Univerzita Karlova v Praze

Filozofická fakulta

Ústav anglického jazyka a didaktiky

Bakalářská práce

The just noticeable difference for English melodic prominence elicited on Czech listeners

Diference limen pro melodickou prominenci v angličtině zjištěná na českých posluchačích

Kateřina Brabcová

Praha, 2015

Vedoucí práce: Mgr. Pavel Šturm

Acknowledgements

I would like to express my sincere acknowledgements to the supervisor of my thesis, Pavel Šturm, for the willingness in helping me with the choice of a suitable topic and in providing me with invaluable advice whenever needed, both in the process of carrying out the experiment and in the writing itself. Furthermore, his comments on the actual structure of the work cannot be overlooked, and neither can his readiness to assist with the technology-based tasks. Kind words of thanks should be also given to my classmates who agreed to become the listeners of the perceptual test, Mrs Gina Hearn, who provided her voice for the recording, Dr Šaldová, who kindly advised me on some structural details of the work, and to my parents and friends for their constant support and encouragement.

Poděkování

Na tomto místě bych ráda poděkovala vedoucímu své práce, Pavlu Šturmovi, za ochotu pomoci při výběru vhodného tématu a za celou řadu neocenitelných rad, jak při tvorbě experimentu, tak v procesu samotného psaní. Opomenuty nesmí být ani jeho návrhy a připomínky, co se týče struktury práce, a pomoc s více technologicky založenými úlohami. Velký dík patří též mým spolužákům, kteří souhlasili s tím, že se stanou posluchači mého percepčního testu, paní Gině Hearn, jež poskytla svůj hlas pro nahrávání, paní doktorce Šaldové, která byla tak ochotná a poradila mi ohledně některých formálních úprav práce, a samozřejmě mým rodičům a kamarádům za neustálou podporu.

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 11. 5. 2015 ………………………………..

Souhlasím se zapůjčením bakalářské práce ke studijním účelům.
I have no objections to the BA thesis being borrowed and used for study purposes.

ABSTRACT

The aim of the thesis is to ascertain whether and how the modification of pitch in the acoustic signal influences what we hear, i.e. whether we perceive a difference between the original, unmodified stimulus and the one with a modified melodic contour. The theoretical part of the thesis explains terms such as: the fundamental frequency and its contour, the just noticeable difference, intonation and the structure of intonational phrases, and the difference between English and Czech intonational structure. The practical part uses recordings of Standard British English. The contour of the fundamental frequency (F0) is changed at different places of the intonation phrase (the nuclear syllable, stressed and unstressed syllables of the head) and the pitch is either lowered or raised by 1.5 semitones with respect to the original production. The main goal is (1) to determine the influence of the melodic prominence of a syllable on the pitch manipulation detection (judged by 20 listeners of Czech origin, all students of English) and (2) to find out which other factors (such as the already mentioned direction of the change) might be significant. The results are analysed and their statistical significance is evaluated.

Key-words: British English, intonation, fundamental frequency, just noticeable difference, perception

ABSTRAKT

Cílem bakalářské práce je určit, zda a jakým způsobem změna výšky tónu akustického signálu ovlivní to, co slyšíme. Záměrem je tedy zjistit, zdali vnímáme rozdíl mezi původním, nijak nezměněným, podnětem a zvukovým podnětem s upravenou melodickou konturou. V teoretické části práce jsou vysvětleny pojmy jako: základní frekvence a její kontura, diference limen, intonace a struktura intonačních frází a rozdíly mezi českou a anglickou intonační strukturou. Praktická část pracuje s nahrávkami standardní britské angličtiny. Kontura základní frekvence (F0) je změněna na různých místech intonační fráze (melodém, přízvučné a nepřízvučné slabiky tvořící hlavu intonační fráze) a výška tónu je buď snížena, nebo zvýšena o 1,5 půltónu s ohledem na původní verzi nahrávky. Cílem je (1) určit, jakou roli hraje melodická prominence slabiky na vnímání změny výšky tónu a (2) stanovit, jaké další faktory (jako například již zmíněný směr změny) mohou percepci ovlivnit. Odpovědi na tyto otázky jsou formulovány na základě percepčních testů, jichž se zúčastnilo 20 českých posluchačů, kteří byli zároveň studenty angličtiny. Výsledky jsou v práci analyzovány a statisticky vyhodnoceny.

<u>Klíčová slova</u>: Britská angličtina, intonace, základní frekvence, diference limen, percepce

LIST OF TABLES AND FIGURES

TABLE OF CONTENTS

1 INTRODUCTION

The aim of the bachelor thesis is to determine whether, and if so then how, the melodic prominence of different syllables within an intonational phrase influences our perception of the variation in pitch. Our work thus follows previous studies on the just noticeable difference (the amount of pitch change given in semitones necessary to be perceived by listeners) and tries to either confirm or disprove some of the hypotheses suggested before by various authors. Therefore, besides focusing on the melodic prominence only, it also tries to examine if the direction of the pitch change or the type of the intonational movement can influence the process. The material used is all in English, the listeners tested in the experiment are speakers of Czech.

The *Theoretical part* of the thesis (chapter 2) explains what the term *melodic prominence* means and how it is realised; several sections are also dedicated to the pitch perception and the studies on it. The theoretical part is divided into two subchapters, the first of them clarifies the meaning of the terms *the fundamental frequency* (F0) and *the just noticeable difference* (the JND). The most important concept, the fundamental frequency, is commented on right at the beginning (section 2.1.1) as it is directly connected with the second term and also with intonation. Section 2.1.2 focuses on the perception of pitch and sections 2.1.3 to 2.1.6 concentrate more specifically on the JND and the results of the previous studies on it.

The second subchapter of the theoretical part studies the fundamental frequency and its impact on the changes in intonation. The possible definitions of intonation in English are given in section 2.2.1, section 2.2.2 summarises its functions and sections 2.2.3 to 2.2.8 deal with the description of intonation in terms of the intonational tones, the structure of an intonational phrase, the intonation contours, and the transcription of intonation. Section 2.2.4 further specifies the term *stress*, which is closely connected to the study of intonation and pitch perception, and describes *the syllable* as the unit on which stress and the intonational movements are realised. Section 2.2.9 comments on the perception of intonation, and section 2.2.10 serves as a brief summary of intonation in Czech.

Chapter 3, *Method and material*, formulates our main hypothesis and also mentions the relevant results of the previous studies as those will be later compared to ours. The manipulation of the acoustic material is described and the last section is dedicated to the actual experiment and the way it was carried out.

The *Results* (section 4.1) provide several figures either confirming or disproving some of the hypotheses suggested before. They analyse the listeners´ responses mainly with respect to (1) the pitch change perception for the syllables differing in melodic prominence, (2) the direction of the intonational movement on the nucleus, and (3) the change in pitch in the upward versus the downward direction. Additionally, the musical training, as one of the possible factors which might have had an impact on the perception of pitch, is commented on.

In the *Discussion* (section 4.2), we try to analyse the possible reasons for our findings. Our primary hypothesis is confirmed while some of the results appear to be in contrast with the outcomes of the previous experiments performed by various authors on various languages. Again, possible causes are explained, one of these could be the fact that we used an English material while the listeners were all speakers of Czech. Suggestions for some potential future studies are also included.

2 THEORETICAL BACKGROUND

2.1 Fundamental frequency and the just noticeable difference

This chapter will serve as a basic introduction of terms such as *the fundamental frequency* and *the just noticeable difference*. The first part will define the relations between the period, wavelength, and frequency of a sound wave and will continue focusing on the frequencies of the human voice with a special emphasis on the fundamental frequency, marked as F0. The F0 formation together with its different realisation by various speakers will be mentioned as well. The second part will mainly deal with the term *pitch* and its perception, which is dependent on F0 or its multiples known as *harmonics*. The third part will define *the just noticeable difference* understood as the minimal difference between two sounds (given in semitones) which can be perceived by listeners. The use of a new unit – semitones (ST) – and its relation to hertz (Hz) will also be explained. Lastly, this chapter will show the difference between pitch perception of a pure tone, two separated tones, and complex sounds in speech.

2.1.1 Fundamental frequency

Any kind of sound can be defined as a series of fluctuations in air pressure (Reetz and Jongman, 2009: 120). Such fluctuations can be measured by an oscillograph, a device which records periodic motion about a mean value, and analysed in the form of sound waves recorded in an oscillogram or stored on computer. In order to describe a sound wave, we have to define its period, wavelength, and frequency.

A sound wave is called periodic when it comprises a certain number of events which occur repeatedly and at equal intervals. A *period* of a sound wave therefore marks the interval between any two points in the same phase position within a wave in the oscillogram. The word *frequency* indicates the rate of occurrence of events, previously defined as periods, per unit of time. The unit that is used to measure frequency is known as hertz (Hz) and it expresses the number of periods per second. As Reetz and Jongman remark: “An oscillation has a frequency of one Hz if something swings back and forth once a second.” (Reetz and Jongman, 2009: 116).

The last criterion needed for the description of a sound wave is the distance between its two phase points, which is called a *wavelength*. A wavelength can be easily calculated once we know the period and the speed of sound (340 m/s) by the use of the following formula: $\lambda = c \cdot T$, where c expresses the speed of sound, T the period, and λ is a symbol used for wavelength measured in meters (Reetz and Jongman, 2009: 117).

Human speech, however, does not have a form of a simple sound wave. Voice is a complex of more signals, each having its own frequency. The period of such signals is called *a complex period*, which cannot be transcribed as a single contour but rather as having a complex shape (Reetz and Jongman, 2009: 119). The lowest periodic frequency of human speech is known as the fundamental frequency (F0), which is produced by vibrations of the air coming out of the lungs. These vibrations are caused by the vocal folds located in the larynx, that periodically close and open again. The other frequencies which comprise the complex signal are referred to as *harmonics*; they are multiples of F0 (´t Hart, Collier and Cohen, 1990: 18).

Fundamental frequency and its harmonics are perceived as *pitch*; the faster the vocal folds vibrate, the higher the pitch (Reetz and Jongman, 2009: 119). The range of F0 frequencies differs for every single speaker as it is dependent mainly on the mass and tension of the vocal folds, their elasticity, and the amount of pressure below the glottis (´t Hart et al., 1990: 12). Thus, if the tension on the vocal folds increases, the fundamental frequency becomes higher; if it decreases, the frequency is lowered. For a male voice it usually fluctuates between 80 to 200 Hz, for a female voice it varies between 180 and 400 Hz (Noteboom, 1999: 2). The average voice range of an adult is said to be two octaves, professional singers can increase it to five (Palková, 1994: 58). Speakers are partly able to control some of their pitch movements but the rest are side-effects of various speech processes. One of these is known under the term *declination* which marks the universal tendency of F0 to become gradually lower in a phrase pronounced without a pause. This phenomenon is said to be mainly caused by the decrease of air pressure below the glottis (Ladd, 2008: 1).

During the pronunciation of voiceless consonants, the vocals folds do not produce any vibrations but such events are not interpreted as perceptual pauses in the pitch. As listeners, we still recognise speech and the changes of its pitch as continuous, without interruptions; these would be perceived only if the silent intervals of voiceless consonants were longer than approximately 200 ms (Noteboom, 1999: 3).

2.1.2 Pitch perception

´t Hart and his colleagues (1990) introduce a general concept of a central processor, located somewhere in our brain, which, according to their own words, “tries to detect the greatest common divisor of the frequencies of a number of harmonics” (´t Hart et al., 1990: 25). The most effective ones are said to be the third up to the sixth harmonic that are the multiples of the fundamental frequency, which does not have to be present itself. The pitch perception is thus dependent either on F0, its multiples, or the combination of them both. For the perception of pitch, two harmonics are needed for the calculation of the greatest common divisor. F0 has to lie in the interval between 40 Hz and 4 kHz (as it does in the case of the human voice) and pitch can be perceived even in very short stimuli of only 30 ms (´t Hart et al., 1990: 26).

To measure the frequency of vibration of the vocal folds, computer integrated extractors of F0 are used. These enable phoneticians to analyse the F0 contour and to study the changes in pitch (Volín, 2009: 228).

2.1.3 The just noticeable difference

In this section a new term, *the just noticeable difference* (JND), needs to be introduced. Since listeners, with the exception of those in possession of the absolute pitch, are generally unable to recognise the exact pitch of a tone, researchers usually focus on the frequency difference between two successive ones. The term *the just noticeable difference*, or *the differential threshold*, is used to mark such relative minimal difference (´t Hart et al., 1990: 27). Some authors, for example Harris and Umeda (1987), also use the term *difference limens,* in which case *limen* is treated synonymously with *threshold* which is defined to be “the level or strength of a stimulus which begins to produce a physiological or psychological effect” (Longman Dictionary of the English Language, 1991, p. 1565).

As the emphasis is given more on the difference between two frequencies than their absolute values, a new unit of measurement, semitones (ST) instead of hertz, is used for such comparison. A very useful example is provided by Reetz and Jongman (2009) explaining the relationships between Hz and ST. They state that “[a] 200 Hz tone is perceived as being twice as high as a 100 Hz tone” (Reetz and Jongman, 2009: 242). This difference is in musical terms known as *an octave* which is divided into twelve semitones. If we want to increase the pitch again to be an octave higher, we nevertheless do not change it to 300 Hz but double it to

400 Hz. Instead of differences, we use proportions. This means that “the perceptual distance between two octaves is always the same but the physical distance is not” (Reetz and Jongman, 2009: 242).

The use of semitones instead of hertz enables us to compare F0 contours from different speakers and their different voices (´t Hart et al., 1990: 24). Semitones express distance D between two frequencies marked as F1 and F2, which can be calculated by the use of the following formula on the logarithmic semitone scale (Reetz and Jongman, 2009: 242): $D = 12log_2\left(\frac{F1}{F2}\right)$, or its more convenient variant: $D = 39.863\ log_{10}\left(\frac{F1}{F2}\right)$.

2.1.4 Perception of the JND between two tones and in speech

The JND is generally believed to be lower at frequencies below 1,000 Hz, or frequencies higher than 10 kHz. Palková (1994) in her work states that for the stimuli with the frequency of 50 Hz, there is a need of 60 ms for the pitch to be perceived. In case of 200 Hz, it takes about 25 ms. The signals are perceived fastest in the range of 500 and 5,000 Hz; the time needed is said to be 15 ms. It takes 25 ms again when the stimuli reach 10 kHz (Palková, 1994: 100). Some authors claim that the minimum of 16 periods of a sound are needed to be present. In the case of complex signals, the ability to recognise the difference between two stimuli also decreases if they are shorter than 30 ms regardless of their frequency (´t Hart et al., 1990: 28).

One of the experiments concerning duration of a stimulus and the perception of the pitch change was performed by Sergeant and Harris (1962; cited in ´t Hart et al., 1990: 29–30). They did not focus on a difference between two pronounced syllables but rather on the absolute threshold of the pitch change within one record. Their study proved that if a recording played back to listeners had the duration of 75 ms and the initial pitch was 1,500 Hz, the change of 30 Hz was needed in order to elicit a different pitch perception. If the duration increased, both the recognition speed and the amount of necessary frequency change decreased. It was concluded that in the case of long stimuli, the sensitivity to variation in pitch seems to be pretty high (usually about 1 Hz/s at the duration of 10 s); when the stimulus lasted only a few milliseconds, the ability to perceive a change decreased (the difference of at least 150 Hz/s at 100 ms is needed). However, the stimulus cannot be too long since it might be our memory which may cause some problems then.

Only very few people, usually musically trained, are able to recognise the pitch of a single pure tone with the accuracy of 0.25 to 0.125 of a semitone. The majority usually proclaim the same tone to be 5 to 9 ST higher or lower (´t Hart, 1981: 811). In the case of two successively presented pure tones, the best listeners can tell the just noticeable difference of about a half of a semitone (Plomp, Wagenaar and Mimpen (1973) cited in ´t Hart et al., 1990: 28). Nevertheless, even people who are unable to determine the frequency of one single tone are much better at distinguishing two different ones from each other (´t Hart, 1981: 811). However, the perception of musical intervals is not relevant for the identification of pitch in speech (´t Hart et al., 1990: 28).

There are many difficulties in the perception of the human voice in real speech. First, we do not concentrate on the pitch only but we also pay attention to the linguistic information presented. Second, speech perception does not comprise just the perception of F0 but also involves formant frequencies and amplitude which all influence what we hear. As ´t Hart and his colleagues (1990) explain: "In our experience, a drop in amplitude of 10 to 20 dB in several tens of milliseconds, as can often be found in transitions from vowels to consonants, can entirely obscure the presence of changes in F0 of up to half an octave." (´t Hart et al., 1990: 36). Palková (1994) further states that with the increasing intensity of a sound our ability of perception increases till the sound reaches 60 dB; from there on, it remains the same (Palková, 1994: 100). Additionally, the clarity of periodicity in the signal may also influence the accuracy of the pitch perception (Nooteboom, 1999: 3). It includes factors such as the absence of periodicity in silent gaps or voiceless consonants, or different quality of voice, which can be too hoarse or breathy. And finally, our perception of the pitch in whole utterances differs from the pitch perception of isolated speech sounds (´t Hart et al., 1990: 36). It is therefore evident that the JND of neither pure tones and complex tones nor isolated tones and sounds in speech can be compared.

2.1.5 Measurement of the JND

There are several methods used to determine the JND perceived by listeners. One of these called "a three-tone paradigm" was used by Stevens and Volkman (1940; cited in ´t Hart et al., 1990: 28). In this experiment, the participants were supposed to adjust a tone between two provided ones so that its perceptual difference from each was equal. Burns (1974) provided his subjects with a set of four tones and asked them if they perceive the difference between tones 1 and 2 to be of the same distance as the difference between tones 3 and 4 (cited in ´t Hart et al., 1990: 28–29). Another method is known as "pair comparison" which asks the listeners to compare two pitch changes in an utterance and to determine whether they are of the same distance or not (´t Hart (1981) cited in ´t Hart et al., 1990: 29).

2.1.6 Previous studies of the JND

There were several experiments trying to determine the JND both in pure tones and in speech. Flanagan and Saslow published their study on the JND for F0 of synthetic vowel stimuli with steady F0 in 1958. Their results, unfortunately given only in hertz and not in semitones, suggested that there was the need for a 0.32-Hz change so that the listeners would perceive the manipulation. This value was calculated as the mean of different vowel qualities elicited on six highly trained listeners (Flanagan and Saslow (1958) cited in Harris and Umeda, 1987: 1139). In 1973, Klatt used the synthetic vowel /ɛ/ in his experiment, again trying to determine the JND. He worked with a steady F0, ramp F0, and F0 with a steep rate of change. His results were as follows: for the steady F0, the JND of 0.3 Hz was discovered; for the ramp F0, the change of at least 2 Hz was needed; and for the steep rate of change of F0, the JND proved to be 4 Hz. The last result of the three was considered the most important as it is the closest to real speech (Klatt (1973) cited in Harris and Umeda, 1987: 1139).

´t Hart in his experiment from 1981 wanted his listeners to compare two different pitch movements and to judge which of the two contained a larger movement. In his experiment, he worked with six different four-syllable Dutch number names and changed F0 in the range from 1 to 6 ST. The initial frequencies were 115, 135 or 160 Hz (´t Hart, 1981: 812–813). His results proved the mean overall JND to be approximately 2 ST (´t Hart, 1981: 818-820). In a different study of his, he also claims that the JND between two syllables in speech appears to be 1.5 to 2 semitones, and only the difference of 3 semitones is expected to be recognised in proper communication (´t Hart et al., 1990: 29).

Harris and Umeda (1987) performed four various experiments always using the method of pair comparison. In the first two studies, they worked with five pairs of sentences of about 2,000 ms read by a male speaker. They always changed F0 in a part, usually of the duration of 625 to 825 ms, of one of the sentences in the pair, and the JND was said to be 5 to 16 Hz or even higher (Harris and Umeda, 1987: 1139-1142). The JND in semitones is not directly given in the text but if we consider that the average F0 of the male speaker used (called MH) was 126 Hz (as stated by the authors), then we know that the difference limens of 5 Hz equals approximately 0.6 ST, and the difference limens of 16 Hz equals approximately 2.1 ST. The second of the first two experiments was based on a F0 change where the end and the beginning of the manipulation occurred on a stop consonant because it was suggested that voicing in the previous study might have provided an additional clue for the listeners to perceive the change. This time, only four pairs of sentences of 2 to 3 s in duration were used, read by two different speakers. The JND proved 20 times higher than that for the steady synthetic vowels of Flanagan and Saslow (1958) (that means about 6.4 Hz) and two to four times larger than the difference limens for steady synthetic vowels with descending F0 that Klatt (1973) worked with. The possible reason for this was claimed to be the fact that the listeners were not told where in the sentence to compare the F0 (Harris and Umeda, 1987: 1141–1143). The other two experiments by Harris and Umeda used shorter sentences and changed the whole F0 contour. Again, they compared the original and the manipulated version. This time, the JND reached 4.8 Hz (Harris and Umeda, 1987: 1143). Again, provided that we know that the speaker´s average F0 was 126 Hz, the JND can be converted to semitones and will equal 0.6 ST.

Micheyl and his colleagues (2006) in their work primarily investigated the influence of musical training of the listeners and their experience with psychoacoustic testing on the JND. They used pure and complex tones of 200 ms in duration and the initial frequency of 330 Hz. These were compared with tones the frequency of which was changed to be higher. As listeners, a group of thirty musicians playing an instrument for more than ten years, and a group of non-musicians with no musical training were used. The JND of these two was measured, again the technique of pair comparison was used, and it proved to be six times smaller when judged by the musicians compared to the musically untrained people. This means that for the musicians, a change of 0.13% of the original pitch height was required, for the non-musicians the manipulation needed equalled 0.86%.

The last study which should be only briefly mentioned was done by Vongpaisal and Pichora-Fuller (2007) who compared the ability of young (approximately 26 years old) and old (74 years old) listeners to perceive the changes in pitch. These were to listen to the pitch changes of the vowel /a/ of 120–150 Hz. For the young participants, the JND (when converted to semitones) equalled 0.09 ST (0.6 Hz) and for the older ones, it reached 0.26 ST (1.8 Hz). This means that the differential threshold of the young subjects was 3 times smaller than that of the older adults.

2.2 Intonation

This chapter starts with a basic introduction of intonation and its connection to F0. Various definitions of intonation together with its functions will be mentioned first, followed by a basic description of the five tones used in English. Stress placement and its realisation on syllables will be also commented on. The structure of an intonational phrase with its most important component, the nucleus, will be discussed before we introduce the two major approaches towards an intonation contour – *the overlay* and *the linear approach*. These will be connected with various types of symbols which can be used for the transcription of intonation; and lastly, the perception of intonation and the difference between British English and Czech intonation will be dealt with.

2.2.1 Definition of intonation

Intonation is said to be "one of the most universal and one of the most language specific features of human language" (Hirst and Di Cristo, 1998: 1). There is not a single language that does not use any intonational features.

Studying intonation means dealing with prosodic or the so-called suprasegmental aspects of speech. Very often, it might be only prosody according to which different languages can be distinguished one from another. Prosodic features are also said to be learnt first during native language acquisition, and are last to be lost when learning a new language (Hirst and Di Cristo, 1998: 2).

Intonation had not been given much attention before 1970s. There were two strictly divided approaches – *the instrumental* and *the impressionistic*. The instrumental approach focused mainly on speech perception and dealt with acoustic parameters. The impressionistic treated intonation as a small number of categorically distinct elements. Modern approaches usually try to combine both (Ladd, 2008: 10–11). Each author uses a slightly different definition of intonation. As Roach (2000: 150) remarks: "No definition is completely satisfactory, but any attempt at a definition must recognise that the pitch of the voice plays the most important part." Ladd (2008: 4) explains that intonation "refers to the use of *suprasegmental* phonetic features to convey ‘postlexical’ or *sentence-level* pragmatic meanings in a *linguistically structured way*.” He puts emphasis on the three words which are italicised and comments on them in the following way. *Suprasegmental* are the features of prosody which include changes in F0, intensity, and duration. *Sentence-level* characteristic signifies the property of intonation to convey meanings that are connected with phrases and utterances as a whole, and the fact that this is done in a *linguistically structured way* means that intonational features are somehow organised (Ladd, 2008: 6).

A slightly better definition is provided by Hirst and Di Cristo (1998). They clarify that intonation can be understood in two ways – in broad, or lexical sense, and in narrow, or the so-called non-lexical sense. In the broad sense, intonation is often synonymous with prosody as it includes features such are word-stress, tone, and quantity. In the narrow sense, we are dealing with the “intonation proper” which is defined as the changes in pitch caused by the variations in F0 which happen within a sentence or its part. This variation, which could be also referred to as *melody*, is dependent on vocal folds vibration, which can be adjusted by the change of their length and tension controlled by muscles. The resulting F0 is also influenced secondarily, mainly by the pressure below the larynx or by the identity of the speech sound. Apart from F0, some other parameters which may also be involved in the “intonation proper” are intensity, duration, and rhythm (Hirst and Di Cristo, 1998: 4–6).

2.2.2 Functions of intonation

As has already been mentioned, languages may differ considerably with respect to their prosodic features. Additionally, languages are divided into two major groups: *tone languages* and *intonation languages*. In the case of tone languages, a specific type of intonation (tone) is used in connection with particular words or morphemes: if we alter the pitch, we can change the meaning of the word since the tone is treated as a feature of the lexicon (Cruttenden, 1997: 8). On the other hand, pitch variations in intonation languages such as English are not bound to concrete words. If we realise it in a different way, the lexical meaning of the word will remain the same; the major function of intonation in intonation languages is thus signalling the degree of melodic prominence of a certain word or syllable.

Besides the lexical function and the function of giving a melodic prominence to a syllable, intonation can have various other functions (Wells, 2009: 11–12). These are: (1) attitudinal function when intonation is used to express specific emotions or affective states, (2) grammatical function when intonation serves for identification of grammatical structures, (3) accentual function when it establishes focus on a part of an utterance, (4) pragmatic function when specific type of intonation is used to emphasise something, (5) discourse function which explains relations between clauses, (6) psychological function which helps to organise our speech so that it can be easily memorised, and finally (7) indexical function which marks the identity of the speaker.

2.2.3 Tones

Intonation is mainly dealt with in terms of *intonational phrases* (sometimes referred to as *intonational groups*) into which speech can be divided. This division more or less corresponds to punctuation in writing (Wells, 2009: 189), but not always. Each phrase carries one piece of information and each has its own intonation contour which represents a sequence of tones. An intonational phrase usually lasts 1–2 seconds but in conversation, there are even shorter ones (Wells, 2009: 191–192). *Tone*, here used in connection with intonation (and not tone languages), marks a specific intonation pattern – the differences in pitch. The basic division of a tone is into high (H) and low (L) levels (Wells, 2009: 3). When a speaker uses a high tone, it means that his or her vocal folds vibrate rapidly, whereas in the case of a low tone, the rate of vibration is slow.

Tones can be further divided into static and kinetic (Roach, 1990: 114). The static tone is known as *the level tone* and it marks an intonation pattern caused by the vocal folds vibrations which are kept at a constant rate. The vibrations used to produce a kinetic tone change their speed and the pitch can then be either falling or rising (Wells, 2009: 3). Five basic intonation patterns in English are thus said to be: *the level tone, the fall, the rise, the rise-fall* and *the fall-rise* (Roach, 1990: 114). More complex patterns are possible.

The fall in English usually suggests finality of a sentence or utterance and is mainly used in statements, exclamations, some types of questions, and commands. The pitch of voice starts high and then moves downwards. In *the rise*, voice pitch begins relatively low and moves upwards. It is the typical intonation pattern used in yes/no questions. The pitch of *fall-rises* first starts high, moves downwards and then upwards again. Unlike the fall, the fall-rise suggests non-finality of the speaker´s utterance which means that something else might be implied (Wells, 2009: 27). *Rise-falls* are used to indicate completeness and can be found in declarative sentences, mainly in connection with sarcastic comments, or utterances expressing impression or strong disapproval (Cruttenden, 1997: 92–3).

2.2.4 The syllable and its prominence

Before we introduce the most prominent part of an intonational phrase – *the nucleus* – the term *stress* and its importance for the study of intonation need to be discussed. Nevertheless, stress can be explained only in connection with the units which carry it, the syllables, and these should be dealt with first. A *syllable*, defined by Palková (1994), is the fundamental unit of connected speech and the smallest unit of speech production and perception (Palková, 1994: 152). Its parts are called *onset* and *rhyme* where rhyme is further divided into *nucleus* and *coda* (note that the term *nucleus* here is not synonymous with the same term used for describing a part of an intonational phrase). The nucleus, the only obligatory part of a syllable, carries the prosodic features of a language and the highest degree of sonority, which means that it appears louder compared to the onset and coda. It is therefore the nucleus on which the most prominent pitch movement can be performed. In English, most syllabic nuclei are realized by vowels where the open vowels proved to be the most sonorous ones. A consonant (an approximant, a nasal or a liquid) can also appear in the position of a nucleus of a syllable but always attached to another syllable formed by a vowel in its prosodic centre. There are around twenty types of syllables in English with a varied number of consonants in the onset and coda forming the so-called *consonant clusters* (Gut, 2009: 77). Syllables can be also divided into different subgroups. One of the divisions is into open and closed syllables. Open syllables lack a consonant in the coda whereas closed syllables have one or more consonants in that position. Another way of categorising syllables is into heavy and light. Rhymes of heavy syllables consist of a long vowel or a diphthong, or possibly a short vowel with a coda consonant. Light syllables comprise a short vowel or a syllabic consonant in the rhyme.

Now, to return to the term *stress* mentioned at the beginning of the previous paragraph, it should be noted that stress can be explained both in terms of speech production and perception. It is realised (or not realised) on syllables which can therefore be either stressed or unstressed. From the point of view of production, a stressed syllable is created by greater air pressure caused by the increased effort on the muscles of our respiratory system, and a higher pitch produced by tensioning the vocal folds. This means that the variations in pitch (therefore the change in intonation) and stress are directly connected. The articulatory movements are also performed to a larger extent if the syllable is stressed (Gut, 2009: 83). Such a syllable thus appears to be of a higher degree of prominence compared to those which are unstressed.

In English, the stressed syllable is indicated by variation in four acoustic cues. These are: the fundamental frequency, intensity, duration and the vowel quality. Their perceptual correlates are then: pitch, loudness, length and a different timbre of the vowel (Frost, 2011: 68). Phoneticians and phonologists have been trying to determine which of the four is the most important factor in stress production and perception and although not all of them agree, they usually conclude that variations in F0 (and therefore the changes in intonation) play the most important role (Frost, 2011: 68). Furthermore, F0, formant structure and duration seem to influence the stress production and perception more than intensity (Frost, 2011: 71).

Stress is often confused with another term – *accent*. While stress is only an abstract property of a syllable, accent can be measured, which means that the syllable needs to be said aloud. As Ulrike Gut explains, “a stressed syllable is the one a speaker knows to be stressed, whereas accents can be measured in real speech” (Gut, 2009: 84). Not every syllable in English can be stressed. Those which are defined to be light are always unstressed while stressed syllables are always heavy (Gut, 2009: 77).

The interplay of stressed and unstressed syllables contributes to the speech rhythm (Gut, 2009: 87). The unit containing one stressed syllable and a various number of the unstressed ones is called a *stress group*. It starts with a stressed syllable and spreads over all the following unstressed ones until it reaches another stressed syllable which is then excluded. It is also a unit which is smaller than an intonational phrase; one intonational phrase can comprise of more stress groups.

The last distinction which has to be clarified is that between the *word stress* and the *sentence stress*. Every lexical word in English contains a stressed syllable. The word stress in English is described as free as it is not always the first, the second or the last syllable which is stressed but the placement varies depending on the word itself. If a word contains more than two syllables, the most prominent one carries the primary stress and then the secondary and even tertiary stress follow in the degree of prominence. A similar labelling system can be, as an analogy, applied on the syllables which create an intonational phrase (Chamonikolasová, 2007: 13). Here, the primary or the so-called *sentence stress* (also known as *pitch accent*) is carried by the unit known as the *nucleus,* which will be described in the next section.

2.2.5 The nuclear tone

The previous section has dealt predominantly with stress and its placement. The term *sentence stress* referred to as *the nucleus* was introduced as the most prominent part of an intonational phrase. It is the syllable which carries the most prominent pitch movement and it is the place where the realisation of *the nuclear tone,* the most prominent tone in a phrase, begins (Wells, 2009: 7). The nuclear tone then stretches over the other parts of the intonational phrase following the nucleus up to the very end of a phrase (Cruttenden, 1997: 50). It is not always easy to recognise such a syllable but there are several pieces of advice which might help.

The nucleus goes on a stressed syllable. This syllable carries the pitch movement, is usually longer in duration, and carries a rhythmic beat (Wells, 2009: 93). In each intonational phrase, the nucleus bears the last stress. Additionally, the nucleus is usually to be found towards the end of an intonational phrase and in most cases goes on lexical (content) words (Wells, 2009: 95–97). Furthermore, the nucleus occurs in words which present a new piece of information as opposed to what has already been mentioned (Wells, 2009: 109).

A few examples should be added to support the short explanation; the syllable which carries the nucleus is underlined.

1) "This is my cat." The word "cat" carries the nucleus since it bears the most important information. It is a lexical word and occurs at the end of the intonational phrase.

2) "I will eat it." The verb "eat" in the previous sentence is where the nucleus appears. Again, it is a content word which is the most prominent one in the sentence. Personal pronoun "it" is a grammatical word which is not accented.

3) "Would you like some sausages and eggs?" – "No, I´d like some bacon and eggs."

In the short conversation which serves as example 3, the nucleus is to be found in the word "bacon" in the second intonational phrase. It is not the word "eggs" which was already mentioned in the previous sentence as it only repeats information which is already known.

4) "Who gave you the orange T-shirt?" – "Mum gave me the orange T-shirt."

In this case, the noun “mum” carries the nucleus as it is the most important part of the sentence. The rest is only repeated.

2.2.6 The structure of an intonational phrase

The structure of an intonational phrase usually follows the grammatical structure. This means that speech is divided into chunks based on the syntactic structure (Wells, 2009: 187). The boundary can be found between sentences, clauses or parts of a clause. The example will be: “He <u>came</u> │and she <u>left</u>.” (The symbol│ marks the division of the two phrases which create two different clauses.)

As was already said, the most important part of an intonational phrase is the nucleus, the syllable where the realisation of the nuclear tone begins. Let us now examine the other parts of the intonational phrase.

The first accented syllable in the part of an intonational phrase before the nucleus is called *onset*. The word *head* is used to mark all the syllables which extend from the onset to the last syllable before the nucleus. If there are any unstressed syllables before the onset, they build the so-called *prehead*. An example will be a sentence: “Some ˈpeople would ˈlike to ˈ<u>see</u> you.” The symbol (ˈ) marks each stress in the sentence. The nucleus is realised on the word “see.” This is where the falling nuclear tone starts. The head begins with the word “people” and stretches up to the nucleus, and the prehead is created by the word “some.”

There can be some pitch patterns already recognised in heads and preheads. Both preheads and heads can be realised as high or low; and in the case of heads, we can also distinguish some pitch movement. It can be either high falling, or low rising tone (Wells, 2009: 209). The part directly following the nucleus is known as *the tail*. If there is a tail present in an intonational phrase, the pitch movement of the tone is not completed on the nucleus but stretches till the end of the phrase (Roach, 1990: 125). If there is a rising tone beginning on the nucleus, the pitch of the tail moves upwards. In the case of a falling tone, it continues falling in the tail (if not at the bottom already). If it is a complex tone, for example a fall-rise, the first pitch movement is usually realised on the nucleus, the second on the tail (Roach, 1990: 126–128). In our example “Some ˈpeople would ˈlike to ˈ<u>see</u> you,” the tail is thus realised by the word “you” and the pitch of it will be constantly moving downwards.

2.2.7 Approaches to intonation contours

Intonation is usually treated in terms of *intonation contours*. These are defined as sequences of tones in a segmental string (Ladd, 2000: 37). There are two approaches towards these contours which differ very much. The first is called *the overlay approach*, which treats an intonation contour as a series of domains, each having its characteristic pitch features. Every domain consists of various phonological elements creating a so-called *tier*. These tiers can spread over words, phrases, sentences, or even paragraphs and what is typical for them is that one can be superimposed on another of a different size (Ladd, 2000: 39). According to Ladd (1995) there are several problems of this approach. First, each domain does not have to necessarily have its own slope or shape of the pitch, and contrarily, some pitch features may extend over the boundaries of one domain. Another deficiency of this approach is its inability to describe the so called "hat pattern" which is defined as "a rising pitch movement on one accented word, followed by a high level stretch, followed by a falling pitch movement on the next accented word" (Ladd, 1995: 2). The problem of this pattern is that it contains two accented words, meaning two domains, but only one pitch pattern is used to describe its shape. Ladd (1995) therefore prefers another kind of approach known as *the linear (*or *levels) view* which treats an intonation contour as one single string of individual phonological events, not as domains. This approach is based on Pierrehumbert´s and Bruce´s theories which describe an intonation contour as "linguistic events in a sequence that happen at fairly well-defined points in the utterance" (Ladd, 2000: 38). The third possibility is the *metrical/hierarchical approach* which combines features of both of the preceding ones. It accepts the existence of domains which spread over phrases, sentences, or even paragraphs but in the same time argues that intonation should be analysed in terms of local phonological properties belonging to individual accents (Ladd, 1995: 5).

2.2.8 Transcription of intonation

Even though authors generally tend to follow the linear approach to deal with an intonation contour, then notation systems often vary. The British authors such as Roach, Wells, or Cruttenden, who have been previously mentioned, traditionally tend to use an iconic system, where the shape of the marks which are used suggests the pitch movement these marks refer to (Cruttenden, 1997: 260). The symbol ↘ clearly indicates a fall, whereas the symbol ↗ marks a rise. The high and low levels are simply distinguished by placing the symbol slightly up or down; underlining indicates an emphasis. The marks used in this system are summarised in Table 1; this system of notation is also going to be used in this work.

Basic markup	Symbol	**Finer distinctions**	Symbol
fall	ˎ	high fall	ˋ
		low fall	ˎ
		rise-fall	^
rise	ˏ	high rise	ˊ
		low rise	ˏ
		wide rise	/
		mid level	>
fall-rise	v	mid fall–rise	v
		rise–fall–rise	^v
		high fall–rise	ˇ
Other symbols			
accent	ˈ	high head	ˈ
		high falling head	ˋ
		low level head	ˌ
		low rising head	ˏ
prehead	(no mark)	high	¯
		low	_
rhythmic stress	(no mark)		° • °

Table 1: Intonation symbols used by the British authors (Wells, 2009: 260).

There is another system of notation which can be presented as a counterpart to that of the British school. It is known as ToBI (Tones and Break Indices) and it is based on Pierrehumbert´s studies described by Ladd (2008: 87–103). It treats intonation as a string of pitch accents and edge tones and not as one single contour. In an intonational phrase there is always a high (H) or low (L) level tone which is considered to be the central one (the pitch accent), marked with an asterisk (H* or L*). It is also referred to as *starred tone*. In some phrases there might be the combination of both of these tones in which case one is treated as *central* and the other as a *leading* (preceding) tone or *trailing* (following) tone. For example the symbols H + L* mark a low starred tone with a leading high tone. The edge tones are divided into two categories – the *phrase accents* and *boundary tones*. The phrase accents are defined as unstressed tones which do not serve as leading or trailing ones; they occur after the last pitch accents and are marked as H or H- and L or L-. The boundary tones (marked as H% or L% standing for a final or initial rise, or final or initial fall) occur at the beginning or, more frequently, at the end of an intonational phrase. Pierrehumbert originally worked with many different tone types but ToBI, which developed later, reduced these to only five basic ones (H*, L*, L+H*, L*+H, H+!H*) and it additionally uses a diacritic /!/ for down-step (Ladd, 2008: 105). Table 2 compares the notation system of Pierrehumbert and “the British school” version.

Pierrehumbert	British-style
H* L L%	fall
H* L H%	fall–rise
H* H L%	stylised high rise
H* H H%	high rise
L* L L%	low fall
L* L H%	low rise (narrow pitch range)
L* H L%	stylised low rise
L* H H%	low rise
L+H* L L%	rise–fall
L+H* L H%	rise–fall–rise
L+H* H L%	stylised high rise (with low head)
L+H* H H%	high rise (with low head)
L*+H L L%	rise–fall (emphatic)
L*+H L H%	rise–fall–rise (emphatic)
L*+H H L%	stylised low rise
L*+H H H%	low rise
H+L* L L%	low fall (with high head)
H+L* L H%	low fall–rise (with high head)
H+L* H L%	stylised high rise (low rise?) with high head
H+L* H H%	low rise (high range)
H*+L H L%	stylised fall–rise (‘calling contour’)
H*+L H H%	fall–rise (high range)

Table 2: Intonation symbols used by Pierrehumbert compared to those of “the British school” (Ladd, 2008: 91).

Lastly, one more system of transcription can be mentioned. It is an iconic system called INTSINT (International Transcription System for Intonation) which also uses symbols which suggest the pitch movement. Unlike the one used by many British authors, it places the marks bellow the orthographic text (Hirst and Di Cristo, 1998: 15). An example can be seen in Figure 1.

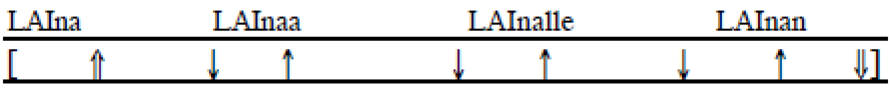


Figure 1: Transcription of a typical pitch pattern in the Finnish sentence: *Laina lainaa Lainalle lainan.* (Laina lends Laina a loan.) (Hirst and Di Cristo, 1998: 16).

2.2.9 Perception of intonation

We should not confuse the F0 contour and the intonation contour and treat them as equal. As we said before, F0 is expressed in Hz and can be directly measured by the use of an extractor, whereas the changes in intonation are given in semitones, which have to be calculated. However, intonation is not simply converting hertz values into semitones. As Volín (2009) explains, it is easy to do such a transformation but it is not the same when we need to create the intonation contour by the use of the contour of F0 (Volín, 2009). The general problem comes with the pitch perception of a syllable. As has been proven, the pitch of a syllable cannot be described as one unchanging tone, since the value of F0 constantly fluctuates, but it is usually perceived as such. Phoneticians therefore had to deal with the question of which parts of the F0 contour play the most important role in the perception of the pitch of a syllable. According to some experiments, the pitch of an average syllable is determined by the frequency measured in the middle of its core which is usually formed by a vowel or possibly an approximant, a nasal or a liquid. The other consonants creating the syllable seem to influence the pitch perception only partially. The exact frequency in the core of the syllable that would determine what intonation is perceived is still unknown.

2.2.10 Intonation in Czech

As the listeners used in our experiment are all native speakers of Czech, Czech intonation and its difference from the British English intonation need to be briefly commented on.

As well as in English, Czech intonation is dealt with in terms of intonational phrases. Each phrase has a specific intonation contour, which is known as *cadence* (in Czech *kadence*), its counterpart in the British English labelling system for intonation would be *nuclear tone*. As Daneš (2009) explains, an intonational phrase consists of one or more stress groups, which are described as segments of speech containing one stressed syllable. In Czech the length of the phrases is dependent mainly on the number of syllables involved (it is the so-called syllable-timed language) whereas the length of an English phrase is defined by the number of stress groups (English therefore ranks among the stress-timed languages). An intonational phrase in Czech usually consists of 2 to 3 stress groups, the average number of syllables in one phrase is said to be 9 to 11 (Daneš (1957) cited in Chamonikolasová, 2007: 18); Palková in her studies states 6 or 7 syllables (Palková (1994) cited in Chamonikolasová, 2007: 18).

Three specific nuclear tone types (*cadence types*, sometimes also *melodemes*, or *melodies*) are distinguished in Czech. They are known as *terminal falling melody* (M1), *terminal rising melody* (M2), and *non-terminal rising melody* (M3) (Chamonikolasová, 2007: 19). As described by Daneš (2009), M1 is the most commonly used melodeme occurring in declarative or imperative sentences and wh-questions. This tone is characterised by a fall on the nuclear syllable from which the voice either remains low or is gradually falling. It also includes the counterpart of the English fall-rise, where there is a rise on the nucleus followed by a fall. M2 is typical for yes-no questions, this melodeme represents either a rise or a rise-fall. M3 indicates further continuation of a sentence or clause in the subsequent tone units. It has two forms – in the first, the voice is rising, in the second, it falls. The common feature of both is a fall on the syllable which precedes the nucleus. It should be noted that all of the three contour types in Czech can be further divided into marked and unmarked subtypes.

As for the factors which may cause differences in intonation in distinct languages, these are said to be the functional sentence perspective and word order, speech rate, the average pitch the speaker uses, and stress-placement (Chamonikolasová, 2007: 21; Palková, 1994: 317–318). The last of the four should be emphasised the most in connection with Czech which, unlike English, places the word stress on the first syllable of a word (Palková, 1994: 338).

An interesting study comparing the intonation of the two languages was written by Chamonikolasová (2007). For her experiment she used Václav Havel´s play *Protest* and its translation into English. She worked with two types of dialogues – scripted and non-scripted. She proved that intonational phrases in English seem to be longer (by 0.75 words) compared to their Czech counterparts. This is due to the fact that the two languages are of a different type – English being an analytic, isolating language and Czech a synthetic, inflectional one. As for the position of the nucleus in the phrase, its placement on the syllable of the last word is slightly more frequent in Czech than in English. It proved to be 82.2% in the scripted and 72.0% in the non-scripted version of the Czech text in contrast with 69.1% and 66.0% in its English version. Another difference was that unlike in English, the nucleus may go on a preposition in Czech. Lastly, her results demonstrated that falls were at least four times more frequent than rises in English, and falling tones were generally used more often in the English translation of the play than in the Czech original (69.5–70.8% of falls in the English version, 50.9–65.1% in the Czech version). Rises were much more common for Czech as well as rise-falls, but fall-rises occurred more in English.

3 MATERIAL AND METHOD

3.1 The hypothesis of the present study

The present study primarily follows the experiments performed by ´t Hart (1981) and Harris and Umeda (1987). Their main task was to determine the pitch change necessary to make the listeners perceive the difference in various types of speech stimuli. Our study uses the data provided by the previously mentioned authors and further investigates whether the type of the syllable – stressed, unstressed or nuclear – may also influence the JND. The hypothesis is based on the question already asked by Harris and Umeda (1987), which has not been answered yet. The experiment will thus determine if melodic prominence should be considered one of the main factors influencing the differential threshold.

´t Hart´s experiment (1981), using Dutch number names, revealed that the listeners performed better in the case of speech rises than falls when judging the JND, and the author gave the following reason: "This ability could originate from the more intensive exposure to rises owing to the preference, in Dutch intonation, for the use of rises for pitch accents. If this were true, the reverse effect should occur in speakers/listeners of British English, which has a preference for falls" (´t Hart, 1981: 820). The present study is thus going to try to either confirm or disapprove ´t Hart´s assumption since it works with English spoken material and the listeners are the speakers of Czech who, similarly to the English, use speech falls more frequently than speech rises. Harris and Umeda (1987) further proved that there was no difference between the upward and the downward change of the stimulus; our study will therefore determine if this claim might be considered generally valid or not. The musical training of the listeners and their possible improvement in the course of the test will also be considered.

3.2 Acoustic material

The experiment used four intonational phrases pronounced by a native speaker of British English, which were marked as *fall1*, *fall2*, *rise1* and *rise2* (listed in Table 3). Their duration ranged between 2.7 and 3.6 seconds. They consisted of four stress groups of one stressed syllable followed by two unstressed ones. The final stressed syllable was always the nuclear one, where the major intonational movement started. The nuclear tone was twice realised as a rise, twice as a fall and it always spread over the final three syllables. As an example, one of the intonational phrases (*rise2*) is analysed; the other three are structured in the same way:

Could **any**one **give** me a **syn**onym for ↗**won**derful?

PREHEAD HEAD NUCL. TAIL

These phrases were integrated into a larger piece of text of 4 to 8 clauses (included in Appendix 1), which was read out aloud naturally by a female native speaker of Standard British English and recorded in a recording studio of the Institute of Phonetics at the Faculty of Arts at Charles University in Prague. The speaker familiarised herself with the text before the recording; the purpose of the experiment was unknown to her.

The pitch contour of the four phrases was then checked in the programme Praat (Boersma and Weenink, 2015); any mistakes were corrected. These four intonational phrases were used as the basis for the later manipulation.

fall1	He **would** have re<u>**mem**bered the</u> **mar**vellous ↘<u>**liv**</u>ing-room.
fall2	**Car**oline <u>**dreamed**</u> <u>of be</u>**com**ing a ↘<u>**min**</u>ister.
rise1	The **book**shelves are <u>**load**ed with</u> **books** about ↗<u>**gard**</u>ening,…
rise2	Could **an**yone <u>**give**</u> <u>me a</u> **syn**onym for ↗<u>**won**</u>derful?

Table 3: The four intonational phrases used in the experiment. The stressed syllables are in bold, underlining indicates which syllables were changed in pitch, and an arrow precedes the nuclear syllable, its direction indicates either a falling nuclear tone or a rising one.

3.3 Manipulation

The above described structure of the phrases was used to prevent the listeners from predicting the place of manipulation. The second stressed syllable, both unstressed syllables in the second stress group, and the nuclear syllable of each of the four phrases were used for the manipulation. The two unstressed syllables were treated as a group so that these together would be approximately of the same duration as the stressed syllable and the nucleus (considering the stress-timed nature of English rhythm). These two unstressed syllables thus constitute the "unstressed position" (further referred to as "the unstressed syllable"). The duration of all types of syllables ranged between 163 ms and 361 ms. Only syllables consisting of voiced speech sounds were used to avoid the disruption of the F0 contour.

As the pitch perception of a syllable is dependent on the frequency of the middle part of its peak (Volín, 2009: 230), only the pitch of this part of the vowel of each relevant syllable was changed by 1.5 ST. This particular amount of pitch manipulation, as the previous experiments suggested, should be close to the starting point of the pitch change discrimination in speech perception. In the case of the two unstressed syllables, the middle part of both vowels plus the pitch contour between them was changed. The manipulation was both in an upward and in a downward direction; each time the frequency of only one syllable in a phrase was increased or decreased. The remaining pitch points of the pitch contour in the given syllable were erased before the manipulation so that this part of the contour also underwent a change in pitch in the course of the main vowel change. This change was determined by quadratic interpolation. The whole process of manipulation was again done in Praat. Figure 2 presents an unchanged pitch contour contrasted with its manipulated version.

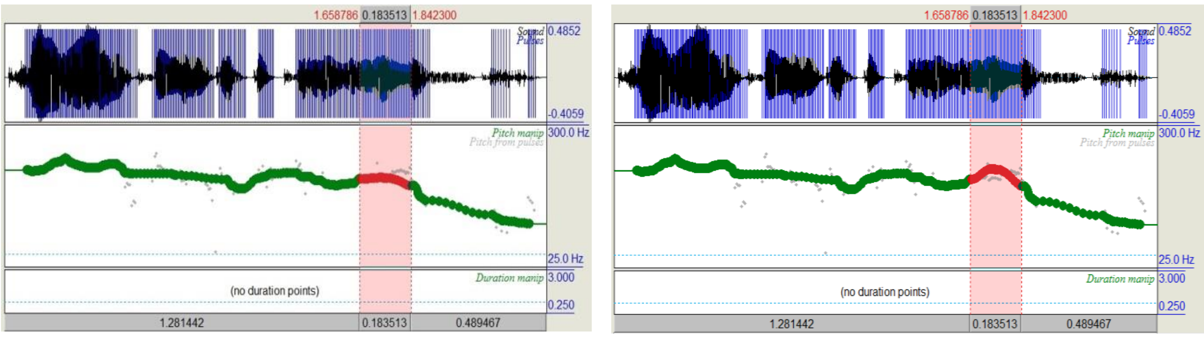


Figure 2: The pitch contour of the phrase *fall2* ("*Caroline dreamed of becoming a minister*."). The first picture shows the original contour, the second displays a pitch contour whose nuclear syllable /mɪn/ was changed by 1.5 ST in an upward direction. The nuclear syllable is marked in red.

3.4 The experiment

20 university students of English in their third year of Bachelor studies (6 men and 14 women, aged 22–27) were used as listeners for the experiment. They were all native speakers of Czech and reported having no hearing problems. Most of them had no previous experience with psychoacoustic tests. They were invited individually to the recording studio of the Institute of Phonetics at Charles University in Prague and used headphones when listening to the perceptual test, which was administered in Alvin (Hillenbrand and Gayvert, 2003). The method of pair comparison was used in the test. Each time, a pair of phrases was played to the listeners and their task was to determine whether they perceive a change in pitch between the two stimuli or not. They were not told where to expect the change. The test consisted of four

blocks of 20 items (i.e. pairs of phrases) plus an additional training block was added to the beginning so that the listeners could familiarise themselves with the nature of the test.

In each of the regular trials of the test, two intonational phrases with identical text separated by a 500-ms pause were played to the listener. There were three types of items which could be played. In the first type, both of the phrases in the pair were left unmodified (marked as SAME items); in the second type, the pitch was changed in the same way in both phrases (again marked as SAME items); and in the third type, the first of the two phrases did not undergo any manipulation and the second did (these pairs were marked as DIFFERENT items). The test included 40 items in total (16 of the SAME type, 24 of the DIFFERENT type), each was repeated twice during the test (the repeated items never occurred in the same block, otherwise, the order of the items was random; the order of the test blocks also varied for each listener). Table 4 shows all possible combinations of phrases used as the second (manipulated) phrase within the pairs marked as DIFFERENT. All items used in the test are included in Appendix 2.

phrase type	*fall1*						*fall2*						*rise1*						*rise2*					
syllable type	U		S		N		U		S		N		U		S		N		U		S		N	
direction of the 1.5 ST manipulation	*u*	*d*	*u*	*d*	*u*	*d*	*u*	*d*	*u*	*d*	*u*	*d*	*u*	*d*	*u*	*d*	*u*	*d*	*u*	*d*	*u*	*d*	*u*	*d*

Table 4: The calculation of the final 24 DIFFERENT items. Four different phrase types were used in the test (marked as *fall1*, *fall2*, *rise1* and *rise2*). Three syllables in each of them (unstressed marked as U, stressed marked as S, and the nuclear marked as N) underwent manipulation and this change was either by 1.5 ST in a downward or in an upward direction (in the table therefore labelled *u* or *d*). This gives 24 different items which were compared by the listeners with their original, unchanged, version.

A short desensitisation sound was played after each experimental item. The listeners had the opportunity to replay each item once and their task was to click on one of two "buttons" on the computer screen labelled as either SAME or DIFFERENT. Their responses and the number of replays needed were recorded. There was a brief pause after each trial.

The initial block of the test consisted of two additional intonational phrases pronounced by the same speaker. It was also composed of 20 items, 8 of them were of the SAME type, the remaining 12 were of the DIFFERENT type. The pitch of 9 syllables was changed by 2.2 ST to enable the listeners to perceive the change better; the other ones of the DIFFERENT type were treated in the same way as those of the regular trails. The results of this part were not analysed. The listeners were also asked to fill in a brief questionnaire predominantly regarding their musicality (Appendix 3); its contents will also be examined in sections *Results* and *Discussion*.

4 RESEARCH PART

4.1 Results

The first step in the analysis of the data acquired from the perceptual test was to determine the reliability of the listeners. Their responses to the pairs of phrases of the SAME type were analysed and, as a result of this, 5 of the 20 listeners had to be excluded as they reported to have heard the two items (falsely) as different in more than 40% of the cases. The subjects were judged unreliable or unable to perform the task and their answers were thus not further worked with. Additionally, only the pairs of phrases presented as DIFFERENT underwent close examination; therefore, the label "correct" is used for phrase pairs of the DIFFERENT type really judged by the listeners as DIFFERENT. The total number of these items, which were analysed, was 720 (24 items repeated twice judged by 15 listeners). The results were statistically evaluated by the use of a binomial test. Error bars created for each figure indicate 95% confidence intervals.

4.1.1 Correctness

Figure 3 compares the results of the JND for the three syllables under examination, which differ in the degree of melodic prominence. The listeners perceived the change in the nuclear syllable 118 times out of 240 (therefore in 49% of the cases), the change in the stressed syllable was noticed 95 times (40%) and in case of the unstressed syllable, the manipulation was perceived 88 times (37%). The results thus confirm our main hypothesis, suggesting that the melodic prominence (the degree of stress and the intonational movement) does indeed influence our pitch perception; the differences between the unstressed and the nuclear syllables proved to be statistically significant. Figure 4, however, indicates that the judgement of the pitch change varied for the four intonational phrases, and the possible reasons for this variation will be discussed in the following subchapter.

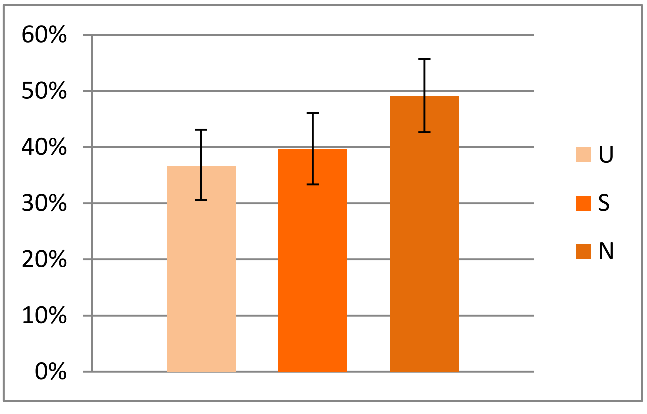


Figure 3: The correctness of answers for the perception of the ± 1.5-ST change in pitch in the unstressed, stressed and the nuclear syllable. Error bars indicate 95% confidence intervals.

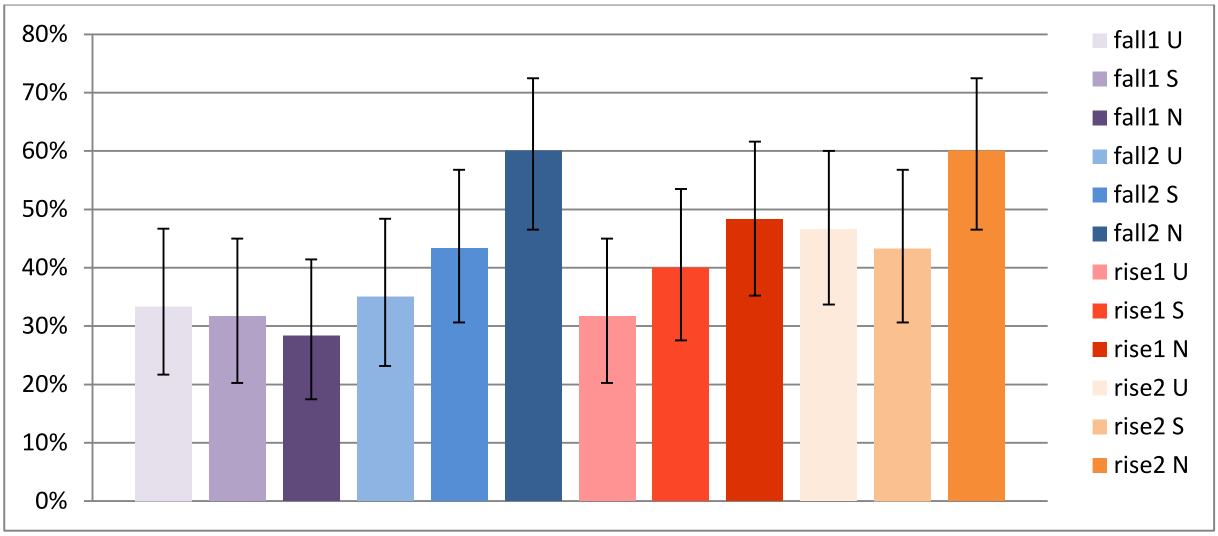


Figure 4: The correctness of answers for the perception of the ±1.5-ST pitch change in the four intonational phrases (for the unstressed, stressed and the nuclear syllable) used in the experiment. Error bars indicate 95% confidence intervals.

The differences in pitch proved to have been noticed more often in the intonational phrases marked as *rise1* and *rise2* (= *rises*), containing the final rising nuclear tone, in comparison with the phrases marked as *fall1* and *fall2* (= *falls*) (as shown in Figure 5). These results were, however, not statistically significant. Additionally, if we consider the pitch manipulation perception in the nuclear syllables only, it seems that it is the phrase *fall1* which significantly differs from the rest. Figure 6 shows that the manipulation in pitch of the falling tone in the phrase *fall2* and the rising tone in the nuclear syllable in the phrase *rise2* were both perceived 36 times (out of 60, therefore in 60% of the cases); it was 29 times (48%) in the nucleus of *rise1*, and only 17 times (28%) in the nuclear syllable of *fall1*. The possible reasons are again explained in the *Discussion*.

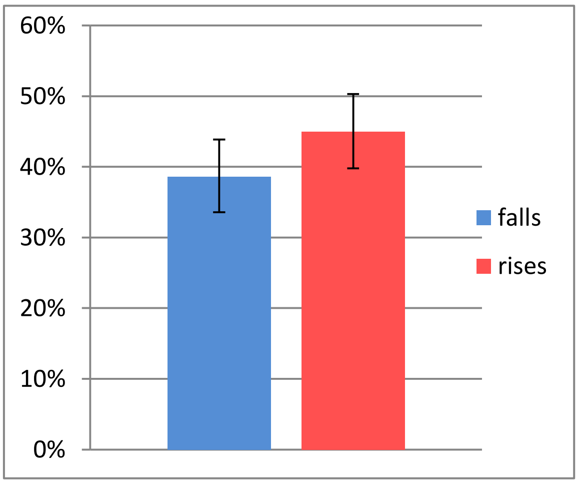


Figure 5: The correctness of answers for the perception of the ±1.5-ST pitch change in the phrases marked as *fall1* and *fall2* (= *falls*) compared to phrases *rise1* and *rise2* (= *rises*). Error bars indicate 95% confidence intervals.

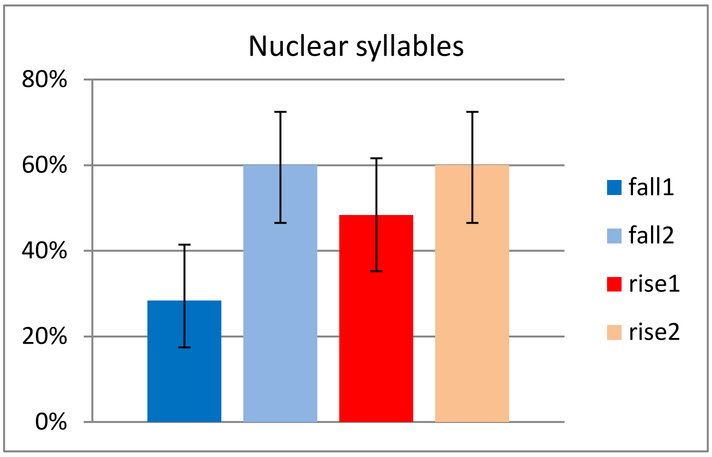


Figure 6: The correctness of answers for the perception of the ±1.5-ST pitch change in the nuclear syllable of the phrases marked as *fall1*, *fall2*, *rise1* and *rise2*. Error bars indicate 95% confidence intervals.

Furthermore, the pitch change in the upward direction proved to have been perceived slightly better than the change downwards (as can be seen in Figure 7). Nevertheless, there were differences among the phrases. The pitch manipulation in the upward direction was better noticed in the phrases marked as *rise1* and *rise2* (Figure 8), while the pitch changes in the downward direction appeared more salient in phrases *fall1* and *fall2* (Figure 9).

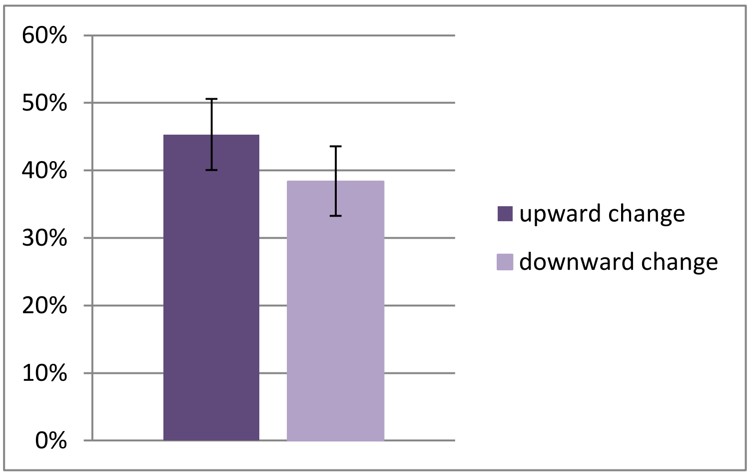


Figure 7: The correctness of answers for the perception of the pitch change in the upward (+1.5 ST) and in the downward direction (-1.5 ST). Error bars indicate 95% confidence intervals.

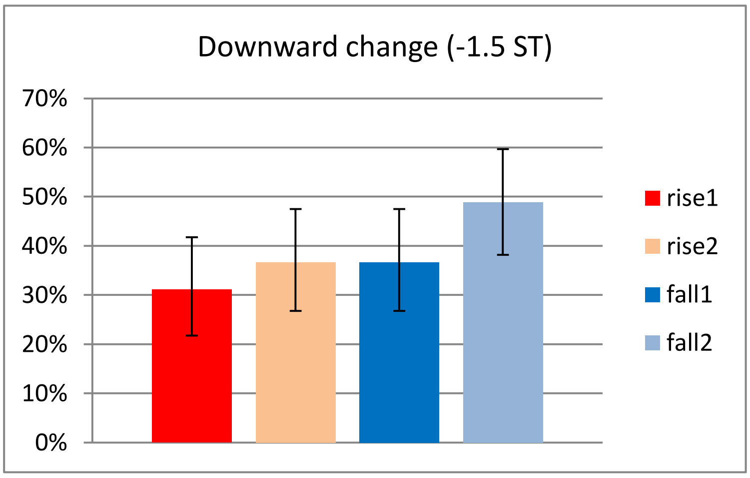


Figure 8: The correctness of answers for the perception of the pitch change in the downward direction (-1.5 ST) in the four intonational phrases. Error bars indicate 95% confidence intervals.

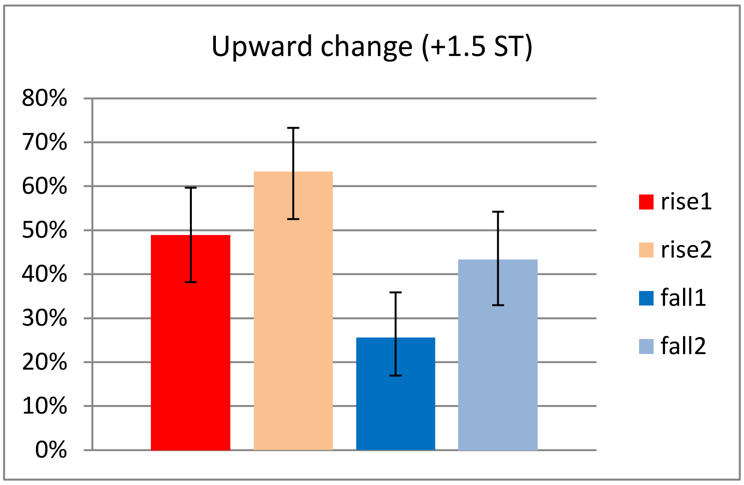
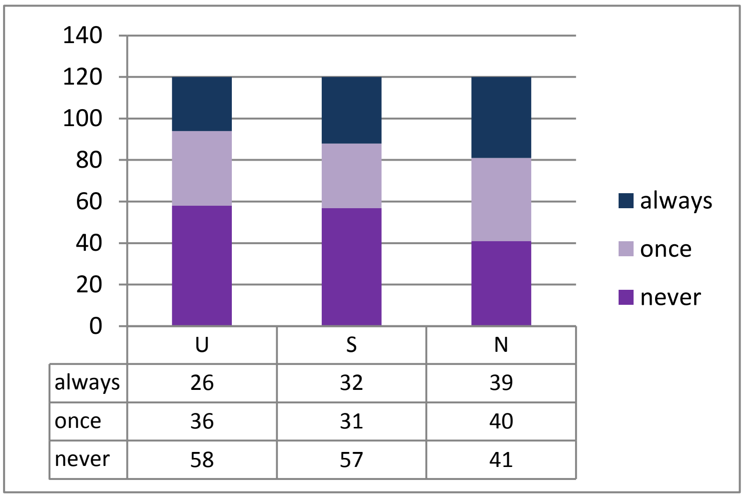


Figure 9: The correctness of answers for the perception of the pitch change in the upward direction (+1.5 ST) in the four intonational phrases. Error bars indicate 95% confidence intervals.

4.1.2 Consistency

The analysis of the items (each pair occurred twice in the test) revealed that the two items were consistently perceived as DIFFERENT in 39 cases (33%) when the change was in the nuclear syllable, in 32 cases (27%) when in the stressed syllable, and only in 26 cases (22%) when the change was located in the unstressed syllable (see Figure 10). The opposite consistency effect was also present, i.e. cases when the pitch was perceived in neither of the two items. The listeners were consistently unaware of any change in 58 cases (in 48% of the cases) when judging the unstressed syllable, in 57 cases (48%) when judging the stressed syllable and in 41 cases (34%) in the case of the nuclear syllable.



	U	S	N
always	26	32	39
once	36	31	40
never	58	57	41

Figure 10: The number of cases (out of 120 = 1 type of syllable × 2 directions × 4 phrases × 15 listeners) when two repeated items (identical test pairs) in the test were *always* (n = 2), *once* (n = 1) or *never* (n = 0) judged as DIFFERENT.

4.1.3 Notes on the listeners

The musical training of the listeners was also considered but no significant differences were revealed. Additionally, those subjects who used the optional replay of the pair of phrases more often did not prove more successful when judging the change in pitch compared with those who used this option only a few times. Finally, the listeners seem to have improved their perception skills during the test. Most of them correctly judged the pairs of phrases in the seventh and eighth tenth of the test, the deterioration towards the end is most likely the result of the loss of attention and of the listeners´ weariness.

4.2 Discussion

The majority of the listeners found the task very demanding; only four of them answered correctly when judging the types of phrases labelled DIFFERENT in more than 50% of the cases. Five of the listeners had to be considered unreliable as they proved to have been merely guessing or actually trying to hear the difference in the pairs of phrases where there was none.

The results of the pitch change judgements for the three syllables differing in the degree of melodic prominence confirmed our primary hypothesis. The suggestion of Harris and Umeda (1987) that the JND varies for different types of syllables was thus found valid as the difference in pitch was perceived most frequently in the nuclear syllable and least often in the case of the unstressed syllable. Nevertheless, as was already suggested in the *Results*, there were considerable differences among the four intonational phrases used in the experiment. Let us now discuss the possible reasons.

The phrase marked as *fall1* (*He would have remembered the marvellous living-room*.) appeared to have caused most problems to the listeners, who were judging the differences in pitch. Especially the pitch manipulation in the nuclear syllable of this phrase was perceived less frequently than the change in any other syllable examined in our study. One of the possible explanations might be the difference in duration as the nuclear syllable of *fall1* was the shortest compared to the nuclei in the other three phrases (164 ms in *fall1*, 184 ms in *fall2*, 287 ms in *rise1* and 221 ms in *rise2*). Another reason might be the fact that this syllable was of the lowest frequency, again compared with the other three (166–88 Hz in *fall1*, 182–197 Hz in *fall2*, 170–212 Hz in *rise1* and 226–239 Hz in *rise2*). Other syllables in the phrase *fall1* also proved more prominent in terms of F0. The fundamental frequency of the syllable /wʊd/ reached 300 Hz, which means that it differed from the highest frequency in the nuclear syllable /lɪv/ by 112 Hz (8.1 ST). In the case of the syllable /mɑːv/, F0 reached 211 Hz and this syllable contained a long vowel, where the movement in pitch might have been more prominent so that the nuclear syllable actually became less dominant. Moreover, the overall pitch span within the phrase was 15.6 ST – the highest of all – and the speaker used creaky phonation towards the end, which might have caused further difficulties to the listeners when judging the difference in pitch. Contrarily, when we examine the phrase *fall2* (*Caroline dreamed of becoming a minister*.), where the pitch manipulation was perceived more often, we can find that it had a smaller overall pitch span (13.4 ST) and there were no other syllables significantly higher in pitch than the nuclear one (the syllable /kær/ differed from the nucleus only by 3 ST).

We should also briefly comment on the phrase *rise2* (*Could anyone give me a synonym for wonderful*?) as it appeared the easiest for pitch change detection among the four phrases. The overall pitch span within the phrase was only 6 ST and all of the syllables in the phrase did not differ much in pitch. The pitch manipulation in its unstressed syllable proved more noticeable than the manipulation in its stressed syllable but this might have been caused by the duration of the latter, which lasted only 142 ms and was the shortest of all syllables under examination. The change in the unstressed syllable of *rise2* was perceived most frequently when compared with the unstressed syllables of the other three phrases, and this might have been caused by the presence of a long vowel (/iː/ in "me"), which did not occur in any of the other phrases.

The variability among the phrases therefore suggests that pitch change detection is not only dependent on the melodic prominence of the syllable itself but the neighbouring syllables and the overall pitch span in the phrase might also be significant.

To return to our original hypothesis, we should notice that the number of cases where the listeners perceived the change consistently in both repetitions of the items was the highest for the nuclear syllables and the lowest for the unstressed ones. This again confirms that the manipulation in pitch was better heard (therefore more salient) in the most prominent syllable.

´t Hart (1981) in his experiment with Dutch number names suggested that in the case of English, falls should be less difficult to judge than rises. He proved that the opposite was valid for Dutch, which has a dominance of rises in speech. Although the listeners were of the Czech origin, who also use speech falls more often than speech rises, this hypothesis could not be generally confirmed. When comparing the whole intonational phrases, the ones marked as *rise1* and *rise2* were easier for the pitch manipulation perception (however, this difference cannot be judged as statistically significant). When looking only at the nuclear syllables of the four phrases, the ones in *fall2* and *rise2* were perceived more frequently than those in the other two phrases. The possible reasons for this are discussed in the previous paragraphs and we should emphasise that no conclusion can be formed on the basis of these results as there were only four phrases studied. Considering the study of Chamonikolasová (2007), there actually might be an influence of the Czech language because, as Chamonikolasová pointed out when examining the original and the English translation of Václav Havel´s play *Protest*, despite the fact that both languages prefer speech falls, rises were much more common for Czech than for English.

Harris and Umeda (1987) determined that there was no difference in the JND for changes in the upward and in the downward direction. Our experiment indicated that generally, the change in the upward direction was only slightly easier to be noticed (our listeners perceived it correctly in 54% of the cases) but there were differences among the four phrases. The change in pitch of the phrases *fall1* and *fall2* was better recognised in the downward direction (in 49% of the cases in phrases marked as *fall2* and in 37% of the cases in phrases labelled as *fall1* as opposed to 37% of the cases of the *rise2*-type and 31% of the cases of the *rise1*-type). The change in the upward direction had the reversed effect. It was noticed in 63% of the cases in pairs in the case of *rise2* and in 49% of the cases in *rise1* while for the *falls*, it was 43% of the cases of the *fall2*-type and 26% of the *fall1*-type. Again, further investigation would be needed.

Micheyl and his colleagues (2006) in their work studied whether musical training influences our ability to detect the pitch change. They proved that the JND for pure and complex tones was six times smaller when determined by the musicians contrasted with the non-musicians. Although the three listeners of our perceptual test who were the best at distinguishing the pitch change had some experience with music (as they stated in the questionnaire), it does not seem that there were any significant differences between the two groups (musicians and non-musicians). However, the subjects of the study of Micheyl et al. were professional musicians and thus the results of the study cannot be directly compared to our experiment. Nevertheless, Micheyl and his colleagues proved that the non-musicians improved during the test (they reached the JND of the musicians after 4 to 8 hours of training in the case of pure tones). When we divide the test into 10 sections of 8 pairs of phrases, we can see that the majority of our subjects had the most correct answers in the seventh and the eighth section. This also suggests some kind of an improvement if we consider the deterioration in the last two sections to be a result of the listeners´ growing fatigue.

Finally, we have to emphasise the fact that although our listeners were all university students of English, they were, nevertheless, speakers of Czech listening to the English material. It would be interesting to compare our results with those for listeners who would be the native speakers of English and also those with no knowledge of English at all. Such results might then significantly differ.

5 CONCLUSION

The main goal of the thesis was to ascertain whether the pitch change perception varies due to the different melodic prominence of syllables within an intonational phrase. Previous studies on the just noticeable difference (mainly the experiments performed by ´t Hart (1981), Harris and Umeda (1987) and ´t Hart et al. (1990)) revealed that the change of approximately 1.5–2 ST was needed so that the listeners would perceive the manipulation in pitch in complex (speech) signals. Harris and Umeda (1987) suggested that there might be some differences based on the original pitch movement but they have not worked on the hypothesis any further. Our experiment thus follows their work and suggests that the pitch change perception is dependent on many other factors, not only the change caused by the increase or decrease of pitch expressed in semitones.

In the first subchapter of the *Theoretical part* of the thesis, we defined terms such as *the fundamental frequency* (F0) and *the just noticeable difference* (the JND), which were crucial for our work. The term *frequency* in general was described as one of the qualities of a sound wave, and F0 in connection with the human voice was examined in greater detail. The perception of F0 as pitch was commented on and the just noticeable difference (the minimal difference between two discriminated sounds) and the previous studies on it were described together with their results creating the basis for our work. The various ways to measure the JND were also mentioned; our perceptual test was based on one of them known as the method of "pair-comparison."

The second subchapter was dedicated to intonation and the various approaches towards it. Our study worked primarily with the definition which stated that intonation stands for the changes in pitch caused by the variation of F0 (Hirst and Di Cristo, 1998). The term *stress* and its realisation in syllables were also explained as all of these terms are needed for the preparation, realisation and analysis of our experiment itself. The structure of an intonational phrase was introduced together with different systems of transcription of intonation. Intonational tones and the perception of intonation within a syllable were commented on, and the differences between the intonation in English and Czech were briefly mentioned as we used an English material for the experiment but the listeners were all speakers of Czech.

The chapter *Material and method* introduced our main hypothesis and mentioned several other possible factors influencing the JND such as the direction of the manipulation in pitch, the type of the nuclear tone, or the musical training of the subjects tested. The recording of the material in Standard British English and the subsequent acoustic manipulation performed on it were depicted in detail. The testing of the 20 Czech listeners and the phrases used in the test were described.

Our main findings were summarised in the chapter *Results*. It was revealed that our hypothesis proved correct as the listeners had generally noticed the change on the nuclear syllable most frequently (in 49% of the cases) in comparison with the syllables marked as *stressed* and *unstressed* (perceived in 40% and 37% of the cases), which were lower in the degree of melodic prominence. Phrases containing a final rise were judged more easily in comparison with those containing a final fall but these results cannot be marked as statistically significant and they also differed for the four intonational phrases used for the testing. Changes in the upward direction were noticed more often in the case of phrases labelled as *rise1* and *rise2*, manipulation in the opposite direction was better heard in the case of both *falls*. The repetition of the same items was also considered – and this again confirmed our primary hypothesis. The number of replays of the items, which could be optionally used by the listeners in the course of the test, did not reveal any differences, nor did the musical training of the subjects. The participants, however, seemed to have improved during the test (apart from the last two sections but at that point, we cannot rule out an influence of growing fatigue).

In *Discussion* we tried to explain various reasons for our findings and especially to analyse the four intonational phrases more accurately. It was revealed that the phrase marked as *fall1* proved the most difficult for the pitch change detection and the possible reasons could be its overall great pitch span, the lower frequency of the nuclear syllable in comparison with the rest of the phrase, or the creaky voice at the very ending. *Rise2*, on the other hand, was the easiest for the listeners, and the possible cause might be the fact that the overall pitch span was no more than 6 ST. The other results were discussed and compared to other outcomes and assumptions of the previous works, some of which were confirmed. It was also emphasised that more material would be needed for more precise conclusions to be formed, and that our results would most certainly be worth comparing with those using either English material and

English listeners only, or English material and listeners without any knowledge of the language.

REFERENCES

Boersma, P., and Weenink, D. (2015). Praat: Doing phonetics by computer (Version 5.4.08.). Downloaded on 4 January 2015 from http://www.praat.org/.

Chamonikolasová, J. (2007). *Intonation in English and Czech Dialogues*. Brno: Masarykova Univerzita Brno.

Cruttenden, A. (1997). *Intonation*. Cambridge: Cambridge University Press.

Daneš, F. (2009). *Kultura a struktura českého jazyka*, pp. 355–364. Praha: Karolinum.

Frost, D. (2011). Stress and cues to relative prominence in English and French: A perceptual study. *Journal of the Intonational Phonetic Association, 41*(1), 67–84.

Gut, U. (2009). *Introduction to English Phonetics and Phonology*. Frankfurt am Main: Peter Lang.

Harris, M.S., and Umeda, N. (1987). Difference limens for fundamental frequency contours in sentences. *Journal of the Acoustical Society of America, 81*(4), 1139–1145.

Hillenbrand, J.M., and Gayvert, R.T. (2003). Open source software for experiment design and control. *Journal of Speech, Language, and Hearing Research, 48*, 45–60.

Hirst, D., and Di Cristo, A. (1998). A survey of intonation systems. In D. Hirst and A. Di Cristo (Eds.), *Intonation Systems: A Survey of Twenty Languages,* pp. 1–44. Oxford: Oxford University Press.

Ladd, D. R. (1995). “Linear” and “overlay” descriptions: An autosegmental-metrical middle way. *Proceedings of the 13th ICPhS 2,* pp. 116–123. Stockholm.

Ladd, D. R. (2000). Bruce, Pierrehumbert, and the elements of intonational phonology. In G. Bruce (Ed.), *Prosody: Theory and Experiment*, pp. 37–48. Dordrecht: Kluwer Academic Publ.

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

Micheyl, C., Delhommeau, K., Perrot X., and Oxenham A.J. (2006). Influence of musical and psychoacoustical training on pitch discrimination. *Hearing Research, 219,* 36–47. Available on-line at www.sciencedirect.com.

Noteboom, S. (1999). The prosody of speech: Melody and rhythm. *The Handbook of Phonetic Sciences*. Hardcastle: Blackwell Publishing. Blackwell Reference Online. 28 Dec. 2007. Available on-line at http://www.blackwellreference.com/subscriber/uid=532/tocnode?id=g9780631214786_chunk_g978063121478621 (accessed: 21 Aug. 2014).

Longman Dictionary of the English Language (6th impr.). (1991). Harlow: Longman.

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

Reetz, H., and Jongman, A. (2009). *Phonetics: Transcription, Production, Acoustics, and Perception.* Oxford: Wiley-Blackwell.

Roach, P. (1990). *English Phonetics and Phonology*. Cambridge: Cambridge University Press.

Roach, P. (2000). *English Phonetics and Phonology* (3rd ed.) Cambridge: Cambridge University Press.

Volín, J. (2009). Extrakce základní hlasové frekvence a intonační gravitace v češtině. *Naše řeč, 92*(5), 227–239.

Vongpaisal, T., and Pichora-Fuller, M.K. (2007). Effect of age on F_0 difference limen and concurrent vowel identification. *Journal of Speech, Language, and Hearing Research, 50,* 1139–1156.

Wells, J.C. (2009). *English Intonation: An Introduction*. Cambridge: Cambridge University Press.

´t Hart, J. (1981). Differential sensitivity to pitch distance, particularly in speech. *Journal of the Acoustical Society of America, 69*(3), 811–821.

´t Hart, J., Collier, R., and Cohen, A. (1990). *A Perceptual Study of Intonation. An Experimental-phonetic Approach to Speech Melody*. Cambridge: Cambridge University Press.

RESUMÉ

Cílem bakalářské práce je určit, zda melodická prominence slabiky ovlivní naše vnímání změny ve výšce tónu. Materiálem jsou nahrávky v angličtině, mateřským jazykem posluchačů je čeština. Práce v mnoha ohledech navazuje na předchozí výzkum, především experimenty ´t Harta (1981), Harrise a Umedy (1987) a ´t Harta a kol. (1990), které se zaměřovaly na zjištění tzv. *diference limen*, tedy minimální změny ve výšce tónu, kterou dokážeme rozpoznat sluchem. Předchozí výzkum stanovil přibližnou hodnotu této změny (vyjádřenou v půltónech) pro řečové signály obecně, naše práce se snaží dokázat, že se tato hodnota bude dále měnit v závislosti na melodických vlastnostech slabiky.

V teoretické části práce se věnujeme především představení pojmů, jako jsou základní frekvence, vnímání výšky tónu, diference limen a její percepce a měření, a shrnujeme předchozí studie na toto téma. Termín *základní frekvence* (F0) je pro naši práci klíčový, jedná se o počet kmitů hlasivek za sekundu vyjádřený v hertzech (Hz). F0 a její násobky jsou sluchem vnímány jako výška tónu, což znamená, že pokud zvýšíme rychlost kmitání hlasivek, docílíme tím dojmu vyšší výšky tónu a naopak. Na vnímání výšky tónu má také vliv celá řada dalších faktorů, jako jsou například hlasitost, délka stimulu nebo hodnota základní frekvence (lidské ucho například nerozpozná nízko- nebo vysoko frekvenční zvuky s hodnotami pod 16 Hz nebo nad 20 kHz). Pro určování minimální změny ve výšce tónu, tedy oné diference limen, se používá jednotka *půltón* (anglicky *semitone*), s jejíž pomocí můžeme srovnávat percepční změny výšky tónu pro různé mluvčí. Použití jednotky hertz bylo nevhodné, protože stejná hodnota změny v hertzech v případě nižšího hlasu neodpovídá stejné změně v hertzech v případě hlasu vyššího. Pro měření diference limen se užívá různých metod, jedna z nejčastějších je známá pod názvem *pair-comparison*, při níž posluchači srovnávají výšku tónu dvou různých stimulů.

Diferenci limen se věnovala celá řada studií. Flanagan a Saslow (1958) například zjišťovali hodnotu nejnižší postřehnutelné změny v případě synteticky vytvořených vokálů, Klatt (1973) se zaměřil pouze na vokál /ɛ/ a srovnával rozdíl mezi stálou a měnící se hodnotou F0 daného stimulu. ´t Hart (1981) za použití slov v dánštině nechal posluchače porovnávat dvě různé změny výšky tónu a určil, že diference limen pro řečové signály dosahuje hodnoty 2 půltóny. Ve studii z roku 1990, na níž se podílel s dalšími autory, toto dále upřesnil na 1,5 až 2 půltóny. Harris a Umeda (1987) také použili metodu srovnávání dvou stimulů, tentokrát

v podobě delších vět, a dospěli k hodnotě 0,6–2,1 půltónu. Micheyl a kol. (2006) se zabývali vlivem hudebního vzdělávání na diferenci limen a dospěli k závěru, že skutečně existují výrazné rozdíly mezi skupinami „hudebníků“ a „ne-hudebníků“ (v originále „musicians“ a „non-musicians“). Dokázali též, že se posluchači mohou v průběhu testování zlepšit. Vongpaisal a Pichora-Fuller (2007) navíc určili, že roli hraje i věk testovaných subjektů.

Druhá polovina teoretické části je zasvěcena intonaci, jejíž vliv na percepci změny ve výšce tónu primárně zkoumáme. Intonace je v užším smyslu definována jako řetězec změn ve výšce tónu odehrávající se v řeči. Synonymně může být použito i slovo *melodie*. Intonace má celou řadou funkcí, pro naši studii je důležitá především její schopnost vyjádřit melodickou prominenci určité slabiky. Pro popis intonace jsou důležité především kadence (v angličtině *tones*) vyjadřující určitý druh změny v intonační kontuře. Intonační kadence v češtině i angličtině mohou být několika typů: kadence stoupavá, klesavá, klesavě-stoupavá, stoupavě-klesavá nebo kadence neměnná (tzv. *level tone*). Intonace je také spojena s pojmem přízvuk a je vlastně jedním ze způsobů jeho vyjádření. Přízvuk, neboli prominence slabiky, se dále realizuje za pomoci délky, hlasitosti a kvality hlásek. Pojem přízvuk a přízvukový takt je pro naši práci významný především ve spojení se strukturou intonační fráze, která je popsána následně; prostor je věnován především její melodicky nejvýraznější slabice, nukleu, na níž začíná nejdůraznější intonační změna celé fráze. Na závěr jsou též vysvětleny různé přístupy k intonačním konturám a transkripce intonace (především rozdíly mezi systémem tzv. Britské školy a systémem ToBI). Protože mateřským jazykem posluchačů, kteří se účastnili našeho experimentu, je čeština, je závěrečný oddíl teoretické části věnován její intonaci ve srovnání s intonací angličtiny.

Kapitola *Materiál a metoda* představuje prováděný experiment a naši hlavní hypotézu. Ta předpokládá, že změna ve výšce tónu bude posluchači nejlépe rozpoznána v nejvýraznější slabice intonační fráze, tedy nukleu. Méně přesná bude diskriminace v případě slabiky přízvučné, a ještě méně v případě slabiky nepřízvučné. Hypotéza tím navazuje na otázku zmíněnou ve studii Harrise a Umedy (1987), kteří poukázali na to, že diference limen může záviset na typu slabiky. Experiment tedy pracuje s těmito třemi typy slabik (nukleární, přízvučná a nepřízvučná) ve čtyřech různých intonačních frázích, z nichž dvě jsou ukončeny kadencí stoupavou, dvě klesavou. Jako výchozí materiál slouží nahrávky standardní britské angličtiny namluvené britskou mluvčí za účelem tohoto experimentu (cíl experimentu však předem neznala). Výška tónu na předem stanovených slabikách ve frázích byla změněna o 1,5 půltónu výš, nebo níž a otestováno bylo dvacet posluchačů, studentů angličtiny na vysoké škole, kteří měli za úkol určit, zdali změnu ve výšce tónu slyší, nebo ne. V experimentu byla též použita metoda párového srovnávání (*pair-comparison*).

Kapitola věnovaná samotnému výzkumu prezentuje statisticky ohodnocené výsledky percepčního testu. Ten potvrdil naši hlavní hypotézu. Bylo tedy dokázáno, že posluchači nejčastěji sluchem rozpoznali změnu v slabice nukleární (v 49 % z 240 použitých dvojic frází s touto změnou), méně často v slabice přízvučné (v 40 %), a ještě méně v případě slabiky nepřízvučné (v 37 %). Dále se ukázalo, že úkol byl pro posluchače jednodušší při určování změny směrem nahoru (kdy výška tónu byla o 1,5 půltónu zvýšena) a v případě frází s finální stoupavou kadencí. Tyto výsledky však nebyly statisticky významné, použité fráze se od sebe navíc značně lišily. Dalšími analyzovanými proměnnými bylo hudební vzdělání posluchačů, jejich zlepšení, četnost použití možnosti volitelného opakování frází a shodnost odpovědí v případě identických párů frází, které v testu zazněly vždy dvakrát.

Část *Diskuse* se k výsledkům opět vrací a snaží se je interpretovat z různých hledisek. Odhaluje rozdíl mezi jednotlivými frázemi a vysvětluje například, jaké byly možné příčiny toho, že posluchači měli velké problémy při rozeznávání změn v melodému fráze označené jako *fall1* s finální klesavou kadencí. V tomto oddílu se znovu zmiňujeme o předchozím výzkumu a ukazujeme, že ne všechny předchozí výsledky mohly být potvrzeny. ´t Hart (1981) například odhadoval, že v případě angličtiny bude změna ve výšce tónu postřehnuta častěji ve frázích s klesavou kadencí, což náš experiment nepotvrdil. Harris a Umeda (1987) zase tvrdili, že není rozdíl mezi změnou ve výšce směrem nahoru a změnou směrem dolů. Naše studie naopak ukázala, že v případě frází se stoupavou kadencí bylo pro naše posluchače snazší poznat změnu směrem nahoru, v případě frází s kadencí klesavou tomu bylo naopak. V této části práce též zdůrazňujeme, že pro stanovení dalších závěrů by bylo třeba více materiálu i více testovaných posluchačů; zajímavé by bylo také srovnání experimentů, které by použily anglický materiál a anglické subjekty, nebo anglický materiál a posluchače s minimální znalostí tohoto jazyka.

APPENDIX

1 The original (full) text for the recording – Four intonational phrases were chosen for the experiment from those marked in bold with always the same intonational structure. Stressed syllables are in grey.

1) Three days before Christmas and the supermarkets are crowded with people, desperately searching for presents which will fit perfectly under the Christmas tree. What are the bestsellers of this season? Forget about cookery books and computer games! This year, **the bookshelves are loaded with books about gardening**, filmmaking and acupuncture. And to surprise your children, you´d better buy them a huge box of loom-bands.
2) Judy claimed they had visited the house together but her husband was of a different opinion. Surely, he would have remembered the carpet on the stairs and the high windows, too. **And he would have remembered the marvellous living-room!** It was so beautiful! If only they could afford to own one like that. He could almost see himself sitting on that large sofa drinking a glass of champagne.
3) Our neighbours had three children, John, Caroline and Susan. John spent most of his days playing football, **Caroline dreamed of becoming a minister,** and Suzie, the youngest, joined a theatre group and is now a famous actress.
4) Writing an article is always so difficult for me. I never know how to start, and whenever I decide to really concentrate on my writing, I can´t find the proper words to express myself. I feel useless. My head is completely empty. **Could anyone give me a synonym for "wonderful"**? Because I just can´t think of anything.
5) Daniel couldn´t watch the game the day before. So he found his best friend John, who told him the unpleasant news. The team they both supported had lost, and what was more, **their favourite player had then been humiliated.** It was awful. The other guys left him on the pitch and locked the door to their cabin[1], so that all the fans could laugh at him.
6) What he was telling me was simply wrong. **That can´t be my mother´s opinion on snowboarding**! She would never let me have such a dangerous hobby.
7) My mother and I had a big fight that day. Eventually, I escaped from the house with the intention to drive away but she followed me, leaving the pots on the stove. She jumped in front of my old car and started screaming in a desperate attempt to stop me. I had to laugh. **Was she planning to damage my car with a wooden-spoon?** That's what she was holding in her hand. I simply started the engine and backed out.
8) I was surprised: **Caroline dreamed of becoming a minister**. She had decided a long time ago, though. She said it was her dream job.

[1] I am aware of the fact that the word "cabin" cannot be used in this context; I should have used the term "changing room" instead. The word "cabin" was used for the recording but the whole paragraph 5 was not further worked with and so the inappropriate choice of the word could not have influenced the results of the perceptual test.

2 The list of items used in the test

The table on the left provides all phrases marked as SAME used in the perceptual test. The table on the right shows all phrases marked as DIFFERENT. Letters S, U, N indicate the type of a syllable (stressed, unstressed and nuclear), words *down* and *up* stand for the direction of the change in pitch. All items were repeated twice during the test.

SAME	
first phrase	**second phrase**
fall1	
fall1 S up	fall1 S up
fall1 U down	fall1 U down
fall1 N up	fall1 N up
fall1 N down	fall1 N down
fall2	
fall2 unmodified	fall2 unmodified
fall2 S down	fall2 S down
fall2 U up	fall2 U up
fall2 N up	fall2 N up
rise1	
rise1 unmodified	rise1 unmodified
rise1 S down	rise1 S down
rise1 U down	rise1 U down
rise1 N down	rise1 N down
rise2	
rise2 S up	rise2 S up
rise2 S down	rise2 S down
rise2 U up	rise2 U up
rise 2 N down	rise 2 N down

DIFFERENT	
first phrase	**second phrase**
fall1	
fall1 unmodified	fall1 S up
fall1 unmodified	fall1 S down
fall1 unmodified	fall1 U up
fall1 unmodified	fall1 U down
fall1 unmodified	fall1 N up
fall1 unmodified	fall1 N down
fall2	
fall2 unmodified	fall2 S up
fall2 unmodified	fall2 S down
fall2 unmodified	fall2 U up
fall2 unmodified	fall2 U down
fall2 unmodified	fall2 N up
fall2 unmodified	fall2 N down
rise1	
rise1 unmodified	rise1 S up
rise1 unmodified	rise1 S down
rise1 unmodified	rise1 U up
rise1 unmodified	rise1 U down
rise1 unmodified	rise1 N up
rise1 unmodified	rise1 N down
rise2	
rise2 unmodified	rise2 S up
rise2 unmodified	rise2 S down
rise2 unmodified	rise2 U up
rise2 unmodified	rise2 U down
rise2 unmodified	rise2 N up
rise2 unmodified	rise2 N down

3 The questionnaire for the listeners

DOTAZNÍK

Prosím, zakroužkujte:

1) Byly vám zjištěny nějaké sluchové vady?

ANO NE

2) Máte zkušenost s psychoakustickými testy?

ANO NE

Pokud ano, s jakými, s kolika:

3) Máte hudební vzdělání?

ANO NE

Pokud ano, prosím, napište jaké
(hra na nástroj, zpěv, tanec…):

Doplňte:

4) Z jakého regionu pocházíte?

5) Studiu angličtiny se věnuji ...... let.

Commentary on the questionnaire: The answers of the individual listeners are not provided as they do not form an essential part of our study. All listeners have reported having no hearing problems and being the subjects of none or at most two perceptual tests, which could not have possibly influenced our experiment. Twelve of them said to be experienced in music and this finding is worked with in the analysis but again, the individual answers did not seem to play an important role. The question about the part of the republic the speakers are from did not reveal any interesting information, and as they were all university students of English at approximately the same level, the exact number of the years they had been studying the language was not relevant either.