkují mluvčí češtiny delší prozodické fráze s užším intonačním rozpětím a méně výraznou melodickou variabilitou než mluvčí angličtiny. Zároveň byly prozodické fráze v obou jazycích velmi krátké – nejčetnější délka fráze v češtině byla 4 slabiky, v angličtině 3 slabiky, v obou jazycích bylo jen málo frází delších než 10 slabik. To může souviset se stylem veřejných projevů. Zdá se, že dobří mluvčí v obou jazycích člení proud řeči častěji a na kratší úseky za účelem dosáhnout lepšího dojmu na posluchače. Projev, který je jasně a výrazně strukturován, se posluchačům lépe zpracovává a napomáhá jim snadno porozumět obsahu sdělení. Předchozí výzkumy ukázaly, že větší intonační rozpětí koreluje s vnímáním mluvčího jako charismatického, a tedy má také pozitivní efekt na posluchače. To by mohlo znamenat, že dobří řečníci strategicky využívají ve veřejných projevech většího intonační rozpětí než v běžné konverzaci. Bylo by však nutné srovnat výsledky tohoto výzkumu s daty z jiných oblastí komunikace, např. právě běžné, neformální konverzace, abychom mohli stanovit významnost vlivu mluvního stylu na prozodické členění.