FILOZOFICKÁ FAKULTA
UNIVERZITY KARLOVY
V PRAZE

ÚSTAV ANGLICKÉHO JAZYKA
A DIDAKTIKY

BAKALÁŘSKÁ PRÁCE

Dagmar Hanzlíková

Perceptual evaluation of rhythmic features in Czech English

Percepční hodnocení rytmických parametrů české angličtiny

PRAHA 2013

Vedoucí práce:
doc. PhDr. Jan Volín, Ph.D.

I would like to thank most sincerely my thesis supervisor doc. Volín for all his helpful comments and advice he was always ready to provide, for his willingness to offer consultation whenever I needed it, and for his most valuable assistance with calculating statistics for the study. I would also like to thank other members of the Institute of Phonetics, namely Mgr. Weingartová for her help with calculating statistics for the study, and doc. Skarnitzl for providing me with useful literature and allowing me to use the rooms at the Institute of Phonetics for the experiment. Lastly, I would like to thank all the people who participated in the experiment.

I declare that the following BA thesis is my own work for which I used only the sources and literature mentioned, and that this thesis has not been used in the course of other university studies or in order to acquire the same or another type of diploma.

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, 13.8.2013 …….

Dagmar Hanzlíková

Abstract

The aim of this thesis is to investigate the effect of changes in speech rhythm on the perception of personality traits which are detectable in speech. The theoretical part is concerned with the description of some important theories in the field of psychology of personality, then with summarizing research done in the effects of foreign-accented speech on perception, and finally with presenting the issue of speech rhythm with a focus on representative studies of speech rhythm and its effect on speech perception. In an independently designed experiment, the changes in speech rhythm were created by manipulating the durational patterns and pitch contours. Subsequently a perceptual test was prepared. Twenty subjects were asked to evaluate individual speech samples and rate to what extent they would characterize the speaker by the given personality trait. The results from the perceptual tests were analyzed from multiple points of view: the effects of the manipulations on the perception in general, and then in relation to the nationality of the speaker (native vs. non-native speaker), to the personality trait in question, and in relation to the individual speakers and items. The results showed some tendencies for example in perception of honesty in relation to durational patterns, or in perception of nervousness in connection with changes in pitch.

Keywords: Czech English, foreign accent, judgement, perception, prosody, rhythm of speech, second language acquisition

Abstrakt

Cílem této bakalářské práce je prozkoumání vlivu změn rytmu řeči na vnímání osobnostních rysů, které jsou patrné v řeči. Teoretická část se věnuje popisu některých důležitých teorií z oblasti psychologie osobnosti, dále také shrnutí výzkumu v oblasti dopadu cizineckého přízvuku na vnímání řeči a nakonec se zabývá problémem rytmu řeči se zaměřením na reprezentativní studie rytmu řeči a jeho vlivu na vnímání řeči. V rámci experimentu byly prováděny změny v rytmu řeči manipulací temporální struktury a intonace. Následně byl připraven percepční test. Dvacet respondentů posuzovalo jednotlivé testové položky a hodnotilo, do jaké míry by mluvčího charakterizovali daným osobnostním rysem. Výsledky percepčního testu byly analyzovány z několika úhlů pohledu: obecný vliv manipulací na vnímání, a dále potom ve vztahu k národnosti mluvčího (rodilý vs. nerodilý mluvčí), k daným osobnostním rysům a ve vztahu k jednotlivým mluvčím a položkám testu. Výsledky ukázaly určité tendence například ve vnímání upřímnosti v souvislosti s temporální strukturou nebo ve vnímání nervozity v souvislosti se změnou intonace.

Klíčová slova: česká angličtina, cizinecký přízvuk, úsudek, percepce, prozodie, rytmus řeči, osvojování druhého jazyka

Table of Contents

1. Introduction

There have been many studies investigating speech perception; however, most of these were only concerned with evaluating how modifying certain acoustic features affects the intelligibility of speech, thus which aspects are important for speech perception. There have also been some papers that studied evaluation of personality based on speech only; however, almost none of these works attempted to relate the evaluation to particular acoustic aspects of speech, and those which tried to do so usually used such broad concepts as foreign accent, but they did not investigate which particular aspects of foreign accent cause the changes in perception. The present study will examine whether there is any relationship between changes in speech rhythm and changes in perceptual evaluation of personality traits, and whether the manipulations in rhythmic properties of speech cause comparable changes in perception for native English speakers and Czech learners of English.

The paper is divided into two parts. The first part of the study presents a theoretical background for the subsequent experiment and, because the work is interdisciplinary, the theoretical part is further divided into multiple sections. The first section (2.1) is concerned with psychology of personality, which is relevant because the present study investigates the perceptual evaluation of personality traits that are detectable in speech, and this section will describe the major theories in the development of psychology of personality, including that of Hippocrates, Jung, or the modern five-factor model. The following section (2.2) relates the issue of foreign accent in speech with particular attention being paid to the research in durational patterns of non-native speech. The third section (2.3) describes the problem of rhythm in speech, the way in which rhythm is measured and how it may possibly affect speech perception. Finally, section 2.4 introduces an original experiment and presents the hypothesis relevant to it.

The second part describes the experiment which was performed in order to see whether there is a link between changes in speech rhythm and perceptual evaluation of speech concerning personality traits. Chapter 3 presents the materials which were used in the experiment and the method in which they were handled. Section 3.1 describes the sound materials, the way they were selected, and the way the rhythm of speech was manipulated,

and then how the speech samples were allocated to blocks for the perceptual test. The next section (3.2) presents the perceptual test, namely the way the particular personality traits for the present study were selected and how the individual speech samples were allocated to the individual traits. The last section of this chapter (3.3) introduces the group of subjects who participated in the experiment and took the perceptual test, and also describes how the test was presented to the listeners and the conditions in which it was taken.

Chapter 4 presents the results of the experiment. In the first section (4.1) the effect of individual manipulations on speech perception is described. The next section (4.2) shortly introduces the general differences, caused by the manipulations, between native and non-native speakers of English. The section 4.3 relates the changes in perception of the personality traits to the individual manipulations, and the section 4.4 describes how the manipulations affected the individual speakers. The following section (4.5) is concerned with consistency of answers among the listeners, and the last section (4.6) focuses on the individual items analyzed in groups by each personality trait. Chapter 5 then discusses the issues encountered during the analysis of results in more detail and attempts to explain the results of the experiment.

2. Theoretical background

2.1 Psychology of personality

People have always been interested in psychology and it still seems to be one of the important branches of today's science. A general knowledge in psychology also appears to be useful in various kinds of jobs, be it marketing, where you need to sell products to the customers, or in media where they need to present information in a trustworthy manner. Moreover, not only psychologists use theories of personality to attempt to classify people based on their behaviour, but most people do it every day without even being aware of doing so: When somebody sees or talks to someone, they unconsciously label the person as being kind, patient, or cheerful etc. without knowing anything about psychology of personality. Many psychologists are also concerned with investigating on what basis people evaluate others and what can possibly influence the judgement. Among the obvious aspects influencing one's judgement is undoubtedly the physical appearance, but it is also the speech, in which case not only the contents of the speech matter, but also the voice, for instance pitch, timber, or rhythm of speech.

Nevertheless, people have always been attempting to classify human behaviour and categorize various personality types into groups. The first comprehensive theory in psychology of personality appeared already in Ancient Greece about 400 BC. Morton Hunt (1993) mentions Hippocrates among the first thinkers in the field of psychology. Hippocrates was originally a medical doctor, but unlike his predecessors, he refused to see illnesses as a manifestation of the will of gods, but rather he thought that they had natural reasons. He believed that good health is a result of bodily fluids (which he called humours) being balanced and he named the four fluids: *black bile*, *yellow bile*, *phlegm*, and *blood*. If one of the humours was lacking or exceeding, it resulted in a disease. He also applied this theory on the human mind to explain psychological condition. He claimed that people who are influenced by phlegm suffer from anxiety and fear, and they are unusually calm, or people who are influenced by bile are angry and noisy. Galen later extended this theory in his work *On the Temperaments* where he explained differences in temperaments based on the four humours defined by Hippocrates and laid basis to the modern understanding of the four temperaments. Galen explained that the sanguine temperament was related with the

excess of blood, the choleric temperament with the excess of yellow bile, the melancholic temperament with the excess of black bile, and the phlegmatic temperament with the excess of phlegm (Hunt, 1993).

Another psychologist who was very important for the development of psychology of personality was an analytical psychologist Carl Jung. Among his most important contributions to psychology of personality are the concepts of *introversion* and *extraversion,* which are still being used, and his theories of collective unconscious and the archetypes. He defined extraversion as the orientation of the individual on what is outside of the self, extroverts being talkative, assertive, and enthusiastic, and introversion as the orientation of the individual on one's self. He combines these two attitudes with four types of psychological functions: *sensation*, *intuition*, *thinking*, and *feeling*, in order to describe the main personality types (Drapela, 1997). A psychometric questionnaire which is still used today called Myers-Briggs Type Indicator (MBTI) was based on Jung's theory of personality types. For the description of personality, Jung's concept of collective unconscious is also important. He agreed with Freud on the idea of unconscious mind, however, Jung's concept of the unconscious was broader. Unlike Freud, who related the unconscious mind to an individual only, Jung claims that there is a collective unconscious which does not develop individually, but is inherited from the experience of all human race and it consists also of the archetypes (Jung, 1969). Jung defined the archetypes as patterns and images that are inherited and related to the collective unconscious, he compared them to the instincts of animals, and he claimed that they influence people's behaviour. Among these archetypes are for example *the Hero* or *the Martyr* (Drapela, 1997).

The newest and most influential model describing personality is called a five-factor model of personality. The beginnings of this model date back to 1940's when Cattell first attempted to study personality based on analyzing individual traits of behaviour (Drapela, 1997). He collected data from three sources: *Life data*, which tell about the behaviour of the individual in the society, and the data are collected for example from school or court records; the second source are *experimental data*, which inform about the individual's reaction to situations created in a lab; and finally, *questionnaire data*, which collect information from self-rating questionnaire where the subjects rate to what extent a statement applies to them. After collecting the data, Cattell applied a method called factor

analysis and derived sixteen fundamental factors, among these being for example emotional stability, dominance, or reasoning (Cattell, 1957).

Many researches in the area of factor analysis of personality have been conducted since Cattell's definition of the sixteen fundamental factors, eventually deriving the five-factor model of personality. Hřebíčková (2011) describes the personality characteristics from multiple points of view. First, she divides them into characteristics derived from observing and describing an individual person, and then to psychological determiners of behaviour, which are in fact hypothetical constructs. Then she also defines dispositions (also called personality traits) to be inborn or acquired preconditions to behave in a certain way; however, scientists have not yet fully agreed on whether the concept should cover personality characteristics or describe the inborn predispositions to certain behaviour. Since the 1990's, a group of scientists agreed that personality is best described by a five-factor model of personality which covers both the above mentioned concepts of personality traits. In the study of five-factor model of personality, two basic approaches are distinguished: lexical and questionnaire-based. The lexical approach analyzes words that could be suitable for describing personality and it is based on the assumption that the most important differences would be encoded in the language under single distinct words. The questionnaire approach does not deal with single words but rather attempts to describe the motives and inner functioning of individuals. McCrae and Costa (1983), following the questionnaire approach, derived three main personality dimensions: *neuroticism*, *extraversion*, and *openness to experience*, to which they later added two more dimensions, derived from the lexical approach: *agreeableness* and *conscientiousness*. Based on their research they created a questionnaire based NEO personality inventory which is now widely used in personality description. It is also necessary to mention that the five factors are not equally important. Norman (1963) arranged them based on their importance using Greek numerals: (I) *extraversion*, (II) *agreeableness*, (III) *conscientiousness*, (IV) *neuroticism*, (V) *openness to experience*.

Most of the research conducted in the five-factor analysis was done in English and it was discovered that simple translations of the English taxonomy does not work for other languages; therefore, it was necessary to design a specific taxonomy for specific languages, based on the five-factor model. For the purposes of the experiment in this study, a Czech

taxonomy needed to be used, because the respondents were Czech students. Taxonomy in the Czech language was studied and eventually compiled by Hřebíčková (2011). First, a list of all personality-relevant words had to be selected from a dictionary, among these being adjectives, e.g. *líný* (lazy), attributive nouns, e.g. *lenost* (laziness), descriptive nouns, e.g. *lenoch* (a lazy person), and verbs, e.g. *lenošit* (to laze). Subsequently, the list needed to be reduced by removing archaic, dialectal or rarely used words. Eventually, after further reductions, 366 adjectives describing personality traits were selected and presented to 429 respondents to verify whether all the adjectives are clear and easily understood, so that they can be used for further analysis. Finally, the adjectives were analysed and divided into five groups, matching the five factors of the English five-factor model: *extraverze* (extraversion), *přívětivost* (agreeableness), *svědomitost* (conscientiousness), *emocionální stabilita* (neuroticism), and *intelekt* (openness to experience) (Hřebíčková, 2011). These adjectives were then used for the personality description in the practical part of the study (see section 3.2).

2.2 Non-native accents of speech

Undoubtedly, the first impression judgement plays an important role in human life, and not only for people who are evaluating somebody, but also for people who are being evaluated by others. The first impression is created mainly by the physical appearance, be it clothing, the person's figure, or facial expression; however another aspect significantly influencing the first impression is undeniably the speech. Among the most important aspects of speech that influence our judgement are certainly the pitch and the timber of the voice, but perhaps even more importantly any kinds of speech impediments, or foreign accent. In general, if a person is difficult to understand, other people will very likely avoid talking to this person. Edwards (1999) summarizes that there have been numerous researches in the area of language attitudes and he claims that speech characteristics can be related to different social evaluations. Further, he says that the most probable reason for different social evaluations of different accents and dialects is that the variation in speech-evaluations reflects social perceptions of the speakers of the given dialect rather than intrinsic qualities of the dialect itself. He also emphasized the role of cultural stereotypes in perception of language dialects and he pointed out the fact that very little research has been done in relating the speech evaluations to particular speech attributes.

Munro (2003) investigated discrimination based on foreign accent in the Canadian context; however, his conclusions might be applicable in other contexts as well. In his study he presented three types of discrimination based on accent. First, a stereotyping, which he explains as that if a person is prejudiced against people of certain nationality, then this person might discriminate somebody based on the assumption that he recognized the foreign accent of the said nationality. Munro (2003) illustrates this point on an example: If someone is biased against the Iraqis, he may recognize a Middle Eastern accent on the phone and deny the speaker a service or employment, assuming that the speaker is from Iraq. The second type he calls harassment, which might occur for example at work, if a person is mocked for his foreign accent by his co-workers. The third type, he says, occurs when a potential employee is denied a job being told that his accent is unacceptable for the job, even if the speaker is intelligible or the job does not require language skills. However, Derwing and Munro (2009) emphasize that the foreign accent itself does not cause discrimination but rather the intolerant listeners are to be blamed.

Derwing and Munro (2009) offer an interesting view on foreign accents. They claim that listeners are extremely sensitive to the presence of a foreign accent in speech; however, they think that foreign accent does not necessarily cause problems in communication. Derwing and Munro (2009) stress the importance of intelligibility (which they define as the degree of a listener's understanding of an utterance) because they demonstrated that some speakers who were perceived as having heavy accent were still perfectly intelligible. Although they present some examples of foreign accent being beneficial (e.g. some European accents are associated with sophistication), they still admit that accents may have an unwanted social impact on the speaker, among the most important of these naming the loss of intelligibility. Furthermore, even though they agree on the importance of learning correct pronunciation, they also emphasize that the listener's attitude often plays an important role in communication because the listener may have prejudices against the particular accents or they might be simply convinced that they cannot understand the speech because it is accented which results in their actual not understanding.

Among the studies that are concerned with finding acoustic correlates to differences in speech perception were Tajima et al. (1997). Their research is interesting for the present study because they investigated the relationship between intelligibility of foreign-accented

utterances and the temporal patterning of speech, which is closely related to the rhythm of speech. There are many levels at which the temporal structure works: first, at the segmental level, e.g. a tendency for vowels to be longer before a lenis sound than before a fortis sound, second, at the level of syllables, e.g. different phonotactic constraints on syllable shape or preference of certain syllable types over others, and third, beyond the level of syllables, languages have been described as being stress-timed, syllable-timed, or mora-timed (Tajima, K. et al., 1997). It has been demonstrated that learners of the second language tend to apply phonological knowledge of their native language onto the second language; therefore, the temporal structure of their utterance deviates from that of a native speaker (e.g. Flege & Port, 1981, or Tarone, 1980), and it also has been shown that native listeners are sensitive to such deviations. In their study, Tajima et al. (1997) used the speech transformation method and applied it on short English phrases produced by a Chinese speaker of English and by a native English speaker. The Chinese speaker's productions were temporally modified so that the duration of acoustic segments matched the duration of corresponding segments in the native speaker's utterances, and the native productions were temporally distorted to match the Chinese-accented speech. Using this method, the effects at the segmental level could not be separated from the effects at the higher levels; therefore, the combined effect of temporally defined properties on intelligibility was analyzed. The results showed that the temporally corrected Chinese English was significantly more intelligible than the original Chinese English, while for the temporally distorted English speech samples the intelligibility significantly decreased. These results mean that the native listeners' ability to recognize English phrases is significantly influenced by whether or not the phrases have appropriate native-like properties, as established by a native speaker; however, the intelligibility of the temporally corrected Chinese English was still much lower than that of the temporally distorted English spoken by a native speaker, which suggests that there are other factors apart from the temporal structure that affect the perception and intelligibility of speech (Tajima et al., 1997).

Another study that deals with the non-native durational patterns in speech in relation to the speech intelligibility is that by Quené and Van Delft (2010). While Tajima et al. (1997) were unable to evaluate the relative contributions of the individual factors of temporally modified speech due to methodological difficulties (see above), Quené and Van Delft

(2010) attempted to overcome these difficulties by using the Speech Reception Threshold method (described by Plomp and Mimpen, 1979). They tried to assess the relative contribution of native vs. non-native durational patterns to intelligibility, relative to the native vs. non-native speech sounds, and to quantify the relative contributions of segmental errors and of durational patterns to intelligibility. For the purposes of their experiment, the patterns of segmental durations between the native and non-native realizations of the sentences (the native language in this experiment being Dutch) were exchanged, and differences in tempo and pitch contours were removed not to affect the results. In agreement with the research of Tajima et al. (1997), the results of this study showed that intelligibility decreases if the native speaker's speech is manipulated to have non-native durational patterns, and that intelligibility of non-native speaker's speech manipulated to have native durational patterns increased. Moreover, quantifying the effects suggests that if the speech contains more non-native speech segments, then durational patterns are less relevant for intelligibility. On the other hand, the effect of temporal manipulations proved to be larger if the non-native durations were more deviant from the native durations. To summarize, because Quené and Van Delft (2010) performed their experiment on the Dutch language and used a different method than Tajima et al. (1997), yet they reached the same conclusion, it is possible to generalize that durational patterns are indeed important for the perception of the speech as being native or foreign-accented, even though there are other factors influencing it too.

2.3 Speech rhythm

Rhythm is, without any doubt, very important for people in general, not only in speech. Volín (2010) comments that there are some regularly occurring types of human behaviour which can reveal notable facts about the role of rhythm in our lives, such as the infants who take pleasure in nursery rhymes, or crowds that demonstrate their ideas in chanting. In speech, rhythm is important not only for producing speech but especially for its perception. For the process of perception, as Volín (2010) claims, to recognize a linguistic unit, a neural assembly in the brain which is assigned to the given object has to perform an act of resonance and if the speech has natural rhythm, it is easier for the brain to decode the message. It has also been shown (Buxton, 1983) that reaction times in tasks with natural speech rhythm were considerably shorter than the same tasks performed on speech with manipulated temporal structure.

To begin with, languages are in general classified into three groups based on their rhythmic features: syllable-timed, stress-timed (first described by Pike, 1945), and later discovered mora-timed. As Mok and Dellwo (2008) say, the original suggestion was that there are quasi-isochronous durational units in speech to support such classification, which are: length of syllables for syllable-timing, inter-stress intervals for stress-timing, and mora duration for mora-timing. However, no acoustic evidence for such isochronous units could be found. Nevertheless, this classification seemed appropriate because there are several important aspects in which stress-timed and syllable-timed languages differ from one another, that is: syllable structure, vowel reduction, and stress. Unlike syllable-timed languages, stress-timed languages show more variation in syllable length and structure, they have more reduced unstressed syllables, more stress-related rules and more variation in phonetic realization of stress (Mok & Dellwo, 2008). All these aspects are combined to create the impression of the stress-timed language as opposed to syllable-timed language. In order to support the rhythmic classification of languages, further research in measuring rhythm in language and finding acoustic correlates that would reflect the auditory impression of different rhythmic classes was necessary.

An important study in speech rhythm measurement was done by Ramus et al. (1999) who based their research in phonetic definition of language rhythm on the assumption that infant speech perception is centred on vowels because vowels have more energy than consonants and they also carry stress. Therefore, they assumed that infants perceive speech as a succession of vowels alternating with periods of unanalyzed noise (consonants). Their hypothesis was that segmenting speech into consonantal and vocalic intervals can account for the standard stress-timing or syllable-timing dichotomy and possibly discover other types of rhythm, and that it may clarify how rhythm can be extracted from the speech signal. From their experiment, they derived three variables: First, the proportion of vocalic intervals within a sentence marked as %V, second, the standard deviation of the duration of vocalic intervals within each sentence marked as ΔV, and third, the standard deviation of the duration of consonantal intervals within each sentence marked as ΔC. They found out that out of these three variables, %V and ΔC seemed most suitable for description of the standard rhythm classes because they appear to be directly related to syllabic structure. Their further experiments showed that out of these two variables, %V is more important

for perception of speech rhythm. To sum up, Ramus et al. (1999) found measurements of the speech signal that support the traditional classification of languages into rhythmic classes.

Mok and Dellwo (2008) draw on this research in their study of native and non-native speech rhythm. Apart from using the measures by Ramus, i.e. %V and ΔC, they also used Pairwise Variability Index (PVI) of vocalic and consonantal durations described by Grabe & Low (2002), and a standard deviation of syllabic durations (ΔS). They claimed that although American and British English were typical stress-timed languages, other accents of English may belong to a different rhythmic class. They compared English with Cantonese and Beijing Mandarin, and with Cantonese English and Mandarin English. While Mandarin and Cantonese were both categorized, using the acoustic measures, as syllable-timed languages, the experiment showed that some of the measures, ΔC for example, group Cantonese English and Mandarin English with stress-timed languages; however, some other measures, namely VarcoC: (ΔC / mean consonantal duration) x 100, group the two accents of English with syllable-timed languages, which highlights the issue of using the acoustic measures to determine speech rhythm of non-native speakers. Mok and Dellwo (2008: 426) suggested that slower speaking rate and selective lengthening in Cantonese English and Mandarin English contributed to the discrepancy between the results based on rhythmic measures and the subjective impression of their rhythm because both accents of English gave the impression of being syllable-timed, although some rhythmic measures would classify them as stress-timed.

Further research in validating acoustic measures of rhythm for non-native speakers was conducted by Kinoshita & Sheppard (2011). Unlike Mok & Dellwo (2008), who considered various acoustic measures for their research (see above), Kinoshita and Sheppard (2011) were concerned only with the Pairwise Variability Indexes (PVI) in their study. In their experiment, they asked expert native speakers to rate non-native speakers' rhythm, and compared their perceptual evaluation with the results based on PVI. On the basis of their results it was again confirmed that mainly vowels (their duration, in particular) were important for perception of the speech rhythm. Furthermore, the experiment showed that different raters often chose different aspects of speech rhythm when they were evaluating the same speech samples and that a single rater would change

the criteria used to evaluate speech rhythm depending on the speech extract; however, there seemed to be some underlying common factor. Kinoshita and Sheppard (2011) also demonstrated that there was a link between the ratings and variability in pairwise vowel lengths (nPVIv), which indicated that native speakers used variability in vowel length for judging non-native speakers' speech rhythm; however, the result was not replicated in all the data, which they explained by the fact that when non-native speakers attain a native-speaker level, the PVI measures no longer contribute to variability in the ratings. Nevertheless, their results suggest that PVI is an effective measure of non-native speech rhythm and can be useful in second language acquisition research; even though further investigations need to be done to discover other possible acoustic rhythm correlates which could account for the variance in the ratings.

Another study in speech rhythm was carried out by Quené and Port (2005) who were concerned with regularity of inter-stress timing. They claimed that previous studies had suggested that speech rhythm only had a weak effect on spoken-word perception; however, none of these studies controlled the timing of inter-stress intervals. They mentioned as well that not only the speaker's behaviour is periodic, but that the listener is sensitive to resultant speech rhythms too (as already mentioned above) because the listeners seem to pay more attention to the points where the rhythmic beats are strongest. In their research they found out that timing regularity, which rests upon the difference between rhythmically and arrhythmically arranged onsets of stressed syllables, contributed significantly to spoken-word perception and that, if words in a list were temporally aligned to regular inter-stress intervals, the words were relatively easy to perceive; therefore, they confirmed that rhythm in speech is indeed important for the listeners' perception. Another variable, apart from the timing regularity, which was described, was metrical expectancy which was concerned with variation of iambic and trochaic words in the chain of words. This variable, however, did not prove to be important for speech perception and it did not matter whether the words before the target were of the same metrical pattern or not. Volín (2010) comments that this result should not be surprising because people do not communicate in poetry; therefore, the metrical pattern is not as important for the speech perception as the regular onsets of stressed syllables.

Volín (2010) describes speech rhythm as a perceptual phenomenon, rather than a property of the acoustic signal. In his article, Volín acknowledged the global rhythm metrics (described above); however, he claimed that although these metrics correlate with the impressionistic evaluation of speech, they cannot explain the nature of the problem. He pointed out that the problem of the global rhythm metrics is that they rely wholly on raw durations and do not take into account that the percept of speech rhythm is also strongly influenced by the fundamental frequency of the speaker's voice and spectral properties of phones, which was demonstrated for example by Donovan and Darwin (1979) to be important (see below). Volín, therefore, concluded that future research in rhythm needs to take into account the role of spectral properties of speech segments and fundamental frequency as well as duration of speech units.

As mentioned above, not only raw durations are important for perceiving speech rhythm, but also intonation. An important study in this field was presented by Donovan and Darwin (1979), who investigated a possible contribution of intonation to perceived rhythm. In their experiment they used three sentences that differed in their prosodic and syntactic structure:

1) *Tim's in Tuscany's Training Troops.*
2) *Tim's in Tuscany Training Troops*.
3) *Tim's in Tuscany | Training Troops.*

Sentences 2 and 3 contained a major syntactic boundary in the middle foot, but in sentence 2 this was not marked by a tone group boundary. Sentences 1 and 2 were acoustically identical except for the /s/ in *Tuscany's.* The subjects were told the context in which such sentences might occur and were asked to tap to every stressed syllable. The results of this experiment showed that the number of tone groups had a significant effect on perceived rhythm, while the syntactic structure did not (see Figure 1 below).

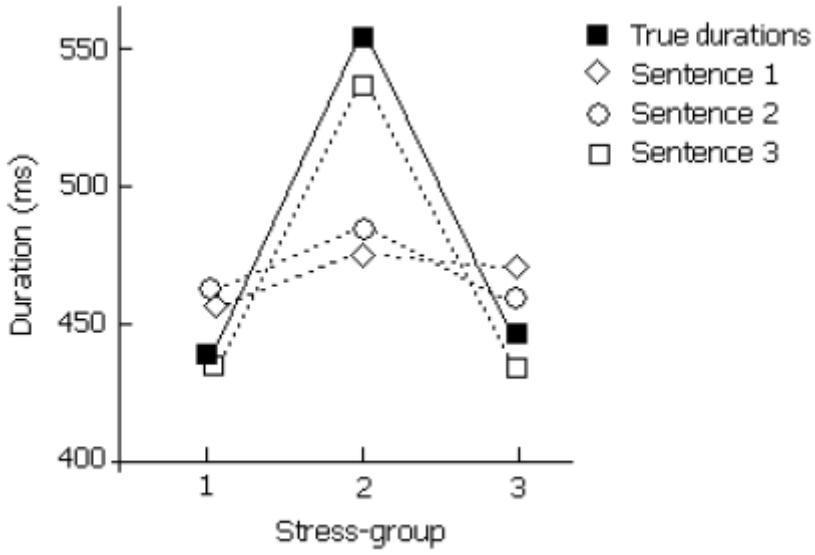


Figure 1. A figure showing physically measured and perceived durations of foot durations for the three sentences. The filled squares refer to actual durations and the empty shapes show the perceived durations (figure taken from Volín, 2010).

No tendency towards perceived isochrony was observed in sentence 3; however, there was a strong tendency towards perceived isochrony in sentences 1 and 2. From this Donovan and Darwin (1979) concluded that subjects perceived one and two tone-group sentences differently, although the foot durations were identical. This result clearly shows that intonation plays an important role in perception of speech rhythm; however, as Volín (2010) adds, the problem is that the acoustic correlate of intonation (fundamental frequency) is not in any straightforward relationship with its perceptual effects.

The conclusion drawn by Volín (2010) is important for the present study, in which the relationship between manipulated speech rhythm and changes in perception of the speaker's personality was examined. Therefore, for the purposes of the experiment, not only the duration of stressed and unstressed syllables was manipulated, but also the fundamental frequency of stressed and unstressed syllables.

2.4 Hypothesis

In the following part of the study, an experiment will be performed to investigate to what extent the speech rhythm is important for speech perception and evaluation of the speaker. Durational patterns and pitch will be modified, and subsequently a group of listeners will

evaluate the speakers on a 7-point scale. Unlike most of the previous studies that investigated primarily the intelligibility of speech or measuring how particular acoustic factors affect speech perception, the present experiment will investigate whether there is any relation between changes in speech rhythm and changes in perception of particular personality traits. The null hypothesis says that the modifications have no effect on the perception of personality traits in speech. However, some alternatives might occur, for example a particular modification might have a different effect on the speech of a native speaker than it has on the speech of a non-native speaker, or some modification might cause the non-native speaker to be perceived like a native speaker.

3. Materials and method

3.1 Sound materials

The sound material for the present study was selected from the Prague Phonetic Corpus (Skarnitzl, 2010). Twelve extracts of speech of eight individual non-professional speakers were used for the analysis; all of the eight speakers were female university students or employees, aged between 20 and 45 years. Four of them were native speakers of English coming from the South of England (AHAN, KBAN, KLAN, MVAN), and the other four were Czech learners of English (KLIA, PLDA, SMRA, VLHA). The Czech speakers were students of English studies and were recorded at the Institute of Phonetics at Charles University. All the speakers were asked to read a news bulletin which had been retrieved from 2002 BBC news, and they had enough time to get acquainted with the text to prevent disfluencies when reading. The recordings were made in a sound treated room and digitized at the sampling rate of 22 050 Hz (Skarnitzl et al., 2005).

3.1.1 Speech extracts

Twelve speech extracts were cut from the recordings taken from the Corpus, so that each of the sequences was approximately 5 seconds long and it was a completed utterance. It was important that none of the extracts contained any hesitations or disfluencies, words which were difficult to pronounce, or words which could evoke emotional response in the listeners, such as *war*, *death*, or *terrorists*. Eventually, two semantically different sequences by eight individual speakers (as defined above) were selected: 'The former United States president, Jimmy Carter, is in the Cuban capital Havana for a five day visit' (further referred to as 'Jimmy') and 'they were repairing the roof of one of the hangars used for assembling and testing space vehicles' (further referred to as 'hangar'). If there were two extracts by the same speaker, they were semantically different. These short sequences were normalized using the Sound Forge editing program to -6 dB in order to remove differences in perception caused by different sound-pressure level of the recording.

Using the sound analysis program Praat (Boersma & Weenink, 2013) the twelve speech samples were annotated with orthographic transcription, and stressed and unstressed vowels were labelled. All vowels which should canonically receive stress were manually

labelled following the guidelines set out by Skarnitzl & Machač (2009), regardless whether the stress was realized or not which was sometimes the case with Czech speakers. The nearest following unstressed vowel was labelled if realized as a vowel sound; if the nearest following unstressed vowel was elided, then the nearest preceding unstressed vowel was marked (see Figure 2 below).

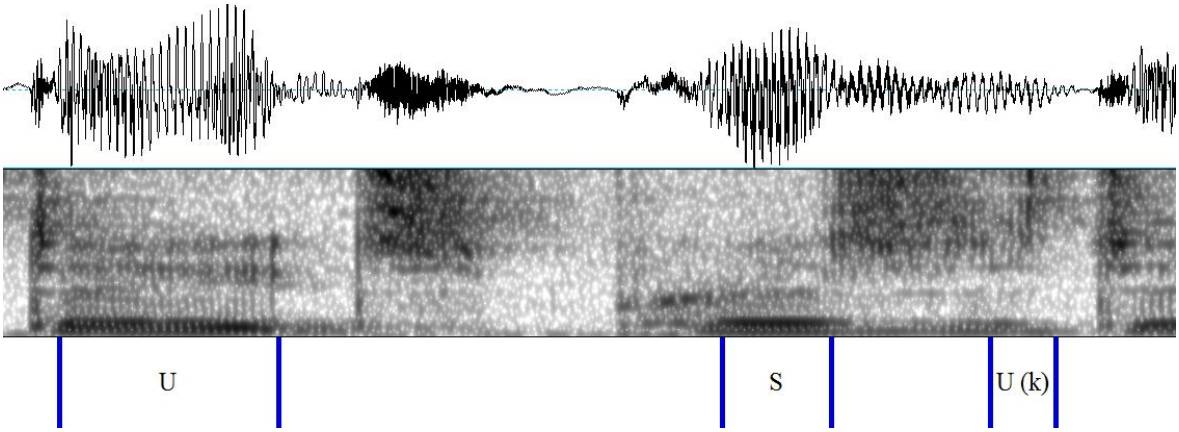


Figure 2. This is an example of a labelled stressed and unstressed vowel. In the word 'president', there is a stressed vowel /e/ marked with S, the nearest following unstressed vowel was elided and marked with U (k), therefore, the nearest preceding unstressed vowel from the word 'states' was used for the analysis, marked with U.

3.1.2 Manipulation

Two following manipulations were carried out on the speech samples using the program Praat (Boersma & Weenink, 2013): manipulations in time and in pitch (fundamental frequency standing for pitch). Because vowels turned out to be more important for perception of speech rhythm than consonants (Ramus et al., 1999), manipulations were carried out on vowels. For the manipulation in time, an object Manipulation was created and the values of duration points at the borders of the stressed vowels were moved to value 0.5 to make the duration of the vowel a half shorter; respectively, the values at the borders of unstressed vowels were moved to value 2 to make the duration twice as long. Additional points at the value 1 were added slightly before the left border and slightly after the right border to keep the rest of the sound without any manipulations (see Figure 3 below).

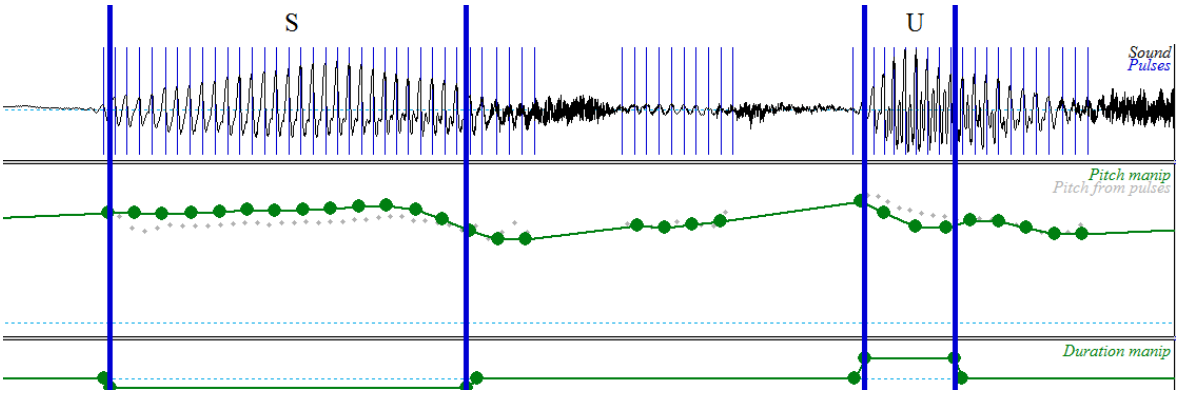


Figure 3. An example of a manipulated sound: The bottom tier shows the moved duration points at the borders of the stressed and unstressed vowels. The middle tier shows the manipulation in pitch where the gray dots show the original pitch and the green dots show the manipulated pitch.

Manipulation in fundamental frequency (F0) was done on the Manipulation object as well. The mean value of frequency in the middle third of the stressed, respectively unstressed, vowel was measured because only the middle third is important for pitch perception; in case the vowel was longer than 100 ms, a longer part was considered (Volín, 2009). The values were exchanged between the stressed and the nearest unstressed vowel and the value of frequency at the beginning and at the end of the vowels was manually adjusted to create a smooth transitional area in order to avoid any unnaturally fast changes in pitch (see Figure 3 above). The frequency values of the last pair of stressed and unstressed vowels were not exchanged because the effect of the end of an intonation phrase is strong and such manipulation could attract too much of the listeners' attention.

When the manipulations in time and frequency were done, the pitch tier and the duration tier from the Manipulation object were extracted. Then the Manipulation object was re-synthesized using the 'overlap-add' method in Praat (Boersma & Weenink, 2013), first with only the duration tier replaced, second, with only the pitch tier replaced, and third, with both the duration and the pitch tiers replaced. Thus we got the twelve extracts in four variations: with zero modification (further referred to as 0-modification), modification in duration (further referred to as T-modification), modification in pitch (further referred to as F-modification), and modification in both duration and pitch (further referred to as TF-modification).

3.1.3 Allocating extracts to blocks

For the purposes of this experiment, the forty-eight extracts were divided into four blocks (A, B, C, and D), following multiple rules. First, each of the four variations of one extract was in a different block. Second, if there were two text versions of the extracts by the same speaker in one block (for example AHAN-Jimmy and AHAN-hangar), they always had a different modification, and the distance between the two within one block was maximised. Third, the distance between the extracts by the same speaker (regardless of the text version) across the block boundary was maximised. Fourth, none of the blocks starts with the same speaker. Fifth, there was never the same sequence of two text versions of extracts in any of the blocks; for example, if KLAN-hangar follows MVAN-hangar in block A, this sequence never appeared in any other block. Sixth, alternating British and Czech speakers, and different modifications within one block were attempted, this being the weakest rule. In addition to the twelve extracts in one block, three more extracts were added, following the above defined rules, in order to measure consistency of answers in the perceptual test. Therefore, each of the resulting blocks contained fifteen extracts.

After the extracts were organised into the four blocks, the individual items of each block were connected into one sound file, so that there was a 9 seconds long silence between the individual items, then approximately 4 seconds long desensitisation sound and a short signal tone 0.5 second before the next item so that the subject starts to concentrate on the following item. Only two seconds of silence and the signal tone were placed before the first item of the block. The desensitisation sounds and the tone were normalized to -18 dB not to be intrusive and irritating for the subject.

3.2 Perceptual test

Instructions and data collection sheets were prepared in the mother tongue of the subjects, that is Czech, in order to avoid misinterpretation of the meaning of words used in the test. Data were collected from two kinds of tests: a unipolar adjective-rating scale for blocks A and C, and a questionnaire scale for blocks B and D. For both the adjective scale and the questionnaire scale, a 7-point scale was used. At first, four personality traits needed to be chosen for the analysis. The personality traits were based on the Five-Factor Model of Personality (Costa & McCrae, 1987); however, as the test needed to be in Czech, an inventory of Czech adjectives representing the five factors of Costa & McCrae (1987)

formed by Hřebíčková (2011) was used. To select the four personality traits, five people, who were not to take the test, were asked to listen to the non-modified speech samples and note down between five and ten adjectives from the inventory by Hřebíčková (2011: 93) according to what personality traits they think they can detect in the voices. From the five lists, four adjectives with the highest frequency of occurrence were selected: *upřímný* (honest), *nervózní* (nervous), *dobrosrdečný* (kind), and *prosazující* se (assertive). For each of the traits, three statements expressing the particular trait were formed based on the NEO Personality Inventory mentioned by Costa & McCrae (1987); for example, *'Tento člověk působí důvěryhodně'* (This person seems trustworthy) for *upřímný* (honest).

After selecting the traits and creating the statements, three sound items were allocated to each trait, so that there were two extracts by the same speaker and one extract by a different speaker of a different nationality (as shown in Table 1 below). Each item was tested for one personality trait only. At the questionnaire part of the test, one statement was randomly allocated to one item, the only criteria being that the statement expresses the particular trait measured for the item, and it was used for the item in both the block B and D (see Appendix A for the statements and their allocation to individual items); in blocks A and C, antonyms were sometimes used for the adjectives, usually if there would be the same adjective used twice in a row: *zdrženlivý* (reserved) for *prosazující* se (assertive), and *necitelný* (cold) for *dobrosrdečný* (kind). (See Appendix B for the example of the testing sheet).

nervózní	dobrosrdečný	prosazující se	upřímný
Cz VLHA-Jimmy	Cz SMRA-Jimmy	Br KLAN-Jimmy	Br AHAN-Jimmy
Cz VLHA-hangar	Cz SMRA-hangar	Br KLAN-hangar	Br AHAN-hangar
Br MVAN-hangar	Br KBAN-hangar	Cz KLIA-Jimmy	Cz PLDA-Jimmy

Table 1. Showing an example of the allocation of speakers to the personality traits. For translation of the names of personality traits see text.

3.3 Subjects

Twenty subjects, university students or graduates, with their age ranging between 20 and 30 years were asked to participate in the test. The subjects were Czech speakers of English who had to prove their knowledge of English, especially of the sound of the English

language (students of the English and American Studies, holders of the FCE certificate etc.). Each subject was tested individually in a sound treated room. They were acquainted with the contents of the speech extracts and they were advised not to pay attention to the content but rather concentrate on the feeling the speaker's voice evokes in them. The subjects were also told that there was no correct or incorrect answer. By stressing this aspect we wanted to relief the subjects' anxiety of being assessed. For the blocks A and C, they were asked to rate on a 7-point scale to what extent they would characterise the speaker by the given adjective; in the blocks B and D, they were asked to rate on a 7-point scale to what extent they think the statement holds true about the speaker, regardless of the content of their speech. Each block was approximately 5 minutes long and there were short breaks between the individual blocks, so that the subject could read the following test carefully before the next recording started to play.

After the perceptual tests were collected from the subjects, the data were entered into a MS Excel table. If there was a statement of a different polarity than the personality trait measured, or if there was an antonym used for the personality trait, the scores were inverted. Subsequently, a mean assessment score was calculated for each item using the arithmetic mean.

4. Results

In this chapter, the data collected from the perceptual test are analysed. In the results, the mean assessment scores were used for the analysis (for the table with mean scores and standard deviations see Appendix C). In section 4.1, only the manipulations are analysed, comparing whether the manipulations in general had any effect on the perception of the speaker. Section 4.2 links the differences of results in manipulations with the nationality of the speakers. The following section (4.3) elaborates on the differences found among the four personality traits and section 4.4 links these to the individual speakers. Section 4.5 is concerned with the consistency of answers among listeners and the last section (4.6) presents an analysis of the individual items.

4.1 Manipulations

Figure 4 (below) shows the mean assessment score (further referred to as mean score) for individual modifications, collected in the perceptual test across the speakers in order to see whether any of the scores of the three modifications shows in general any considerable difference compared to the 0-modification.

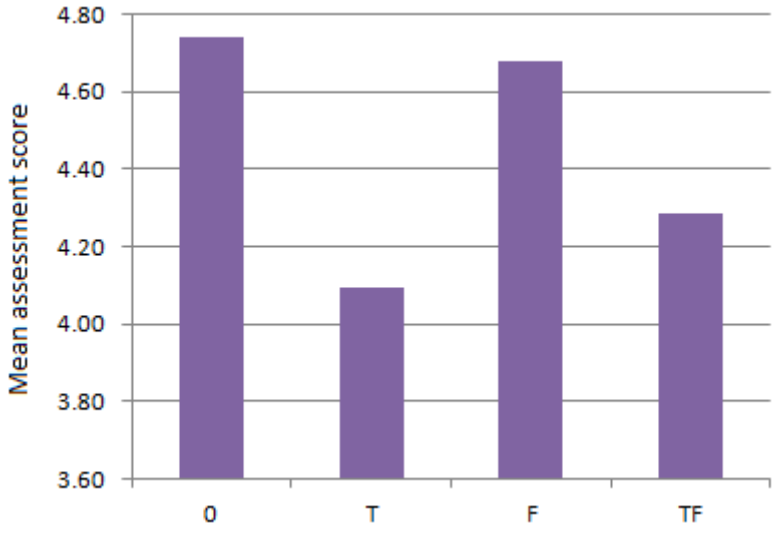


Figure 4. The mean assessment score of each of the modifications across the speakers. The ratings were measured on a 7-point scale. In the figure, 0 stands for 0-modification, T for T-modification, F for F-modification and TF for TF-modification.

While the mean score of the F-modification does not differ significantly from the mean score of the 0-modification, the T-modification shows the biggest difference compared to the mean score of the 0-modification. The TF-modification shows a decrease of the mean score but not as big as it is for T-modification only. However, neither of these differences proved to be significant in one-way analysis of variance: $F(3, 44) = 1.33$; $p = 0.28$.

4.2 Nationality

Figure 5 (below) presents the mean scores for individual modifications (as shown already in Figure 4 above) and adds the information about the nationality of the speakers.

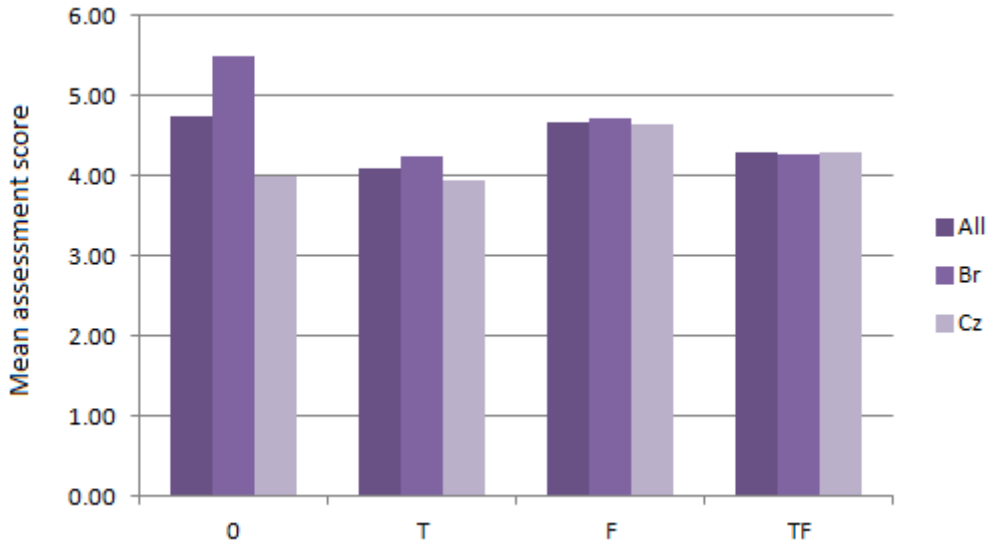


Figure 5. Differences of the mean scores between Czech (labelled Cz in the figure) and British (labelled Br in the figure) speakers of English, compared to the mean score of all items. In the figure, 0 stands for 0-modification, T for T-modification, F for F-modification and TF for TF-modification.

There is a marginally significant difference between the mean score of 0-modification between Czech and British speakers of English (post-hoc Tukey HSD carried out after one-way analysis of variance: $p = 0.072$). However, when we had modified the speech samples, the difference between Czech and native speakers disappeared. Furthermore, the mean scores of the native speakers of English always decreased compared to the 0-modification, regardless of the individual speaker or personality trait. The situation is rather different with the Czech speakers. The differences between the scores of the individual

modifications are small, and while the T-modification causes a slight decrease in the mean score of the personality traits, the F-modification and TF-modification cause an increase in the mean score compared to the 0-modificaton.

4.3 Personality traits

This section is concerned with the differences between the four personality traits caused by the modifications. Figure 6 (below) shows the connection between changes in manipulation and the mean score for the individual personality traits, regardless of the nationality of the speaker.

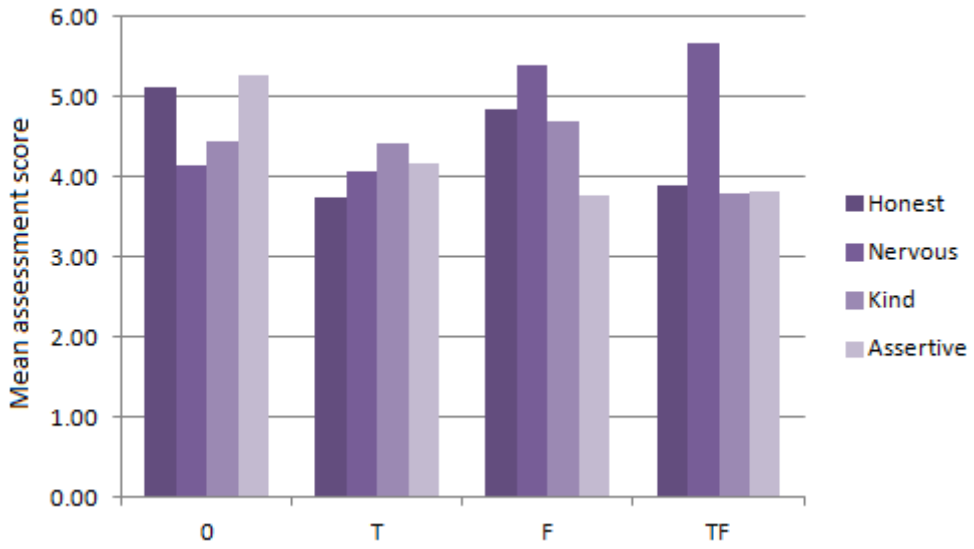


Figure 6. Differences of the mean score of the four personality traits as dependent on the manipulations. In the figure, 0 stands for 0-modification, T for T-modification, F for F-modification and TF for TF-modification.

As regards the attribute *upřímný* (honest), we can see that, in general, all three modifications caused a decrease of the mean score of the attribute, meaning that people seem more dishonest if any kind of manipulation was conducted. While the manipulation in frequency caused only a slight decrease of the mean score, the manipulation in time seems to be more important for perception, possibly causing the manipulation in both time and frequency to reach a lower mean score as well. On the other hand, manipulation in frequency seems to cause an increase of the mean score for the attribute *nervózní* (nervous), the scores for both the F-manipulation and the TF-manipulation being over 5.40

compared to the mean score 4.13 for the 0-modification (see Appendix C for standard deviations). T-modification causes no significant difference in the perception of the personality trait *nervózní*. The attribute *dobrosrdečný* (kind) does not prove to be perceptually significant in speech even though three out of the five people who were selecting the adjectives in the initial stage of preparing the perceptual test (see section 3.2) agreed on recognizing some of the speakers as being kind or cold. The mean score of all four versions of the speech sample centres about the score 4, which means that the listeners who participated in the perceptual test considered the speakers neither cold, nor kind, regardless of the manipulations conducted. The personality trait *prosazující se* (assertive) seems to be in general high for all speakers with the 0-modification. All of the three modifications cause a decrease of the mean score for this personality trait.

In general, the TF-modification seems to influence negatively all of the four personality traits (for the attribute *nervózní* (nervous), the increase of the mean score signals negative influence because the listener perceives the speaker as more nervous). The speakers seem to be colder, less assertive, less honest and more nervous, compared to the 0-modification. The F-modification alone seems to influence negatively only the attributes *nervózní* (nervous) and *prosazující se* (assertive). Although *upřímný* (honest) is influenced negatively as well, the difference is not big enough to be important. The attribute *dobrosrdečný* (kind) shows only a slight increase for the F-modification, compared to the 0-modification. The modification in time seems to influence negatively the traits *upřímný* (honest) and *prosazující se* (assertive), but it has no significant effect on the other two attributes (nervous and kind).

In conclusion, two-way analysis of variance showed only a marginal significance of the interaction between the modifications and personality traits: $F(9, 32) = 1.97$; $p = 0.08$. The tests of follow up comparison did not produce any interesting outcomes, but it is clear that this result is mainly due to the behaviour of the factor of nervousness.

4.4 Individual speakers

It is also important to link the general observations about modifications, nationality, and personality traits to individual speakers. Figure 7 (below) presents the eight individual speakers and how each of the modifications affected the perceptions compared to the

original recording (0-modification). If there were two semantically different speech samples for one speaker (the case of AHAN, KLAN, SMRA, and VLHA), the arithmetic mean of the two scores was calculated.

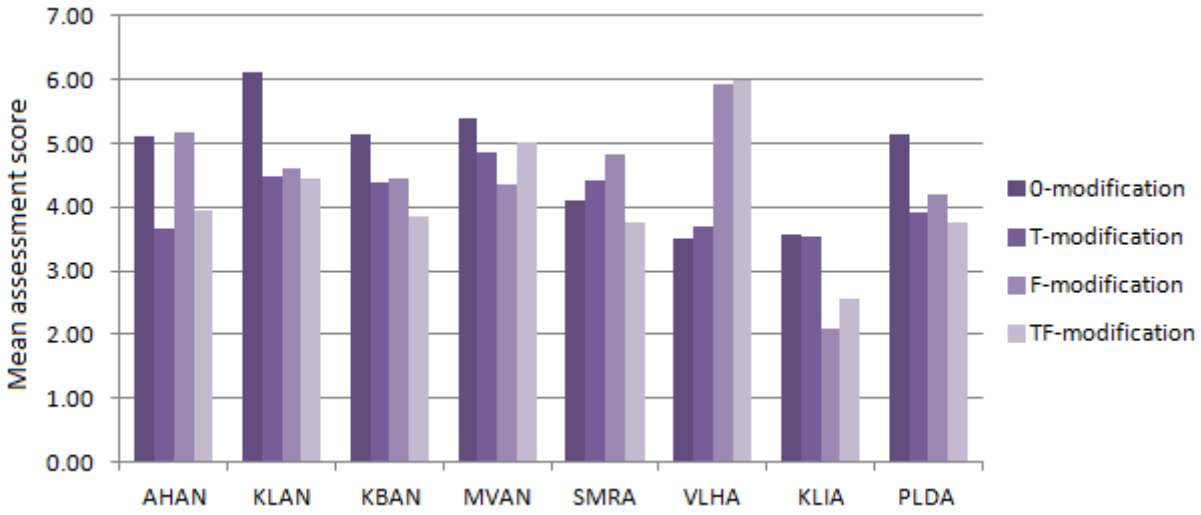


Figure 7. The mean scores of each of the eight individual speakers broken according to the type of speech modification (T = time, F = fundamental frequency).

As can be seen, it is clear that the manipulations have considerably different effects on each of the speakers; therefore, we cannot generalize as to what modification causes what, but rather we have to consider each of the speakers an individual case. The only thing to notice is that for five speakers (KLAN, KBAN, MVAN, KLIA and PLDA) out of eight the score for the 0-modification is the highest compared to the other modifications (even though only slightly for the speaker KLIA), which supports the theory that the modifications affect the speaker negatively. One-way analysis of variance confirmed significant differences only for the speaker KLIA against KLAN, MVAN and VLHA: $F(7, 40) = 2.54$; $p < 0.05$. It is important to remember that different personality traits were analyzed for different speakers.

4.5 Consistency of answers

Before analyzing the individual items in detail, attention must also be paid to the consistency of answers of the listeners. Apart from the two semantically different items for four of the speakers (mentioned in section 4.4), twelve additional items were added to the

perceptual test (see section 3.1.3) in order to measure consistency of answers of the listeners about the same item (the items and the differences are shown in Table 2).

Speaker	1	2	Diff.	Speaker	1	2	Diff.
AHAN-Jimmy (T)	2.70	3.30	-0.60	SMRA-hangar (TF)	3.80	3.60	0.20
AHAN-hangar (F)	5.45	4.80	0.65	VLHA-hangar (TF)	5.65	5.85	-0.20
KLAN-Jimmy (T)	4.20	4.60	-0.40	KLIA-Jimmy (0)	3.75	3.40	0.35
KBAN-hangar (0)	4.90	5.40	-0.50	KLIA-Jimmy (TF)	2.90	2.20	0.70
KBAN-hangar (T)	4.45	4.30	0.15	PLDA-Jimmy (T)	3.80	4.05	-0.25
MVAN-hangar (T)	4.35	5.35	-1.00	PLDA-Jimmy (F)	4.20	4.20	0.00

Table 2. The differences between two mean assessment scores measured on the same item. In the column labelled 1 there are mean assessment scores of the items being heard for the first time, in the column labelled 2, there are mean assessment scores of the items being heard for the second time, and in the column Diff. there are differences between the two mean scores. In the brackets behind the speakers there are signalled modifications of the items (0 for 0-modification, T for T-modification, F for F-modification, and TF for TF-modification).

As it can be seen from the table above, for most of the items there were only small differences in perceptions of the same item heard twice. Indeed, t-test for independent samples revealed insignificant differences for ten of the twelve items (p ranging from 0.12 to 1.00). For the item AHAN-hangar (F) the t-test for independent samples showed a marginally significant difference ($p = 0.08$) and for the item MVAN-hangar (T) the difference was significant ($p = 0.04$); therefore, the consistency of answers for these two items must be considered with caution. We can see no tendency in the answers of the respondents to evaluate the item with a higher or lower score when heard for the second time (for six items the mean score raised, for five items the mean score lowered, and for one item the mean score did not change).

The differences between two semantically different items by the same speaker (listed in Table 3 below) occasionally showed a significant difference. The t-test for independent samples showed marginally important differences for the items SMRA (F): $p = 0.07$, and VLHA (TF): $p = 0.08$. For three items the t-test for independent samples revealed a significant difference: AHAN (T): $t(58) = 3.38$; $p < 0.01$, then KLAN (F): $t(38) = 3.32$; $p < 0.01$, and VLHA (F): $t(38) = 3.44$; $p < 0.01$. The significant differences might have been caused by the semantic content of the sound extract, or the particular realisation of the

speech sample by the speaker might have been different (to be discussed in section 5.1). Otherwise the differences between the two samples by one speaker proved to be insignificant (*p* ranging from 0.12 to 0.82).

Speaker	Jimmy	Hangar	Diff.	Speaker	Jimmy	Hangar	Diff.
AHAN (0)	5.45	4.75	0.70	SMRA (0)	4.35	3.85	0.50
AHAN (T)	3.00	4.30	-1.30	SMRA (T)	4.20	4.65	-0.45
AHAN (F)	5.20	5.13	0.08	SMRA (F)	4.45	5.20	-0.75
AHAN (TF)	3.80	4.10	-0.30	SMRA (TF)	3.80	3.70	0.10
KLAN (0)	6.15	6.05	0.10	VLHA (0)	3.15	3.85	-0.70
KLAN (T)	4.40	4.55	-0.15	VLHA (T)	3.85	3.50	0.35
KLAN (F)	5.40	3.80	1.60	VLHA (F)	6.60	5.25	1.35
KLAN (TF)	4.15	4.75	-0.60	VLHA (TF)	6.25	5.75	0.50

Table 3. Mean assessment score of two semantically different items by the same speaker. In the brackets behind the speakers there are signalled modifications of the items (0 for 0-modification, T for T-modification, F for F-modification, and TF for TF-modification). The columns Jimmy and Hangar refer to the two semantically different speech samples of the same speaker and the column Diff. shows the difference between the two mean scores.

4.6 Item analysis

In this section, the individual items will be analysed in relation to their modifications, nationality of the speaker, and the personality traits allocated to them. Each of the four personality traits will be treated separately along with the three items allocated to it.

4.6.1 Personality trait: Honest

For the personality trait *upřímný* (honest), one native speaker (AHAN) and one Czech speaker (PLDA) were analysed (selection of the items for each trait described in section 3.2).

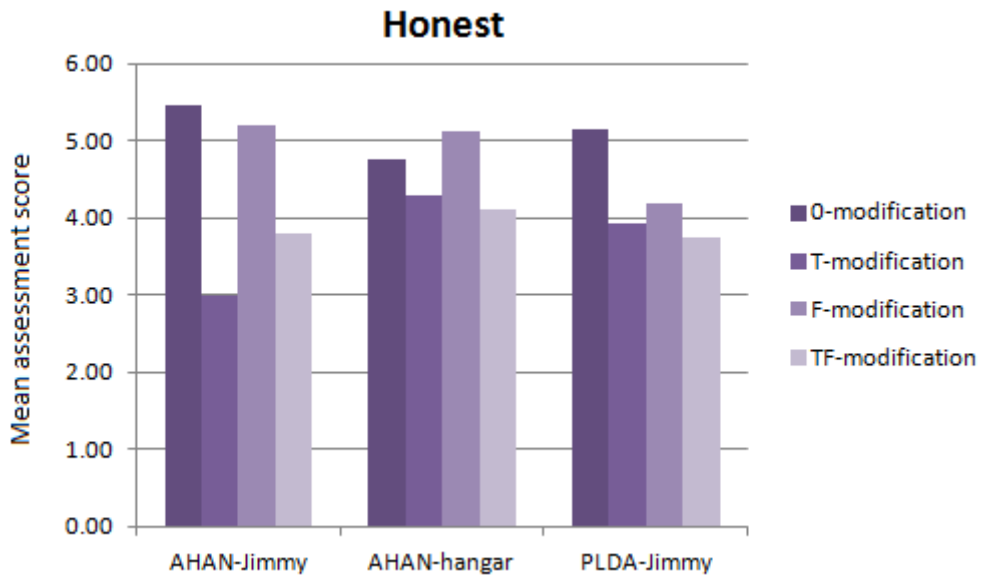


Figure 8. This figure shows the mean score of the items allocated to the personality trait *upřímný* (honest), regarding the modifications.

As can be seen from Figure 8 (above), both speakers are considered honest by the listeners, the mean score reaching over 5 (see Appendix C for the standard deviations). However, it is interesting to notice that the speaker AHAN reached different mean scores for the two speech samples. While AHAN-Jimmy has the mean score of the non-modified speech sample reaching almost 5.50, the mean score for 0-modification AHAN-hangar does not even reach 5 (this issue to be addressed in section 5.2). The mean score for *upřímný* (honest) for the non-modified extract by the Czech speaker is comparable to that of the native speaker. For the item AHAN-Jimmy, the T-modification causes a significant decrease in the mean score compared to the 0-modification (post-hoc Tukey HSD after two-way analysis of variance: $p < 0.001$). The decreased mean score for this item reached 3, which is below the arithmetic mean on the 7-point scale, implying that the speaker is not only perceived as less honest, but we could even say dishonest, being closer to the other end of the scale. In the case of the item AHAN-hangar, the T-modification also causes a decrease of the mean score, however, it proved to be insignificant. For the Czech speaker PLDA, the T-modification also causes a decrease of the mean score; bigger than for AHAN-hangar but smaller than for AHAN-Jimmy, which might be surprising, considering that the other two items are by the same speaker, only of a different semantic content. Modifying the frequency alone does not seem to affect the speech of the native speaker greatly, regarding the personality trait *upřímný* (honest). For the item AHAN-Jimmy, there

is only a slight decrease of the mean score, whereas for the item AHAN-hangar there is even a slight increase of the mean score. On the other hand, the mean score for the F-modification exhibited a marginally significant decrease for the speaker PLDA (post-hoc Tukey HSD: $p \sim 0.06$). Modifying both time and frequency appears to cause a remarkable decrease (below the score 4) in perceived honesty for the item AHAN-Jimmy, however, not as big as in the case of T-modification alone. For AHAN-hangar, the mean score of the TF-modification also decreases and reaches a comparable mean score as the AHAN-Jimmy; however the decrease is not statistically important. For the speaker PLDA, the decrease of the mean score of the TF-modification is comparable to the decrease caused by other modifications and it is also comparable to the mean scores of TF-modification of the other two items in this category. In general, for the speaker PLDA, the mean scores of the modified speech samples are all centred around the arithmetic mean of the 7-point scale, while the mean scores of the item AHAN-Jimmy are dispersed for the T-modification and F-modification. Unlike for the items AHAN-Jimmy and PLDA-Jimmy, all four mean scores are unusually balanced for the item AHAN-hangar, regardless of the modification; there is no extremely high or extremely low score.

Indeed, two-way analysis of variance (with factors item and modification) showed that modifications, regardless of the item, exhibit a significant difference: $F(3, 308) = 18$; $p < 0.001$. The post-hoc Tukey HSD test revealed that the significance of the difference is caused by the T- and TF-modifications compared to 0- and F-modifications (T vs. 0, T vs. F, TF vs. 0, and TF vs. F: $p < 0.001$). The interaction between the items and modifications also revealed a significant difference: $F(6, 308) = 4.1$; $p < 0.001$. The tests of follow up comparison showed that the difference between the groups 0 & F, and T & TF is mainly caused by the item AHAN-Jimmy (post-hoc Tukey HSD: $p < 0.001$) and marginally also by the item PLDA-Jimmy (post-hoc Tukey HSD: p around 0.06). The differences between individual modifications of the item AHAN-hangar proved to be insignificant.

4.6.2 Personality trait: Nervous

Figure 9 below shows the mean assessment scores of one native speaker (MVAN) and one Czech speaker (VLHA), allocated to the personality trait *nervózní* (nervous).

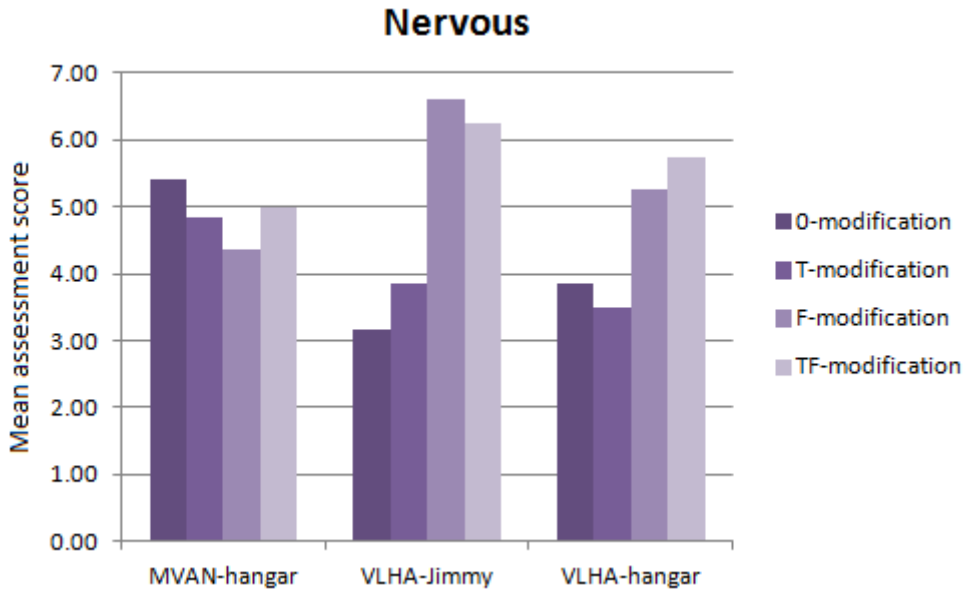


Figure 9. The mean scores of the items allocated to the personality trait *nervózní* (nervous), regarding the modifications.

It is obvious that the mean scores are much more unbalanced for the two items by the Czech speaker, than for the item by the native speaker. The native speaker reaches a rather high mean score of nervousness for the non-modified speech extract, especially compared to the Czech speaker who seemed to be perceived as rather calmer than nervous. The manipulation in time causes an insignificant decrease of the mean score for the item MVAN-hangar, and the manipulation in frequency causes a greater decrease of the mean score than the T-manipulation; however, the difference is still statistically insignificant. Surprisingly, if the two manipulations combine, they cause a decrease in the mean score, however, not as big as for the T-modification or F-modification alone. For the item VLHA-Jimmy, manipulation in time, contrasted to what it produced for the native speaker, causes an increase in the perceived nervousness. For the second item by the speaker VLHA, the manipulation in time causes a slight decrease of the score; however for both items by the speaker VLHA, the mean scores for the 0-modification and T-modification are rather balanced; therefore, we may claim that the manipulation of durations of stressed and unstressed syllables for the speaker VLHA had almost no effect on the perception of nervousness by the listeners. On the other hand, the manipulation in frequency affected the perception of the speech significantly for the Czech speaker for both items by this speaker (two-way analysis of variance, followed by post-hoc Tukey HSD: $p < 0.001$); especially for the item VLHA-Jimmy, where the mean score of the non-modified speech is below the

arithmetic mean of the 7-point scale, while the mean score for the F-modification reaches almost the highest score possible. For the item VLHA-hangar the increase of the mean score for the F-modification is not as big as for the item VLHA-Jimmy, however, it is still significant (post-hoc Tukey HSD: $p < 0.001$). Combining the manipulations in time and in frequency also causes a huge increase of the mean score for both items by the Czech speaker (post-hoc Tukey HSD: $p < 0.001$).

To conclude, two-way analysis of variance (with factors item and modification) revealed a significant effect of modifications (regardless of the item) on the perception of nervousness: $F(3, 268) = 23.2$; $p < 0.001$. The tests of follow up comparison showed that the significance of the difference is mainly caused by the F- and TF-modifications as opposed to 0- and T-modifications (post-hoc Tukey HSD tests: $p < 0.001$). Furthermore, testing the interaction between the items and modifications revealed that both items by the speaker VLHA contribute more to the overall significance of the difference between 0- and T-modifications, and F- and TF-modifications (post-hoc Tukey HSD tests: $p < 0.001$, and in the case of 0-mod. vs. T-mod. for the item VLHA-hangar, $p = 0.08$), while the differences between the individual modifications of the item MVAN-hangar proved to be insignificant.

4.6.3 Personality trait: Kind

As it was already mentioned, the personality trait *dobrosrdečný* (kind) does not seem to be very prominent in speech. Figure 10 (below) shows the mean scores for kindness for a native speaker (KBAN) and for a Czech speaker (SMRA).

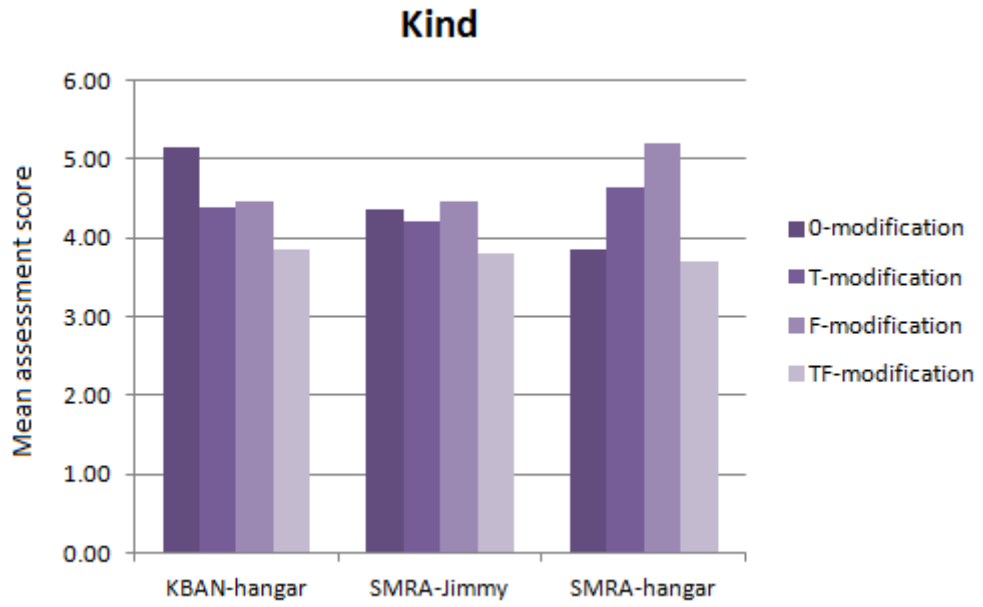


Figure 10. The mean scores of the items allocated to the personality trait *dobrosrdečný* (kind), regarding the modifications.

In general, the modifications seem to have a different effect on different speakers. Among the non-manipulated speech extracts, the native speaker is perceived as being kind, while the score for the Czech speaker, in both items, centres around the arithmetic mean of the 7-point scale (see Appendix C for standard deviation), meaning that the speaker is perceived as neither kind, nor cold. For the speaker KBAN, the score for both the T-modification and the F-modification decreases approximately to the same score, compared to the 0-modification. Combining the manipulation in frequency and in time, the mean score decreases even more; nevertheless, the difference is statistically insignificant. However, the scores of all three modifications centre around the arithmetic mean of the 7-point scale, therefore, we may only conclude that the speaker seems less kind, but not actually cold-hearted. There are some differences among the two items by the Czech speaker; while the T-modification causes a slight decrease of the mean score for the item SMRA-Jimmy, it causes an increase for the item SMRA-hangar, yet this difference did not prove to be statistically important. Manipulating frequency does not affect the perception for the item SMRA-Jimmy greatly; however, it is interesting to notice that it reached precisely the same mean score as the F-modification for the native speaker. On the other hand, the mean score for the F-modification of the item SMRA-hangar rises seemingly greatly; however, statistical analysis did not confirm the difference as significant. The TF-modification

reaches approximately the same mean scores for all three items allocated to this personality trait; while it causes a comparable larger decrease of the mean score for the speaker KBAN, it makes only a small difference for the speaker SMRA for both items.

Although two-way analysis of variance (factors: item and modification) showed a significant difference caused by the modifications (regardless of the item): $F(3, 288) = 4.2$; $p < 0.01$, the post-hoc Tukey test revealed that the significant difference was caused by the difference of the item KBAN-hangar (0-modification) as opposed to TF-modifications of the items SMRA-Jimmy and SMRA-hangar (p ranging from below 0.01 to 0.04), which is an irrelevant result for the present analysis.

4.6.4 Personality trait: Assertive

For the personality trait *prosazující se* (assertive), one native speaker (KLAN) and one Czech speaker (KLIA) were used in the experiment (see Figure 11 below).

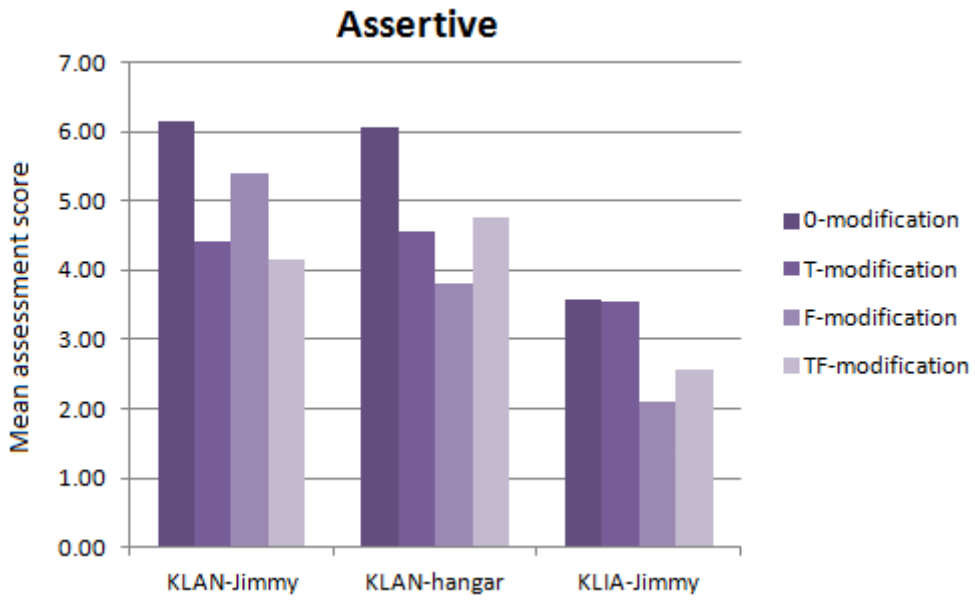


Figure 11. The mean scores of the items allocated to the personality trait *prosazující se* (assertive), regarding the modifications.

Both items by the native speaker reached a very high score with a very small standard deviation (see Appendix C) for the non-manipulated speech samples; therefore, we may say that the speaker is perceived as very assertive. On the other hand, the speaker KLIA was perceived as very little assertive, the mean score not even reaching the arithmetic

mean on the 7-point scale. For both items by the native speaker, the T-modification caused a decrease of the score, the mean scores of the two items being approximately the same. Modifying frequency caused only a small decrease of the mean score for the item KLAN-Jimmy (not even by one point), but on the other hand, F-modification caused a significant decrease of the mean score for the second item by the speaker KLAN (two-way analysis of variance followed by post-hoc Tukey HSD: $p < 0.001$), making the item comparable to the non-modified speech sample by the Czech speaker. TF-modification caused a decrease for both the items KLAN-Jimmy and KLAN-hangar, the decrease being bigger for the item KLAN-Jimmy. For the speaker KLIA, manipulating durations of stressed and unstressed syllables had almost no effect on perceiving the speaker as being assertive or reserved. On the other hand, manipulating frequency caused a decrease of the mean score, the mean score dropping to 2. TF-modification caused a decrease of the mean scores as well, although not as big as manipulating the frequency only.

Two-way analysis of variance (with factors item and modification) revealed significant difference for individual speakers: $F(2, 288) = 58.28$; $p < 0.001$, for the modifications: $F(3, 288) = 15.24$; $p < 0.001$, and for the interaction of the two factors: $F(6, 288) = 4.21$; $p < 0.001$. Analyzing the individual speakers (regardless of the modification) showed a significant difference between the speaker KLIA contrasted to both items by the speaker KLAN (post-hoc Tukey HSD tests: $p < 0.001$). For the modifications in general, the 0-modification significantly differed from all the others (post-hoc Tukey HSD: $p < 0.05$); moreover, the TF-modification showed a significant difference as opposed to T-modification (post-hoc Tukey HSD tests: $p < 0.01$). Furthermore, analyzing the interaction of the two factors for individual items showed that for the item KLAN-Jimmy, the statistically important differences are between T- and TF-modifications as opposed to 0-modification (post-hoc Tukey HSD tests: $p < 0.01$), while for the item KLAN-hangar, the significant difference was between F-modification compared to 0-modification (post-hoc Tukey HSD: $p < 0.001$). For the item KLIA the tests of follow-up comparison revealed a significant difference between 0-modification as opposed to F-modification (post-hoc Tukey HSD: $p = 0.02$).

5. Discussion

It is not very surprising that no statistically significant difference occurred while comparing the effects of the modifications in general because different personality traits were measured for different people (see section 4.1). However, the matter needed to be investigated this way because there is not much previous research that would attempt to measure the effects of individual measurable aspects of speech on perception. On the other hand, there is a marginally significant difference between Czech and native speakers for their 0-modification; however, no such difference appears for the other modifications (see section 4.2). The difference between the 0-modifications might have been caused by the foreign accent of the Czech speakers, and this effect could have been suppressed by the modifications; however, we have to bear in mind that these results do not differentiate between the individual personality traits either. In the following two sections of the discussion, the significant differences will be discussed. In section 5.1, the discrepancies in consistency of answer of the listeners will be related, and the following section (5.2) will cover the issues related to the individual items and personality traits.

5.1 Consistency of answers

Measuring the consistency of answers among listeners (t-test for independent samples) revealed that in three cases out of sixteen, there was a significant difference between perceptions of two semantically different speech samples by the same speaker (see section 4.5). The first of the significant differences is that between the items AHAN-Jimmy and AHAN-hangar for T-modification: $t(58) = 3.38$; $p < 0.01$. A possible reason for the discrepancy in this case is that the realization of the utterance by the speaker is different. In the case of the item AHAN-Jimmy, the speech rate was slower and the durational differences between stressed and unstressed syllables were larger; therefore, the effect of the T-modification was more prominent for this item and caused a significant decrease in perceived honesty (see section 4.6.1 for the results). On the other hand, for the item AHAN-hangar, the T-modification did not cause any significant difference compared to the 0-modification.

Another speaker that showed statistically important difference between the two semantically different speech samples is KLAN and the speech samples modified in frequency ($t(38) = 3.32$; $p < 0.01$). This difference might be a result of a slightly creaky voice at the end of the speech sample KLAN-hangar which was caused by manipulating frequency.

The last significant discrepancy between ratings of two different speech samples by the same speaker appeared for items VLHA-Jimmy and VLHA-hangar modified in frequency (t (38) = 3.44; $p < 0.01$). In this case there appeared to be wider pitch range for the item VLHA-Jimmy, therefore, the modification in frequency was more prominent than for the item VLHA-hangar. Moreover, the modification in frequency also caused that the voice gained a slightly creaky quality, especially for the item VLHA-Jimmy where the modification was more prominent. Therefore, the discrepancy between these two samples might be a result of both the above mentioned factors combined.

5.2 Personality traits in relation to individual items

Apart from the difference between the T-modifications of the items AHAN-Jimmy and AHAN-hangar, which was described in the section above, there appeared other interesting aspects to mention in relation to the personality trait *upřímný* (honest). In general, modifying durations of stressed and unstressed syllables seems to be important for perception regarding the quality of honesty. For the items AHAN-Jimmy and PLDA-Jimmy the difference between T-modifications and 0-modifications proved to be significant, respectively marginally significant, and although the difference for the item AHAN-hangar was not statistically significant, it shows a decreasing tendency as well (see section 4.6.1 for the results). It is therefore very likely that the difference between TF-modifications and 0-modification is caused by the temporal aspect of the modification. Furthermore, it is interesting to notice that while some of the modifications caused significant differences for items AHAN-Jimmy and PLDA-Jimmy, the differences between individual modifications of the item AHAN-hangar proved to be insignificant. A possible explanation for this discrepancy is a different semantic content of the speech sample; the listeners' own beliefs about the truth value of the speech content might have affected their evaluation of the speakers and this effect might be stronger than the effect of the

manipulations. Another possibility is that the listeners were influenced by a faster speech rate of the item AHAN-hangar.

The first thing to notice about the personality trait *nervózní* (nervous) is the difference between the native speaker and the Czech speaker (0-modifications). While it seems logical to assume that the native speaker should be perceived as being less nervous speaking her native language, the results revealed that the Czech speaker is perceived as much calmer than the native speaker. In this particular case, the difference might be a result of a slow and very monotonous speech of the speaker MVAN which might have influenced the listeners to perceive the speaker as more nervous compared to the speaker VLHA, who tries to pronounce words carefully and also the timbre of her voice suggests self-confidence. It is obvious from the results (see section 4.6.2) that manipulating pitch (represented by fundamental frequency) affects the speaker VLHA greatly, resulting in the speaker being perceived as more nervous. The difference between the 0-modifications and TF-modifications for the items by the speaker VLHA also proved to be significant and it is very likely that the aspect contributing to the significant difference is the manipulation in frequency because the manipulation of durations alone had almost no effect on the perception. Based on the results for the speaker VLHA we may assume that pitch plays an important role in perception of the quality of nervousness; however, the results for the speaker MVAN do not support such hypothesis. As it was already mentioned, the speech of the speaker MVAN is very monotonous; therefore, the manipulation in frequency might not have been prominent enough in the speech to cause any significant changes in perception. Furthermore, as it was already mentioned in section 5.1, manipulating frequency for the speaker VLHA resulted in the voice gaining a slightly creaky quality, which might have also contributed to the perception of the speaker as being nervous. However, more samples and deeper analysis will be necessary to draw general conclusions.

As we already mentioned in section 4.6.3, the personality trait *dobrosrdečný* (kind) did not prove to be prominent in speech for the two speakers (KLAN and SMRA). This was largely due to considerably big standard deviations (see Appendix C). It is very likely that each of the listeners used different criteria for his judgement, such as timbre of the voice, speech rate, or the degree of foreign accent. It might also be possible that people need more

information about the speaker, such as gesticulation or facial expression, to be able to qualify the speaker as kind or cold.

For the personality trait *prosazující se* (assertive) the results revealed a striking difference between the speakers KLAN and KLIA, regardless of the modifications (see section 4.6.4 for results). Such difference is very likely caused by a heavy foreign accent and a hesitant speech of the speaker KLIA which possibly resulted in the speaker being perceived as more reserved. Apart from the difference between the items KLAN-hangar and KLAN-Jimmy modified in frequency (mentioned in section 5.1), the results also showed another discrepancy between the two items. While for the item KLAN-Jimmy, the statistically important differences appeared for the speech samples with T- and TF-modifications, it was manipulation in frequency that exhibited a significant difference for the item KLAN-hangar. Moreover, for the Czech speaker it was also the F-modification that caused a significant difference. This result suggests that pitch might be important for perceiving people as being assertive or reserved; however, more data need to be analyzed before we may draw general conclusions. Another interesting difference is between the perceptions of the items manipulated for durations because while it caused a significant decrease for the item KLAN-Jimmy and a very strong tendency to decrease for the item KLAN-hangar, it had almost no effect on the Czech speaker. A possible explanation for this discrepancy is that the Czech speaker did not make differences in duration between stressed and unstressed syllables under the influence of her native language that does not differentiate stressed and unstressed syllables by their length; therefore, manipulation the durations was not prominent enough in the speech to cause differences in perception.

6. Conclusion

The present study belongs to a pioneering work in the field of relating particular measurable aspects of speech to changes in perception of personality. The theoretical background presents information about psychology of personality, foreign accent and its social impact on a speaker, and about rhythm of speech, and also summarizes some representative studies in these fields. However, there was no research connecting these branches of science in a comprehensive study of the relationship between particular aspects of speech and their effects on speech perception and evaluation of personality.

The purpose of this paper was to conduct an experiment which attempts to discover whether there is any connection between changes in speech rhythm and perceptual evaluation of several personality traits. Previous research demonstrated that not only durational patterns in speech are influencing our perception of rhythm, but also intonation, represented by fundamental frequency (summarized by Volín, 2010). Therefore, for the purposes of the experiment, durational patterns and intonation were manipulated for several speech samples. For the manipulation in the temporal domain, the stressed syllables were shortened while the nearest following unstressed syllables were prolonged. For the manipulation in frequency, the value of frequency was exchanged between the stressed and the nearest following unstressed syllables. Subsequently four personality traits were selected to be measured and a perceptual test was prepared. Twenty subjects rated the speech samples and were evaluating to what extent they would characterize the speaker by the given personality trait. After collecting the perceptual tests from the subjects, the data were analyzed from multiple points of view.

The results revealed some significant tendencies for three personality traits while they showed no tendency for the personality trait *dobrosrdečný* (kind). First of all, the durational patterns appeared to be important for perceiving people as honest or dishonest; however, for one of the three samples the difference was statistically insignificant. We hypothesized that the difference might have been caused by a different semantic content of the sample, or by a different speech rate. Further research might investigate whether the tendency might be generalized to all speakers or whether durational patterns were only

significant for the speakers analyzed in this research. A similar tendency appeared for frequency in relation to the personality trait *nervózní* (nervous).

To be able to generalize, a deeper analysis of more data is necessary. There are also other factors which influence the listener's perception of the speaker, such as a degree of foreign accent of Czech speakers, or timbre of the voice. These factors could either support the effect of the manipulations or operate against it. The effects of these factors could not be separated from the effects caused by the manipulations by the method used in this experiment, which also presents another possible direction for future research.

References

Boersma, P. & Weenink, D. (2013). Praat: doing phonetics by computer [Version 5.3.41]. Retrieved 19 February 2013 from http://www.praat.org/.

Buxton, H. (1983). Temporal predictability in the perception of English speech. *Prosody: Models and Measurements*. A. Cutler & D. R. Ladd (eds.). Berlin: Springer-Verlag, pp. 111-121.

Cattell, R. B. (1957). *Personality and Motivation Structure and Measurement*. Yonkers, NY: World Book.

Costa, P. T. & McCrae, R. R. (1987). Validation of the five-factor model of personality across instruments and observers. *Journal of Personality and Social Psychology*, 52 (1), pp. 81-90.

Derwing, T. M. & Munro, M. J. (2009). Putting accent in its place: Rethinking obstacles to communication. *Language Teaching*, 49 (4), pp. 476-490.

Donovan, A. & Darwin, C. J. (1979). The perceived rhythm of speech. *Proceedings of IXth International Congress of Phonetic Sciences*, pp. 268-274. Copenhagen: University of Copenhagen.

Drapela, V. J. (1997). *Přehled teorií osobnosti.* Balcar K. (transl.). Praha: Portál.

Edwards, J. (1999). Refining our understanding of language attitudes. *Journal of Language and Social Psychology*, 18 (1), pp. 101-110.

Flege, J. E. & Port, R. (1981). Cross-language phonetic interference: Arabic to English. *Language and Speech*, 24, pp. 125-146.

Grabe, E. & Low, E. L. (2002). Durational variability in speech and the rhythm class hypothesis. *Laboratory Phonology VII*, Gussenhoven C., Warner, N. (eds.), pp. 515-546. Berlin: Mouton de Gruyter.

Hřebíčková, M. (2011). *Pětifaktorový model v psychologii osobnosti: Přístupy, diagnostika, uplatnění.* Praha: Grada.

Hunt, M. (1993). *The Story of Psychology*. New York: Doubleday.

Jung, C. G. (1969). *The Archetypes and the Collective Unconscious*. NJ: Princeton University Press.

Kinoshita, N. & Sheppard, C. (2011). Validating acoustic measures of speech rhythm for second language acquisition. *International Congress of Phonetic Sciences XVII*, pp. 1086-1089.

Machač P. & Skarnitzl R. (2009). *Principles of Phonetic Segmentation*. Praha: Epocha.

McCrae, R. R. & Costa, P. T. (1983). Joint factors in self-reports and ratings: Neuroticism, extraversion, and openness to experience. *Personality and Individual Differences*, 4, pp. 245-255.

Mok, P. P. K. & Dellwo, V. (2008). Comparing native and non-native speech rhythm using acoustic rhythmic measures: Cantonese, Beijing Mandarin and English. *Speech Prosody 2008*, pp. 423-426. Brazil: ISCA.

Munro, M. J. (2003). A primer on accent discrimination in the Canadian context. *TESL Canada Journal*, 20 (2), pp. 38-51.

Norman, W. T. (1963). Toward and adequate taxonomy of personality attributes: Replicated factor structure in peer nomination personality ratings. *Journal of Abnormal and Social Psychology*, 66, pp. 574-583.

Pike, K. L. (1945). *The Intonation of American English.* MI: University of Michigan Press.

Plomp, R. & Mimpen, A. M. (1979). Improving the reliability of testing the speech reception threshold for sentences. *International Journal of Audiology*, 18 (1), pp. 43-52.

Quené, H. & Port, R. F. (2005). Effects of timing regularity and metrical expectancy on spoken-word perception. *Phonetica*, 62, pp. 1-13.

Quené, H. & Van Delft, L. E. (2010). Non-native durational patterns decrease speech intelligibility. *Speech Communication*, 52, pp. 911-918.

Ramus, F. et al. (1999). Correlates of linguistic rhythm. *Cognition*, 73, pp. 265-292.

Skarnitzl, R., Volín, J. & Drenková, L. (2005). Tangibility of Foreign Accents in Speech: the Case of Czech English. In: A.Grmelová, L.Dušková & M.Farrell (Eds.) *2nd Prague Conference on Linguistics and Literary Studies Proceedings*, pp.11-20. Praha: UK PedF.

Skarnitzl, R. (2010). Prague Phonetic Corpus: status report. *Phonetica Pragensia*, 12 (1), pp. 65-67.

Tajima, K. et al. (1997). Effects of temporal correction on intelligibility of foreign-accented English. *Journal of Phonetics*, 25, pp. 1-24.

Tarone, E. (1980). Some influences on the syllable structure of interlanguage phonology. *International Review of Applied Linguistics in Language Teaching*, 18, pp. 139-152.

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

Volín, J. (2010). On the significance of the temporal structuring of speech. ...*for thy speech bewrayeth thee (A Festschrift for Libuše Dušková)*, M.Malá & P.Šaldová (Eds.) pp. 289-305. Praha: UK FF.

Appendix

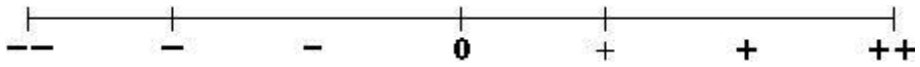
Appendix A: The table below shows all the statements and their allocation to individual items for the purposes of the perceptual test. The statements are in Czech, because the perceptual test was prepared in the Czech language and translation is not available.

Upřímný	
AHAN-Jimmy	Tento člověk působí důvěryhodně.
AHAN-hangar	Tento člověk při hře zásadně nepodvádí.
PLDA-Jimmy	Tento člověk často říká, že se mu něco líbí, i když to není pravda.
Nervózní	
MVAN-hangar	Tento člověk se často cítí nesvůj v neznámé situaci.
VLHA-Jimmy	Tento člověk bývá i před těžkou zkouškou v klidu.
VLHA-hangar	Tento člověk při rozhovoru často nervózně přešlapuje.
Dobrosrdečný	
KBAN-hangar	Tento člověk je milý a dobře vychází s druhými.
SMRA-Jimmy	Tento člověk se nezajímá o pocity druhých.
SMRA-hangar	Tento člověk by vždycky pomohl kamarádovi v nouzi.
Prosazující se	
KLAN-Jimmy	Tento člověk se často rád chopí iniciativy a řídí ostatní.
KLAN-hangar	Tento člověk si většinou nechává své názory pro sebe.
KLIA-Jimmy	Tento člověk se nebojí obhajovat vlastní názor.

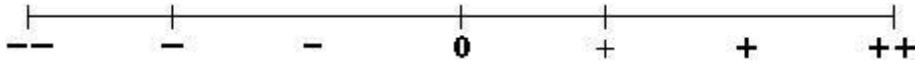
Appendix B: An example of the perceptual test used in this study. (There are only examples of the blocks A and B because the versions of the blocks C and D are almost identical, the only difference being the order of the items on the test.)

BLOK A Zkuste zachytit svůj **první dojem** z mluvících osob. Bez ohledu na to, co říkají, do jaké míry si myslíte, že je můžete charakterizovat následující vlastností?
Odpovídejte na stupnici: – – znamená **vůbec**, ++ znamená **velmi**

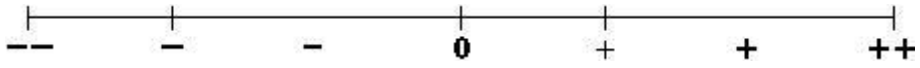
1. Nervózní



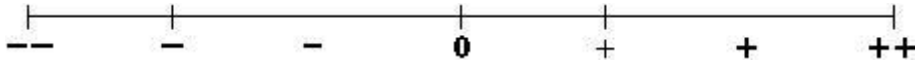
2. Upřímný



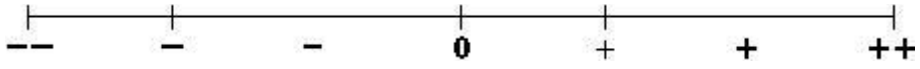
3. Dobrosrdečný



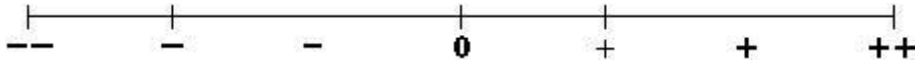
4. Nervózní



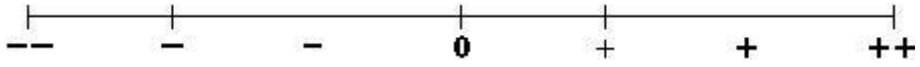
5. Prosazující se



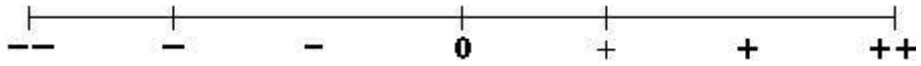
6. Zdrženlivý



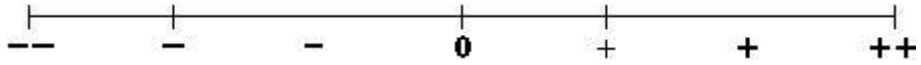
7. Upřímný



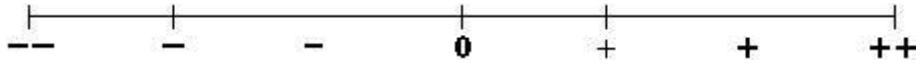
8. Nervózní



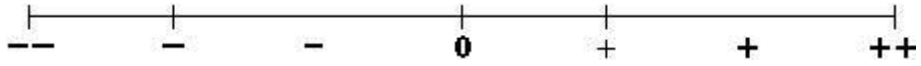
9. Dobrosrdečný



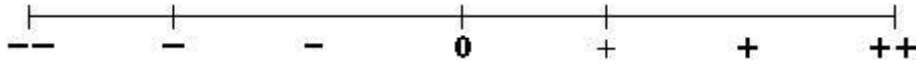
10. Upřímný



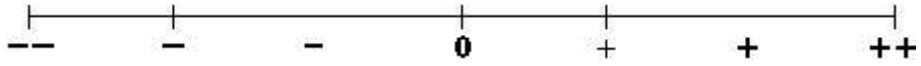
11. Prosazující se



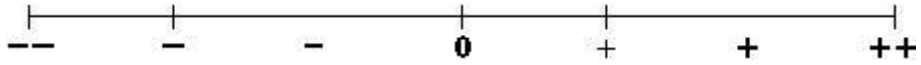
12. Necitelný



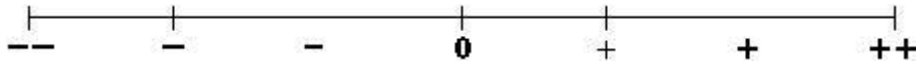
13. Nervózní



14. Upřímný

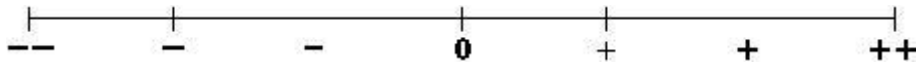


15. Prosazující se

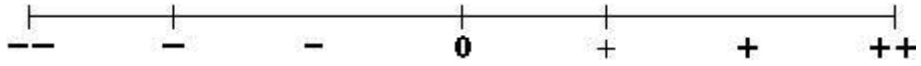


BLOK B Zkuste zachytit svůj **první dojem** z mluvících osob. Bez ohledu na to, co říkají, nakolik si myslíte, že o nich platí následující výroky?
Odpovídejte na stupnici: – – znamená **vůbec**, ++ znamená **velmi**

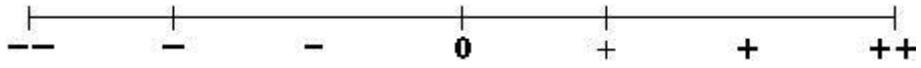
1. Tento člověk je milý a dobře vychází s druhými.



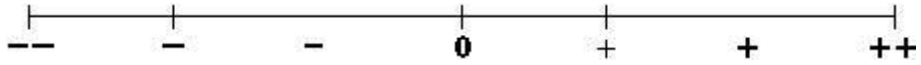
2. Tento člověk bývá i před těžkou zkouškou v klidu.



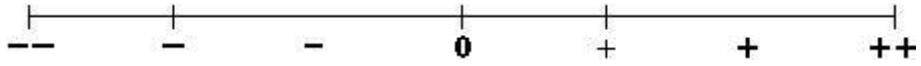
3. Tento člověk často říká, že se mu něco líbí, i když to není pravda.



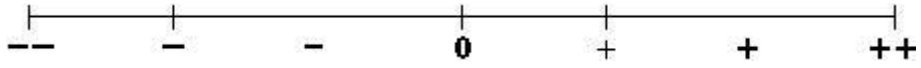
4. Tento člověk by vždycky pomohl kamarádovi v nouzi.



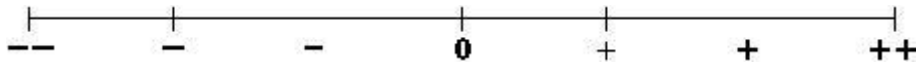
5. Tento člověk se často rád chopí iniciativy a řídí ostatní.



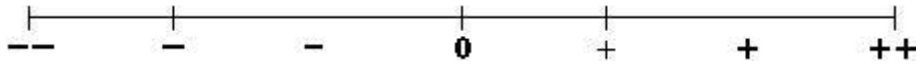
6. Tento člověk působí důvěryhodně.



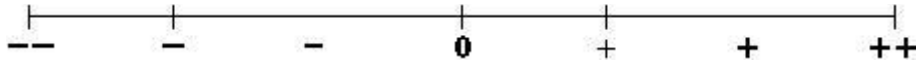
7. Tento člověk se nebojí obhajovat vlastní názor.



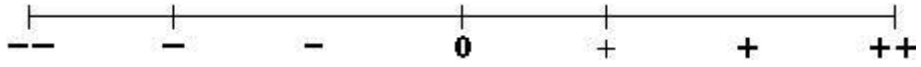
8. Tento člověk se často cítí nesvůj v neznámé situaci.



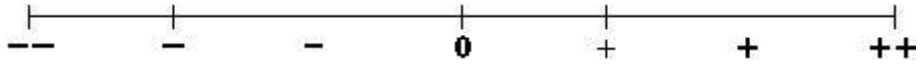
9. Tento člověk se nezajímá o pocity druhých.



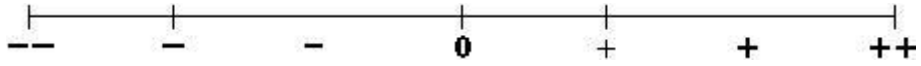
10. Tento člověk při rozhovoru často nervózně přešlapuje.



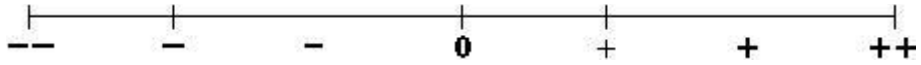
11. Tento člověk si většinou nechává své názory pro sebe.



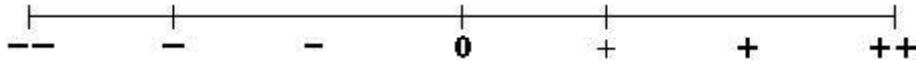
12. Tento člověk při hře zásadně nepodvádí.



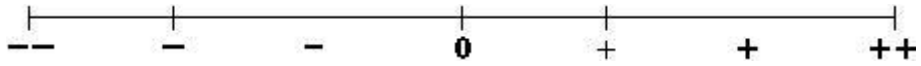
13. Tento člověk je milý a dobře vychází s druhými.



14. Tento člověk se nebojí obhajovat vlastní názor.



15. Tento člověk často říká, že se mu něco líbí, i když to není pravda.



Appendix C: Mean scores and standard deviations for all the items on the test. The left-hand part presents the results of the native speakers and the right-hand part of the table presents the results of the Czech speakers (**Mod**. = Modification of the item, **Mean** = Mean score for the item, **SD** = standard deviation, **0** = 0-modification, **T** = T-modification, **F** = F-modification, **TF** = TF-modification).

Speaker	**Mod.**	**Mean**	**SD**		**Speaker**	**Mod.**	**Mean**	**SD**
AHAN-Jimmy	0	5.45	1.10		SMRA-Jimmy	0	4.35	1.93
AHAN-Jimmy	T	3.00	1.32		SMRA-Jimmy	T	4.20	1.28
AHAN-Jimmy	F	5.20	1.24		SMRA-Jimmy	F	4.45	1.36
AHAN-Jimmy	TF	3.80	1.54		SMRA-Jimmy	TF	3.80	1.67
AHAN-hangar	0	4.75	1.68		SMRA-hangar	0	3.85	1.53
AHAN-hangar	T	4.30	1.56		SMRA-hangar	T	4.65	1.93
AHAN-hangar	F	5.13	1.16		SMRA-hangar	F	5.20	1.24
AHAN-hangar	TF	4.10	1.55		SMRA-hangar	TF	3.70	1.56
KLAN-Jimmy	0	6.15	1.04		VLHA-Jimmy	0	3.15	1.42
KLAN-Jimmy	T	4.40	1.68		VLHA-Jimmy	T	3.85	1.66
KLAN-Jimmy	F	5.40	1.35		VLHA-Jimmy	F	6.60	0.68
KLAN-Jimmy	TF	4.15	1.84		VLHA-Jimmy	TF	6.25	1.07
KLAN-hangar	0	6.05	0.69		VLHA-hangar	0	3.85	1.76
KLAN-hangar	T	4.55	1.39		VLHA-hangar	T	3.50	1.43
KLAN-hangar	F	3.80	1.67		VLHA-hangar	F	5.25	1.62
KLAN-hangar	TF	4.75	1.52		VLHA-hangar	TF	5.75	1.01
KBAN-hangar	0	5.15	1.39		KLIA-Jimmy	0	3.58	1.62
KBAN-hangar	T	4.38	1.21		KLIA-Jimmy	T	3.55	1.70
KBAN-hangar	F	4.45	1.32		KLIA-Jimmy	F	2.10	1.21
KBAN-hangar	TF	3.85	1.42		KLIA-Jimmy	TF	2.55	1.43
MVAN-hangar	0	5.40	1.39		PLDA-Jimmy	0	5.15	1.53
MVAN-hangar	T	4.85	1.55		PLDA-Jimmy	T	3.93	1.31
MVAN-hangar	F	4.35	1.76		PLDA-Jimmy	F	4.20	1.29
MVAN-hangar	TF	5.00	1.56		PLDA-Jimmy	TF	3.75	1.62