

**Univerzita Karlova**  
**Filozofická fakulta**

# **Bakalářská práce**



Praha 2022

Adéla Hertlová

**Univerzita Karlova**

**Filozofická fakulta**

ÚSTAV ANGLICKÉHO JAZYKA A DIDAKTIKY

# **Bakalářská práce**



Adéla Hertlová

## **Melodic patterning in advanced Czech students of English**

Melodické členění u pokročilých českých studentů angličtiny

Praha 2022

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

## **Acknowledgements**

I would like to sincerely thank doc. Mgr. Radek Skarnitzl, Ph.D., for his very patient and kind guidance, helpful advice, and all the time spent on this thesis.

**Prohlášení:**

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

V Praze, dne 26. května 2022

Adéla Hertlová

Souhlasím se zapůjčením bakalářské práce ke studijním účelům.

**Abstract and key words:**

This thesis is aiming to observe the process of acquisition of intonation in native Czech students of English on an advanced level, who were regularly and intensely exposed to English. The students were recorded reading a simple text at the beginning of their studies, and for the purpose of this thesis, they were re-recorded two years later, and compared to two native speakers of English. The theoretical part of this thesis first explains the key terminology of intonation together with its functions. The following part of the theoretical background focuses on the acquisition of intonation, the differences between Czech and English in regards to intonation, and is concluded with research questions. The following chapter consists of the methodology employed in this thesis: the particularities regarding the students, the text, and the recording process, together with how the analysis was performed and the data extracted. This is followed by the results and their interpretation, which is further elaborated in the following chapter as well. The data generally indicated a decrease in the melodic range of the students recorded two years later. Yet, some improvements were observed in the distribution of tones, where the later recordings indicated an increase in complex tones. A slight positive development was also noted in another prosodic aspect analysed – phrase length, where the students recorded two years later made slightly shorter prosodic phrases. It should be however stressed that the students recorded in their first years of studies have already performed similarly to the two native speakers. It is, therefore, possible that intonation may be already acquired in some of the earlier stages of language learning.

**Key words:** intonation, acquisition of intonation, L2 English, prosody, students of English, Czech speakers

### **Abstrakt a klíčová slova:**

Tato práce se zaměřuje na osvojování melodických kontur angličtiny u českých pokročilých studentů angličtiny v intervalu dvou let, při kterých byli s tímto jazykem v intenzivním kontaktu. Studenti byly nahráni při čtení jednoduchého textu na začátku svého studia, a poté znovu po přibližně dvou letech. Pro srovnání byly ještě nahrány dvě rodilé mluvčí angličtiny. Teoretická část této práce se zaměřuje na klíčové pojmy v rámci studia intonace společně s její funkcí. Následující část teorie popisuje osvojování intonace na obecné rovině, dále pak konkrétní melodické rozdíly mezi češtinou a angličtinou a je zakončena výzkumnými otázkami. Další kapitola popisuje užitou metodologii především ohledně povahy mluvčích, textu, procesu nahrávání, analýzy a extrakce dat. Následují výsledky a jejich interpretace, která je pak dále rozvedena v další kapitole. Data této práce vykazují celkový pokles v intonačním rozpětí u mluvčích nahrávaných po dvou letech, ovšem zároveň došlo ke zlepšení v poměru použitých tónů, kde došlo k nárůstu komplexních kontur. Další prozodický aspekt, který byl analyzován, byla délka prosodických frází, kde studenti nahrání o dva roky později tvořili kratší fráze. Mělo by ale zároveň být zdůrazněno, že nahrávky studentů už z prvního roku jejich studia měly velmi podobná data v porovnání se skupinou rodilých mluvčích. Proto je možné, že k osvojování intonace by mohlo docházet už ve dřívějších stádiích studia cizího jazyka.

**Klíčová slova:** intonace, osvojování intonace, L2 angličtina, prozodie, studenti angličtiny, čeští mluvčí

## Table of contents

1. Introduction .....	7
2. Prosodic Structure .....	9
2. 1. Word Level.....	9
2. 2. Phrase-level Intonation.....	10
2. 3. Phrasing .....	11
2. 4. Nucleus .....	13
2. 5. Functions of Intonation.....	15
3. Acquisition of Intonation.....	17
3. 1. Dimensions of Intonation Acquisition.....	17
3. 2. L2 Acquisition of the Intonation Features.....	18
3. 3. Intonational Differences between English and Czech.....	20
3. 4. Research Questions .....	22
4. Method.....	23
4. 1. Speakers.....	23
4. 2. Text.....	23
4. 3. Recording Process .....	24
4. 4. Analysis .....	24
4. 5. Data Extraction.....	26
5. Results and Discussion .....	27
5. 1. Phrase Length .....	27
5. 2. General Distribution of Tones .....	29
5. 3. 80-percentile Range of Pre-Nuclear Areas and Tone.....	31
5. 4. 80-percentile Range of F0 in Nuclear Area for Individual Speakers .....	36
6. General Discussion.....	39
7. Conclusion.....	42
8. References .....	43
9. Resumé .....	45
10. Appendix .....	49

## Table of Figures:

Figure 1: Structure of a prosodic phrase (Levis & Wichmann, 2015: 139).....	12
Figure 2: Possible variation of pitch movement on nuclear syllables relating to the onset in comparison to speakers baseline pitch (Cruttenden, 1997: 45).....	13
Figure 3: Tonal development with more than one accented syllable in the head. (Roach, 2009: 139).....	14
Figure 4: Presumable tonal movement of fall-rise in polysyllabic tails (Roach, 2009: 137)...	14
Figure 5: Tendency of rise-fall tonal development in polysyllabic tails (Roach, 2009: 138)..	15
Figure 6: Marginal rise-falls encountered during analysis that were labelled as level .....	25
Figure 7: Length of prosodic phrases in syllables per phrase produced by all speakers based on the groups. The boxplot shows the median value and inter-quartile range (25%-75%) for each group.....	27
Figure 8: Length of prosodic speakers and comparison with median value and inter-quartile range .....	28
Figure 9: Overall length of prosodic phrases in words per phrase, showing the median and inter-quartile range for each group.....	28
Figure 10: Length of prosodic phrase produced by individual speakers in words per phrase with median and inter-quartile range for each speaker.....	29
Figure 11: Distribution of tones in total numbers .....	29
Figure 12: Tone distribution among the groups in percentage.....	30
Figure 13: Overall 80-percentile range in phrases - including also the pre-nuclear area, with median value and 80-percentile range .....	31
Figure 14: Median and 80-percentile range in phrases divided by speakers .....	32
Figure 15: Overall median values and 80-percentile range in tones for each of the groups ....	32
Figure 16: Median and 80-percentile range in the tonal area divided by individual speakers.	33
Figure 17: Median and 80-percentille range in phrase for each tone.....	34
Figure 18: Median and 80-percentile range in tones for each of the nuclear tones .....	35
Figure 19: Comparison of median values of individual speakers in the early and late group regarding the 80-percentile range of tonal movement in fall .....	36
Figure 20: Comparison of median valued of individual speakers in the early and late group regarding the 80-percentile range of tonal movement in rise.....	36
Figure 21: Comparison of median values of individual speakers in the early and late group regarding the 80-percentile range of tonal movement in fall-rise .....	37
Figure 22: Comparison of median values of individual speakers in the early and late group regarding the 80-percentile range of tonal movement in rise-fall .....	38

## **1. Introduction**

Learning a foreign language has become a basic education standard in the majority of countries of the world. Predominantly, the language children in Europe start learning as a part of elementary school education of foreign languages is English – mostly for its wide global use. Although language learning starts usually considerably early, it does not guarantee to master the foreign language in all targeted features. Especially in intonation, the chances of targeting native speaker production may be considerably low since intonation is often scarcely taught in the education system. This deviation of the native-like production of intonation may add to the sense of foreign-accentedness of L2 speakers and can possibly result in social stigmatization as well – especially for students who chose English as their field of academic studies and therefore would be expected to have a very high level of English.

This thesis focuses mostly on the acquisition of intonation features in advanced students of English in the interval of two years, although the analysis also included the phrase length, which is another prosodic parameter apart from intonation. The students were native speakers of Czech, who have been accepted as students of English and American Studies at Charles University, Faculty of Arts, with the assumed level of English to be B2 (Common European Framework of Reference for Languages). They were first recorded in 2019 reading a simple text, which they later analysed by self-evaluating their pronunciation. Two years later, this text was re-recorded, and the intonation developments were analysed. As a comparison, one speaker of American English and one speaker of British English were also recorded and analysed. This thesis then aimed to prove that after a two-year interval, the students would expand their range of melody in phrases and tones, and target the native production of intonation, as well as make shorter phrases, again fairly matching the natives.

Apart from this introductory section, this thesis consists of chapters on Prosodic Structure, Acquisition of Intonation, Methodology, Results and Discussion, General Discussion, Conclusion, References, Resumé, and Appendix. The second and third chapter, Prosodic Structure and Acquisition of Intonation, present a theoretical background of this thesis. The prosodic structure explains the realization of intonation on word level and phrase level, as well as some of the key concepts of prosody – prosodic phrases, tones, and intonation with its functions. This chapter also establishes the terminology this thesis uses. The following chapter of the theoretical background focuses on the process of language acquisition with an emphasis on the acquisition of intonation. By the end of this chapter, the intonational differences between Czech and English are analysed, precluding which aspects of English intonation may be the most

difficult to acquire, and the research questions are presented. Another chapter explains the methodology of this thesis; basic information about the speakers, the process of recording, the features of the text as well as the analysis and data extraction. This is followed by the presentation of data and their brief interpretation in regards to the length of phrases and tones, and their 80-percentile range of F0. General Discussion then presents a summary of the previous interpretation of the data, discusses the possibilities of such outcome, and presumes future developments in the study of the acquisition of intonation. The results of the thesis together with their influence on the study of intonation acquisition are presented in the Conclusion. Lastly, the paper concludes with a list of references, a resumé in Czech, and an appendix.

## **2. Prosodic Structure**

### **2. 1. Word Level**

Prosodic features can be analysed on two levels: word and phrasal. This passage will discuss how prosodic features, such as pitch, loudness, length, and sound quality, operate particularly on the word level. Each phoneme has a specific realization: some phonemes are inherently more prominent than others and some are reduced in order to make the others stand out. The most prominent syllable is then denoted as the primary word stress. Length, loudness, and pitch are prosodic features “most consistently used for linguistic purposes” (Cruttenden, 1997: 2) in analysing prosody: these subjective elements are represented by the objective measures of duration, intensity, and fundamental frequency (F0), respectively. Although there is a correlation between the objective acoustic measure and the subjective prosodic feature, they often differ in the realization and perception: the objective duration is fundamentally dependent on the length of time the listener hears certain units, which does not have to correlate with the objective duration of a certain speech unit, loudness corresponds to the speaker’s breath force used to produce speech, and the pitch generally involves the development of melody realized by variation of fundamental frequency (Cruttenden, 1997: 2-4). From these features, the most essential one determining the location of the word stress in English is the pitch. The production of pitch is related to the frequency of vibration of the vocal folds and thus creates what we hear as the melody – the musical notion of pitch notes (Roach, 2009: 74). Perception-wise, those vibrations then seem like a continuous melody, which however does not match the actual production. Primarily, not all phonemes are able to carry a melody, solely sonorants have this function (vowels, nasals, and liquids), while on the contrary, plosives or fricatives usually have little influence on the continuity of melody (Levis & Wichmann, 2015: 142). Analysis through a spectrogram then reveals even more inconsistencies – there are often visible spikes in pitch or sequences that are significantly higher or lower than the surrounding pitch pattern, none of which is audible in continuous speech (Levis & Wichmann, 2015: 142). Loudness can be especially misleading when it comes to detecting the stressed syllable, since multiple other features than prominence can manipulate the impression of loudness – first of all, the production is frequently not linear to the actual perception (twice as loudly produced by the speaker does not mean twice as loudly perceived by the listener) (Cruttenden, 1997: 3). Secondly, loudness is able to cover a sequence of syllables rather than solely the one accented, and lastly, the loudness is also very much related to an affective response – uttering something loud can merely signal the speaker’s emotion, for example anger or frustration, and thus even unstressed

syllables may be loud (Cruttenden, 1997: 3). As for length, the production of certain syllables is also bound to their location, as many languages prolong the final syllable, regardless of the position of the stressed syllable – one of them being also Czech (Skarnitzl & Volín, 2019: 2891), although the objectively produced duration does not seem to clearly correlate with the perceived length.

For English, Cruttenden (1997: 44) distinguishes overall four degrees of word stress in English:

- primary stress, which involves the most prominent pitch and is also referred to as a nucleus or a nuclear syllable, e.g.: *participation* (Wells, 2008, 3rd edition)
- secondary stress, which has a subsidiary prominence (other prominent, syllables around the nucleus), e.g.: *participation* (Wells, 2008, 3rd edition)
- strong unstressed, which is produced primarily by loudness and/or length, e.g.: *participation* (Wells, 2008, 3rd edition)
- weak unstressed, which bears no acoustic prominence e.g.: e.g.: *participation* (Wells, 2008, 3rd edition)

The placement of the word stress in English is very difficult to predict, although generally, the stress is dependent on the word class, number of syllables, phonological structure of syllables, or on the presence of affixes (Roach, 2009: 76). In Czech, the word stress is placed on “the first syllable of the prosodic word” (Skarnitzl & Eriksson, 2017: 3221), regardless of affixation.

## 2. 2. Phrase-level Intonation

Even though there is no single and clear definition of intonation (Vaissiere, 2005: 238), it could generally be described as a variation of the speaker’s fundamental frequency “for conveying information at levels higher than the word” (Vaissiere, 2005: 236). As already stated, intonation is concerned with both production and perception: the analysis involves how certain intonation cues can modify the perception of information relevance, how certain syntactical functions are delivered through intonation, or how the production and perception of certain prosodic cues differ on both the word level and the beyond-word level, which is called the “phrase-level”. It should be also said that there is a certain inconsistency in the use of the terms “accent” and “stress”, this thesis will use the word “stress” when denoting the most prominent syllable on the word level, and an “accent” referring to a prominent realization of stress in an utterance.

As for the phrase level, intonation can be one of the elements that help the listener to identify and interpret the meaning of speech. There seems to be a close relationship between the communicative importance of a phrase element and the degree of prosodic prominence (Chamonikolasová, 2017: 35). Among the forms of intonation, one of the most essential elements to consider during an analysis of speech is the realization of tone – the melodic

movement starting on the stressed syllable of a word or a phrase that frequently also carries the highest degree of communicative dynamism (Firbas, 1974: 112). In other words, the tone usually serves as an indicator of the rhematic part of a syntactic element containing a new or important piece of information. Tones then divide speech into prosodic units – this thesis will refer to those units with the term “prosodic phrase”, although there is a general confusion in the terminology. The same phenomena can be found under many synonymous terms, such as an intonation-group, intonation phrase, or tone unit.

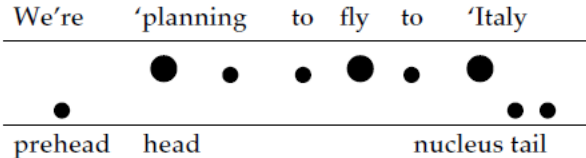
### **2. 3. Phrasing**

Prosodic phrases are not inherently bound to any grammatical units, although in the majority of cases (about 40%), the distribution of prosodic phrases matches the clause division (Cruttenden, 1997: 69). In English, there is generally a high tendency to divide one sentence into multiple prosodic phrases; we often encounter prosodic phrases that are matching the syntactic phrase division, and in fewer cases, one prosodic phrase can only consist of one word (Cruttenden, 1997: 69). The distribution of phrases also frequently corresponds with the speaker’s employment of adverbials, and particularly disjuncts and conjuncts, which semantically imply a separation, coordination, or evaluation of the rest of the phrase (Cruttenden, 1997: 69). Some adverbials that tend to be prosodically isolated are “those labelled message-attitudinal (e.g. *unfortunately*), message-likelihood (e.g. *obviously*), viewpoint (e.g. *officially*), speaker/listener-oriented (e.g. *seriously*), style (e.g. *briefly*)” (Cruttenden, 1997: 69), or appositions (Cruttenden, 1997: 71). Some adjuncts also tend to be uttered in an isolated prosodic phrase especially when they are not a part of the verb valency. This is frequently projected in written discourse by the use of punctuation, which to a certain extent matches the division of prosodic phrases.

Predicting the boundaries of prosodic phrases can be considerably problematic, since it is largely dependent on the speaker’s choice (Cruttenden, 1997: 72,73). There is an implicit tendency to divide the phrase by length; it has been observed that on average, one prosodic phrase usually consists of five words and rarely exceeds seven words (Cruttenden, 1997: 72). Each prosodic phrase then may involve a “pitch discontinuity and changes of loudness and tempo (“final lengthening”) and increased vocal fry (“creak”)” (Levis & Wichmann, 2015: 144), which indicates prosodic boundaries and creates a clearer distribution of spoken discourse – thus making speech more comprehensible. Occasionally, a pause can function as a boundary as well, usually indicating a more significant text division, such as the beginning of a new sentence or paragraph. This pause can be either unfilled (meaning silence) or filled (Cruttenden, 1997: 30); the filled pause is often realized by a central vowel [ə], a bilabial nasal [m]

(Cruttenden, 1997: 30), or a combination of those two [əɪm] in various length. The pause can also function as an indicator of an error or a word-finding difficulty, which is denoted as a hesitation pause (Cruttenden, 1997: 30, 31).

The prosodic phrase can be further divided into a prehead, head, nucleus, and tail, as demonstrated in *Figure 1* (Levis & Wichmann, 2015: 140). Prehead is an unaccented part of a phrase that carries no acoustic prominence; thus it often serves as an introduction and can alert the listener that more significant information is to come. Head may have a certain melodic prominence; however, it will likely be perceived as a sort of melodic variation before the nucleus and will likely merely serve as a means to avoid monotonousness. The relationship between prehead and the head is explained by Levis & Wichmann (2015: 140); the head is “the stretch from the onset to the nucleus,” therefore prehead entails “any preceding unstressed syllables.” The nucleus consists of one syllable, and it is the vital part of the utterance, informatively mostly relevant and in many languages also acoustically the most prominent. The tail then serves as a continuation of the tonal development initiated on the nuclear syllable. In this example, the most prominent words are those that are also most essential meaning-wise; *planning, fly, Italy*. *We’re* serves as an introductory element, and therefore will most probably not carry an accent, and therefore serve as a prehead. The most informatively vital element of this prosodic phrase is the word *Italy*, it is therefore the nucleus, or the nuclear syllable. The following syllables are then labelled as a tail and help to create the falling contour of this nuclear tone.

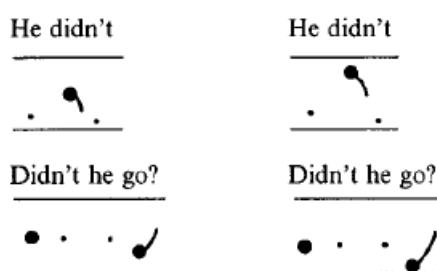


*Figure 1: Structure of a prosodic phrase (Levis & Wichmann, 2015: 139)*

As stated previously, the most prominent syllable of the prosodic phrase will be the one conveying some new or important piece of information (Levis & Wichmann, 2015: 148), or in other words, the rhematic part of a clause – as should be now obvious, this syllable will likely be the longest, loudest and with the most prominent pitch pattern. As has been implied, this thesis will refer to the primary accent of a prosodic phrase as a nucleus, although again there is no consistency in the English terminology; the same term is also found under other denotations such as “tonic” syllable.

## 2. 4. Nucleus

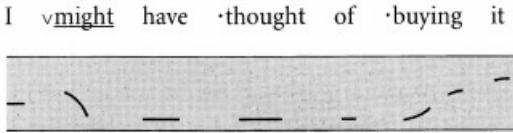
The nucleus is the vital element of the prosodic phrase, and it is very closely related to pitch – in many languages it has the most prominent pitch pattern of the prosodic phrase and creates what the listener perceives as a melody (Cruttenden, 1997: 4). There are generally 5 nuclear tones in English – fall, rise, fall-rise, rise-fall, and level (Levis & Wichmann, 2015: 140), nuclear tones in Czech in relation to those of English will be explained in section 3. 3. As previously stated, these tones occur on the nuclear syllable, although they often merely initiate there and spread across the tail until the end of the prosodic phrase, affecting the following unaccented syllables as well (Levis & Wichmann, 2015: 140). This movement then indicates the final realization of the nuclear tone contour. Some melodic movement may also occur in the head as well, though the term “tone” refers solely to the melodic pattern defining the prominence of the nucleus. Each phrase can also have only one nucleus and therefore only one tonal movement, the movement in the head is denoted as melody. The initial position of tone is then related to the speaker’s inherent pitch range. A significant variation of pitch height, in comparison to the speaker’s baseline pitch, often indicates the speaker’s emotional attitude; the emotion of surprise or strong disagreement would be projected in broader pitch movement, as demonstrated in the upper section of *Figure 2*. Those onsets are referred to as high (H) or low (L). As the second set of examples in *Figure 2* illustrates, the uncertainty of the speaker can be reflected in the tonal realization as well: in this instance, the higher the speakers tend to end the prosodic phrase the more they suggest uncertainty or surprise (Cruttenden 1997: 45).



*Figure 2: Possible variation of pitch movement on nuclear syllables relating to the onset in comparison to speakers baseline pitch (Cruttenden, 1997: 45).*

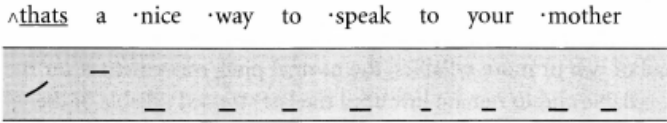
As already mentioned, it is possible to find some melodic elements already in the head, and the tonal movement itself then is not solely bound to the nucleus, but frequently expands onto the tail. The Head functions as the speaker’s introduction of the discourse before delivering the key message of the phrase, the nucleus. The Head itself has very generally merely two pitch possibilities; either there may be a high pitch on the accented syllable at the beginning of the head, which tends to be higher in pitch than the tone of the nucleus, or the first accented syllable

of the head initiates in a low pitch, which is simultaneously also frequently lower than the pitch of the nuclear syllable (Roach, 2009: 138). The following unaccented syllables of the head then tend to “continue the pitch of the stressed syllable that precedes them” until the nucleus (Roach, 2009: 139). In longer heads where there is more than one accented syllable, the melodic movement is much more variable (Roach, 2009: 140). However, there can be observed a general tendency that the melody will likely continue towards the beginning pitch of the nucleus; for instance, a high head followed by a falling nucleus will presumably progressively step downwards to approach the beginning of the nucleus (Roach, 2009: 139), as *Figure 3* illustrates. It should be however stressed that the reality is often much more complex and the distribution of high and low head is based mostly on the perception – more elaborate development is rarely recognized by the majority of speakers (Roach, 2009: 140).



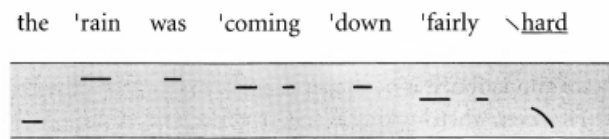
*Figure 3: Tonal development with more than one accented syllable in the head. (Roach, 2009: 139)*

As for the pitch development in the tail, it has been already said that the melody is an extension of the tone initiating on the nuclear syllable. With complex tones, rise-fall, and fall-rise, the pitch movement is “often broken up or distorted by the structure of the syllables they occur on” (Roach, 2009: 136). In other words, tonal development depends on the number of syllables of the tail. If there is no tail, the tone is realized merely on the nucleus. In two syllabic words, the initial tonal movement is carried by the nuclear syllable and the following movement occurs on the second syllable (Roach, 2009: 137). If there are more than two syllables in the tail, the tonal movement will likely end with the last syllable of the tail. In a fall-rising realization, for instance, the nuclear syllable will likely begin the falling movement and the other syllables of the tail will remain low up to the last accented syllable, which would begin the rising movement (Roach, 2009: 137), as demonstrated in *Figure 4*. As for the rise-fall tonal development in complex polysyllabic tails, “the syllable immediately following the [nuclear]



*Figure 4: Presumable tonal movement of fall-rise in polysyllabic tails (Roach, 2009: 137)*

syllable is always higher and any following syllables are low” (Roach, 2009: 137), as *Figure 5* illustrates.



*Figure 5: Tendency of rise-fall tonal development in polysyllabic tails (Roach, 2009: 138)*

## 2. 5. Functions of Intonation

The possibilities of speaker’s modification of intonation can be defined upon multiple spheres, accordingly to the division of Lewis & Wichmann (2015: 148) – attitudinal, pragmatic, or grammatical sphere. Intonation often also helps to indicate the information structure (Levis & Wichmann, 2015: 148). Each of those spheres is employed to adapt the utterance by melodic features (intonation in a narrow sense), or by other prosodic elements, for instance changes in loudness, tempo, or pitch register (Levis & Wichmann, 2015: 148) and thus ideally result in correct comprehension. Those spheres then help to indicate the discourse function of the text; they make speech comprehensible and help to indicate the purpose of the utterance.

Information structure in English is to an extent dependent on both the location of the tonal contour and the realization of the tonal contour (Levis & Wichmann, 2015: 150). In other words, it had already been said that the reason the nuclear syllable is the most prominent is mostly due to its informational relevance (Levis & Wichmann, 2015: 150). In English specifically, parenthetical information or elements informationally significant will likely be marked by extremes of pitch range (Levis & Wichmann, 2015: 148). On the contrary, informationally insignificant elements following or preceding the nuclear syllable will likely be de-accented (Levis & Wichmann, 2015: 148). Intonation therefore helps to differentiate between semantically important and unimportant elements within a prosodic phrase.

Grammatical meaning derives from the distribution of prosodic phrases in a text and “indicates a degree of relatedness between the component parts, whether in terms of grammar, e.g., phrase structure, or mental representations” (Levis & Wichmann, 2015: 150). As for grammar itself, there is a general tendency to place the prosodic boundaries accordingly to the syntactic boundaries of phrases or clauses (Levis & Wichmann, 2015: 151). The use of pitch frequently therefore serves as the speaker’s indicator of junction (Levis & Wichmann, 2015: 148), or semantic relatedness of syntactic elements. Intonation can also help to distinguish sentence types – for example indicate a question or an exclamation.

Attitudinal meaning is related to the function of intonation to convey emotions and attitudes (Levis & Wichmann, 2015: 148). There are no clear one-to-one correspondences between the melodic patterns and emotions – many types of emotions are conveyed by similar prosodic features and their correct determination is thus highly context-dependent (Levis & Wichmann, 2015: 149). As for the perception of specific emotions, anger and sadness show the highest rate of the likelihood of a correct identification by the listener (Levis & Wichmann, 2015: 148). On the contrary, fear, disgust or happiness seem to be more challenging to recognize (Levis & Wichmann, 2015: 148). For the correct interpretation of one's attitude, the listener cannot solely depend on the melody of speech, but has to take into consideration some other features as well (Levis & Wichmann, 2015: 149), such as the context or facial expressions, which may play even more essential role than the words (Vaissiere, 2005: 251).

The pragmatic sphere deals with the meaning of intonation arising from the impressions triggered by use of certain pitch contours (Levis & Wichmann, 2015: 149). Levis and Wichmann (2015: 149) state that “the most pervasive is the meaning ascribed to final pitch contours:” ending the utterance in rising indicates openness, while falling contours suggest a conclusion. This feature also has an impact on politeness: ending with a rise is perceived as more polite, while ending with a falling contour may be regarded as dominant or assertive (Vaissiere, 2005: 253). It cannot be said that those features are conclusively applicable to all languages, however, as for English and Czech this system is fairly similar. More cross-language differences between Czech and English regarding the pragmatic sphere will be later discussed in section 3. 3.

The discourse function of intonation encompasses all the previously mentioned functions. It is closely related to the difference between written and spoken text: it might be challenging to clearly determine the purpose of the text and intonation is one of the vital cues in decoding the intended meaning and structure of speech, similarly as what punctuation does in written discourse. It often lies in the speaker's use of pitch range, tempo, and loudness, which may help the addressee to analyse the structure of the text or the aim of the speaker (Levis & Wichmann, 2015: 151-2). “Pitch range, for example, is exploited to indicate the rhetorical relationships between successive utterances” (Levis & Wichmann, 2015: 151); the pitch variation may signify the beginnings of new paragraphs or topic shifts, which “tend to be indicated by an extra-high pitch on the first accented syllable of the new topic” (Levis & Wichmann, 2015: 151). Parenthetical sequence or a new theme in a speech may be indicated by an expansion or compression of pitch range, as those passages are typically “lower in pitch and slightly faster than the surrounding speech” (Levis & Wichmann, 2015: 151).

### **3. Acquisition of Intonation**

#### **3. 1. Dimensions of Intonation Acquisition**

Intonation is a vital, although often disregarded, element of successful language learning. As all languages differ in their employment of prosodic features, it is frequently considerably challenging for learners of a foreign language (L2) to acquire those features authentically (Mennen, 2015: 171), more so if they extensively differ from their native language (L1). According to Mennen (2015: 171), the cross-language differences can be divided into four dimensions of intonation: the systemic, realizational, semantic, and frequency dimensions. Describing the L2 differences among those dimensions then may play a vital role for learners to be able to achieve some native-like features and thus reduce the degree of the foreign-accentedness, which is often connected with stigmatization.

The systemic (or phonological) dimension describes the phonological inventory of the L2 and the distribution of those structural phonological elements (Mennen, 2015: 171). Specification of this dimension involves “typological similarities or differences in the inventory of structural phonological elements” (Mennen, 2015: 174): those may include the inventory of nuclear tones and how they structurally differ between L1 and L2. This dimension can also involve how the structural features can be possibly combined in the targeted L2, or the semantic associations the melody patterns may imply (Mennen, 2015: 174). Most importantly, this dimension serves as a defining feature for the other three dimensions, which then primarily compares the differences of employment of the systemic features in L1 and L2.

The realizational (or phonetic) dimension is concerned with the phonetic implementation of intonational systemic elements (Mennen, 2015: 171). For instance “how they are scaled (i.e. what their relative height is) or what their shape or slope is (e.g. shallow versus steep rising or falling pitch accents” (Mennen, 2015: 174). This dimension may also involve the timing of those features, their location, or distribution. This includes differences in the proximity between the prenuclear melodic accents and the nucleus (Mennen, 2015: 174), or possible realizations of the nuclear tones.

The semantic dimension deals with the ability of systemic elements to convey intonation function and thus the help to decipher the intended meaning (Mennen, 2015: 171). In other words, it indicates possible variation in the function of intonation; for instance, how the languages differ in marking focus or interrogativity (Mennen, 2015: 175), or different implications the nuclear tones may have. The focus elements can also be accented by different means, such as by a distinct pitch accent in one language, or by word order in another (Mennen,

2015: 175). As for the variation in marking interrogativity, this might be particularly crucial if one of the languages can solely use a specific intonation pattern to indicate a question, thus does not require an inversion or auxiliary verb. Problems may also occur in the information structure, which seems to be especially problematic for learners of L2 English from various L1 backgrounds, who have experienced difficulties in producing the pitch marking new information in contrast to the given information (Mennen, 2015: 177). This may be the result of the strict word order of English, where a new piece of information will be likely marked with a higher pitch rather than a linear structure.

Lastly, the frequency dimension is concerned with the density of employment of these features among languages (Mennen, 2015: 171). Even if identical tonal elements are used, their frequency of use may significantly differ (Mennen, 2015: 175). For example, the frequency of the distribution of certain nuclear tones or variation in the frequency of phrasing may be considerably distinct among certain languages or even language varieties (Mennen, 2015: 175). This dimension may also involve wrong employment of tones generally in comparison to native speakers; for instance, a substitution of tone-types or exceeding use of certain tones (Mennen, 2015: 177).

Outlining the variation of L1 and L2 based on those dimensions “along with some general assumptions and hypotheses on L2 intonation learning, can predict where L2 deviation is likely to occur” (Mennen: 2015, 171). Some major deviations in L2 intonation can possibly even cause a critical misunderstanding, which might be especially problematic in tonal languages, such as Mandarin Chinese or Vietnamese. The degree of incorrect use of intonation may differ; disturbances from L2 features can occur among merely on one dimension, or can involve multiple of them. Mennen (2015: 177) finally mentions that the dimensions often interact, and failure in one of them can result in disturbance of the others; “a deviance in the realizational dimension of intonation may result in a semantic or [semantic] deviation” (Mennen, 2015: 177). Language-learning based on those dimensions can therefore potentially hasten the learning process, increase chances of correct comprehension, and decrease the level of stigmatization of L2 learners based on intonational differences.

### **3. 2. L2 Acquisition of the Intonation Features**

Generally, the pronunciation features are difficult to produce authentically and for a number of speakers the acquisition of accurate intonation patterns in the L2 is an unattainable aim (Mennen, 2015: 171). The chances of achieving native-like features of intonation seem to be getting generally slimmer the later learners start with the language-learning process (Mennen, 2015: 171). Therefore, some factors that highly determine the success of acquiring a

native-like intonation are the age of arrival (AOA) and the age of learning (AOL) (Mennen, 2015: 180). The background of the learning process may also play a role (Mennen, 2015: 180), as the chances of an accurate acquisition may be greater with an L2 native teacher or with visiting the L2 country and using the L2 in native-like environment. Presumably there is a bigger chance of success in acquisition with childhood AOA, although the intonation features among all four dimensions may not be acquired to an equal degree (Mennen, 2015: 180-1); some students may achieve a native-like production of pitch height while fail in accurately distributing prosodic phrases. No age factor may also be observed if the L2 exposure is intense and persisting.

As implied above, deviations in intonation between L1 and L2 also play a significant role: it has been indicated that the acquisition of native-like features is especially difficult when the L1 systemic elements differ from those of L2 (Mennen, 2015: 179). Naturally, learners will first have a tendency to rely on the L1 features in the L2 production of intonation and only after expanding their knowledge of the foreign language, they might start with the perception of the differences in L2 (Mennen, 2015: 181). This fragmentary acquisition often results in “merging effects” (Mennen, 2015: 183), where some L2 features of a certain dimension are assimilated but not to a full extent and some L1 elements of intonation still remain – the final production is then mediating between features of both languages. This effect was for instance supported by research of Flege & Hillenbrand (1984), where the values of voice onset time on /t/ produced by French speakers were inter-mediating between the values typical for French and English. The L2 systemic segments that are similar to those in L1 are generally easier to assimilate, although deviation is likely to occur (Mennen, 2015: 179) as those features in the L2 can differ in their frequency or realization. If the L2 intonational deviations are substantially different, they often tend to create a new mental category in learners’ mind (Mennen, 2015: 179). Interestingly, it has also been observed that students sometimes tend to dissimilate or polarize the L2 features to an extent, where they significantly exceed the realization from the typical norms of production in both L1 and L2 (Mennen, 2015: 183). This can be due to some degree of anxiety from using a foreign language (Volín et al., 2015: 121), although there is a hypothesis this might change with registers, as other research has found much better intonation performance in spontaneous speech rather than in scripted speech (see Taniguchi et al., 2011).

Ideally, the cross-language differences should be described to learners among all the four dimensions; the learners should be generally introduced to different nuclear tones compared to their L1, but their frequency of use, semantic meaning, and their scaling or length should be stressed as well. This systematic education could then increase the chances of clearer

comprehensibility of the meaning, intention and emotional attitude of the learners, as well as reduce the potential stigmatization. This is supported by a recent research by Van Maastricht et al. (2020) where the results indicated that intonation has a vast effect on comprehensibility, and reduces the impression of accentedness. It has also been proven that by incorporating the cross-language intonational differences into the process of language learning, the chances of targeting a native-like production of intonation get higher (see Verdugo, 2006; Taniguchi et al., 2011).

### **3. 3. Intonational Differences between English and Czech**

This chapter will demonstrate the differences in intonation production and perception between Czech and English, where the main focus should be on Standard Southern British English. The variations will be based on the dimensions of cross-language differences (explained in section 3.1.) with an emphasis on problems relating to Czech learners of English.

On the systemic dimension, Czech and English principally differ in the number of nuclear tones. English, as already said in section 2.4., has overall 5 nuclear tones: fall, rise, fall-rise, rise-fall, and level. The Czech system is slightly more complex; it differentiates between the function of the tone and the contour, meaning the realization in connection to the function. There are three nuclear tones (in Czech called: “melodém”): a falling conclusive tone, rising conclusive, and non-conclusive. Each of those functional tones then may have a different realization, as will be presented in the following paragraph. Czech learners of English especially tend to struggle with authentic production of fall-rise, as this tone is missing in the Czech tonal inventory.

Regarding the realizational dimension, what seems problematic for Czech learners of English is the realization of the nuclear tone contours. Primarily, it seems that tone syllables in Czech do not bear any prominence in the acoustic domains, and it is rather the post-stressed syllable that is occasionally realized with higher F<sub>0</sub> (Skarnitzl & Eriksson, 2017: 3223). This is in direct opposition to English, where the nuclear syllable is also the most prominent one. The tonal inventory in Czech is generally much flatter than that of English. Regarding fundamental frequency, the 80-percentile range of F<sub>0</sub> in speech produced by Czech professional speakers has been measured by Volín et al. (2015: 114) to be around 5-6 semi-tones (ST), while for professional English speakers the results showed a variation around 7-8 ST. Interestingly, for Czech speakers of foreign-accented English, the 80-percentile range seems to be even flatter – only around 3 ST (Volín et al., 2015: 118), which would not support the presumption the merging effect as described by Mennen (2015), although only for this feature specifically. Similar results were recorded in a study by Skarnitzl & Hledíková (2022), where the F<sub>0</sub> standard deviation in English was larger than in Czech by 0.78 ST. Due to this difference in production

of intonation, Czech learners of English are often perceived as monotonous or uninterested. As for the contours implied in the previous paragraph, Czech speakers frequently use falling, rising, level or rise-falling contours with both high and low onsets (Skarnitzl et al., 2016: 136-7). In Czech, the realization of rise-falling contour is dependent on the functional nuclear tone; a falling conclusive tone often starts with a minor rise and the falling part is substantial (Skarnitzl et al., 2016: 137). In rising conclusive tone, the rising part is much more significant, while fall often ends in the middle of speaker's intonational span (Skarnitzl et al., 2016: 137). Although both languages use rise-falling tonal movement, in English it usually spreads smaller number of syllables than in Czech.

As for the semantic dimension, the primary difference is the meaning of the complex tones: rise-fall and fall-rise. In English, rise-fall tone most commonly indicates some sense of finality, conclusiveness, or separateness (Cruttenden, 1997: 92). When this tone is used as a response the addressee may perceive it as "challenging" (Cruttenden, 1997: 93), while in some other uses, it might imply the speaker is impressed or sarcastic (Cruttenden, 1997: 93). In Czech, rise-fall contour of the falling conclusive tone seems to be perceived as an unmarked realization (Skarnitzl et al., 2016: 135). Non-conclusive tone may also be realized by rise-falling contour, although here it seems to be associated with a dramatic effect (Skarnitzl et al., 2016: 137). As for fall-rises, in English they may suggest speaker's reservation to the topic (Levis & Wichmann, 2015: 149-150). Semantically, the flatness of Czech could be related to the means of realizing information structure in Czech, where the essential text elements can be stressed by both intonation and inversion of word order. Due to the much stricter word order in English, intonation could presumably play a more important role in marking semantically important elements.

The cross-language differences among the frequency dimension concern mostly the distribution of prosodic phrases. In English, the number of prosodic phrases per utterance is generally higher than in Czech; according to Volín (2019: 145), Czech prosodic phrases made by professional speakers are about 40% longer if counted in syllables per phrase than prosodic phrases produced by English professional speakers (10.78 against 7.56). Skarnitzl & Hledíková (2022) recorded a shorter length of prosodic phrases produced by English speakers by approximately 40% (0.89 syllables). Since the number of prosodic phrases is intertwined with the number of nuclear tones, the lower number of tones per utterance may add to the effect of monotonousness Czech speakers of English seem to have on native speakers.

### **3. 4. Research Questions**

This thesis will focus on the acquisition of English melodic features in advanced Czech students of English and will compare the development of their melodic capabilities in a two-year interval. The subjects were regularly and extensively exposed to English by both native speakers and very advanced Czech speakers of English; therefore some improvements should be observed. The working hypotheses include a presumption that the students will make shorter prosodic phrases in the texts recorded after two years of studying in English. They could also produce a broader 80-percentile of F0 in tones and as well as the pre-nuclear area. A relative growth of the number of complex tones is also expected, especially in fall-rises, which are not typical for Czech but are frequent in English. Presumably, it could also be expected to see decline in level tones and its substitution for another tone.

## **4. Method**

### **4. 1. Speakers**

For this experiment, 15 Czech students (4 males and 11 females) of English and American studies of the Faculty of Arts, Charles University, were recorded reading a text in English at the beginning of their first year of studies and then re-recorded approximately two years later. The official obligation for acceptance to this study program is the level of English of at least B2 of the Common European Framework of Reference for Languages. All the students started studying this program in 2019, and were re-recorded in their third year of studies, the academic year of 2021-22. In 2019, after the first recordings were made, the students were obliged to complete a two-semester course on the phonetics and phonology of English, where they were also educated on the basic features of, among others, the English melodic system mostly based on terminology by Roach (2009). The part of the course focused specifically on intonation was approximately five to six weeks long and the students were educated on the structure of English prosodic phrase, tones, and were encouraged to practice the realizations by given assignments. The text itself was also encountered multiple times during this course, as a self-analysis of the recorded text was a part of the obligatory assessment for completing this course. The students thus worked with the recording for multiple hours in the time period of nine months and got well acquainted with it. During their studies, they also were regularly and intensely exposed to English by both native speakers and advanced Czech speakers of English, and most of their courses were conducted in English.

For comparison, two native speakers of English were additionally recorded; one bilingual speaker of Czech and American English (F) and one native speaker of British English (F). None of those had a strong regional dialect that could affect the final results.

### **4. 2. Text**

The text chosen for the analysis was selected from the Prague Phonetic Corpus titled “Holidays from Hell” (see Appendix). In the first year, the text was chosen for the students to self-analyse their production of English phonemes which are not in the inventory of Czech, such as dental fricatives, or aspirated fortis plosives. Yet, this text proved suitable for exploration of intonation as well, since it included many emotionally charged passages: exclamations such as “it was just awful.” or “it looked horrible!”, or contrastive clauses: “they said it was a five star hotel, but I wouldn’t give it one star.” Additionally, the text was written in first person narration, which could have strengthened the speakers’ tendency to express some passages more dramatically or emotionally. Majority of the text contained commas or full stops,

which frequently indicated clause division. Some phrases used an exclamation marks, or ellipsis. However, in some of the exclamations, especially by the end of the text, only full stops were used: “but the I saw that it was moving.” The text was structured into paragraphs separated by lines, therefore clearly indicating that each paragraph was a conclusive piece of text.

### **4. 3. Recording Process**

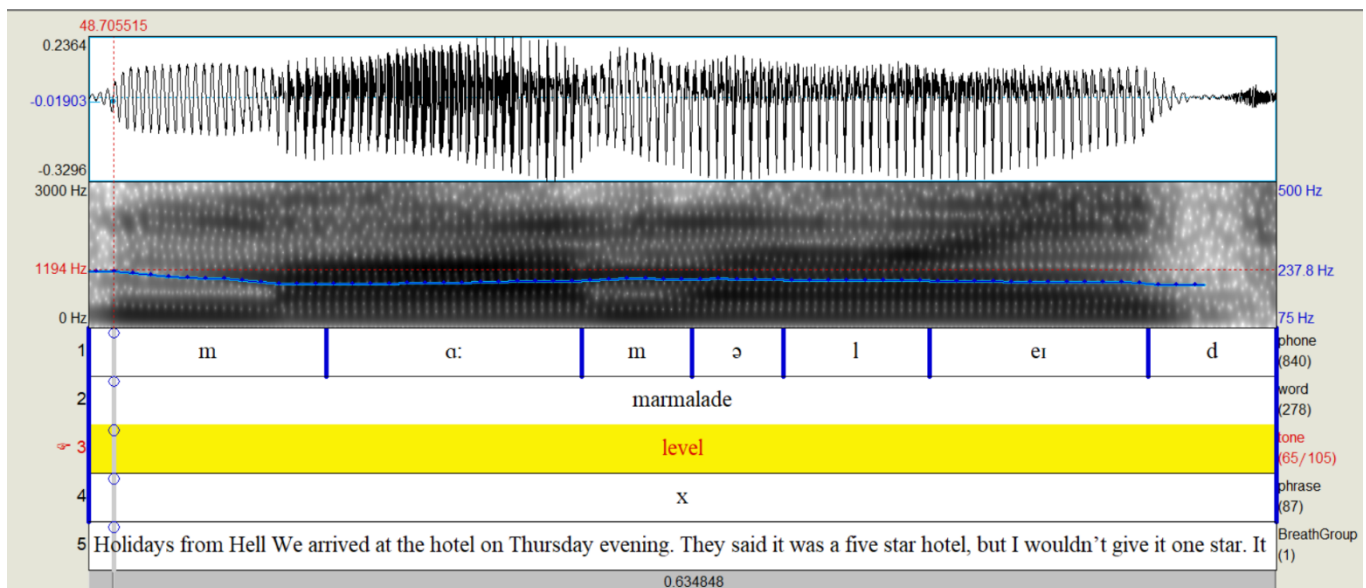
In early October of 2019, all students of English and American studies were recorded in the studio of the Department of Phonetics, Faculty of Arts, Charles University in Prague. They were given a few minutes to read through the text and without any further instructions they were recorded in the Department’s studio room on a condenser microphone AKG C4500 B-BC and sampled at 48 000 Hz. Two years later, the students were contacted with a plead to record a short text, no information specified the re-recording of the text analysed two years ago. All of them were recorded in the same studio with the same conditions and same instructions (only to read through the text). An exception to this was the native British English speaker, who recorded herself on her phone with the identical sampling frequency – 48 000 Hz. In the second recording, all of the students recognized the text. Their questions about the purpose of the re-recording as well as the aims of this thesis were answered once the recording process was finished. Most of the students (12) were recorded almost exactly in the two-year-interval – October 2021. Three students were recorded later, March 2022, mostly due to reasons connected with Covid-19 restrictions and temporary closing of the faculty. However, this fact should not have a significant influence on the results, as the few months are not expected to change the speakers’ intonation patterns significantly.

### **4. 4. Analysis**

The recordings were first edited in Audacity (Audacity Team, 2019) to delete long silent pauses or serious disfluencies. Then a software for an automatic segmentation (Yuan & Liberman, 2008: 3878) was used to separate the text into individual words and phonemes. This was followed with a listening analysis and manual segmentation into prosodic phrases in Praat (Boersma & Weenink, 2019). In each of those phrases the nuclear syllable was found, and the realization of the tone was evaluated. This was also done predominantly by listening, although sporadically the pitch pattern displayed by Praat was used to help. Problematic passages were then consulted with the thesis supervisor. Phrases with disfluencies were ignored and not included in the analysis. Most phrases also required further editing of the boundaries of words or phonemes forming the nuclear syllable and tail, which was done also manually with the help of a spectrogram and waveform. The boundaries included even the dissolutions of formant

structures, as those still carried some melody, therefore they were also incorporated into tones. The title of the text was excluded from the analysis, as many students did not read it.

There were however some problems encountered during the analysis. Some of the students occasionally produced a small melodic variation, yet it was barely recognisable and the differences in frequencies were insignificant. This frequently happened with for example rise-falls, where the rise and fall were objectively present, however, they were too insignificant to be considered an English complex rise-fall tone and it did not fulfil the function of rise-fall – it was used as an unmarked variant and not indicating an emotional involvement. Instead, those tones resembled the Czech realization of post-stress rise of F0, and the rising movement was relatively fragmented rather than a being clear continuous movement. Therefore, those realizations were labelled as level, as is illustrated in *Figure 6*. The differences of frequencies between the beginning of the tone and the lowest falling point were about 100-200 Hz. Despite those differences, perceptually the tonal movement was hardly distinguishable. In other cases, the movement again resembled a rise-fall, but the rising movement was much less prominent



*Figure 6: Marginal rise-falls encountered during analysis that were labelled as level*

then the falling movement, therefore it was labelled as fall. It should also be said that the tones were analysed objectively – this is stressed because in some cases the intention clearly did not reflect the realization. Specifically, occasionally the speakers objectively ended in rise, although it was at the last phrase of a paragraph where falling intonation is much more natural to indicate definiteness or conclusion of a section. Some of those ending passages were realized by level intonation, those were specified in the thesis by term “level-f”, meaning level-falling contour. This term includes those tonal movement, where no significant change of pitch contour was found, thus realization as level, but the specific passage indicated a falling intonation. It is

therefore a level realization of functionally falling tone. Lastly, correlation between melodic acquisition and sex will not be analysed due to the disproportional representation of sexes in this thesis.

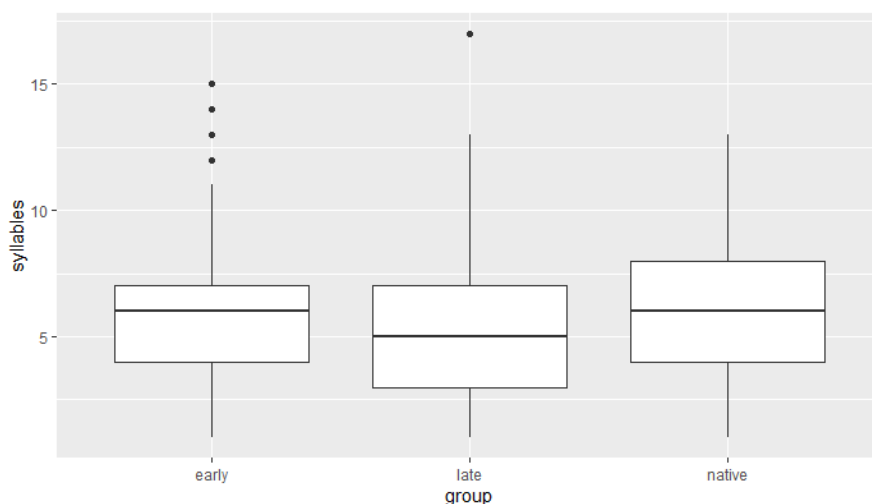
#### **4. 5. Data Extraction**

Once the analysis was finished, the data were extracted with a Praat script into an Excel table. All of the speakers were divided into three groups; “early” group, which consisted of the recordings made in 2019, “late” group, which included recordings made in 2021 and 2022, and lastly, a “native” group that consisted of the two native speakers. All of the speakers were anonymised into AmE native and BrE native for the native speakers, F1-F11 for female speakers and M1-M4 for male speakers, numbers were attached randomly. An R (R Core Team, 2021) script was then made, and the graphs were created with the help of R studio. The limit of significance was chosen to be  $p < 0.05$ . For F0 extraction, an implemented autocorrelation in Praat was used, the lowest frequency was 65 Hz, which corresponds to an analysis window of 46 milliseconds. The measures for analysis included the median values, and 80-percentile range of F0 (a range between 10<sup>th</sup> and 90<sup>th</sup> percentile to eliminate possible extreme values in the minimum and maximum). The length of prosodic phrases was analysed in syllables per phrase and words per phrase. The 80-percentile was examined in STs (semi-tones) with respect to 100Hz. The figures were then generated in R, plots were created with ggplot 2 package (Wickham, 2009).

## 5. Results and Discussion

### 5. 1. Phrase Length

First, the analysis concerned the length of prosodic phrases, where one of the hypotheses was based on a prediction that the late group would make shorter phrases and would match the production of the native group. The overall results for all groups in syllables per phrase are indicated in *Figure 7*. The native group had the median value of 6 syllables per phrase (syll/p). The early group then performed with median of also 6 syll/p, and the median for the late group was 5 syll/p. It seems that there is a certain improvement according to the hypothesis, although the results for general length of phrases among late and early group are not statistically significant; Welch two-sample  $t(1606.1) = 0.52524$ ,  $p > 0,5$ . According to the median values, it seems it was rather the early group that was accurately matching the production of native speakers. It should however be considered that the native group included only two speakers and therefore the conclusions based on this are not strictly definite and more people should be included in the native group to make such claims.



*Figure 7: Length of prosodic phrases in syllables per phrase produced by all speakers based on the groups. The boxplot shows the median value and inter-quartile range (25%-75%) for each group*

As for length of phrases in syll/p in individual speakers portrayed in *Figure 8*, in some cases there seems to be no advancement, such as for instance speakers M01, F10, or F07. In all of the cases however, they seem to be matching the production of the native speakers and no significant deviation is observable between the early and late groups. This could be primarily due to the already high level of English required – as the phrase length values were already

comparatively matching the native production already in the early group, it could be argued that the phrase length was already acquired during previous language learning.

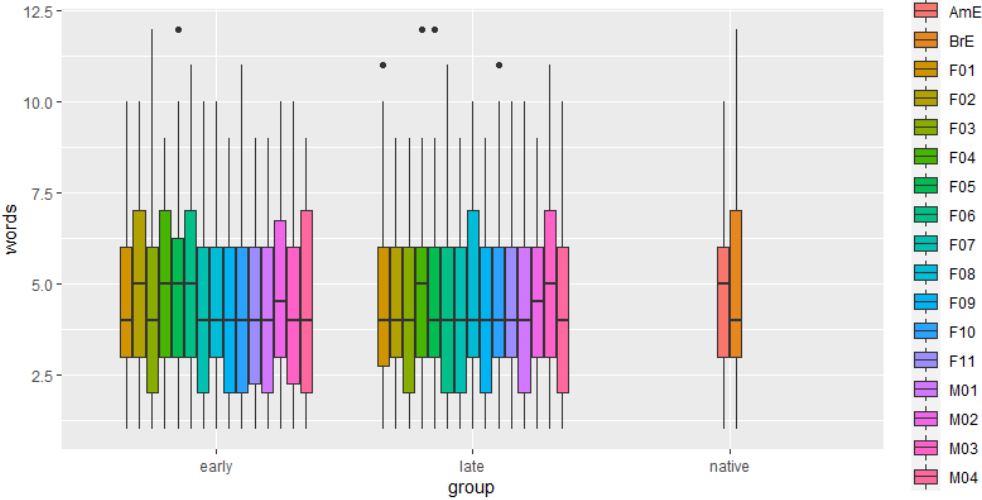


Figure 8: Length of prosodic speakers and comparison with median value and inter-quartile range

The length was also analysed for words per phrase and the overall results are portrayed in Figure 9, where the median value for early and late group was identically 4 words per phrase, while the native group median value was slightly higher – 4.5 words per phrase. Yet again the results are not statistically significant: Welch two-sample  $t(1605,6) = 0.85576, p > 0,3$ . Yet, it is again visible that the production was fairly matching the native speakers. Figure 10 then concerns the same issue for individual speakers of each group. It is apparent that generally there has been a modest decrease in words per phrase as well. Certain improvements are visible especially in the span of speakers M04, or F09, although the median values have stayed identical for majority of the speakers.

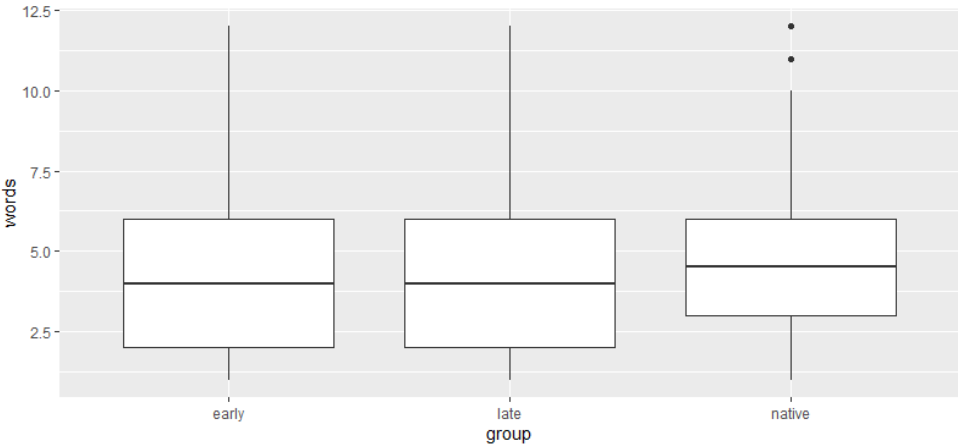


Figure 9: Overall length of prosodic phrases in words per phrase, showing the median and inter-quartile range for each group

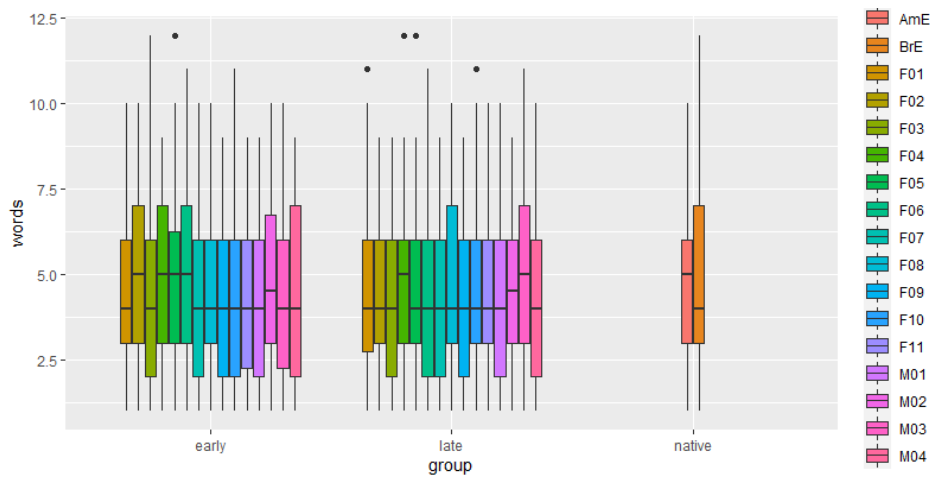


Figure 10: Length of prosodic phrase produced by individual speakers in words per phrase with median and inter-quartile range for each speaker

In both cases, words per phrase and syllables per phrase, it seems that speakers F02, F04, and F06 slightly improved and produced somewhat shorter phrases. On the contrary, speakers F08 or M03 lengthened their production of words as well as syllables per phrase.

## 5. 2. General Distribution of Tones

The following analysis concerned the distribution of English tones among Czech learners in comparison to native speakers. Overall, six tones were analysed: 5 English tones, which were introduced in section 2. 4., and level-f (as has been explained in the previous section, it is a tone realized as level, but functions as a fall). The overall results for all groups are presented in total numbers of tones in *Figure 11* and percentage to observe proportions of each tone is portrayed in *Figure 12*. Chi-square test was done to indicate whether the tone distribution among groups is statistically significant:  $X^2(5, N = 1609) = 24.252, p < 0,0001$ .

The most frequent tone in all analysed groups was fall, and its distribution was highly exceeding production of other tones – overall the numbers were 466 falls in the early group and

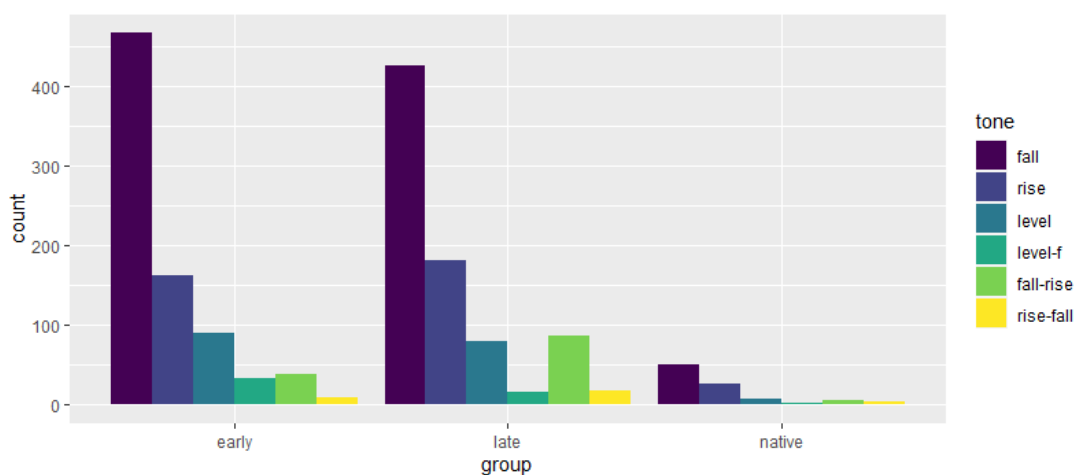


Figure 11: Distribution of tones in total numbers

425 in the late group. This could simply be because fall is presumably the most frequent tone at the end of prosodic phrases, where it indicates conclusiveness. Syntactically, fall was most frequently encountered at the end of sentences, but appeared also at the end of clauses or adverbials. The second most frequent tone in all groups was rise (162 rises in the early group and 182 in the late group), meaning the production of simple tones is much more frequent than complex tones or levels in all analysed groups. Greater variability among groups is observable in the tones level (90 levels for early and 80 for late) and level-f (34 for early and 17 for late group), where the differences between the early and late group seem to be more significant. The early group produced more level and level-f tones, presumably also for narrow melodic variability and therefore the tones were labelled as levels. The acquisition of English melodic features seem to be apparent especially in the case of those two tones; the distribution of level and level-f tones is higher in early group compared to the late group. The late group also fairly matches the production of those tones in comparison to the native group, it could be therefore said the acquisition was achieved in the two-year interval and what was previously realized as level tones gradually changed into growth in other tones, presumably mostly rise, rise-fall and fall-rise. As for the complex tones, some improvement is observable in the production of rise-falls, where the number between early and late group increased (10 in early and 18 in late), however still not matching the production of the native group. This tone also proved to be the most problematic during analysis – as already mentioned in section 5.2., what seemed as a minor rise-fall was labelled as level due to the low variability in pitch. The development in fall-rises was slightly unexpected: the late group also improved significantly in comparison to the

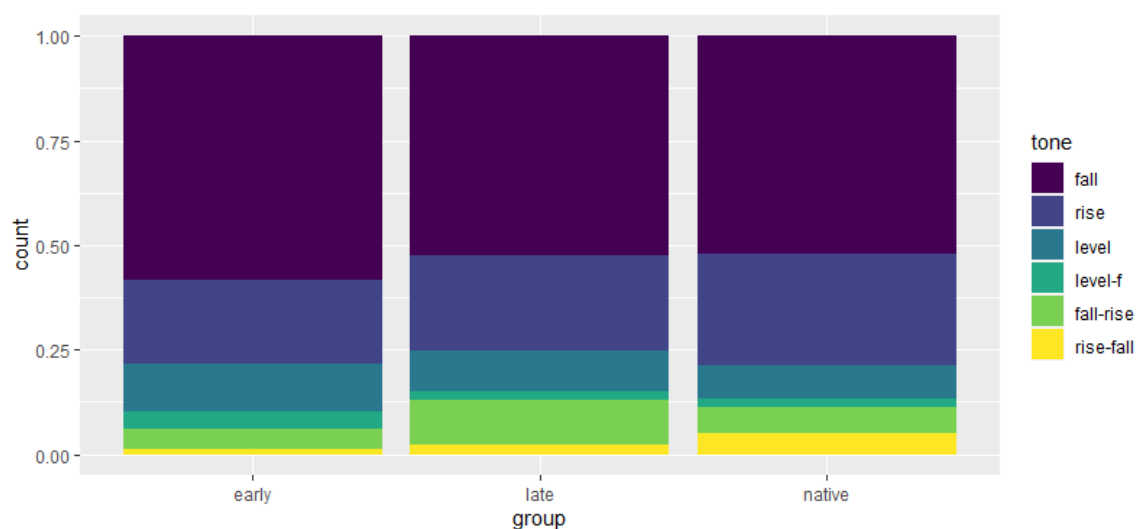


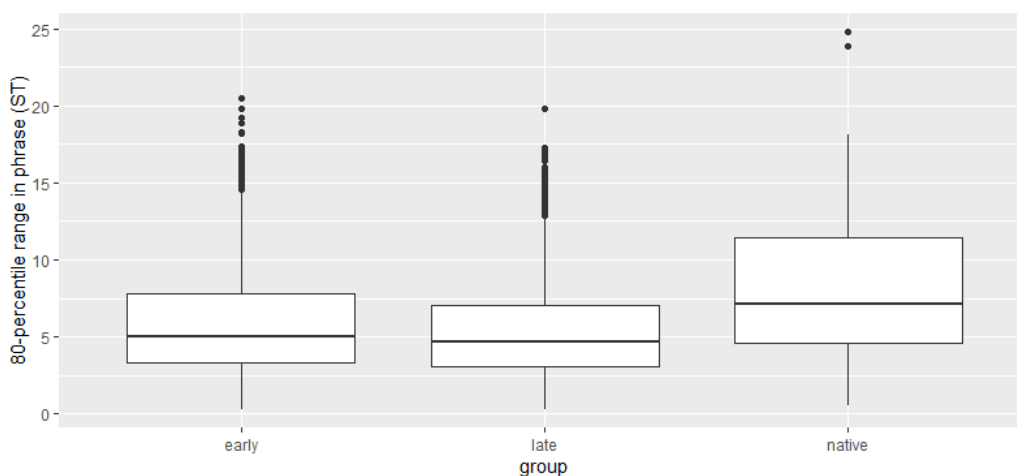
Figure 12: Tone distribution among the groups in percentage

early group (39 in early and 86 in late), although in these cases, they also greatly exceeded the production of the native group.

To summarize, significant variability was observed especially in the increase of complex tones, rise-fall and fall-rise, as well as a decrease of level and level-f tones in the late group.

### 5. 3. 80-percentile Range of Pre-Nuclear Areas and Tone

To see the F0 variability, an 80-percentile range in semi-tones (ST) was chosen to be analysed in tones and also phrases, which could help to see the potential pre-nuclear variability in heads. The overall range for all groups is illustrated in *Figure 13*. The median values between early and late group seemed to be fairly similar; the early group had a median value of 5 ST, the late group had a median value of 4.7 ST. Both groups however differed from the native group, where the 80-percentile range median was 7.15 ST. It is also interesting that the decrease between early and late group was statistically significant, Welch two-sample  $t(1562.3) = 3.6317, p < 0.001$ . The 80-percentile range in the late group was also thinner, which indicates the speakers did not vary the melody as much as during the first reading. This is in contrast with the presumption that the students would perform with greater F0 variability instead of maintaining the “flat” melody of Czech. This outcome is also supported by the results in *Figure 14*, showing the melodic variability of individual speakers. From this figure, it is clearly observable that only a minority of the speakers actually matched the production of the natives. Many speakers then demonstrated significant decrease in performance, especially speakers F01, F02, F06, and M04. Moreover, there seems to be no significant increase, only modest improvements are detectable in F07, and M03. The speakers resembling the native range production in the early group, such as F01, F06, or F10, drastically decreased their performance



*Figure 13: Overall 80-percentile range in phrases - including also the pre-nuclear area, with median value and 80-percentile range*

in the late group. Merely the speaker F10, who already matched the melodic movement of the natives, has maintained the range span, although here, it was the median value that considerably decreased.

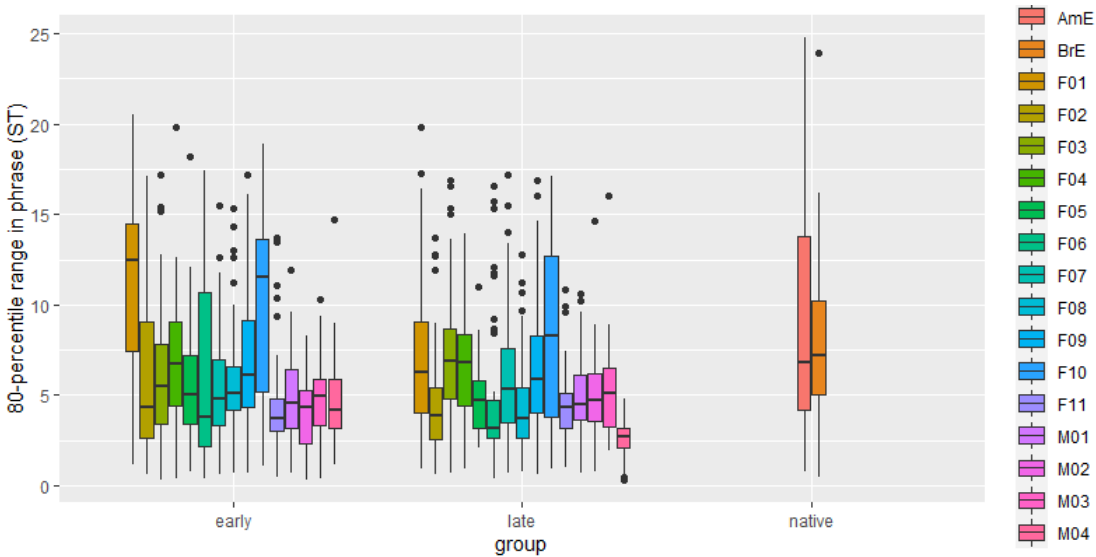


Figure 14: Median and 80-percentile range in phrases divided by speakers

The 80-percentile range of F0 was also analysed specifically for tones, where the expectation was identical – speakers in the late group would expand the range of STs compared to the early group and have closer results to the native group. The results in Figure 15, however, indicate no such development, and the late group rather again slightly decreased in their performance compared to both the native and early groups. The median value for the late group was 3.9 ST, while for the early group it was 4.4 ST and the native group performed with median 4.7 ST. The difference between early and late data in regards to 80-percentile range in tones was also statistically significant; Welch two-sample  $t(1580.1) = 2.8881, p < 0.01$ . Again, the performance of the early speakers already fairly matched the native production, yet after two

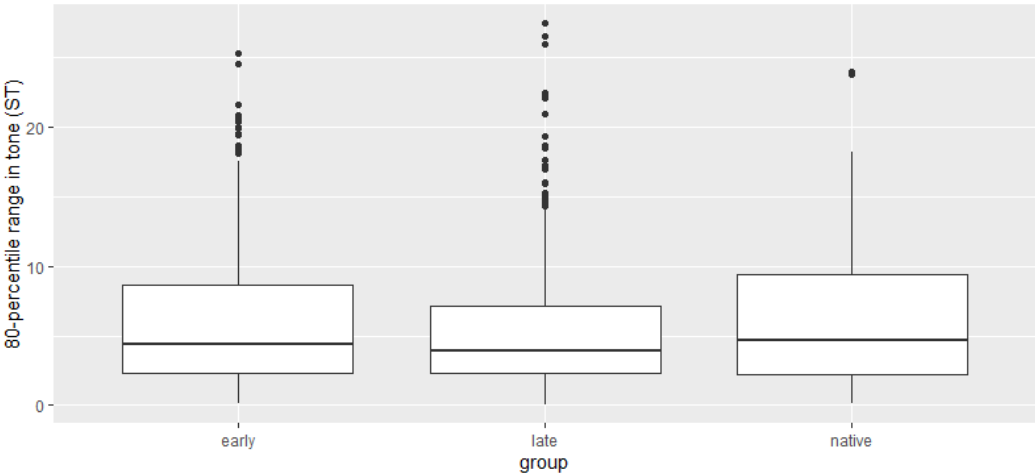
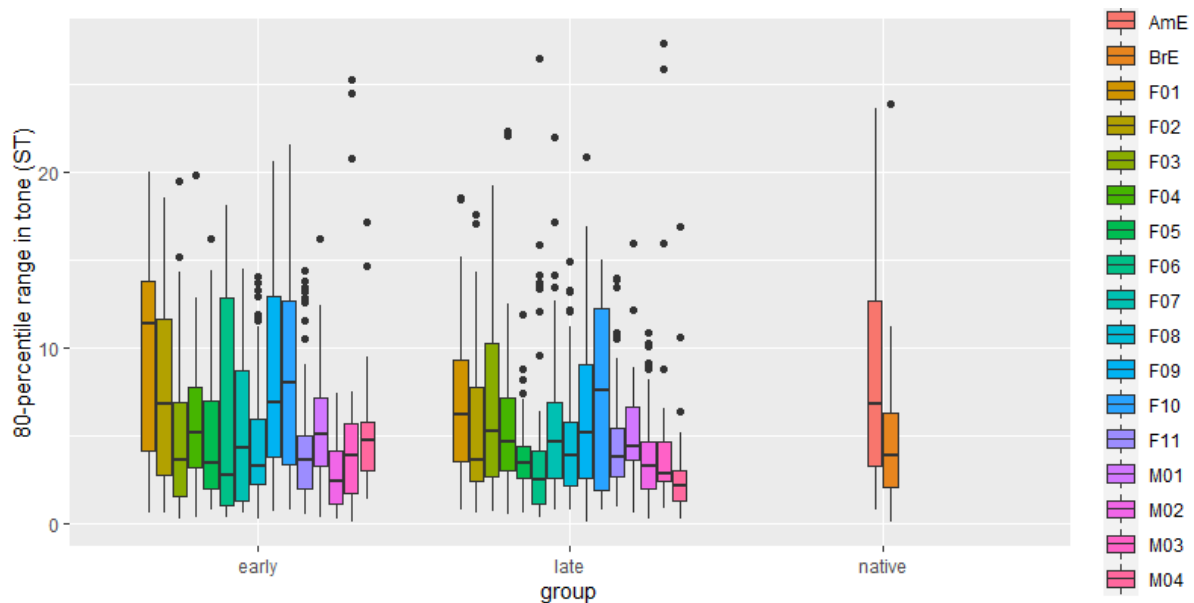


Figure 15: Overall median values and 80-percentile range in tones for each of the groups

years there was a decrease in the performance in the overall median value. Additionally, the whole 80-percentile range again narrowed, meaning the speakers used a smaller span of STs.

As for the results of individual speakers, which are illustrated in *Figure 16*, the decrease in performance is even more obvious. In the majority of the late group, their span of STs used significantly narrowed. A drastic decrease is evident in speaker F06, where the median value has stayed similar, but the range of STs is minimal in comparison to the early performance. Another substantial decline is observable in speakers F01, F05, or M04, where specifically in this case the later production does not even barely match the early range. The only improvement can be seen in speaker F11, or a very minor one in M02. Speaker F10 then performed with fairly similar values in both of the recordings. The results in native group seem to be unusual as well; the range of AmE speaker is significantly wider than that of BrE speaker. More than half of the early group achieved a wider tonal range than the BrE speaker, although the same cannot be said for the late group, where the results show the contrary. Those speakers relatively matching the AmE native speaker's production have all, for one exception of speaker F10, decreased in their performance and rather reflected more the BrE native speaker's production.



*Figure 16: Median and 80-percentile range in the tonal area divided by individual speakers*

Another issue that was analysed was the 80-percentile range of F0 for each of the tones in phrases, which is illustrated in *Figure 17*. Since this analysis portrays the melodic movement in the entire phrase, it also therefore takes into consideration the movement in the pre-nuclear area. In some tones, there are observable different results for native and non-native speakers;

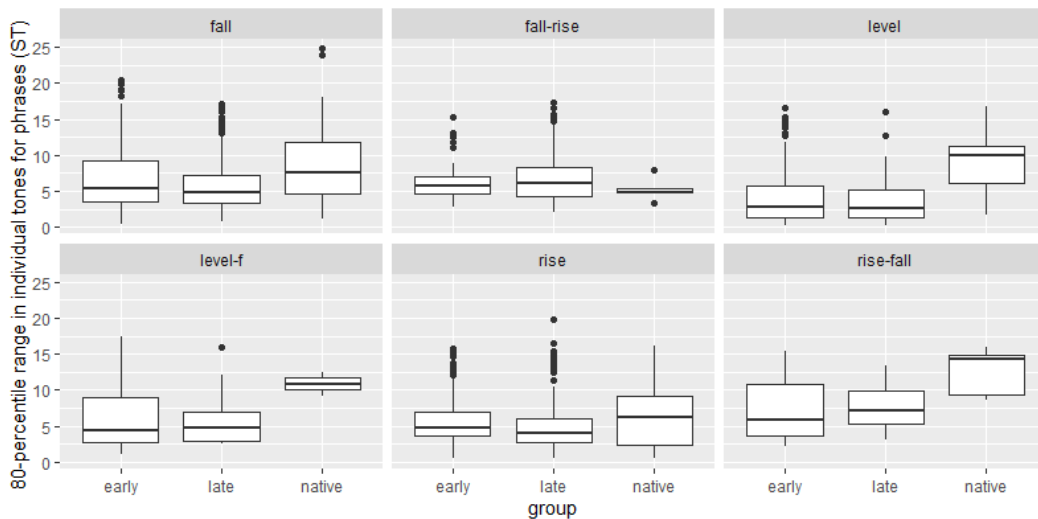


Figure 17: Median and 80-percentile range in phrase for each tone

especially in fall, level, and rise, where the span of native speakers is broader than for both early and late groups. This presumably means the native group has performed with more various melodic movement in the pre-nuclear area as compared to the learners. On the contrary, in fall-rise and level-f, the range of native speakers spread around smaller number of STs compared to both of the groups of students. This could however again be due to the low number of subjects in the native group and therefore the distribution of tones is not reflecting the reality of native speakers generally, mostly because for example level-f or the complex tones were rarer in the recordings than rises or falls. As for the early and late group, the development in rise and level shows almost no progress in acquisition; in level, the span in the early group again got narrower compared to the late recordings, and in rise the number of semitones has again decreased for the late group. For analysing the results, a two-way ANOVA test was used to compare tone distribution among groups,  $F(5, 1689) = 15.544, p < 0.0001$ . Tukey HSD test showed the differences of tones between early and late group was statistically significant ( $p < 0.001, 95\% \text{ C.I.} = [-1.172, -0.286]$ ). As for individual tones, the span of fall, level-f and rise-fall tones got narrower, although in the last two tones the median value was marginally higher for the late group. For the fall tones, Tukey HSD test showed statistically significant results between late and early group ( $p < 0.01, 95\% \text{ C.I.} = [-1.918, -0.144]$ ), native and early fall tone ( $p < 0.05, 95\% \text{ C.I.} = [0.216, 4.118]$ ), and also native and late group fall ( $p < 0.0001, 95\% \text{ C.I.} = [1.238, 5.158]$ ). Tukey HSD test also indicated a significant difference in level tones produced by the native and late group ( $p < 0.05, 95\% \text{ C.I.} = [0.521, 10.330]$ ). The only exception to the decreasing trend is in the case of fall-rises, where the production both exceeded the native production and there is also an apparent improvement between early and late groups. The

narrow movement in the native group could be the result of low representation of fall-rises in the native material.

The 80-percentile range specifically in nuclear and tail area is illustrated in *Figure 18* for each of the tones analysed. Two-way ANOVA was again performed, indicating a significant difference in the tone distribution ( $F(5, 1689) = 27.107, p < 0.0001$ ), and tones among groups ( $F(10, 1689) = 2.351, p < 0.01$ ). Again, there is a decrease in performance in the late group compared to the early recordings, this time in fall and rise-fall. In fall, the early speakers again matched the native production and although the late speakers decreased in the span, the median value has remained on approximately the same level. It has already been pointed out there has been an increase in the distribution of fall-rises in the late group in comparison to both the native and early recordings. This advancement also reflected in the range in tone, where the data of the late speakers slightly surpass the early and native values. As for level and level-f tones, the 80-percentile range in the native level and level-f in early group seem to have a greater span than would be expected for level tones. This could be due to voice creaks, where the analysis of Praat is often slightly problematic. The only statistically significant result in Tukey HSD test was between level tones of native and late group ( $p < 0.05, 95\% \text{ C.I.} = [0.269, 11.663]$ ). Despite witnessing different results among groups in realization of rise in phrases, as for tones, the groups performed very similarly. Rise-falls in the late group seem to moderately match the native values as compared to the early group, although the native group performed with marginally higher median ST values. Again, a large part of the early group data already somewhat matched the natives, and rather than improvement, for many tones there is again a marginal decrease in performance.

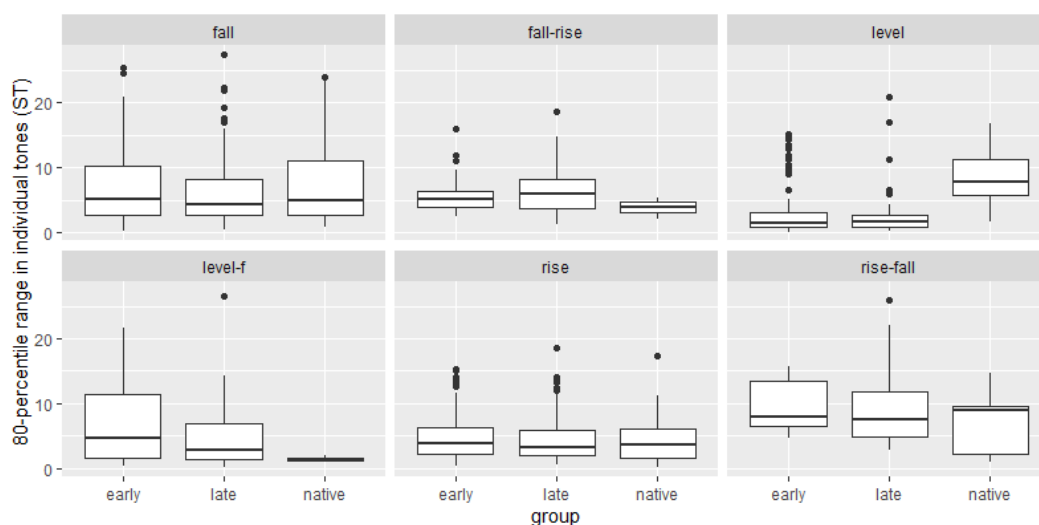


Figure 18: Median and 80-percentile range in tones for each of the nuclear tones

### 5. 4. 80-percentile Range of F0 in Nuclear Area for Individual Speakers

The 80-percentile range of F0 was also analysed for individual speakers in early and late groups specifically in tone to see the tonal development of every speaker over the course of two years. *Figure 19* demonstrates such development for falling tonal movement. Again a general decrease in performance is observable, especially in speakers F01, F02, F06 the decline in ST median seems truly apparent. Many speakers have maintained somewhat similar values, and only in speakers M02, F03 or F07, there seems to be some marginal improvement, although not as evident as the decrease values. The general tendency seems rather to use a narrower span of STs in the nuclear part then in recordings done two years prior.

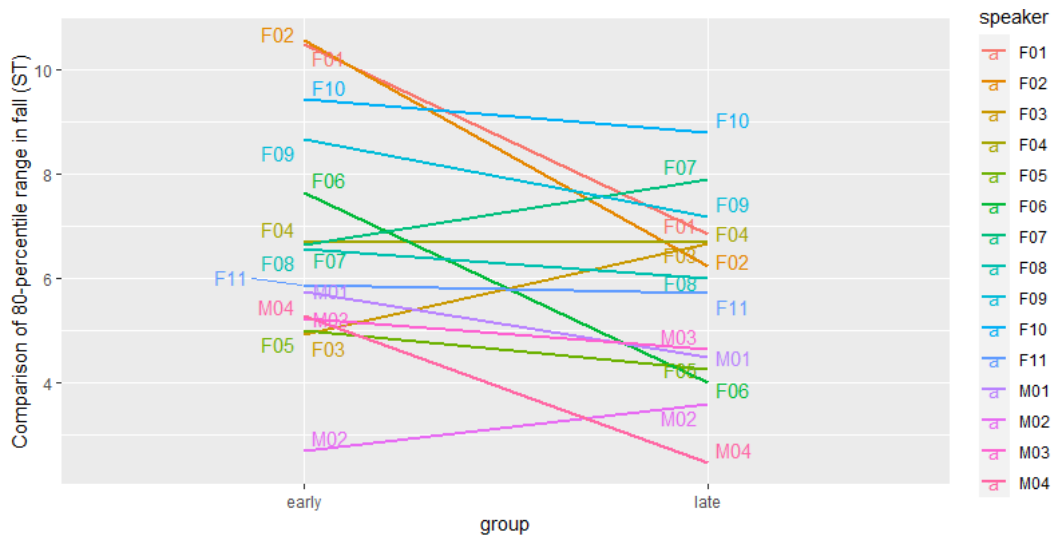


Figure 19: Comparison of median values of individual speakers in the early and late group regarding the 80-percentile range of tonal movement in fall

Similar development is observable in another simple tone – rise, as illustrated in *Figure 20*. Here some increase is marginally present in some cases, such as the majority of male speakers, or most apparently in F03. The decrease in performance however seems to be more

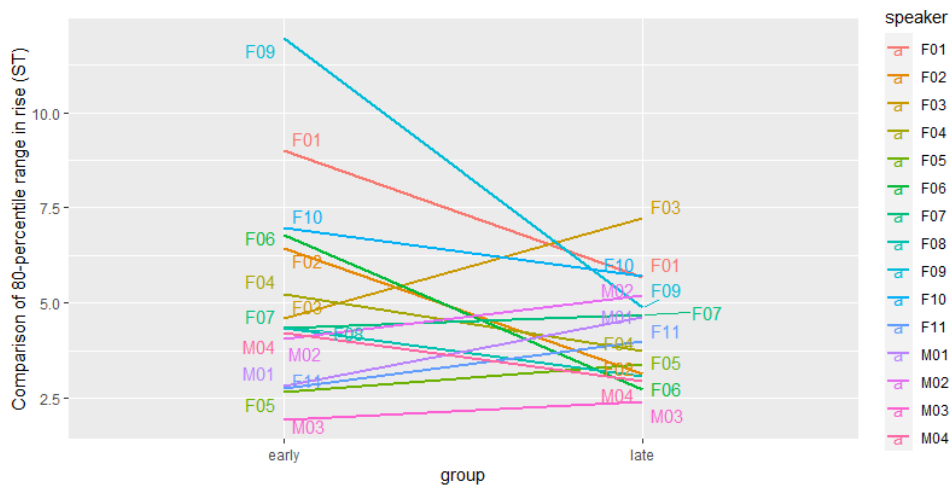


Figure 20: Comparison of median values of individual speakers in the early and late group regarding the 80-percentile range of tonal movement in rise

dramatic, especially in speakers F01, or F09, where the span reduced from values of approximately 12 STs into values around 5 STs. There has been more speakers narrowing their 80-percentile range rather than expanding it.

As for the complex tones, the overall distribution of fall-rises has substantially increased, as has been shown in section 5. 2. The presumption would therefore also indicate an improvement in performance in regards to the 80-percentile range. The results are illustrated in *Figure 21*, where again some decline in performance is apparent, especially in speakers F09, or M04. Interestingly, for the rest of the speakers it seems that either their range was narrowed only marginally, or there was an improvement. Particularly notable progress is noticeable in speaker F02, where the range expanded almost three times in comparison to the 2019 performance. The development of speaker F03 also indicates a considerable improvement. Generally, majority of speakers performed very similarly in the both of the recordings, however, there are some considerable declines, yet even more noticeable improvements.

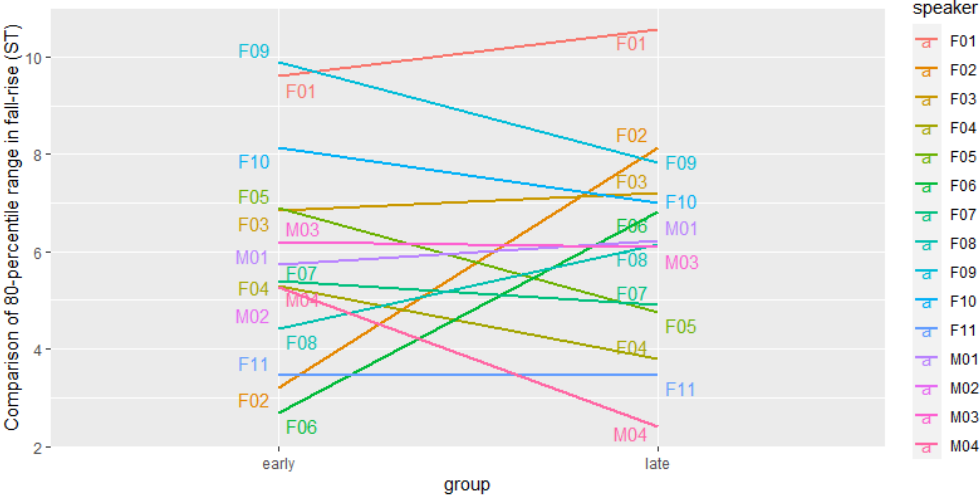


Figure 21: Comparison of median values of individual speakers in the early and late group regarding the 80-percentile range of tonal movement in fall-rise

And lastly, the situation for rise-fall is fairly similar, as can be seen in *Figure 22*. Again, many speakers performed similarly in both recordings. A more significant decrease in STs is visible in speakers F09, and a more marginable one in speaker F05. Most speakers, however, remained stagnant with the exception of speaker M03, where the range improved by five times.

It should also be noted that in the case of this tone, only 7 speakers used rise-fall in both the early and late recordings – thus the limited comparison.

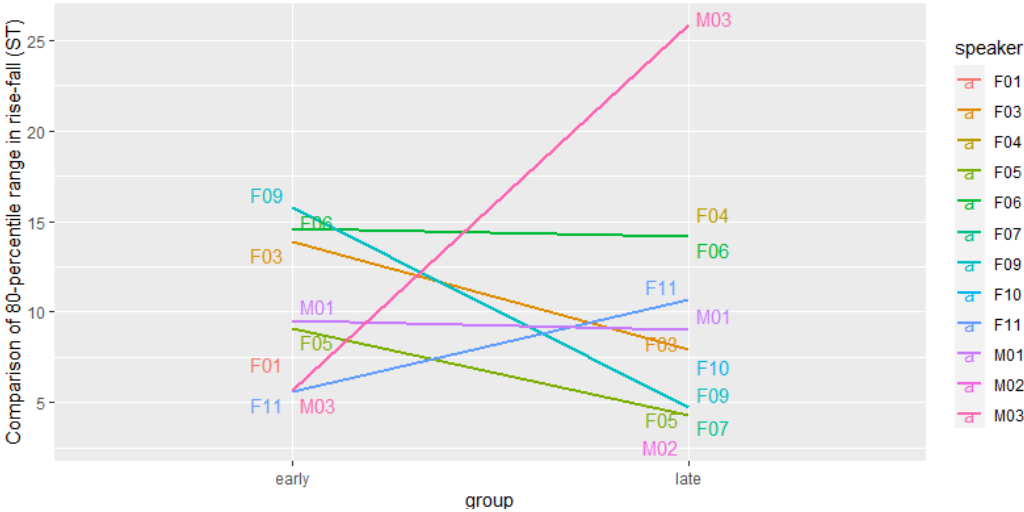


Figure 22: Comparison of median values of individual speakers in the early and late group regarding the 80-percentile range of tonal movement in rise-fall

## **6. General Discussion**

It was mostly presumed that the speakers will acquire the prosodic features typical for English after comprehensively studying the language from various aspects for the period of two years. However, the results do not indicate such a considerable development as was expected. The hypotheses, introduced in section 3. 4., included an assumption that the phrases will be shorter, which has not been proven. Generally, in syllables per phrase the results have shown a slight improvement, but not statistically significant. In words per phrase, the results between the early and late group were to a large extent identical. However, despite no statistically significant improvements, it is apparent that the students already fairly matched the production of the native speakers already back in 2019. The similarity of the results among groups could have also been caused by the nature of the text. First of all, the text was supposed to be read aloud, therefore what presumably highly influenced the results was punctuation. Also, the form of the text was not demanding, and it did not allow much creativity when it comes to phrasing. Only some parts of the text were realized with a greater variability among speakers, usually fluctuating between a division into one or two phrases. Such parts included; “there were big ships travelling past,” “and the sea was all polluted and brown,” or “it was a sort of greeny-black colour.” In other parts of the text, the realization was fairly similar among all subjects. The length of phrases therefore may have been to an extent the result of the choice of the text.

Another presumed development was a relative growth in the number of complex tones together with a decrease in level tones. This was one of the most apparent improvements of the results, where the distribution among groups was statistically significant. Both hypotheses were supported by the performance of the late group, where the number of rise-falls and fall-rises was higher than in the early group, while the number of level tones decreased. In comparison to the native group, it turned out that the production of fall-rises in the late group exceeded the percentage in the native group, while for rise-falls, the number was lower. The choice of tone realization in fall-rise seems interesting; as has been said, it is the only tone that is in English while not in Czech, even at the level of a contour of another underlying functional melodic pattern. It could be interpreted as a success in teaching the students about English intonation. Another conclusion may also be that the complex tones may be easier to acquire than their tonal range, and that acquisition of unfamiliar elements may be easier rather than using a known pattern in a different manner.

The most disappointing results were observed in the 80-percentile range of phrases and tones, where the results indicated a significant decrease in performance among the early and

late group. The results for individual speakers have shown predominantly a decline in 80-percentile range of STs, therefore not supporting the hypotheses predicting a broader melodic range compared to earlier recordings. Some speakers performed worse in multiple categories; for instance, speaker F06's range decreased in both of the simple tones. On the contrary, this speaker also significantly increased in the range of fall-rise. Speaker F09 also got worse in rise, and additionally fall-rise. Some students, however, improved their performance; especially speaker F03, who increased the melodic range in both simple tones. Interestingly, speaker M03 was the only speaker, where the results have indicated a significant expansion of the range of rise-fall, although remained fairly stagnant in other three tones: fall-rise, rise, and fall. It should also be noted that the most relevant data are for the simple tones because they were most frequently used, as was presented in section 5. 2.

Yet, similarly as in the case of length, the speakers again fairly matched the production of the natives already in the early group, so it could be presumed that for some speakers the range of melody was to some extent already acquired. The reason for such a significant decline in the performance of some speakers could have been again the nature of the text. As has been said in section 4. 3., all of the students recognized the text during the second recording. It should however be stressed that the reactions were not exactly pleasant, most students seemed annoyed or slightly distressed. It is possible that if the text had not been a part of the course assessment in the first year, the reactions could have been different and there may have been more excitement projecting in intonation. It is also likely that the students focused more on articulation, which was the element evaluated in their self-analysis of the text. Some speakers indeed occasionally stressed words containing dental fricatives, some of those even being determiners. On the contrary, during the analysis it was very clear that many of the speakers improved in their pronunciation, usually exactly in the elements evaluated during the first year – dental fricatives and velar nasals, in both of which in many cases there were apparent explosions of [d] or [k] on the spectrogram in the early recordings but not in the late ones. The decrease therefore may be a result of the focus on articulation rather than reading the text naturally, which could possibly be prevented in the future by giving some specific instructions.

Based on the results, it could be said that the acquisition of L2 features of English for L1 speakers of Czech include mostly the production complex tones, and therefore according to Mennen (2015), concerning the systemic dimension (those dimensions were introduced in section 3. 1.). This was visible from the influx of the number of fall-rises – tones which are not in the inventory of Czech. Regarding the realizational dimension, the decline of 80-percentile range of F0 indicates a possible failure in acquiring the L2 features of tones, yet it could be

argued this was caused by the text. The semantic dimension was not analysed due to the nature of the text, which did not allow much semantic variability. There have been possibly improvements on the frequency dimensions, as the students tended to make slightly shorter prosodic phrases, and they also used more complex tones. The differences among individual speakers in all the analysed categories are most probably a result of slightly different entrance level of English, or individual tendencies and experience (such as living in an English speaking country for an extended period of time). However, some other elements may also include the studying process itself, or possible consequences of the Covid-19 pandemic. Although the students may have come in contact with native speakers as teachers, the native speaking environment was impossible to visit due to the Covid-19 restrictions. Yet arguably the influence here on the decrease on performance may be debatable, since it is very probable that majority of the students come regularly in contact with English online media. Other elements that may have influenced the process of acquisition could have been connected to the process of studying. The material included both double and single-subject students, where the assessments for each were different. Single subject students generally came more into contact with written language as they often had an additional assessment in the form of an essay. Some speakers may have also attend more classes by native speakers, although it is arguable whether this could be influential since the level of English of all the teachers was very advanced.

As has been implied, many of the results were presumably influenced by the choice of the text. The future research should therefore include a text that has not been previously encountered apart from the recording process, or a text providing more variability in phrasing. However, the material itself may be part of future research, as there are many more elements that can be observed regarding the process of acquisition generally. As has been said, the improvements in articulation and pronunciation were clearly observable, although not analysed.

## **7. Conclusion**

This thesis aimed to portray the development of intonation features in advanced Czech students of English in an interval of two years, and therefore add to the research field on intonation acquisition and language learning. The development was evaluated in the length of prosodic phrases, the distribution of tones, and the 80-percentile range of F0. The comparison consisted of overall three groups: recordings done in 2019, recordings of the same speakers approximately two years later in 2021-22, and a comparison group of native speakers. The hypotheses estimated a wider 80-percentile range as a primary development, but also an influx of the relative number of complex tones, and shorter prosodic phrases.

The results however indicated more of a decline in performance rather than an improvement. The length of phrases in syllables per phrased slightly decreased, although in words per phrase, the results among the early and late group were fairly identical. In regards to the distribution of tones, the results indicated growth in the complex tones and a decrease of level tones, which would indeed indicate a positive development, fairly matching the production of the natives, although not reaching their percentage in the number of rise-falls, while exceeding the number of fall-rises. In the 80-percentile range, the results were opposite to what was expected – predominantly, the speakers were either stagnant, or narrowing their range. Improvements were also observed, although not as often as was anticipated. These results however may have been rather due to the choice of text – since the speakers already analysed the text in their first year of studies, they knew the text somehow better, presumably did not have a positive associations with the text, and also knew what the aim of the previous analysis was and therefore focused more on the segmental pronunciation rather than reading the text naturally. Yet, it should also be mentioned that the values the students had already in the first year already somewhat matched the natives. The conclusion therefore could be that some aspects of intonation may be naturally acquired in previous stages of language learning rather than at university and that the process of intonation acquisition then possibly decelerates in more advanced stages. The mechanisms of acquiring L2 intonation range may thus be more unconscious and connected with the exposure to the language in earlier stages of language learning.

## **8. References**

- Audacity Team (2019): Audacity (Version 2.3.1) [Computer program]. Retrieved April 18, 2019, from <http://audacityteam.org/>.
- Boersma, P., Weenink, D. (2019). *Praat: Doing Phonetics by Computer* (version 6.1.50). Retrieved from [www.praat.org](http://www.praat.org).
- Chamonikolasová, J. (2017). The role of intonation in functional sentence perspective. *Intonation in English and Czech dialogues*, 24-37.
- Cruttenden, A. (1997). *Intonation* (2nd ed., Cambridge Textbooks in Linguistics). Cambridge: Cambridge University Press. doi:10.1017/CBO9781139166973
- Firbas, J. (1964). From comparative word-order studies: (thoughts on V. Mathesius' conception of the word-order system in English compared with that in Czech).
- Flege, J. E., & Hillenbrand, J. (1984). Limits on phonetic accuracy in foreign language speech production. *The Journal of the Acoustical Society of America*, 76(3), 708-721.
- Levis, John. (2015). English intonation: Form and meaning. 10.1002/9781118346952.ch8.
- Mennen, I. (2015). Beyond segments: Towards a L2 intonation learning theory. In *Prosody and language in contact*, 171-188. Springer, Berlin, Heidelberg.
- R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>.
- Ramírez Verdugo, D. (2006). A study of intonation awareness and learning in non-native speakers of English. *Language awareness*, 15(3), 141-159.
- Roach, P. (2009). *English phonetics and phonology paperback with audio CDs (2): A practical course*. Cambridge university press.
- Skarnitzl, R. & Eriksson, A. (2017). The acoustics of word stress in Czech as a function of speaking style. In: *Proceedings of Interspeech 2017*, 3221–3225.
- Skarnitzl, R. & Volín, J. (2019). The effect of durational cues on the reassignment of a syllable in the metrical structure of Czech sentences. In: *Proceedings of the 19th ICPHS*, 2891–2895.
- Skarnitzl, R., & Hledíková, H. (2022). Prosodic phrasing of good speakers in English and Czech. *Frontiers in Psychology*, 13, 857647. [pdf https://doi.org/10.3389/fpsyg.2022.857647](https://doi.org/10.3389/fpsyg.2022.857647)
- Skarnitzl, R., Šturm, P. & Volín, J. (2016). *Zvuková báze řečové komunikace: Fonetický a fonologický popis řeči*. Praha: Karolinum.

- Taniguchi, M., Setter, J., Fulop, S. A., & Golston, C. (2011). Assessing Intonation in the Spontaneous and Scripted Speech of Native and Non-native Speakers of English. In *ICPhS*, 1966-1969.
- Vaissiere, J. (2005). Perception of intonation. *The Handbook of Speech Perception*, 236-263.
- Van Maastricht, L., Zee, T., Krahmer, E., & Swerts, M. (2021). The interplay of prosodic cues in the L2: How intonation, rhythm, and speech rate in speech by Spanish learners of Dutch contribute to L1 Dutch perceptions of accentedness and comprehensibility. *Speech Communication*, 133, 81-90.
- Volín, J. (2019). The size of prosodic phrases in native and foreign-accented read-out monologues. *Acta Universitatis Carolinae Philologica*, (2), 145-158.
- 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.
- Wells, J. C. (2008). *Longman pronunciation dictionary* (3rd edition). Longman.
- Wickham, H. (2009): *ggplot2: Elegant Graphics for Data Analysis*. New York: Springer.
- Yuan, J. & Liberman, M. (2008). Speaker identification on the SCOTUS corpus. *The Journal of the Acoustical Society of America* 123, 3878.

## 9. Resumé

Tato práce se zaměřuje na proces osvojování melodických aspektů cizího jazyka u pokročilých studentů. Konkrétní zaměření bylo na české studenty vysokoškolské angličtiny, kteří byli porovnáváni mezi sebou s dvouletým rozdílem, ale i se dvěma rodilými mluvčími angličtiny. Jelikož české intonační rozpětí se z výzkumů zdá užší oproti angličtině, česká intonace v angličtině by potenciálně mohla způsobovat stigmatizaci. Mezi studenty se porovnávaly faktory jako melodické rozpětí, distribuce tónů nebo délka prozodických frází.

Teoretické pozadí práce je rozdělené do dvou kapitol. V první (celkově kapitola 2.) jsou popsány základní aspekty prozodie především se zaměřením na intonaci, je vysvětlena terminologie používaná v této práci i nastíněné celkové problémy ohledně terminologie v anglické intonologii. Prozodické faktory jsou prvně vysvětleny na slovní úrovni, tedy především jak přispívají k prominenci některých slabik. Prominence ale může být potlačena na úrovni frází, kde slabiky můžou ztrácet svůj důraz ve prospěch důležitějších elementů fráze nebo klauze. Prominence je vysvětlena především s důrazem na melodii, jelikož ta je v angličtině nejdůležitější faktor pro určení přízvuku. Jsou ale zmíněné i dva další koreláty prominence – hlasitost a délka. Dále následuje struktura prozodické fráze, jak se popisuje v anglické terminologii, a popis nukleárních tónů, znovu se zaměřením především na angličtinu. Anglická terminologie dělí prozodickou frázi na 4 části: prehead, head, nukleus a tail. Prehead a head často zahrnují tématickou část výpovědi. Nejprominentnější element prozodické fráze je pak nukleus, který se nachází na přízvukné slabice často informačně důležitého slova. Na této slabice pak začíná tónová kontura, která se pak rozpíná po zbytek prozodické fráze, tedy přes tail. Poslední podkapitola této části (2.5.) popisuje různé funkce intonace v řeči: funkce informační, kdy intonace může posluchač pomoci rozlišit mezi důležitými částmi prozodické fráze a tematickým pozadím, dále pak gramatická, díky které je možné vnímat nejen konce a začátky vět, ale i jednotlivé klauze či fráze v rámci souvětí. Následuje pak funkce afektivní, která umožňuje vyjadřovat postoje, a dále pragmatická, která popisuje, jak melodické kontury mohou přispívat k významu: mluvčí může pomocí intonace naznačovat záměr pokračovat v promluvě, nebo naznačit posluchači, že vyžaduje odpověď. Poslední funkce je diskurzivní, která zahrnuje elementy předchozích elementů funkce a díky které je možné snadněji vnímat a předvídat členění promluvy. Další kapitola teoretické části se zaměřuje na výzkum a popis procesu osvojování intonace. Představuje čtyři dimenze podle Mennen (2015), na kterých je popsán proces osvojování melodických prvků cizího jazyka u studentů. První z nich je systematická, která zahrnuje popis struktury melodického členění

jazyka nebo informace o tom, jaké nukleární tóny jazyk používá. Následuje realizační dimenze, která zahrnuje například realizaci nukleárních tónů – kolik slabik mohou zahrnovat nebo jak konkrétně je jejich pohyb realizován (například u komplexních tónů, klesavo-stoupavých a stoupavo-klesavých – zahrnuje poměr jednotlivých částí v celkovém tónu). Dále je popsána dimenze sémantická, tedy jaké kontury mohou v jazyce indikovat konkrétní významy, a dimenze frekvenční, která zahrnuje popis frekvence jednotlivých prvků v rámci jazyka. Proces osvojování je celkově ovlivněn nejen intenzitou vystavení cizímu jazyku, ale i například vzděláváním ohledně intonačních prvků jazyka nebo podobnost melodických aspektů s jazykem rodilým. Poté následuje popis melodických rozdílů mezi češtinou a angličtinou se zaměřením především na ty aspekty, které potenciálně mohou být potenciálně problematické pro české studenty angličtiny. Tyto rozdíly většinou zahrnují užší melodické rozpětí u češtiny, přičemž použití tohoto rozpětí by mohlo způsobovat stigmatizaci: mluvčí by mohli působit bez zájmu, nebo přímo zrudně. Další rozdíl je pak distribuce tónu, kde čeština na rozdíl od angličtiny nepoužívá klesavo-stoupavou konturu nebo stejný způsob realizace stoupavo-klesavé kontury. Následující podkapitola (3.4.) popisuje výzkumné otázky, tedy předpoklad, že čeští pokročilí studenti angličtiny se po dvou letech intenzivního vystavení angličtině přiblížili k melodickým tendencím angličtiny, spíše než používání českého rozpětí nebo realizace tónu. Dále byl předpoklad, že studenti budou po dvou letech dělat kratší fráze a používat více stoupavo-klesavých a klesavo-stoupavých kontur a v jejich realizaci se budou podobat rodilým mluvčím.

Následující kapitola vysvětluje metodologii použitou při práci – zahrnuje popis mluvčích, textu, procesu nahrávání, analýzy a extrakce výsledků. Studenti, kteří přispěli k této práci, byli studenti programu Anglistika-amerikanistika od roku 2019. Na začátku prvního ročníku všichni studenti nahrávali kratší čtený text, na kterém poté hodnotili vlastní výslovnost v rámci povinného kurzu fonetiky a fonologie angličtiny. V tomto kurzu prošli i v následujícím semestru několika týdenním vzděláváním i ohledně intonace angličtiny. O dva roky později, na podzim roku 2021 a několik studentů i na jaře 2022, byli studenti požádáni o nahrávání kratšího textu, přičemž až na místě se dozvěděli, že se jedná o identický text. Celkově bylo znovu po dvou letech nahráno 15 studentů. Obojí nahrávání proběhlo bez instrukcí a student pouze dostali několik minut na to, aby si text přečetli. Nahrávaný text byl napsaný v první osobě a popisoval negativní zkušenosti během dovolené (viz Appendix), tudíž potenciálně mohl přispívat k barvitější intonaci, jelikož afektivní složka byla bohatě zastoupená. Afektivnost navíc mohla být umocněná i interpunkcí, jelikož text zahrnoval i vykřičníky. Interpunkce pak mohla ovlivnit i dělení do prosodických frází, jelikož variace frázování byla kvůli čárkám a tečkám možná

pouze v několika málo částech. Všichni studenti byli nahráni v nahrávacím studiu Fonetického ústavu Univerzity Karlovy. Pro porovnání byly nahrány i dvě rodilé mluvčí angličtiny, jedna britské a druhá americké angličtiny, obě mluvčí měly stejné instrukce. Všechny nahrávky pak byly automaticky segmentovány na slova a hlásky, kdy některé hranice ještě byly manuálně upravovány. Poté následovala manuální poslechová segmentace nahrávek na prosodické fráze, v každé pak byla poslechově určená nukleární slabika a od ní pak kontura realizovaného tónu. Problematické části pak byly konzultovány s vedoucím práce. Realizace tónu byla hodnocena objektivně a nepřikládalo se důležitosti na zamýšlené realizaci mluvčího. To pak způsobilo, že u některých frází studenti zamýšleli jiný tón, než byl výsledně realizovaný. Toto se občas stávalo i například na úplném konci textu, kde by měla být použita klesavá kontura pro vyjádření konce textu, ovšem ne vždy tomu tak bylo. Častěji se tento problém děl na ploché kontuře, proto byl použitý termín „level-f“, který označoval tón na konci nejčastěji vět, který byl realizovaný jako level, ale funkčně vyjadřoval klesání. Nahrávky byly rozděleny do tří skupin – skupina „early“, která zahrnovala nahrávky z roku 2019, „late“, kde byli studenti znovu nahráni po dvou letech, a „native“, která zahrnovala obě rodilé mluvčí. Data pak byla extrahována a graficky znázorněna.

Další kapitola (5.) zahrnuje prezentaci výsledků a jejich interpretaci. Jako první byla analyzována délka prozodických frází ve slabikách za frází a slovech za frází. Co se slabik týká, medián pozdější skupiny byl menší než skupiny nahrávek z roku 2019, studenti tudíž opravdu mohli mít tendenci dělat kratší fráze, použitý t.test ale nepotvrdil statisticky významný rozdíl. I porovnání studentů individuálně ukázalo spíše nepatrné rozdíly. Ve slovech za frází pak rozdíl mezi skupinami studentů byl sotva postřehnutelný. Je ale důležité také zmínit, že celkový rozdíl mezi studenty v porovnání s rodilými mluvčími byl rovněž velice malý. Další porovnávací faktor byla distribuce tónů. Data zde potvrdily předpoklady, že studenti po dvou letech produkovat více komplexních tónů, a zároveň se zmenšil počet „level“ tónů. Zajímavé je ale porovnání pozdější skupiny s rodilými mluvčími, kde studenti stále nedosahovali počtu stoupavo-klesavých tónů, ale co se klesavo-stoupavého tónu týká, u něj jejich produkce výrazně převyšovala i počet těchto tónů u rodilých mluvčích. Primární zaměření této práce ale bylo na osvojování melodického rozpětí angličtiny, což bylo vyjádřené v 80-percentilu základní frekvence vyjádřené v půl tónech. Toto rozpětí bylo vypočteno pro nukleární tóny (nukleus dohromady s tail) a celé prozodické fráze. Tady výsledky také ukázaly i občasné zhoršení, kde se rozpětí u mnoha studentů po dvou letech zmenšilo, občas i velice výrazně. Ke zlepšení došlo pouze u několika studentů. Přestože se pak někteří studenti svým rozpětím v prvním nahrávání blížili hodnotám rodilých mluvčích, v pozdějších nahrávkách tomu už tak často nebylo.

Výsledky byly také velice podobné v tónech i celých prozodických frázích, což zahrnovalo i melodickou variaci před nukleární slabikou. Poslední podkapitola (5.4.) znázorňuje individuální vývoj melodické rozpětí čtyř nukleárních tónů u studentů v obou nahrávkách. Znovu je zřetelné spíše zúžení v melodickém rozpětí, a to hlavně u jednoduchých kontur (stoupavá a klesavá), kde k výraznému zlepšení příliš často nedocházelo. U komplexních tónů pak bylo vidět větší zlepšení, ovšem u těchto výsledků je třeba brát v potaz fakt, že počet těchto tónů byl v nahrávkách celkově nízký.

Následující kapitola (6.) shrnuje veškeré výsledky a obecně je interpretuje. Z výsledků se zdá, že celkové zlepšení studentů ve zkoumaných aspektech bylo spíše převáženo zhoršením v intonačním rozpětí. Je ale důležité zároveň říct, že jelikož studenti s nahrávaným textem už v prvním ročníku pracovali, je tedy možné, že k textu po sebe-evaluacích neměli příliš pozitivní vztah, a to se tedy mohlo podvědomě promítnout i do realizace intonace. Navíc, sebe-evaluace byla zaměřená na výslovnost, tudíž studenti mohli předpokládat, že se znovu jedná o hodnocení výslovnosti a mohli se tak proto více soustředit na artikulaci a tím znovu podvědomě potlačit intonační variabilitu. Další důležitý faktor je, že se studenti už během prvního roku studia výrazně podobali hodnotám rodilých mluvčích v téměř všech zkoumaných aspektech. Proto je možné, že k osvojování intonace by mohlo docházet už ve dřívějších stádiích studia jazyka. Pro potvrzení této hypotézy by ale bylo vhodné provést měření znovu s jiným textem, na který studenti nebudou mít žádnou potenciální citovou vazbu. Sedmá kapitola pak práci shrnuje a osmá kapitola uvádí seznam použitých zdrojů. Nahrávaný text je zahrnutý v příloze (kapitola 10).

## **10. Appendix**

### **“Holidays from Hell”**

We arrived at the hotel on Thursday evening. They said it was a five-star hotel, but I wouldn't give it one star! It was just awful. The building was next to the sea, but there wasn't a beach! Just a few rocks, and the sea was so dirty you couldn't swim in it. There were big ships travelling past, and the sea was all polluted and brown, it looked horrible!

---

We went to look at the swimming pool, but that was no better. It wasn't a nice blue colour like you would expect. It was a sort of greeny-black colour, and as we looked more closely, we realised that it was full of frogs! So after that, obviously we didn't use it.

---

And then there was the food. I'll never forget it! The first morning, we went downstairs for breakfast, expecting to have the usual things you get in hotels: bread, marmalade, fruit, coffee... Instead, we were very surprised to see lots of different types of vegetables: carrots, peas, cabbage and a big bowl of lettuce! I was really hungry so I decided to have some of the lettuce, but then I saw that it was moving! I looked a bit more closely and saw that the lettuce bowl was full of ants, hundreds of them, and there were so many that they were actually moving the lettuce leaves! After that, we didn't eat in the hotel again.