Univerzita Palackého v Olomouci

Filozofická fakulta

Katedra anglistiky a amerikanistiky

Annotating English intonation in ToBI with a comparison with Czech

Bakalářská diplomová práce

Studijní program: Anglická filologie

Autor: Jan Laštůvka

Vedoucí práce: Mgr. Šárka Šimáčková, Ph.D.

Olomouc 2023

Prohlašuji, že jsem bakalářskou diplomovou práci na téma „*Annotating English intonation in ToBI with a comparison with Czech*" vypracoval samostatně a uvedl v ní veškerou literaturu a ostatní zdroje, které jsem použil.

V Olomouci dne 10. 5. 2023   Jan Laštůvka

I would like to express my sincere gratitude to my supervisor Mgr. Šárka Šimáčková, PhD.
for her help, support and insightful comments and suggestions in the process of writing as
well as providing the data used in this thesis.

V Olomouci dne 10. 5. 2023   Jan Laštůvka

Abstract

The aim of this work is the analysis of English intonation and the models used to describe it. I further use one of the systems and practice applying it. This thesis also includes a comparison of the studied phenomena with Czech. In the first part I provide an overview of the main systems used for intonation annotation in both languages. In the practical part I focus on using one of the systems (ToBI – Tones and Breaks Indices system) on short recordings of read texts. I comment on these annotations in detail and give arguments for my decisions as well as provide alternative solutions in relevant places. I conclude with using the same system on recordings in Czech and draw general conclusions on the differences based on the analysis of my annotated recordings.

Keywords

Anotace

Cílem práce je analýza intonace v angličtině a její způsoby lingvistického popisu a následná osobní aplikace jednoho anotačního systému. Součástí je také porovnání s češtinou daných jevů s češtinou. V první části nejprve poskytnu přehled hlavních postupů používaných k popisu intonace v obou jazycích. V praktické části se zaměřím na využití systému ToBI (Tones and Breaks Indices) na krátké nahrávky ze čtených textů. Tyto nahrávky pak podrobněji okomentuji a poskytnu odůvodnění, případně alternativy pro mé rozhodnutí. Na závěr použiji stejný systém na anotaci příkladů v češtině a popíšu obecné rozdíly mezi intonacemi obou jazyků na základé svých analýz.

Klíčová slova

intonace, anglická intonace, ToBI (Tones and Breaks Indices) system, česká intonace, melodémy, intonační křivky, anotace intonace

Table of contents

1 Introduction

Intonation is one of the major suprasegmental features in phonology. The definition of intonation varies greatly among linguists. Some use the term to describe several features of prosody including rhythm, pitch, tempo, etc. In this thesis I will focus on intonation purely as the realization and variation of pitch in speech. Further, intonation is perceived by language users intuitively, however languages scientists require an objective scientific way of describing it and discussing it.

This thesis focuses on the systems used to annotate and describe intonation. There are two major systems in use for English. The first system stems from the British linguists and considers intonation to consist of contours which are described. This system described in detail in (Crystal 1969) and has been widely used in the past. In 1980s a new approach to studying intonation was started by the work of (Pierrehumbert 1980) which viewed intonation as a series of tones. This system has since been adapted to several languages and further improved and continues to be most common way of annotating English intonation.

The first aim of my thesis is to practice the annotation of intonation using the contemporarily most used system (ToBI system). As a student of English philology at the Palacký University in Olomouc I intended to expand my knowledge from the seminars on phonetics and phonology and get practical skills in linguistics by learning to use this system of annotating speech. The second aim is to investigate Czech examples and compare them with English examples while using the ToBI annotation system conventions to annotate Czech intonation. In Chapter 2 I provide a basic overview of the approaches to studying intonation and describe the two main systems of annotation for English and also a brief overview of the system used for Czech language. In Chapter 3 I describe the methodology and recordings used in the following section. Chapter 4 consists of commentary on the annotations, which can be found in the Appendix section of this thesis. Finally, I commentate on the Czech examples and draw general conclusions from my observations.

2 Literature review

2.1 Approaches to studying intonation

Intonation as a crucial element of human speech has been constantly studied by linguists. The need for objective and scientific description of intonation culminated in the desire to create a system used to transcribe the crucial and most salient aspects that are important in communication. After several attempts by a huge variety of linguists to create such system, two different approaches became widely used. (Beňuš 2021, p.219) names these systems as 'contour based' and 'target-based'.

The 'contour based' system was invented first and is characterized by an analysis of pitch realization and the shape it takes over an utterance. These shapes are then called contours and their conditions and places of occurrences, different shapes and lengths are studied. As this system is mainly associated with the British linguists and the main comprehensive sources written on this system were done by (Crystal 1969) and (Halliday 1967), I will refer to this as the British system and discuss it in detail in the following chapter. Further, due to the Czech tradition also using the similar contour based system, I will discuss the Czech system following the British system.

The 'target-based' system aims to describe pitch at certain phonologically relevant target points. This system has been developed more recently with the main sources being (Pierrehumbert 1980) and (Beckman 1994). As this system was deliberately developed at American universities for usage on general American English, I will refer to this system as American and discuss it in detail in Chapter 2.5 and further use it in the practical part due to the system being the current norm for annotation.

(Ladd 1996, p. 45) reminds that the American system does not render the British system obsolete, however it "casts it in a new light". However, it was also developed with the intention of resolving long-standing issues present in the traditional British taxonomy.

2.2 The British system – traditional intonation studies

2.2.1 The basic intonation unit

In linguistics each field of study often have a smallest unit which is necessary for the studying the subject and requires to be precisely defined (e.g. phoneme in phonetics, morpheme in morphology, lexeme in lexicology, etc.). As for the intonation systems it is also appropriate to define the basic unit of study. Such unit is called differently by linguists of different background. The American linguists of late 20th century, such as Janet Pierrehumbert, tend to use the term *intonation phrase*, while the British tradition with linguists such as M.A.K. Halliday or David Crystal often use the term *tone unit*.

The definition of the basic unit has been discussed thoroughly and continues to be a subject of research. Previous approaches were done by Zellig Harris, who used the morphological approach. He argues in (Harris 1951) that there is an expected correlation with morphemes and the intonation phrase. Such co-extensiveness with specific grammatical structures can be relevant but is not an appropriate approach to studying the intonation phrase. (Halliday 1967, p. 18) also suggests a congruency of tone units with its division into grammatical units and agrees that there is a tendency but continues to state there are cases in which it does not necessarily correspond to any grammatical unit at all.

A semantic definition has also been discussed, where the intonation phrase consists of semantically compact and finished utterances. An extra-linguistic approach can also be of interest, as it can be defined as a part of an utterance belonging to the same breath group. The most common definition however is the phonological definition defined by the boundaries between separate utterances. Such definition is provided by (Garvin 1963, p. 22): *In terms of concrete analytic practice, defining the units means being able to ascertain their boundaries: The definition of a linguistic unit should be such that given a particular analytic input (either a sample of speech or a set of units of a lower order obtained by a previous procedure) and this definition only, the application of this definition should yield unequivocally the boundaries of the units so defined*. (Gimson 2001) provides an alternative to the pause, which can be replaced

either by lengthening of the final syllable before the boundary, or by increasing the speed of articulation of the syllables near a boundary. Both occasions are therefore still treated as a pause. (Crystal 1969, p. 206) also adds that in addition to the boundaries, the internal structure is also relevant for the definition, such as the presence of one peak prominence within each unit and the movement, which will follow this peak. (Brosnahan & Malmberg 1970, p. 154) conclude that the intonation phrases vary in length (number of syllables as well as time length) considerably and the most important markings for the division into separate units is pausation, lengthening or shortening of final syllables, by terminal-marking pitch movements, by cessation of voice and so on.

The internal structure is most commonly divided into the *pre-head*, *head*, *nucleus* and *tail* with the most prominent syllable carrying one of five to seven possible tones: *(high/low) fall, (high/low) rise, fall-rise, rise-fall, and mid (flat/level)* tone (Beňuš 2021, p.219). Lecumberri (1997, p. 104) further states that: *the only essential component of an intonation group is the nucleus (tonic, nuclear tone) which is the last accented syllable of the group*. The other parts are therefore optional and do not occur in every intonational phrase. (Crystal 1969, p. 208) however notes, that the majority of intonation phrases are maximal and usually consist of all the mentioned units. These parts will be further discussed in detail below.

2.2.2   The pre-head of the intonation phrase

The pre-head is a position within the intonation phrase purely defined by the fact that it precedes the head. It can consist of several number of syllables but (Crystal 1969, p. 233) states that the maximum number is five, but most commonly only one to two. These syllables are also usually unstressed, except for the cases where the consist of a verb, noun, adjective or an adverb, in which the stress is slightly noticeable, but never reaches the same level as the peak prominence. As for the pitch, the unmarked case is that it is a level below the onset syllable.

It most commonly features grammatical words such as the articles, quantifiers, conjunctions, prepositions and auxiliary verbs. As mentioned above, it is possibly for it to consist of an open class category, but these cases are the minority.

2.2.3   The head of the intonation phrase

The head is the unit which carries the pattern of the pitch movement. The position of the head within the intonation phrase is after the pre-head and prior to the nucleus. It usually consists of the first pitch prominent syllable and reaches as far as the nucleus but does not include it. (Davy 1968) has studied the percentage of occurrence of heads within the intonation phrase and concluded that around 64 % of spoken text intonation phrases included the head, while rising up to 70 % for written texts. (Crystal 1969, p. 232) also mentions the length of the head, which varies from one up to thirty syllables. Such high range makes the head the most variable element within the intonation phrase as well as usually it being the longest unit.

The head patterns can be classified into two major groups – falling and rising. The definition is based on the pitch at which the begins and the direction it takes. Falling heads therefore have a lower pitch with each following syllable, while rising heads have the opposite. (Crystal 1969, Chapter 5.8, p. 229) introduces even more detailed types of heads with regards to the presence of stressed and unstressed syllables. He also states that falling heads are the most common pattern in English language and represent up to 30% of all heads (Crystal 1969, p. 232). Finally, the heads can also consist of a more complex realization represented by *falling-rising(-falling)* and *rising-falling(-rising)* contours. For these realizations it is typical to include changes of direction within the head, which can even conclude at the same pitch as it started. He notes that these are more rare in English and the simple realizations are more represented.

2.2.4   The nucleus of the intonation phrase

The nucleus is the only obligatory part of the intonation phrase as it carries the nuclear tone, which occurs on the nuclear syllable of the word. This syllable is further required to carry an stress, which is: *"a cluster of phonetic properties that includes increased intensity and duration as well as various spectral correlates."* (Ladd 1996, p. 155). The direction of pitch movement present on this most prominent syllable of the utterance gives the pattern to the whole intonation phrase. It is usually kinetic, which means it includes change of direction, but

it can also occasionally static, and not include any movement. Such nucleus is called level tone[1], which can be distinguished by the presence of a pause after it.

Similarly to the heads, the pitch movement for the nuclei is separated into simple tones, which include only one direction of the movement, complex tones, which include two to three changes of movement and finally compound tones. This distinction is based on (Crystal 1969), but there is great disagreement in the level of simplification of such patterns, detailed description and terminology. I will describe Crystal's division in detail below.

2.2.4.1 Simple tones

Tones consisting of only one direction of pitch are labeled as simple tones. They carry no change of direction and are divided into *falling, rising* and *level*. If the intonation phrase contains a tail, it will continue to follow this direction set by the nucleus. When annotating intonation, these simple tones are usually paired with their starting position, either relatively high, relatively low or medium. Further, they are also paired with pitch range, which can be considered normal, wide (meaning higher range than normal), or narrow (meaning lower range than normal).

The falling simple tone is such nuclear tone, whose nucleus starts high and finishes low, while following a continuous descend over the duration of the nucleus. This is important to note because (Gimson 2001) describes a phenomenon of a declination effect. Such effect is a tendency of intonation phrases, where the pitch at the end is usually always lower than the pitch at the beginning even though they should both be classified on the same level (e.g. both low tones but the low tone at the end of the phrase is even lower than that at the beginning). Such descending tendency is not considered to be a falling tone.

2.2.4.2 Complex tones

Complex tones differ from simple tones in the presence of a change of direction. Contrary to compound tones, they include a maximum of only one prominence peak. The most common occurrences are *fall-rise* and *rise-fall* contours, but those who contain two changes of direction

---

[1] For further discussion on the level tone refer to (Crystal 1969, p. 215)

(*fall-rise-fall* and *rise-fall-rise*) are also included as long as they follow the rule of one prominence peak. For the rise-fall and fall-rise units, the first part is considered to be more prominent than the second, while with contours with two changes of direction, the second one carries the main prominence.

2.2.4.3 Compound tones

Compound tones also contain two directions of contours present, similarly to complex tones. (Crystal 1969, p. 218-220) points to several important differences which need to be distinguished between such units. First, pitch movement in compound tones follows a "trough" movement, which he considers as a continuous evenness of pitch pattern. This is supported by the fact that this stretch is not usually interacted by a pause, which would create two separate intonation phrases. Second, he argues for a tendency of the first part containing an increased prominence than the second, although there still has to be more than one peak of prominence, as opposed to complex tones. Finally, he views the second part as the main functional element, on which the whole tone is labeled. This was further supported by (Quirk & Crystal 1966), where they proved that the first element can be replaced with very little change to overall effect, while the same change being done to the second element resulted in loss of tonicity.

The distinction between compound and complex tones has been discussed by linguists due to its nature of having overlapping cases and similar structure. (Crystal 1969) demonstrates this in Example (1) where the semantic difference is great enough to require two separate categories. In (1a) the complex tones stresses the YOU, which highlights the subject of the sentence and could be followed with a statement "Well, who does, then", while (1b) puts prominence on both the subject and the verb and could be followed by "So why are you saying you do?" He further adds that the double prominence in compound tones is regularly identifiable and should not be overlooked. Example (1) also shows a possible annotation of text to show the realization of intonation within the British system.

(1) a. |YǑU don't 'know|

b. |YÒU don't KNÓW|

2.2.5   *The tail of the intonation phrase*

Another non-obligatory element of the intonation phrase it the tail. The definition of a tail is mainly relating to the position, which is following the nucleus. This position also reflects in their tone, whose direction is automatically determined by the nucleus, namely unbroken fall after falling nucleus and unbroken rise after rising nucleus. These two cases are most common, but the tail can also start by following the initial nuclear direction and then leveling out. Such leveling can occur when the speaker either reaches the limit of their voice-range and physiologically being forced to do so, or the use can be intentional, which is used to mark attitudes of the speaker, such as irony, sarcasm or boredom. The flattening is more common with lowering ones and is rare with rising ones.

Only about 10% of all nuclei have tails with stressed syllables on following words (Crystal 1969, p. 224). Such statistic allows us to generalize the fact that tonicity falls on the last lexical item and it can be deemed to be a reliable statement. There are cases where tonicity carries a grammatical function and is required to not follow this statement, but these scenarios are highly specific. For further information refer to Chapter 6.3 in (Crystal 1969).

2.3 Halliday's view

As mentioned previously Crystal's approach was not the only way of studying British intonation. (Halliday 1967) provides a different view on the units of intonation. He distinguishes two units bigger than a syllable – tone group, which correlates to Crystal's tone unit and but he also introduces the term foot. Foot is a smaller unit and there can be several of them within one tone group. It consists of salient and weak syllables, for which (Halliday 1967, p. 12) uses the terms ictus and remis.

The tone group consists of only two elements of structure. He distinguishes a "tonic" and a "pretonic" part of the unit. Similar to Crystal's nuclear tone, the tonic part is obligatory, while pretonic is optional, although often present. (Halliday 1967, p.13) states that all primary tone contrasts are carried by the tonic. These primary tones have 5 possible realizations. He assigns a number to each tone and uses these numbers in his annotations. The tones are as follows: 1 –

falling; 2 – *rising/falling-rising*; 3 – *rising* with a preceding level tone; 4 – *(rising-)falling-rising* and finally 5 – *(falling-)rising-falling*. He also provides a possibility of double tonics, which combine the numbers used for simple primary tones. Example (2) therefore demonstrates an annotation using his system. An utterance is separated into feet (by a single /) and tone groups (by a double //) with a number representing the tone present and an underline of the tonic syllable.

(2) //2 is that what you / mean //

He further highlights the importance of distinguishing three distinct systems present, which are subsumed under the single heading of "intonation". First, the distribution into tone groups – the number and location of the tone group boundaries; second, the placing of the tonic syllable – the location, in each tone group, of the pretonic and tonic sections; third the choice of primary and secondary tone. He proposes to call these systems tonality, tonicity and tone. (Halliday 1967, p. 18).

2.4 Czech tradition of studying intonation

The main work on Czech intonation was done by (Daneš 1957), who influenced the contemporary work of (Palková 1994). As for the definition of the basic unit, she uses the term *promluvový úsek* in Czech and defines it as: "*a group of tacts[2] (possibly even a single tact) which is considered by the language user to be behaving as a single unit, most commonly by the presence of pauses at their boundaries*"[3] She further describes it as: hierarchically highest unit where rhytmical appropriates of speech are applied. I can therefore conclude that the English criteria of the presence of pauses and presence of an internal structure containing a peak prominence are also used in Czech. (Palková 1994, p. 163) suggests the division into these segments is closely tied to the syntactic structure, which can serve, along with the presence of

---

[2] A tact (takt) is a elemantary linear rhytmical unit defined in Czech phonetics as a group of syllables which being associated with one word accent (Pálková, 1994, p.158).
[3] Unless stated otherwise, all translations from Czech sources are personally done by me.

a pause, to be a great indicator of such division. The definition of the *promluvový úsek* is similar and corresponding to Halliday's definitions of his units.

A difference when compared to English is the functional sentence perspective. Czech grammar allows a more free word order within sentences, which allows more flexibility with highlighting relevant parts of the utterance. (Palková 1994, p. 168) provides a hypothesis that intonation indicates the theme of the sentence. However she immediately follows by saying that research in this area is still limited and does not provide full evidence of such occurrence. She also states that the presence of one of the intonation contours, which will be described below, signal the core information of the utterance.

The term "*melodém*" is used by (Romportl 1973), which correlates to the English realization of pitch at the syllable carrying the "nuclear tone". It is used to describe the pitch pattern and the main direction of pitch within the intonation phrase. He further distinguishes three main types for Czech based on the direction of pitch and also specifically classifies whether it occurs at the end of an utterance or if it is followed by another one. It is defined as the part of the sentence which starts at the main stressed syllable and ends at the end of the sentence. It can therefore be concluded that this term consists of nuclear tone as well as the tail when considering Crystal's terminology. In addition to the direction of pitch, there is a distinction between "ukončující melodém", which occurs at the end of an utterance and signal a completion of it, and "neukončující melodém", which indicates that the utterance will continue. The main patterns in Czech will be discussed below.

2.4.1   *Melodém ukončující klesavý (final falling contour[4])*

The most commonly occurring in Czech is the "melodém ukončující klesavý", which correlates to the English falling tone. This intonation pattern is used in declarative sentences. It is defined a descending contour pattern and also has to specifically be at the end of the utterance. The descend is usually gradual and often reaches the lowest point of speaker's pitch range

---

[4] All the translations of the "melodém" types from Czech to English are unofficial and translated by me for clarity.

(Palková 1997, p. 309). (Daneš 1957) further distinguishes between marked and unmarked variants of the falling tone. The unmarked variant is characteristic by the pitch height of the nuclear tone being lower than the preceding syllable and is typical for neutral speech. The marked variant on the other hand has the nuclear tone higher than the preceding syllable and is typical for highlighting necessary parts of the utterances.

2.4.2   Melodém ukončující stoupavý (final rising contour)

This contour type correlates to the English rising tones and is the main indication of questions in the Czech language. Unlike in English, where questions are mainly indicated by word order, the main indication relevant for Czech is the presence of this intonation pattern (although both languages use rising intonation to mark questions[5]). It is defined by an ascending contour pattern and also occurs and therefore indicates the end of the utterance. Unlike the previous contour type, the ascend is not always gradual. (Daneš 1957) further distinguishes between purely rising contour where the nuclear tone consists purely of a rising contour and contours which can be followed by an additional falling tone (but the syllable immediately following the stressed syllable still has to be higher than it).

2.4.3   Melodém neukončující (non-final contour)

The final type of contour is the sole case of it not being located at the end of an utterance. (Romportl 1973, p.116) describes that it occurs before a pause which is not final and indicates that the utterance will continue. They most commonly occur in complex and compound sentences. The pattern itself can be either raising or falling. This melodém also typically consists of a low pitch syllable preceding the nuclear tone (Palková 1994, p. 313).

---

[5] This is discussed in (Ladd 1996, p. 113) where he provides a universalist view on intonation and discusses the possibility of the existence of intonation universals which are in most languages. The idea of rising intonation marking questions is one of the possible proposed universals.

2.5 The American system – Tones and Break Indices System (ToBI)

The American linguists followed the British tradition of nuclear tones up until 1980's, when a new system started to be developed. The Tones and Break Indices (ToBI) System was created to annotate and describe intonation. The main aim of creating ToBI was to use all prior knowledge of annotating and researching intonation and create a system, which would objectively and precisely be able to describe intonation and be useful in other fields of science, such as text to speech system in computer science, relation of prosody and human language processing in psychology as well as general phonetic annotation for phoneticians. (Jun 2022, p. 151) also highlights that it functions at the level of phonology and aims to describe the phonological properties of intonation.

(Gimson 2001) summarizes ToBI that it: *"decomposes pitch contours into sequences of high and low tones and also detaches a phrase tones and terminal tones at the end of intonational phrases."* Similarly, (Jilka et al. 1999, p.83) describes the theory behind the system to view intonation as: a sequence of discrete local tonal events that constitute the overall contour. In general, this system focuses on objectively labeling each tone to be either low (L) or high (H) relatively to each other or to the local phrasal pitch range of the intonation phrase. It further marks the nuclear accent/pitch accent with * and provides pauses and boundaries. For further details on the origin and usage of ToBI refer to (Beckman et. al 2005).

The system functions on the theoretical basis of the autosegmental-metrical theory. This theory working on the basis proposed by the linguist Janet B. Pierrehumbert is described by (Arvaniti & Fletcher 2020, p. 78): *to make a principled distinction between intonation as a subsystem of language's phonology and F0, its main phonetic exponent … and reflects the connection between two subsystems of phonology: an autosegmental tier representing intonation's melodic part as well as any lexical tones (if part of the system), and metrical structure representing prominence and phrasing.* The system was initially created only for Mainstream American English, but since then it has been developed for other varieties of

English as well as other languages. (Ladd 1996. p. 42) describes the main tenet of this theory, which is important for ToBI, as the idea of linearity of tonal structure and describes it as follows: "*tonal structure is linear, consisting of a string of local events associated with certain points in the segmental string. Between such events the pitch contour is phonologically unspecified and can be described in terms of transitions from one event to the next. In languages like English, the most important events of the tonal string are pitch accents, which are associated with prominent syllables in the segmental string, and edge tones, which are associated with edges of prosodic domains of various sizes.*"

(Beckman & Pierrehumbert 1986) discuss the applicability of the intonation phrase for the contemporary study of intonation. They propose a term called intermediate phrase, which is connected with studying Japanese and try to use such unit for English as well. They consider intermediate phrase to end at the phrase accent, while the intonation phrase ends as far as the boundary tone. Intermediate phrase is therefore a smaller unit and it is a part of an intonation phrase. (Ladd 1996, p. 93) summarizes the intermediate phrase to be: *"a phrase-like prosodic unit in English that is smaller than an intonation phrase, but it can still have more than one accent, and is merely 'intermediate' between the intonation phrase and the prosodic word."*

Silverman et al. (1992) argue that ToBI has succeeded in being useful for other fields of science as follows: *"It has demonstrated high inter-transcriber agreement . . . It is adaptable to various transcription requirements . . . it allows transcribers to represent some uncertainty in their transcriptions . . . and these formats facilitate sharing and comparison of transcriptions across sites and across hardware and software platforms."* Another critical evaluation of the system is presented in (Jilka et al. 1999, p.107), where it is concluded that ToBI is: on the whole very satisfactory and compares very well with other methods.

2.5.1   Application of ToBI

(Beckman et al. 2005, p. 13) state the first main corollary principles when creating ToBI was the easiness of applying it as well as the wide variety of scientists which should be able to

use it. The second principle was the consistency between various transcription sites. The rules for application of ToBI are therefore clearly defined and have to be followed by every scientist, who decides to use such system. In this thesis I will mainly follow the guidelines published in (Beckman 1997) as well as the MIT course on ToBI annotation in (Veilleux et al. 2006), which served as a great tool in teaching me and helping me acquire the skills of using ToBI.

The system consists of several tiers of labelling. (Silverman et al. 1992, p. 868) states that this is due to the natural aspect of prosody, which has multiple components. These tiers are filled in by the annotator, who labels based on his interpretation of the F0 contour. The annotation can be therefore highly variable and subjective. The four main tiers are: the tone tier, the orthographic tier, the break index tier and the miscellaneous tier. The orthographic tier consists of the transcription of the utterance being written in ordinary English orthography. The miscellaneous tier is optional and is used by linguists to comment on occurrences, which do not have to be necessarily directly connected to linguistic phenomena but are relevant in the process of labeling. Such examples might be coughing, laughter, breathing issues, etc. The main tiers relevant for core prosodic analysis are therefore the tone tier and the break index tier.

The whole process of applying the ToBI system starts with segmenting the recording into units (either intonation phrases or intermediate phrases) by filling in the break index tier. During this analysis it is important to note all occurring events relevant for the Miscellaneous tier to be noted. Following is the process of finding prominent syllables. These syllables are then analyzed and the height of the pitch at these syllables is marked at the tone tier. The labels used in these tiers are discussed in detail below.

2.5.1.1 Tone tier

The tone tier serves for identifying the tonal targets (at the places of phrase accents and boundary tones, which were identified when filling in the break index tier, and pitch prominence, which is visible in the F0) (Benuš 2020, p.224). These tones are based on the interpretation of the fundamental frequency F0 contour of the studied utterances. The tones are either high marked by H or low marked by L. Additionally, ToBI uses other symbols such as asterisk * for pitch accent (hence 'starred tone'); % for boundary tones (signaling the end of the

intonation phrase); hyphen - for the phrase accents (signaling the end of the intermediate phrase); and the ! to signal the occurrence of a downstep[6]. (Ladd 1996, p. 80) further notes that the phrase accents are also free standing unstarred tones occurring between the last pitch accent and the boundary tone. Finally, a + is used to indicate pitch accents containing a change of direction. The complete inventory of English pitch accents is listed in (Breen et al. 2012, p. 280) as follows: five basic pitch accent types, which are single-toned (H*, L*) or bitonal (L+H*, L*+H, and H+!H*) with three downstepped variants (!H*, L+!H* and L*+!H).

Further, I occasionally use the symbol <SIL> which stands for *silence* and signifies longer periods of no speech, which is based on the manual offered by (Veilleux et al. 2006).

2.5.1.2 *Break index tier*

The second tier within the ToBI system is the break index tier. It is used to indicate breaks of pitch within the utterances. These breaks indices are defined by (Beckman 1994, p. 31) as a rating for the degree of juncture perceived between each pair of words and between the final word and the silence at the end of the utterance. There are five possible levels of break indices, ranging from 0 to 4. Similarly to the tone tier, the perception of breaks is also highly variable and subjective. There is however an objective rules for break indices as following. (Benuš 2020, p. 221) notes that the most common index is the break index 1, which is used for smooth linking of adjacent words. The break index 0 is used for very weak boundaries. These typically carry the features typical for connected speech such as assimilations. The rules for using indices 3 and 4 are defined by (Beckman 1994, p. 33). The break index 3 (marked by -) is strictly used to mark the end of intermediate phrases. Similarly, the break index 4 (marked by %) is the highest level and represents the most prominent and significant pauses and therefore serves to indicate the end of the intonation phrase. The break index 2 is defined by (Beckman 1994, p. 35) as the index to mark cases of 'mismatch' between the subjective boundary strength and the

---

[6] The downstep is a linguistic phenomenon occurring when two tones of same level follow each other and the second one is lower than the first one. Refer to (Gussenhoven 2004, p.100) for a more detailed overview of the downstep.

intonational constituency, while (Benuš 2020, p. 222) also uses this tier to indicate disjuncture associated with speech errors.

2.5.2 ToBI and Czech

This system has been expanded to other languages and continues to be developed for new ones. Due to the benefits of objective annotation and understandability, it has been considered useful to adapt it to these new languages. So far, Czech does not have a proper official fully developed ToBI system adapted to it. However there have been studies looking into the principles, such as (Duběda 2014), where he looks at Czech intonation with regards to the autosegmental theory by Pierrehumbert. Other linguists who have considered this system for Czech is (Pešková 2017) who presented a brief introduction to this system at a conference in Cologne and continues to use it in her research such as (Pešková et al. 2018). Although she presented a short guide, I found this material to not provide enough information on the topic and was not able to use it in my annotation of Czech examples only based on the paper published on-line.

2.5.3 Rhythm and Pitch System

The rhythm and pitch system is another system used to describe intonation. It was developed by Laura Dilley and Meredith Brown. The aim of this system was to make up for the inefficiencies of the ToBI system and provide more detailed analysis. (Breen et al. 2012, p. 282) provide three main points of criticism. Firstly, it is the discrepancy between the perception of listeners and what the system captures, the second drawback is the lack of consistent, transparent mapping between labeling distinctions and phonetic or perceptual events and finally the inability of ToBI to capture certain distinctions which may be crucial for the researchers.

As the name suggests, instead of focusing on tones and breaks, the main fields of interest are rhythm and pitch instead. It uses similar labels, H for high pitch, L for low pitch and also uses the * symbol for nuclear accent/pitch accents. It uses four tiers of annotation, namely the words tier, rhythm tier, pitch tier and a miscellaneous tier. For further analysis and application of the system refer to (Dilley & Brown, 2005).

3 Methodology

This thesis aims to provide a more complex study of suprasegmental features with the highlight on intonation. Due to the unsatisfactory time allocation in the bachelor level courses at the Palacký University, there is a limited entry level knowledge provided on the topic of intonation. My goal was to use this base and enhance it by the study of additional literature as well as practical application, which is achieved by annotating excerpts of speech recordings.

After presenting the two major approaches of studying intonation within the English speaking world, I will focus on the American based ToBI system onwards in this thesis. As mentioned previously, this system is the contemporary staple of annotating intonation and is being used and developed further for other languages apart from English. In this thesis I will also try to use this system to annotate excerpts from Czech spoken language and look into the adequacy of it for Czech.

Throughout the practical part of this thesis I will do my annotations based on two main sources. Firstly, the official guidelines for the ToBI system provided in (Beckman 1994). The second being the publicly available MIT course on transcribing prosody with ToBI by (Veilleux et al. 2006). This course provides detailed description of every aspect of ToBI annotation with contrasting examples and necessary recordings.

The annotations are done using the publicly accessible Praat program (Boersma & Weenink 2023) which allows to play the recordings and show several linguistically important aspects of each recording. The most important aspect available for this study is the pitch. Using the annotate function of the program I will add the tiers present in ToBI labeling and fill them accordingly.

Finally, I will compare the annotated recordings and look into similarities or discrepancies, for which I will consult the sources give my explanations and analysis. I will also use the Czech recordings to discuss the application of the ToBI system and try to find general differences between English and Czech intonation.

3.1 The recordings

I will be annotating the excerpts of recordings of native English speakers as well as native Czech speakers reading in their respective languages a famous short children's story called "The tiger who came to tea" (Kerr, & McEwan 2006) These excerpts consist of direct speech segments from the read text. I have chosen a total of four passages, which are read by four native English speakers and two native Czech speakers. All participants are female[7] and are reading the text from a book. The English speakers data was collected from the recordings available on the internet, more specifically on YouTube[8], with their age unkown. The Czech speakers were students of Palacký University in their early 20s. Due to the fact that it is a children's book, there may be bigger emphasis on some suprasegmental features including intonation.

I initially intended to annotate passages of the size of one intonation phrase. After examining the data I concluded that this approach would be inappropriate due to different segmenting by various readers of the same sentences into intonation phrases[9]. Therefore I decided to resort to recordings of whole sentences. The selected sentences were*: It can't be the milkman, because he came this morning* (Annotations 1-5*); I think I'd better go now* (Annotations 6-9); *I have nothing for daddy's supper* (Annotations 10-13); *We'd better open the door and see* (Annotations 14-17). The Czech examples are Annotations 18 – 26 and tackle the same sentences as in English. From the available data I the speakers with highest quality recordings and clearly perceivable intonation for each sentence. All annotations are attached in the appendix section, however I will provide tables with relevant parts of sentences, which I are discuss in the practical part section. The annotations also include the pitch range in the misc section, which needs to be set up in praat to achieve similar visualisations.

---

[7] Although I was also provided with recordings done by male speakers, I decided to only focus on female speakers' recordings. Firstly due to the number of male recordings being lower, therefore smaller availability and also in order to lower the amount of variables which would have to be discussed further to provide a comprehensive analysis of differences between intonation in males and females.

[8] The available links are provided in the Appendix section

[9] The various differences in grouping syllables into intonation phrases will be discussed in the practical part.

4   The practical part

In the practical part of this thesis I focused on annotating the recordings with the ToBI system. For the recordings in English I analyzed different segmenting into intonation phrases of each and also looked into different realization of certain passages. Below I will discuss each utterance individually and choose parts, which are valuable for inquiry and comment in detail as well as provide tables with chunks relevant to the discussion.

4.1 Recording 1 - It can't be the milkman, because he came this morning (Annotations 1 – 5)

The first recording is also the longest and consists of two intonation phrases. With other recordings I only chose 4 speakers with best pronunciation, but in this recording I used 5 speakers because there have been several differences and issues worthy of a discussion. This utterance is also unique because it consists of a complex sentence, which is reflected in it carrying two intonation phrases.

The first intonation phrase consists of *It can't be the milkman*. In all recordings it has two prominent syllables at words *can't* and *milk-*. An H tone is used at *can't* unanimously, however we see some variance in pitch at the second prominent syllable as well as the end of the intonation phrase. Speaker 4 has a low pitch throughout the whole word. Although spectrogram in praat shows very high visualization of pitch, this is considered an anomaly as the speaker reaches a very low pitch (creaky voice) and the program struggles with the recording. As for the boundary tones, Speaker 1, 2, 3 and 4 end the intonation phrase with a low pitch rising upwards, which is described using the L-H% marker, while Speaker 2 also reaches a creaky voice and the pitch is difficult to track in praat. Speaker 5 has a very high pitch at the end of the intonation phrase relative to the rest of the sentence. Based on this I decided to use an !H-H% instead of L-H%. Her boundary tone is stable and therefore I used same tones for both phrase accent and the boundary tone and I also decided to use a downstepped H tone instead of L
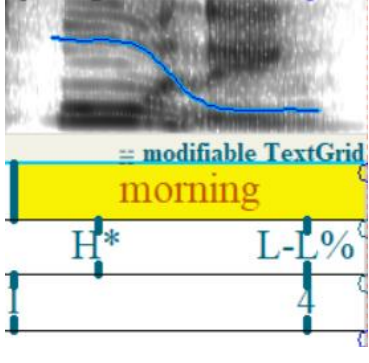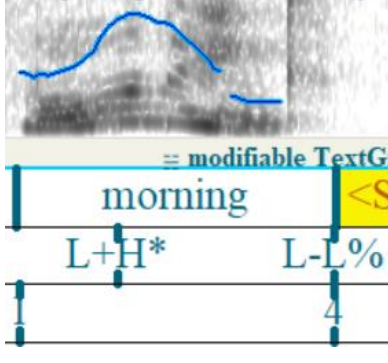
because I still consider this tone to be high to her relative pitch range, especially when compared to her low tone which is visible at the word *morning*, which is noticeably lower.

This recording is also unique with the presence of the 0 break index. All speakers except Speaker 1 have the process of assimilation of place of articulation for the [n] in *can't* to [m]. As per the definition in the literature review, the presence of such phenomenon requires the use of the break index 0 used for very weak boundaries between words.

The second intonation phrase starts with the word *because*. Speaker 3 has the most anomalous realization because she has the first two word separated from the rest of the utterance by a significant pause of 0,33 seconds therefore she creates one more intonation phrase. If we omitted the pause we would see that the pitch is continuous, therefore I suppose this pause to be unintentional and it is a result of poor breath management. This is further supported by the lack of accented syllable in the part prior to the pause. Speaker 1 also introduces an intermediate phrase. Speaker 2 and 4 have two prominent syllables in the whole intonation phrase as they further accent the word *he* with an H tone.

There is a noticeable difference in the pronunciation of the final word in this sentence – *morning*. In the British tradition of studying intonation it would be the final pitch accent in the utterance, therefore it would carry the nuclear tone. I have resorted to using both single tone and bitonal markings. Table 1 is used to show these two possibilities. The clearest use of single tone would be Speaker's 2 realization where we can see a clear single tone at the prominent syllable *mor-* with the gradual lowering to the low phrase accent and boundary tone. Speaker 1, 3 and 4 show a different example of a clear bitonal realization. The prominent syllable carries a visible change of pitch with the usual bell shape of the pitch. Speaker 5 is the most unique and her realization consists of an L tone as the pitch is lower than at the preceding words. All speakers finish the utterance with a falling pitch into an L boundary tone. The exception is again Speaker 5 because her pitch shows a short rise at the very end. Based on perception I would still consider this an L boundary tone but I will leave the H boundary tone to respect the realization as is shown by praat.

Table 1
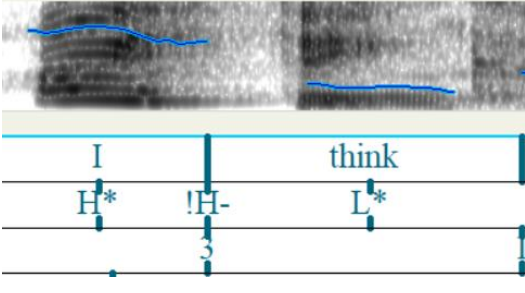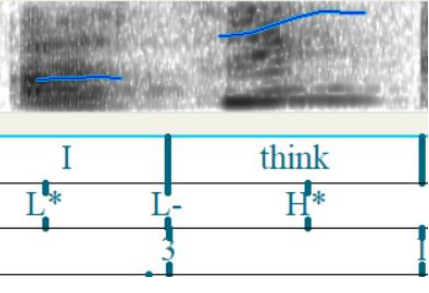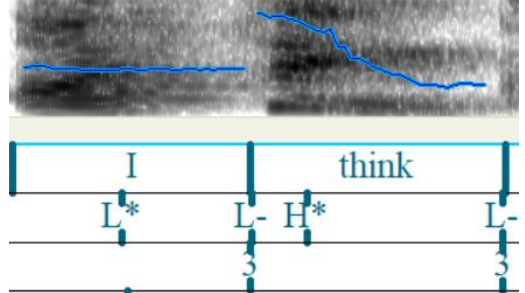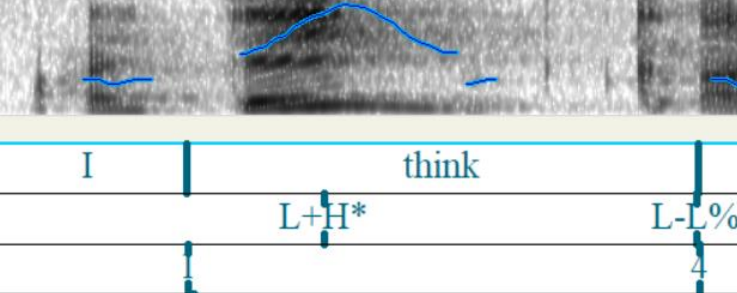
Single tonal realization – Speaker 2	Bitonal realization – Speaker 3

4.2 Recording 2 - I think I'd better go now (Annotations 6 – 9)

The second recording also showed great variety in intonation patterns. Firstly focusing on the first two words – *I think*. The realizations of these two words is presented in Table 2. In speakers 2, 3 and 4 the I was labeled as a separate intermediate phrase based on perception as well as different pitch as shown in praat. Speaker 5 on the other hand shows a smooth continuation of pitch to the following word and belongs in the same unit with it. The chosen unit was a whole intonation phrase marked by boundary 4 based on the length of the word *think*, which is considerably longer in context of the whole utterance than in other speakers, as well as the length of the pause prior to the next word. The inclusion of *I* within this unit is further supported by the relative short length.

Further, Speaker 5 uses a bitonal realization of pitch on the prominent syllable *think*. Although the identification of bitonality can be worthy of a discussion in some cases, both the L tone as well as the H tone are realized over the whole syllable, so I consider this an unambiguous example. Speaker 4 on the other hand emphasizes *think* in her reading and treats it as a separate intermediate unit consisting of H prominent tone falling to L phase boundary. Speakers 2 and 3 group *think* separate of *I* and binds it with following words. However, there are also differences as Speaker 3 uses *think* as a prominent syllable within the unit carrying an H tone while in Speaker 2's reading it consists of L pitch accent.

Table 2

Speaker 2	Speaker 3
Speaker 4	Speaker 5
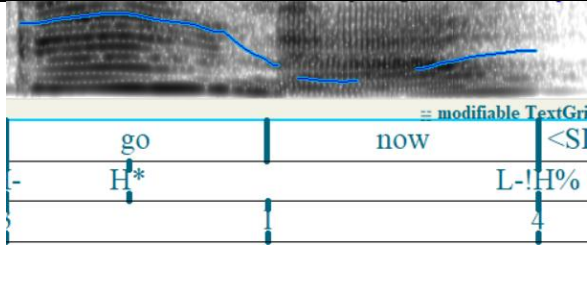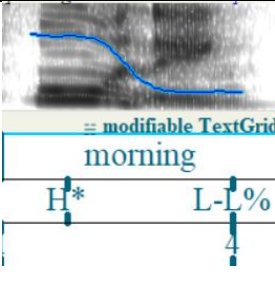
The following sequence of *I'd better* is similar in Speakers 2, 4 and 5 with all consisting of gradual rise of pitch with an H boundary tone. The only difference is seen in Speaker 3, who uses bitonal realization of *better*. This change of pitch is very slight of roughly 100 Hz and can be annotated as a single H*. This passage also contains small errors in praat, which further make the judgment difficult.

Finally, the last two words *go now* also show variance in realizations. With Speakers 3 and 5 I decided to separate each into their own intermediate phrases based on perception as well as relative length of these word. They also share identical realizations as a string of L tones and phrase boundaries ended by a L intonation phrase boundary. Speakers 2 and 4 on the other hand group both words into one intermediate phrase and use H prominent tone over *go* and finish the utterance with L phrase boundary and H boundary tone. Although I provide this version of annotation of these two words, I propose that there is a possibility of discussion over separating them into two separate intermediate phrases similarly to Speakers 2 and 4 based on length.

Throughout my annotations I continue to struggle with the unit called intermediate phrase as I am not able to use my perceptional abilities due to me being a non-native English speaker. The guides I used to learn this system also was not helpful enough in aiding me to learn to identify this unit.

Such realization would be labeled as fall-rise nuclear tone by the British system and such marking with the ToBI system would be the most common. In Table 3 I provide a quick comparison of a fall-rise contour and a falling contour. In Recording 2 we see a clear rise compared to Recording 1.

Table 3

Recording 2 – Speaker 2	Recording 1 – Speaker 2

4.3 Recording 3 – I've got nothing for daddy's supper (Annotations 10 – 13)

According to my analysis, all recordings have the first two words of this sentence – *I've got* – separated into an intermediate phrase. Speakers 2 and 3 end this phrase in an L tone, while Speaker 4 and Speaker 5 end it with an H and a downstepped H tone respectively. I also propose an alternate possible annotation with these two words not having their separate intermediate phrase but rather being a part of a larger phrase along with the following words. My separation is supported by the presence of an H tone over a prominent syllable in *I've* as well as the presence of another prominent syllable in the following words, which requires its own phrase.

The following two words show great variance between the speakers. Table 4 focuses on this sequence of words and shows the discrepancies. In terms of sequencing into units, an

extreme case is with Speaker 4, who has a major pause following the word *nothing* and therefore I used the break index 4 for the end of an intonation phrase. The realization of pitch is by an H pitch accent and an L boundary tone followed by a rise a to a downstepped H boundary tone. Speaker 5 has a similar realization with the exception of it being only a separate intermediate phrase, not a full intonation phrase. Speakers 3 and 2 differ with having the following word *for* grouped with *nothing*. I support this claim by arguing that in these recordings the pitch pattern over the word *for* does not consist of any peaks and only continues in the pattern and direction set by the previous word. Speaker 3 also has a unique realization of pitch by using a bitonal L+H tone. Although all speakers use an H tone followed by a fall into an L tone, I argue that Speaker 3 has the most audible presence of a bitonal realization over the first syllable, while the other speakers realize the first syllable mainly using a single H tone.

Table 4

The final words of the sentence also consist of different sequencing into units. As mentioned previously, the word *for* is grouped with nothing in Speaker 2 and 3, while Speaker 5 and 4 have it grouped with *daddy's*. The final word *supper* is a separate intermediate phrase in all speakers as it is important within the functional sentence perspective. Speakers 5 and 4 have the same realization with the first syllable being stressed and carrying an H tone pitch

accent with a fall to a L phrase as well as intonation boundary. Speaker 3 also consists of same tones, however the H pitch accent is unusually on the second syllable of the word, which should normally occur on the first syllable. I support this claim by auditory perception as well as realization of pitch, which reaches the highest point at the second syllable, unlike other recordings. Speaker 2 further has a different realization as she has an L tone at the prominent syllable and remains the pitch low until the final L phrase and boundary tone.

4.4 Recording 4 – We'd better open the door and see (Annotations 14 – 17)

In this recording the final two words of this utterance are in their separate units, while the first five words show great differences in all the speakers. Speaker 2 has the most unique pronunciation, as she uses three different intermediate phrases. I arrived at such conclusion based on auditory perceptional differences as well as visible gaps in the spectrogram in comparison to other speakers. Further, in other recordings the pitch follows a general pattern such as in Speaker 1 with the rise until the final fall at the end of the intonation phrase, while Speaker 2 has three unique chunks with separate peaks. Table 5 provides an overview on this example.

The second word *better* is stressed in all speakers, but the tone differs as Speaker 1 and 4 have an L tone, while Speaker 1 and 5 have an H tone. This is tied with the following couple of words – *open the* – for which the pitch remains relatively high for Speaker 1 and 4, while Speaker 2 and 5 have them lower relative to the rest of the utterance. Speaker 2 further has an H tone present as it belongs to their own intermediate phrase.

Table 5
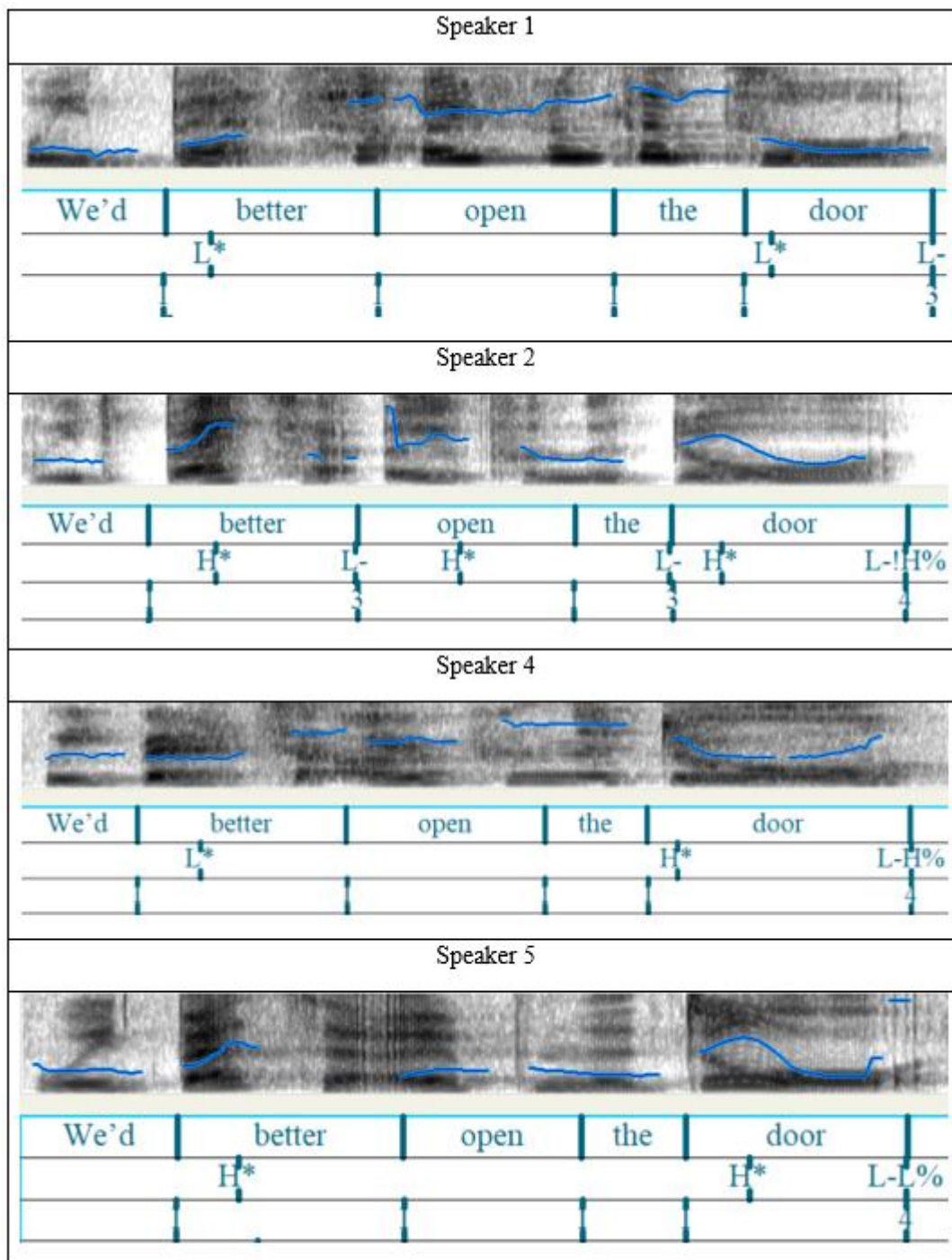
The word *door* is also heavily influenced by the level of pitch from the previous words. Speaker 2 and 5 have a same realization with an H tone at the first prominent syllable. Although the spectrogram in praat suggests a curve usual for bitonal realizations, I decided for a single tone based on the fact that the pitch is low prior to this word and therefore the pitch has to rise to a high level first. This issue is discussed in (Beckmann 1994, p.12) and she states the main

difference lies in what happens before the high tone. The L in the bitonal tone is supposed to mark a rise of pitch to the H tone which cannot be attributed to a previous L phrase or boundary tone. In my example here I do not consider the circumstances of using L+H* tone applicable. Further, we see differences in the boundary tone with Speakers 2 and 4 have the pitch go low and finish with a small rise, therefore the use of L-H%, While speakers 1 and 5 have a low pitch all throughout. Speaker 1 has a different boundary as he only concludes an intermediate phrase, while the other speakers finish one whole intonation phrase.

Finally, the two words *and see* have their own intonation phrase in Speakers 1, 4, 5 and an intermediate phrase in Speaker 2. In Speaker 2 and 5 I decided to use a bitonal tone. (Veilleux et al. 2006) describe in chapter 2.5.2 that the requirement for the L+H* usage is: "a more substantial rising pitch movement leading up to the H* tone." I believe that in these examples it is present, especially when contrasted with Speaker 1, where I decided to only use a simple H tone. Although the preceding pitch for Speaker 2 and 5 is also rising or relatively high, the realization over the stressed syllable *see* shows signs of bitonal realization. The utterance is then finished by a sharp fall to a low pitch and a low boundary tone marked by L%. The only exception is Speaker 5 for which I included a downstepped H boundary tone because the pitch in praat shows a slight rise at the very end.

4.5 Czech examples (Annotations 18 – 26)

The intonation patterns in Czech show different tendencies than in English. Based on my annotations I concluded that Czech intonation differs in the number of prominence peaks within the intonation unit. English can easily have two or more prominence peaks with stressed syllables and also due to the presence of a smaller unit (the intermediate phrase) the intonation phrase is more segmented. ToBI aims to describe pitch at important events for English intonation such as end of intermediate phrases, stressed syllables and boundary tones. In Czech we only see one prominent stressed syllable and one boundary tone with intermediate phrases not being present. This is further supported by the absence of secondary accents in Czech. This leaves the Czech annotations to only feature two places with annotated tones. I would therefore

conclude that although both languages share similar realizations of pitch at the final pitch prominent syllables, with falling intonation finished by L-L% being the most common realization of pitch for indicative utterances, the Czech intonation phrase differs in pitch across the whole rest of the intonation phrase preceding the pitch prominent syllables. In Czech there were not any visible prominence peaks present prior to the main final one.

The general pattern of Czech intonation present in my annotations is of gradual stable direction of pitch throughout the whole intonation phrase with one of the final syllables carrying a sharp change into a falling contour. All the annotations of the chosen examples (Annotations 18-25) therefore end with an L phrase and boundary tone. These realizations of pitch would be described by Czech linguists as melodém ukončující klesavý (final falling contour), which is commonly used for declarative sentences. The clearest example is provided in Table 6a.

Due to these recordings only showing one type of the final contour types in Czech, I looked into the text to find an example of the other type. Annotation 26 shows a question, which is a typical scenario in which Czech speakers use the melodém ukončující stoupavý (final rising contour). Once again the speaker maintains a relatively stable pitch throughout the sentence but at the end there is a sharp dip at the first syllable of *vámi* and the intonation phrase finishes with an H tone. Table 6b illustrates this contour with other contours available for comparison in 4a and 4c.

Finally, the last common type is the melodém neukončijící (non-final contour). It is present in a compound sentences with the utterance continuing afterwards. Table 6c demonstrates this on Annotation 25, where *Pojďme otevřít* behaves as the first sentence with the high pitch realization over the first part of the sentence followed by a fall to low pitch and small rise again at the end of the phrase to signal rising pitch, which would be described by Czech linguists as melodém neukončující stoupavý (non-final rising contour) and a *uvidíme* functions as the second part of the sentence with the usual realization of relatively high pitch followed by a fall into an ending low boundary tone.

Table 6

a)   General Czech final falling contour (Annotation 22)
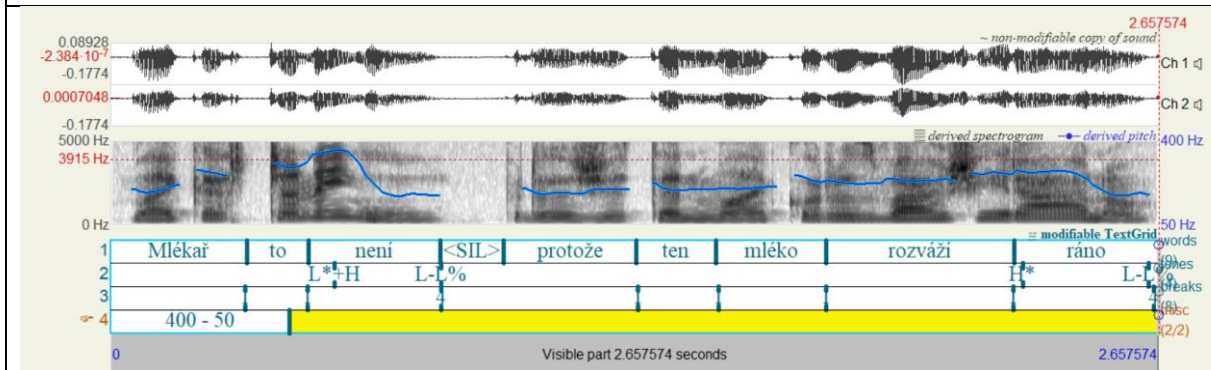
b)   Czech final rising contour (Annotation 26)
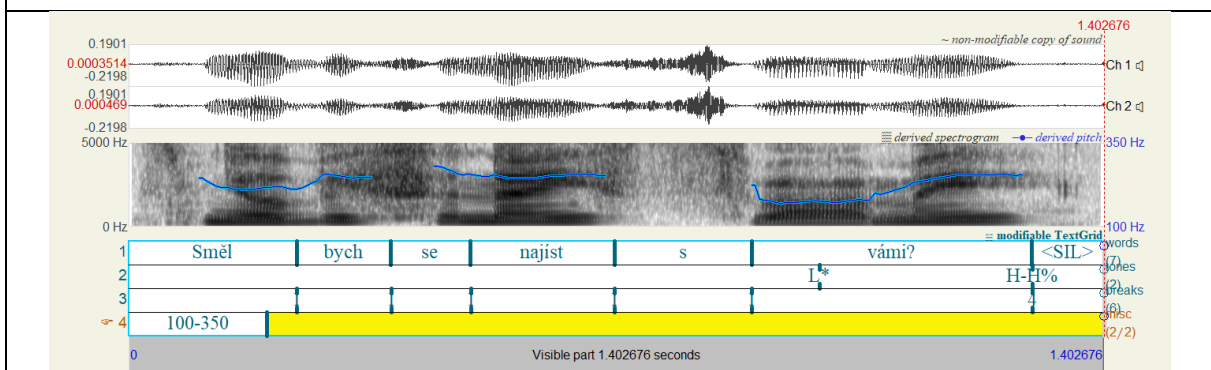
c)   Czech non-final contour (Annotation 25)
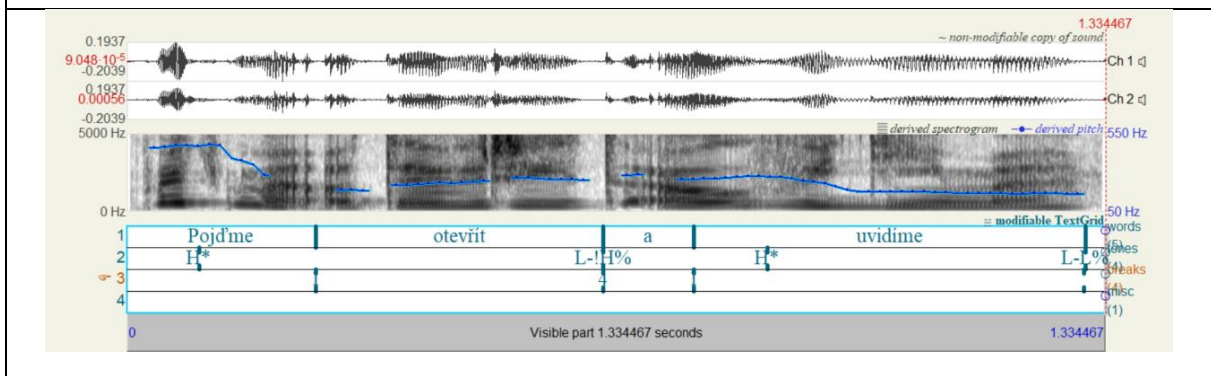
5 Conclusion

The main aim of this thesis to discuss the systems used to annotate intonation in English and Czech and then apply the ToBI system on the recordings of readers of a children's book text The tiger who came to tea. After analyzing the recordings, my goal was to point out differences among the different speakers as well as compare the realizations of same parts of text in English and Czech.

All annotations were done in Praat with the guidelines of ToBI annotation (Beckmann 1994) and the publicly available MIT Course (Veilleux et al. 2006). After comparing the annotations I highlighted parts which showed great variety of different realizations of pitch. In several parts of the recordings I hesitated on the correct way of annotating and provided alternate views which I considered viable. Further, as a non-native speaker of English I found some difficulties in recognizing some English specific aspects, such as identification of the intermediate phrase, which should be easily audibly identifiable by natives.

I continued to compare these recordings with Czech examples and concluded, that in my examples Czech intonation showed to have fewer amount of prominent peaks, different sizes of intonation phrases and different pre-nuclear realizations. The most prominent pitch present at the end of intonation phrases on the other hand showed similarities between the two languages. Overall I would consider the ToBI system application sub-optimal when used with the rules for English. Although the autosegmental-metrical theory approach has the potential to be useful for other languages than English, as was demonstrated by (Duběda 2014) and others, I consider my thesis to prove that the ToBI system requires specific rules for the specific languages it is being used for. When researching this topic I found a small amount of materials discussing it. These Czech specific rules for the ToBI system therefore require more research in order to fully adapt this system to Czech.

6 References

Arvaniti, Amalia, and Janet Fletcher. 2020. 'The Autosegmental-Metrical Theory of Intonational Phonology', in Carlos Gussenhhoven, and Aoju Chen (eds), *The Oxford Handbook of Language Prosody*, Oxford Handbooks (online edn, Oxford Academic, 10 Feb. 2021).

Beckman, Mary E., and Julia Hirschberg. 1994. "The ToBI annotation conventions." Ohio State University.

Beckman, Mary E., Julia Hirschberg & Stefanie Shattuck-Hufnagel. 2005. The original ToBI system and the evolution of the ToBI framework. In Sun-Ah Jun (ed.), Prosodic Typology: The Phonology of Intonation and Phrasing, 9–54. Oxford University Press.

Beckman, Mary E., and Janet B. Pierrehumbert. 1986. "Intonational structure in Japanese and English." *Phonology 3*: 255-309.

Beňuš, Štefan. 2021. *Investigating Spoken English: A Practical Guide to Phonetics and Phonology Using Praat*. Cham: Palgrave Macmillan.

Breen, Mara, Laura C. Dilley, John Kraemer, and Edward Gibson. 2012 "Inter-transcriber reliability for two systems of prosodic annotation: ToBI (Tones and Break Indices) and RaP (Rhythm and Pitch)." Corpus linguistics and linguistic theory 8, no. 2: 277-312.

Boersma, Paul & Weenink, David (2023). Praat: doing phonetics by computer [Computer program]. Version 6.3.10, retrieved 3 May 2023 from http://www.praat.org/

Daneš, František. 1957. *Intonace a věta ve spisovné češtině*. Prague: Nakladatelství Československé akademie věd.

Davy, D. 1968. "A study of intonation and analogous features as exponents of stylistic variation, with special reference to a comparison of conversation with written English read aloud." M.A. Thesis, University of London.

Dilley, Laura C. & Meredith Brown. 2005. The RaP (Rhythm and Pitch) Labeling System, Version 1.0: Available at http://tedlab.mit.edu/tedlab_website/ RaPHome.html

Duběda, Tomáš. 2014. "Czech intonation: A tonal approach." *Slovo a slovesnost*, vol. 75:83–98.

Gimson, A. C. 2001. Gimson's Pronunciation of English. Sixth Edition. 6th edition revised by Alan Cruttenden. London: Arnold.

Halliday, Michael Alexander Kirkwood. 1967. *Intonation and grammar in British English*. Berlin, Boston: De Gruyter Mouton.

Harris, Z. S. 1951. *Methods in structural linguistics*. Chicago: University Press.

Jilka, Matthias, Gregor Möhler, and Grzegorz Dogil. 1999. "Rules for the generation of ToBI-based American English intonation." *Speech Communication 28, no. 2*: 83-108.

Jun, Sun-Ah. 2022. "The ToBI transcription system: Conventions, strengths and challenges." *Prosodic theory and practice*: 151-181.

Kerr, J., & McEwan, G. 2006. The tiger who came to tea. New York: HarperCollins Childrens' Books.

Ladd, D. Robert. 1996. *Intonational Phonology*. Cambridge: Cambridge University Press.

Lecumberri, Mª Luisa García. 1997. "Traditional British analyses of intonation: a review." Atlantis: 103-111.

Palková, Zdena. 1974. *Rytmická výstavba prozaického textu*. Prague: Academia

Palková, Zdena. 1994. *Fonetika a fonologie češtiny*. Prague: Karolinum

Pešková, Andrea. 2017. "Czech ToBI" Paper presented at the Phonetics and Phonology in Europe, Cologne, Germany, June 12-14.

Pešková, Andrea, Laura Colantoni and Trudel Meisenburg. 2018. "Initial and final intonation cues of Czech yes-no questions." Paper presented at the 9th International Conference on Speech Prosody, Poznan, Poland, June 13-16.

Pierrehumbert, Janet Breckenridge. 1980. "The phonology and phonetics of English intonation." PhD diss., Massachusetts Institute of Technology.

Quirk, R. and David Crystal. 1966. "On scales of contrast in English connected speech." *In Memory of J. R. Firth*: 359-369.

Romportl, Milan. 1973. *Základy fonetiky*. Prague: Nakladatelství SPN

Silverman, Kim EA, Mary E. Beckman, John F. Pitrelli, Mari Ostendorf, Colin W. Wightman, Patti Price, Janet B. Pierrehumbert, and Julia Hirschberg. 1992. "ToBI: A standard for labeling English prosody." *ICSLP, vol. 2*: 867-870.

Veilleux, Nanette, Stefanie Shattuck-Hufnagel, and Alejna Brugos. 2006. 6.911 Transcribing prosodic structure of spoken utterances with ToBI. January IAP 2006. Massachusetts Institute of Technology: MIT OpenCourseWare. https://ocw.mit.edu.

7 Appendix

7.1 Links to recordings available on-line:

Speaker 1 – https://www.youtube.com/watch?v=UGqFxN0V-

CA&ab_channel=BooktasticTales

Speaker 2 – no longer accessible on-line

Speaker 3 –

https://www.youtube.com/watch?v=4k87lP1GNrE&ab_channel=MagicOfStoryTime

Speaker 4 –

https://www.youtube.com/watch?v=ggjemuhlmuM&ab_channel=MeadowfieldSchool

Speaker 5 –

https://www.youtube.com/watch?v=RjcxZemAxNI&ab_channel=TeachLikeBenni

7.2 Annotations:

Annotation 1 - Recording 1 - Speaker 1
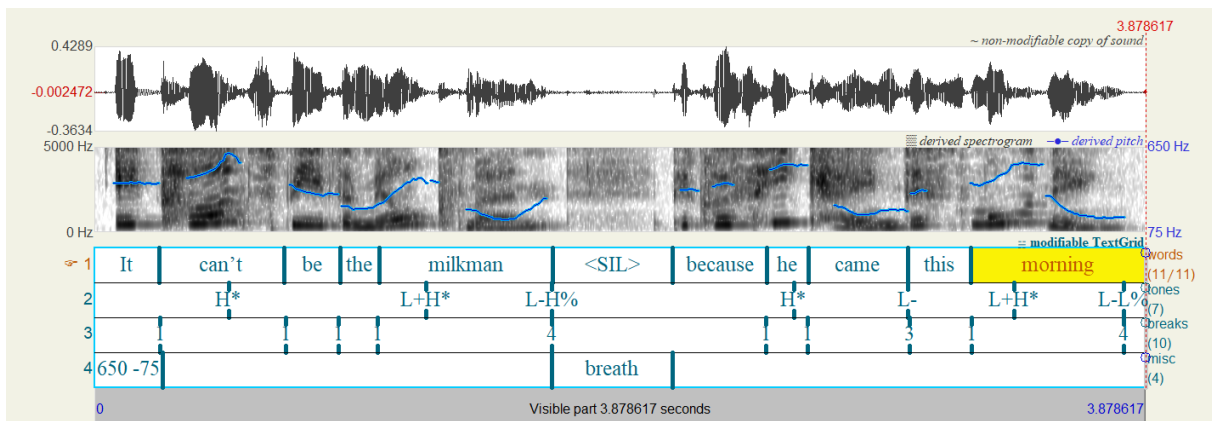
Annotation 2 - Recording 1 - Speaker 2
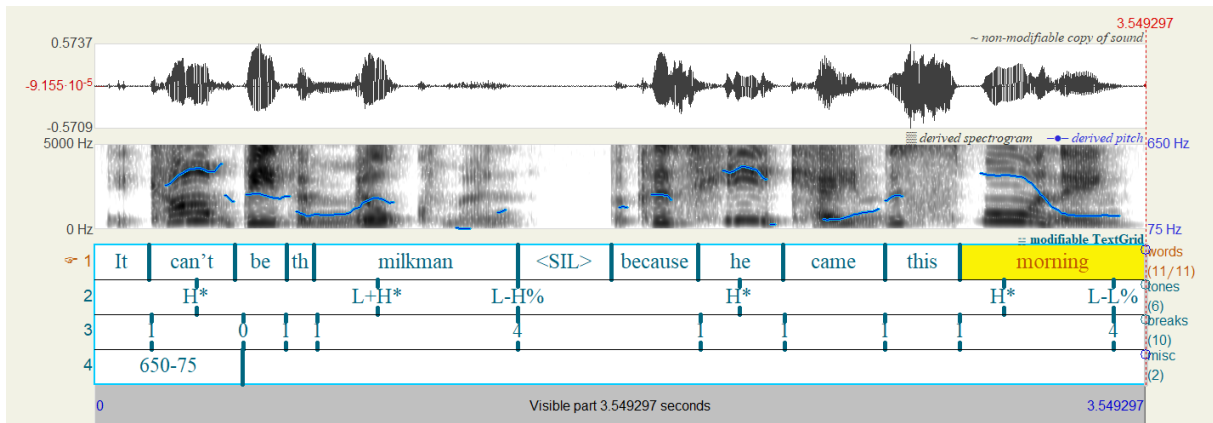
Annotation 3 - Recording 1 - Speaker 3
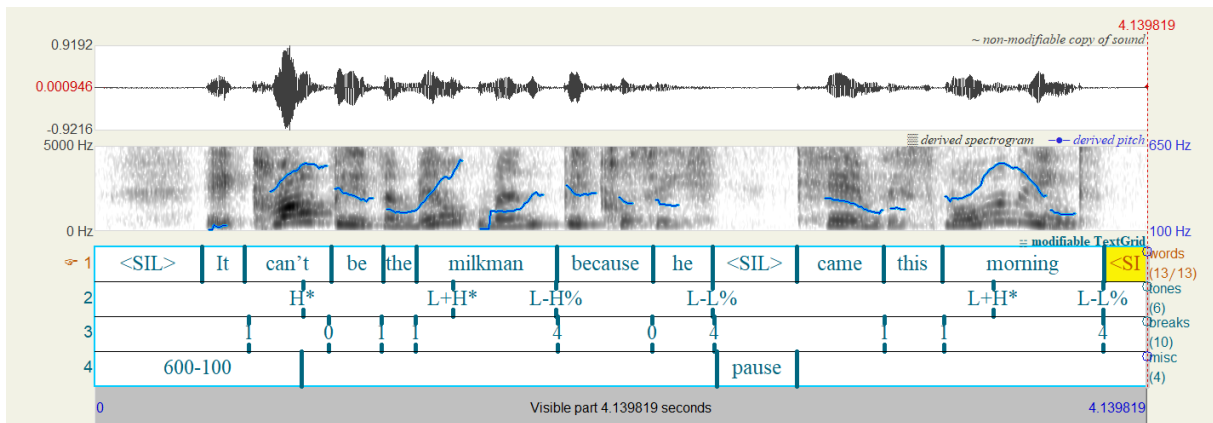
Annotation 4 - Recording 1 - Speaker 4
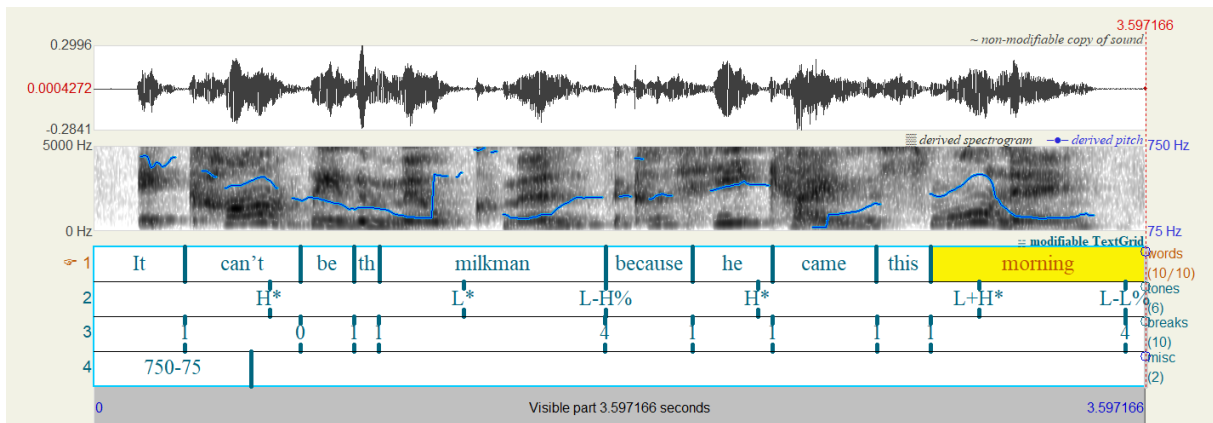
Annotation 5 - Recording 1 - Speaker 5
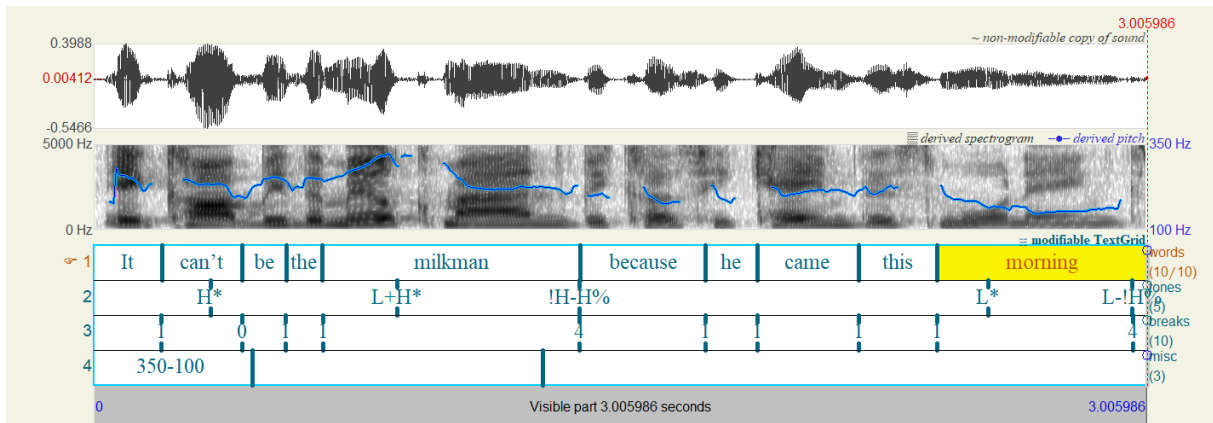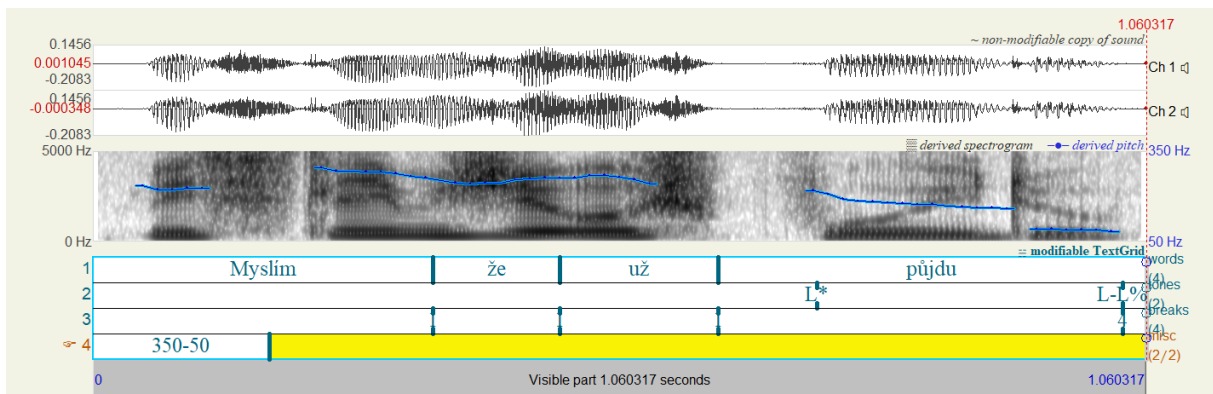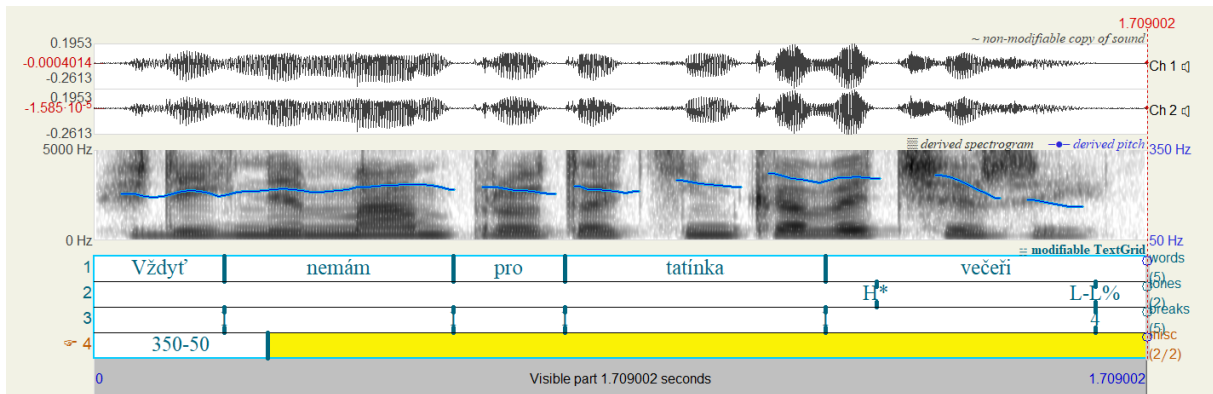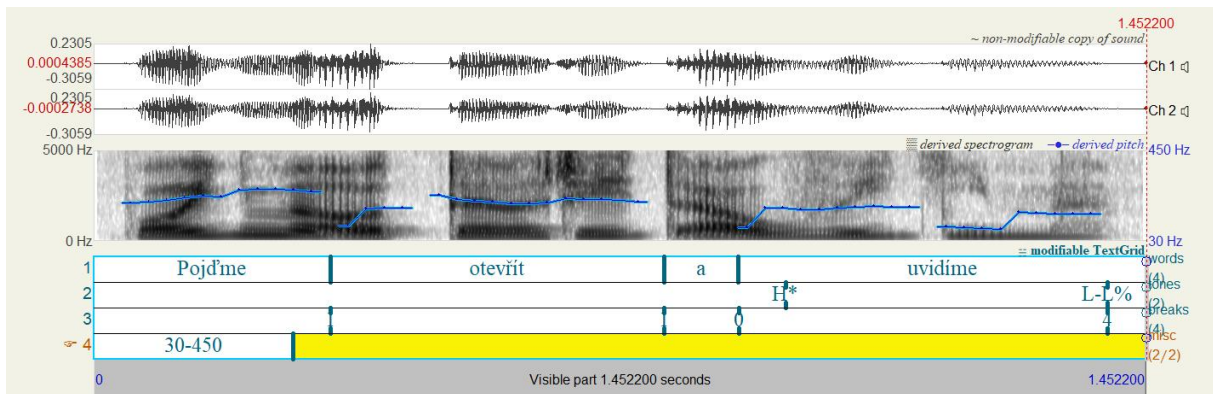

Annotation 6 - Recording 2 - Speaker 2
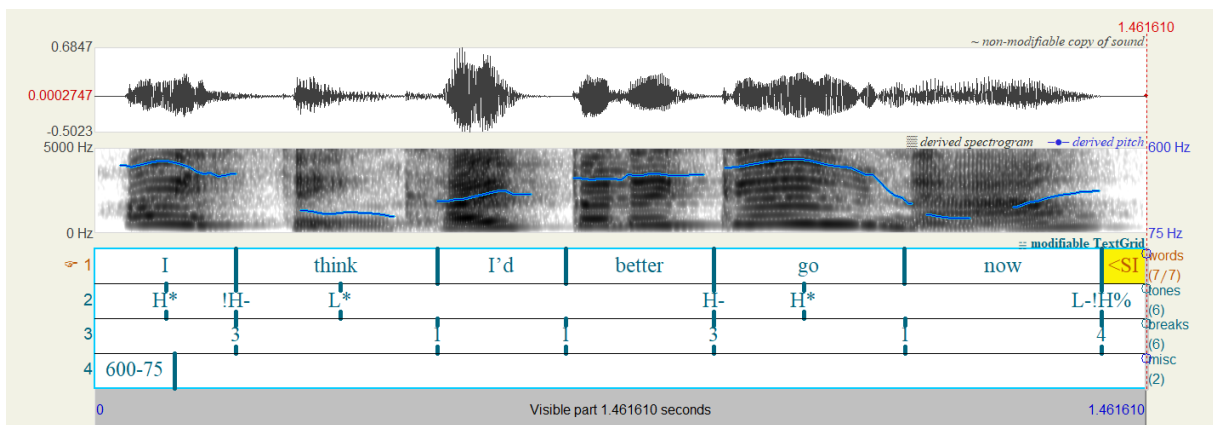
Annotation 7 - Recording 2 - Speaker 3
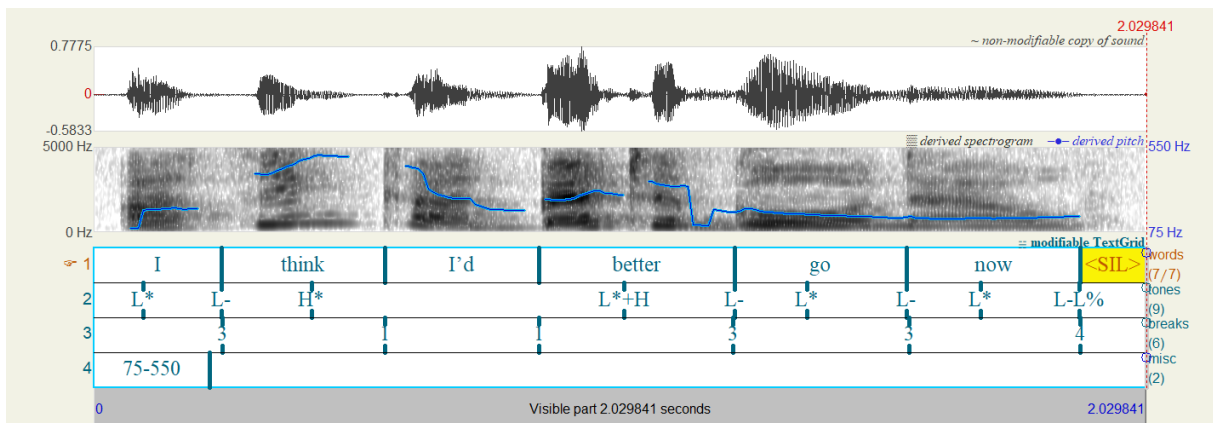
Annotation 8 - Recording 2 - Speaker 4
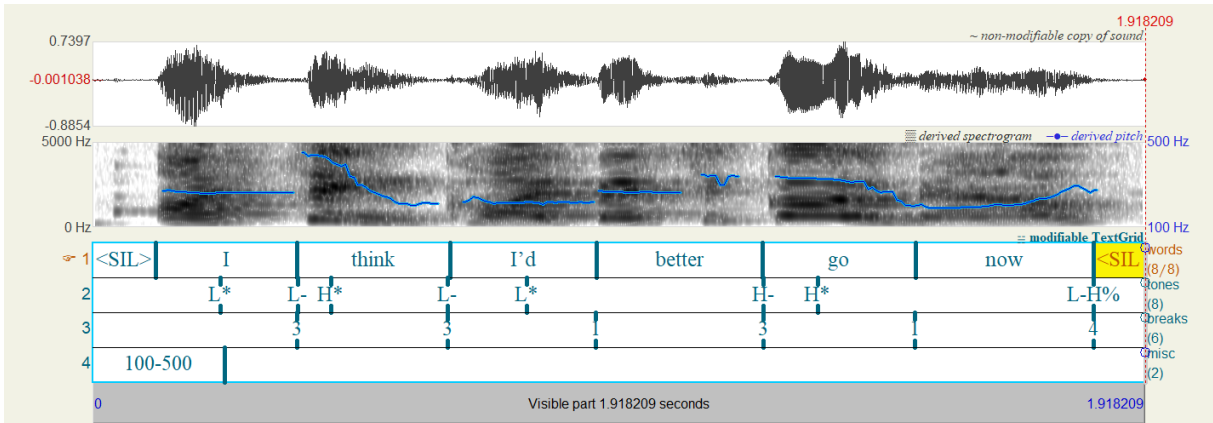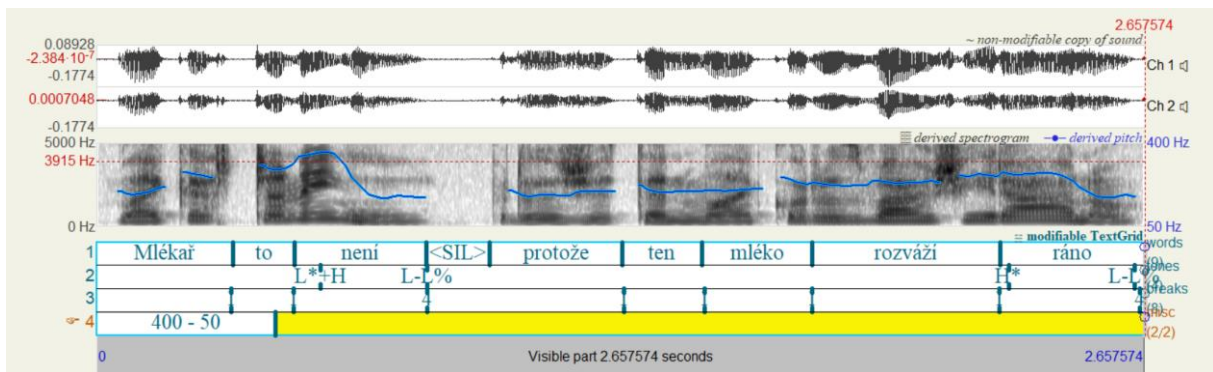

Annotation 9 - Recording 2 - Speaker 5
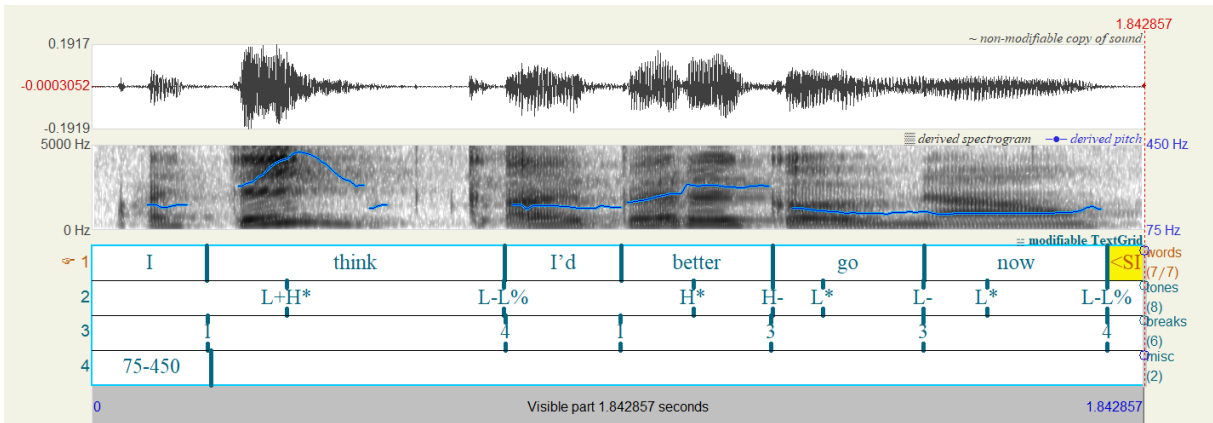
Annotation 10 - Recording 3 - Speaker 2
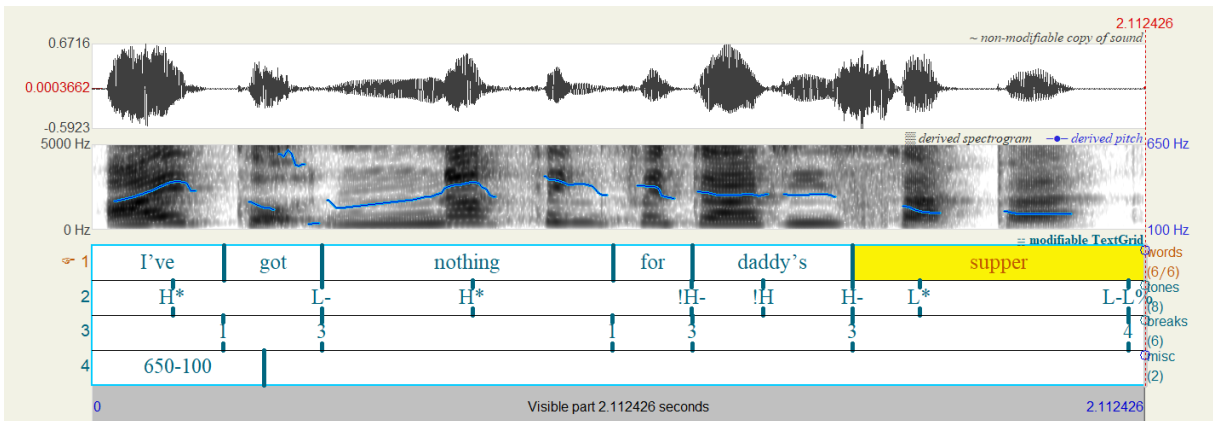
Annotation 11 - Recording 3 - Speaker 3
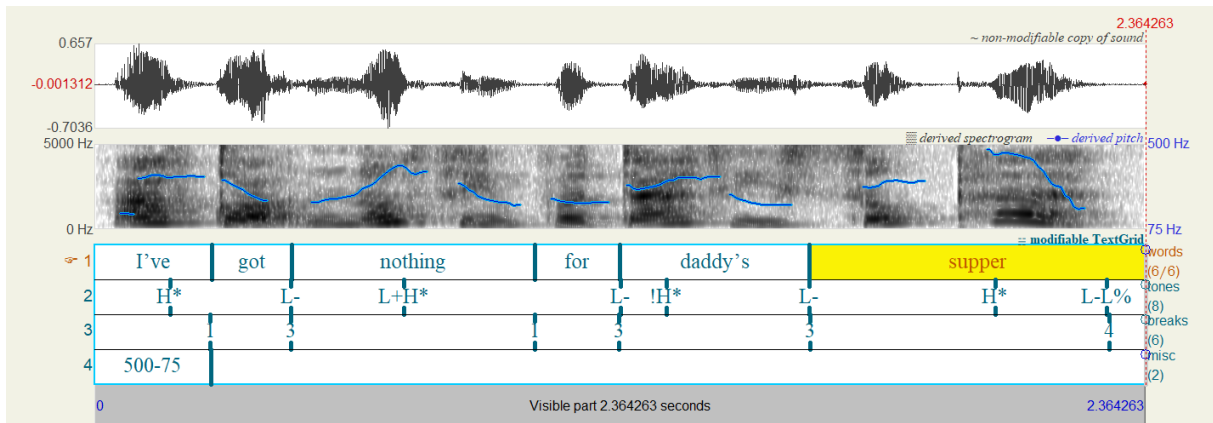
Annotation 12 - Recording 3 - Speaker 4
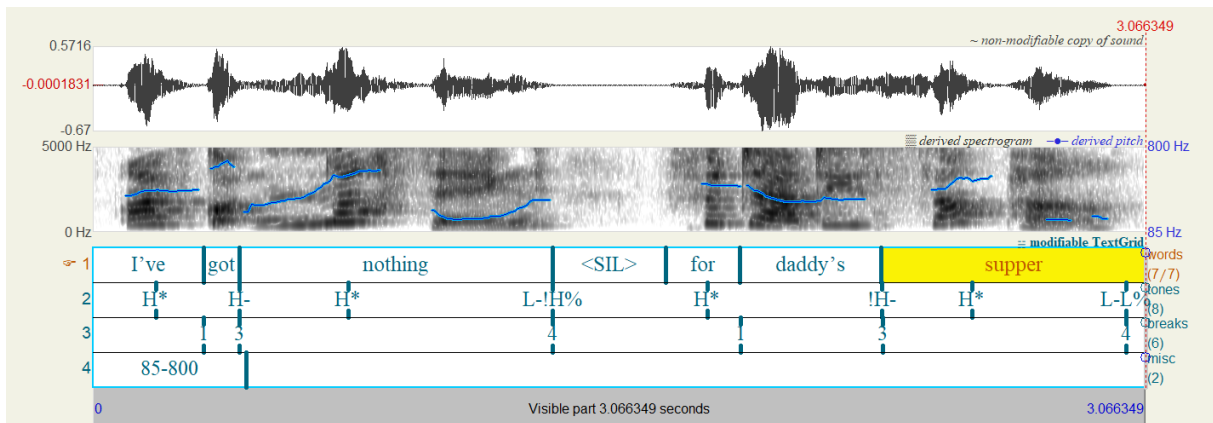
Annotation 13 - Recording 3 - Speaker 5
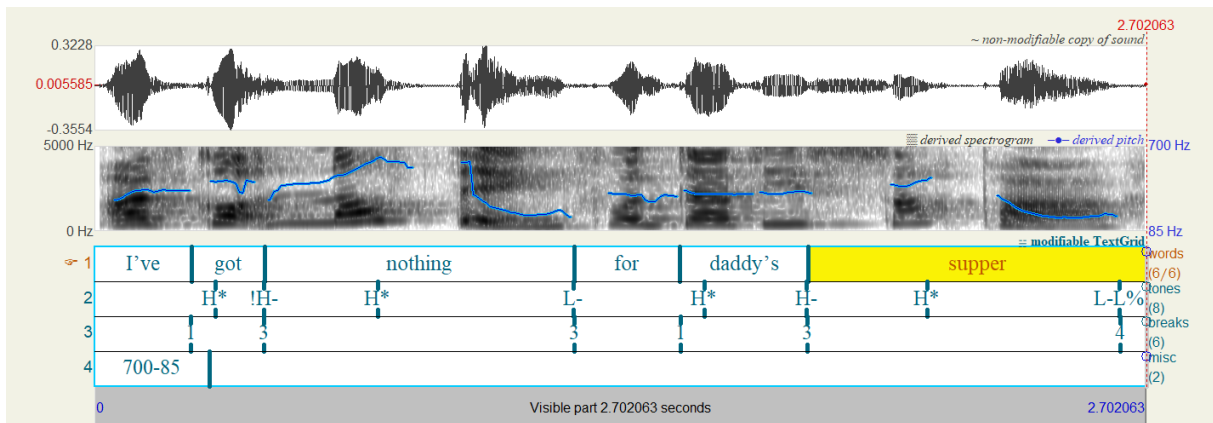
Annotation 14 - Recording 4 - Speaker 1
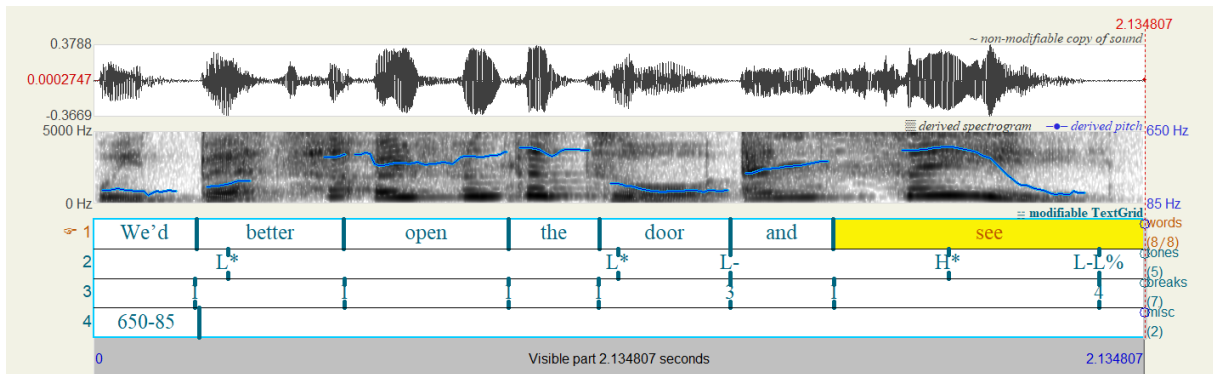
Annotation 15 - Recording 4 - Speaker 2
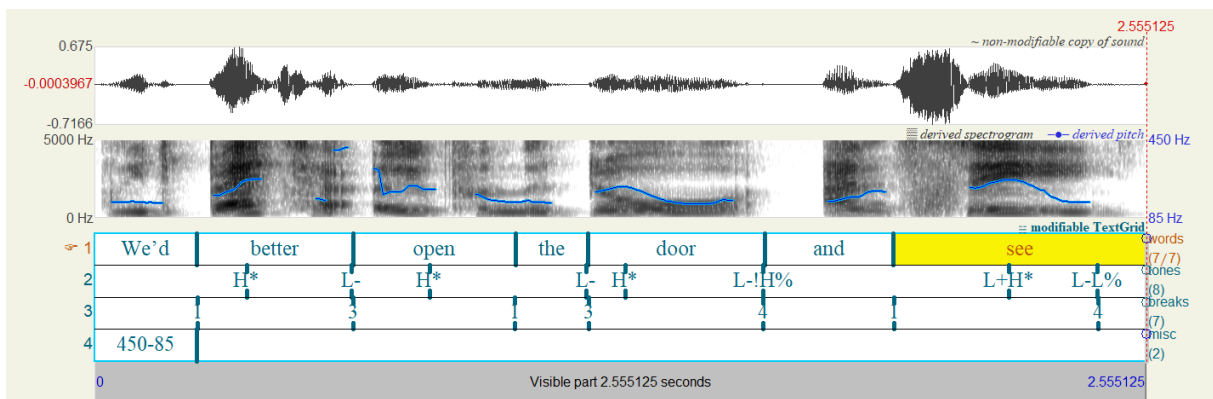
Annotation 16 - Recording 4 - Speaker 4
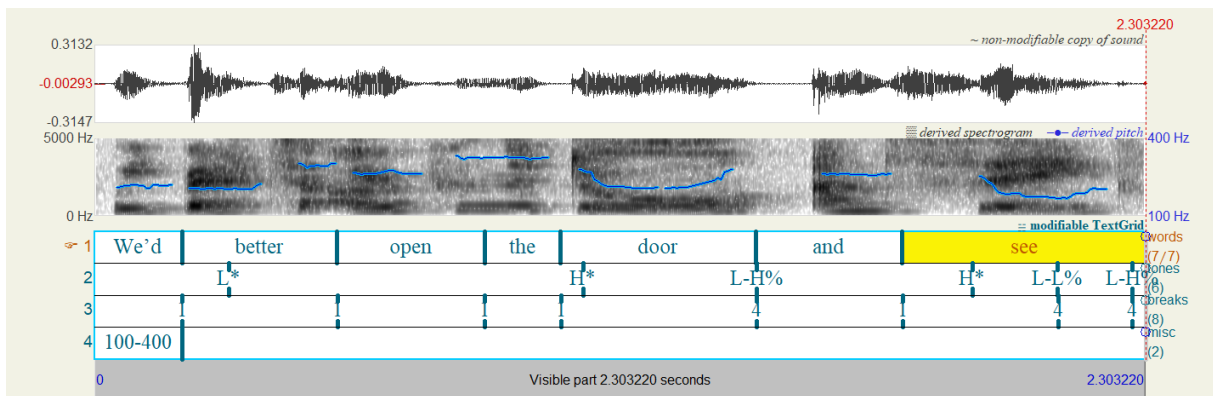
Annotation 17 - Recording 4 - Speaker 5
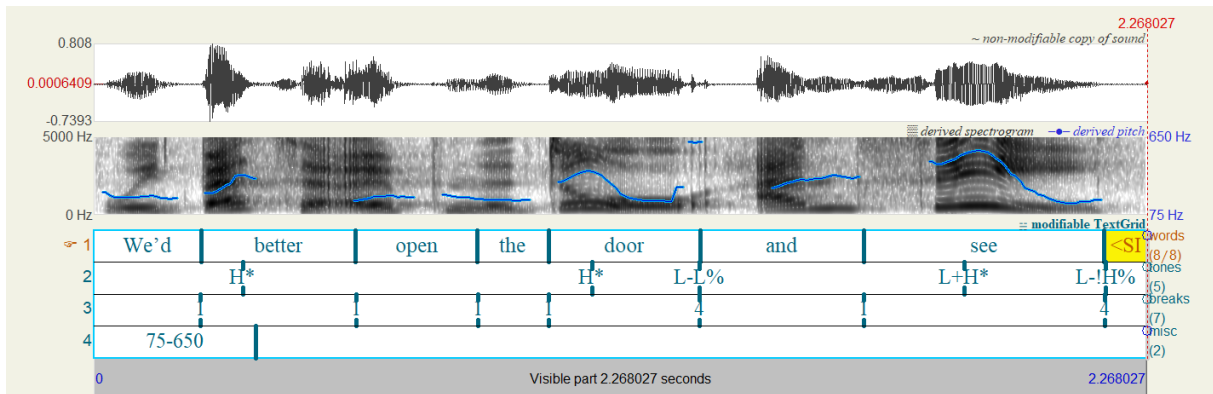
Annotation 18 - CZ Recording 1 - Speaker 1
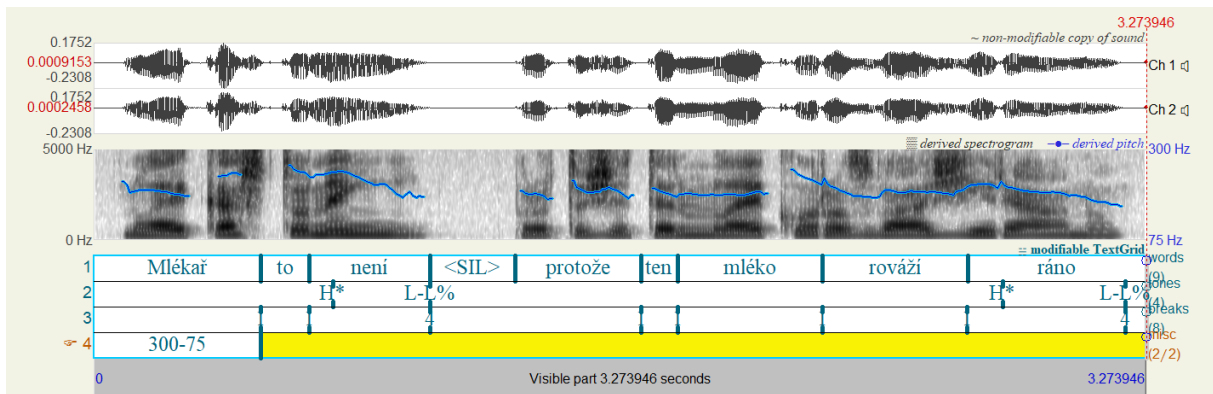
Annotation 19 - CZ Recording 2 - Speaker 1

Annotation 20 - CZ Recording 3 - Speaker 1

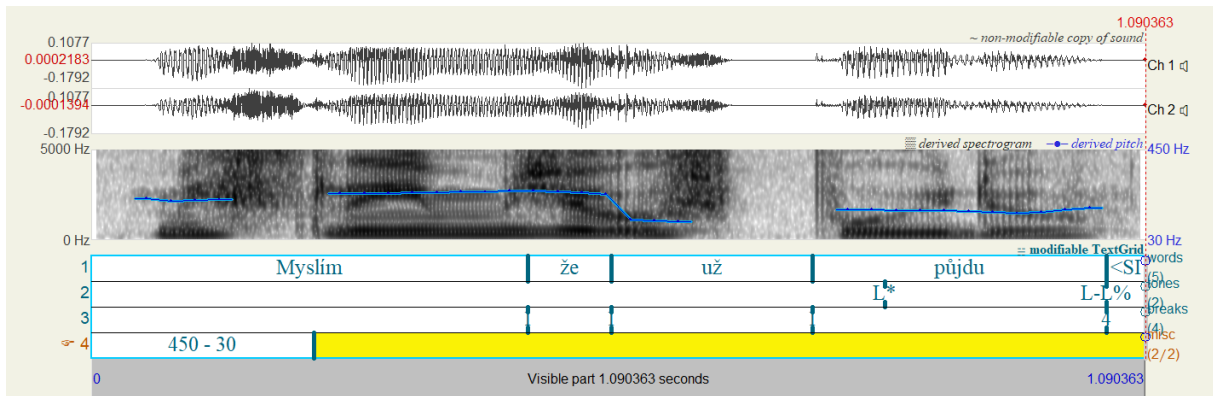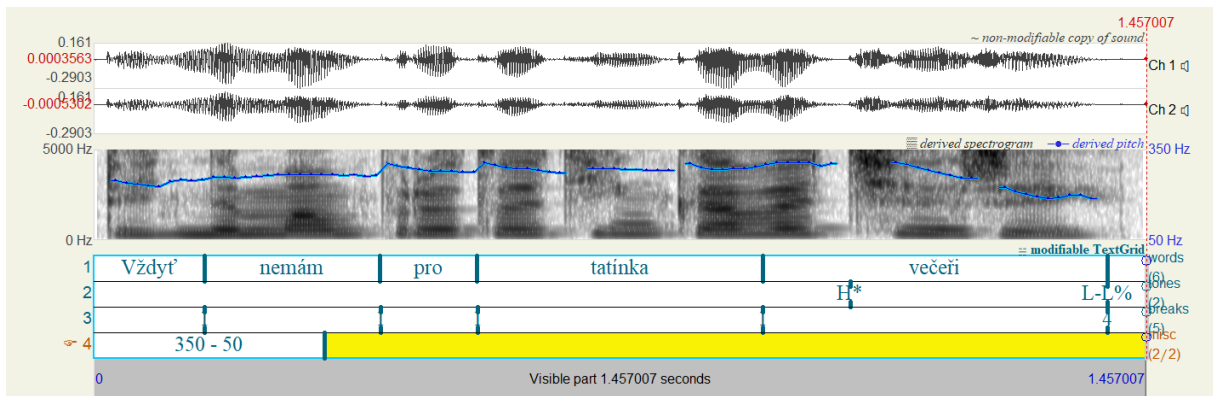## Annotation 21 - CZ Recording 4 - Speaker 1

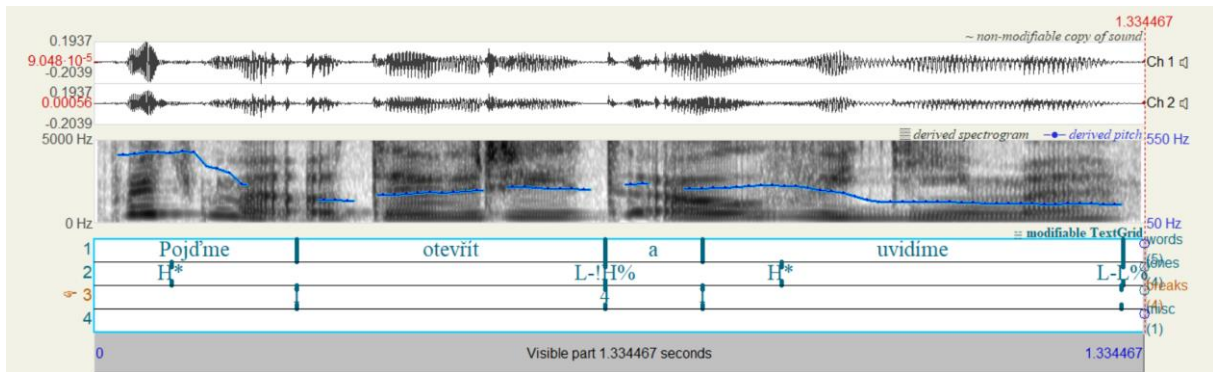## Annotation 22 - CZ Recording 1 - Speaker 2

48

Annotation 23 - CZ Recording 2 - Speaker 2

Annotation 24 - CZ Recording 3 - Speaker 2

Annotation 25 - CZ Recording 4 - Speaker 2

Annotation 26 - CZ Question - Speaker 2