ÚSTAV ANGLICKÉHO JAZYKA
A DIDAKTIKY

Aspirace anglických exploziv u českých studentů anglistiky

Aspiration of English plosives in Czech students of English studies

BAKALÁŘSKÁ PRÁCE

Zpracovala: **Andrea Pospíšilová**

Obor: **Anglistika a Amerikanistika**

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

Akademický rok: **2010/2011**

Declaration of Authorship

Hereby I declare that the following BA thesis is my own work for which I used only the sources and literature mentioned.

Prague, August 12, 2011

Signature

Acknowledgments

I would like to thank my supervisor Mgr. Radek Skarnitzl, Ph.D., for his time, valuable advice and comments, and, in particular, for teaching me how to work in Praat and his patience during the process.

Abstract

Voice onset time (VOT) was shown to provide an effective basis on which to differentiate individual plosives in English. This thesis carries out an analysis of recordings of university students to determine whether specific phonetic instruction improves pronunciation of non-native speakers of English with focus on aspiration of plosives and VOT. In the first part, a theoretical background is provided with definitions of plosives, VOT and its use and measurement. It also reviews the factors that have been shown to affect the VOT values. In addition, a brief overview of second language acquisition is given, with focus on English pronunciation. The second, empirical part describes the material and method used in the analysis and provides figures and results of statistical tests that were run. The results suggest a significant increase in VOT values as a result of one semester of Phonetics and Phonology course. Furthermore, individual plosives and positions in a word were examined separately to determine whether there are any differences as suggested by the theoretical overview.

Keywords: voice onset time, aspiration, Czech English, second language acquisition

Abstrakt

Nástup hlasivkového tónu (voice onset time, VOT) je považován za důlezitý ukazatel pro efektivní rozlišení mezi jednotlivými anglickými explozivami. Tato práce se pomocí analýzy VOT a aspirace v nahrávkách univerzitních studentů snaží rozhodnout, zda cílená fonetická výuka zlepšila výslovnost nerodilých mluvčí v tomto ohledu. V první části je podán teoretický přehled s definicemi exploziv, VOT, jeho použití a měření. Tato část také shrnuje faktory, které se ukázaly jako ovlivňující hodnoty VOT. Z důvodu zaměření práce je do této části začleněn také přehled problematiky osvojování cizího jazyka se zaměřením na anglickou výslovnost. V druhé, empirické části je nejprve popsán materiál a metoda použitá v další analýze, a poté jsou ukázány grafy a výsledky statistických testů. Z těchto výsledků vyplývá, že došlo k významnému zvýšení hodnot VOT po jednom semestru kurzu Fonetiky a fonologie. Kromě toho byly samostatně analyzovány i jednotlivé explozivy, pozice ve slově a dalši jevy, které byly v teoretické části popsány jako ovlivňující hodnoty VOT.

Klíčová slova: nástup hlasivkového tónu (voice onset time), aspirace, česká angličtina, osvojování cizího jazyka

Contents

List of Figures

List of Tables

1 Introduction

The question of whether theoretical instruction in phonetics can improve pronunciation of learners of English, or any other language, is an important one not only for the students themselves, but for their teachers as well. An accented pronunciation may hinder communication even with an otherwise perfect speaker of a second language. This thesis carries out an analysis to try to determine the effect of a first-year mandatory course in Phonetics and Phonology on the aspiration of plosives in various positions within a word.

In the first part, a theoretical background necessary for the following analysis is presented. The first section contains the definitions and categorization of plosives and then introduces the voice onset time (VOT) and the reasons for its use. The following section then reviews factors that have been shown to affect VOT values, such as stress, gender and place of articulation. The next section provides an introduction into second language learning with particular focus on pronunciation. The theoretical part then ends with a section on basic properties of Czech English on both segmental and supra-segmental level.

In the empirical part, two sets of recordings taken before and after one semester of a phonetic course are examined to find out whether there was an improvement between the VOT values. In each recording, VOT values are measured for several plosive positions. A series of statistical tests are run to determine whether the observed change is significant or not. In addition, other aspects influencing the VOT values are analyzed, such as gender, position within a word, stress and place of articulation.

2 Theoretical Part

In this part of the thesis, the theoretical background necessary for the following analyses will be described. The first section introduces plosives, their definition and main categorization and then focuses on description of English plosives. The second part introduces voice onset time (VOT), its use, definition and measurement. In addition, it reviews factors that have been shown to influence VOT values, such as place of articulation, stress or gender. The last section examines second language acquisition with particular focus on English pronunciation, and also the factors influencing the degree of perceived foreign accent. The theoretical part ends with the description of the main features of English pronunciation of Czech speakers.

2.1 Plosives

2.1.1 Definition

Plosives are consonants that are produced by forming a complete stricture in the oral cavity. From the articulatory point of view, the formation of a plosive normally has four stages. First, the articulators move towards each other to form the stricture; this phase is called the closing phase. During the second, compression phase the compressed air is stopped from escaping behind the stricture. The third, release phase occurs at the moment when the articulators move apart to allow the air to escape. Immediately after the release a noise called plosion is produced, because the air is under pressure

when the articulators move apart. This phase is called the post-release phase. (Roach 2000, 32)

Plosives can be distinguished based on their place of articulation. Two plosives $/p, b/$ are called bilabial as they are produced by pressing lips together. During the production of alveolar plosives $/t, d/$, the tongue blade is pressed against the alveolar ridge. The velar plosives $/k, g/$ are produced when the back of the tongue is pressed against the area where the hard palate ends and the soft palate begins. Generally, all plosives can occur in word-initial positions, medial positions between other sounds and in word-final positions. There is another sound that is classified as plosive, the glottal plosive [ʔ], which is produced when the vocal folds are firmly pressed together so that no air can pass between them. In Czech, it is usually pronounced before a word-initial vowel. As Roach (2000, 32) points out, however, in English it is simply an alternative pronunciation of other English plosives in certain contexts; therefore, we will not consider it further.

2.1.2 Categorization

Apart from the place of articulation, the most important categorization of plosives, as with many other speech sounds, is voicing. Lisker and Abramson (1967) note in their article that "among the dimensions which the phonetician finds useful in organizing his description of speech sounds none has a more prominent place than voicing" (1). If the vocal folds vibrate during the production, the resulting sound is voiced. This is the case of voiced English plosives $/b, d, g/$. If it does not, a voiceless sound is produced such as $/p, t, k/$.

While this distinction is said to be quite easily perceived as "buzz emanating from the larynx" (2), it does not fully correspond to the actual voicing of many plosives in different word-positions. In other words, in English, phonological and phonetic voicing do not fully correspond. In Czech, on the other hand, we can talk about the concord of phonetic and phonological voicing because sounds $/p, t, k/$ are pronounced as voiceless, while $/b, d, g/$ mostly

keep their voicing (Skarnitzl, in print). Therefore, for better differentiating among plosives, phoneticians often combine category of voicing with that of aspiration. Aspiration occurs when, in the post-release phase, articulation of a plosive is followed by a period during which air escapes through the vocal folds, making a sound like [h] (Roach 2000, 34). Thus, using this two-dimensional distinction, we can distinguish four types of plosives: voiceless unaspirated, voiced unaspirated, voiceless aspirated and voiced aspirated (Ladefoged 1971, 9). A diagram of the four types can bee seen in Figure 2.1. For a voiceless unaspirated sound the vocal folds start vibrating at about the same time as the stricture is released, while for a voiced unaspirated sound they vibrate during the whole articulation of the sound. For aspirated sounds, "the vocal cords are in the voiceless position during the release" (Ladefoged 1961, 9). Thus, in articulation of a voiceless aspirated sound there is a delay of the start of voicing after the release. On the other hand, for a voiced aspirated sound the vocal folds would be vibrating during the articulation, but then come apart into the voiceless position during the release of the stricture. According to Ladefoged, this type has not been observed in any language yet (Ladefoged 1971, 9). This term is often used for plosives in languages like Hindu, which should, however, rather be called voiced breathy (Skarnitzl, in print).

Individual languages do not necessarily need to use all the four-way opposition. Most of them only use three (such as Thai) or two-way opposition (Czech, English). What is interesting is that voiceless unaspirated plosives are found in almost 92 % of the world languages, voiced unaspirated in about 67 %, and that /b/ and /d/ are more common around the world than velar /g/ (Maddieson 1984, quoted in Skarnitzl, in print).

Apart from voicing and aspiration, there is a third dimension that is often used as a basis for distinguishing individual types of plosives and that is articulatory force. Lisker and Abramson (1964, 385) stress that the assessment of this dimension seems to be a matter of perception, as the loudness of the stop

explosion is connected to higher articulatory force. Thus, $/p, t, k/$ are said to be articulated with higher force and are called fortis, while $/b, d, g/$, or lenis plosives, are said to be articulated less forcefully. This distinction was seen as primary because it alone is operative at all the word positions. However, treatment of articulatory force is quite ambiguous in empirical literature, as it is not clear what exactly should be measured.

Another drawback is the fact that it is not clear whether fortis/lenis opposition is independent of that of voicing and aspiration. As Liker and Abramson (1964, 386) point out, phonetic literature generally fails to present an example of a language that possesses plosive categories that differ only in force of articulation. In most languages, voiceless and aspirated sounds are found to be fortis, while voiced and unaspirated are usually lenis. Moreover, it was also stressed by Lisker and Abramson that "a number of features, *among them voicing and aspiration*, may be taken as manifestations of an underlying division of stops on the basis of a fortis/lenis opposition" (Lisker and Abramson 1964, 386, emphasis in the original). Therefore, as manifestations of individual features are difficult to separate, it is impossible to decide what is the most important distinction between the two sets of plosives.

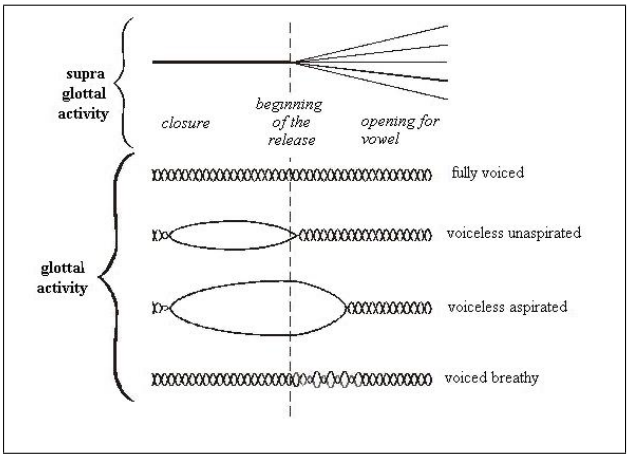


Figure 2.1: Four types of plosives distinguished by the combination of voicing and aspiration.
Source: Skarnitzl (in print).

2.1.3 English plosives

As has already been said, the category of voicing alone is not able to sufficiently describe the system of English plosives, as in certain positions the phonologically voiced sounds lose part of their voicing, and there are other cues than voicing used for differentiating between the respective consonant pairs. In initial positions, for example, voiced plosives lose part of their voicing, while voiceless plosives are aspirated. Therefore, it is not voicing but aspiration that distinguishes the two sets of English plosives in initial positions. In word-final positions, voiced plosives again lose part of their voicing, but the main difference is in the preceding vowel, as vowels followed by voiceless sounds, like $/p, t, k/$, are shortened in English. Therefore, voicing can be used only in a very limited context to separate the two sets of plosives in English.

Even though voicing and aspiration together are able to identify the particular set of plosives, the dimension of articulatory force is used most often in English. However, as has already been pointed out, articulatory force has no agreed-upon physical meaning (Lisker and Abramson 1967, 3), and is thus very difficult to measure. Moreover, even when a certain type of measurement method is chosen, the experimental results do not fully support the theoretical predictions. Already in 1970, an experiment carried out by Lisker (1970) undermined the possibility to use articulatory force (as measured by supra-glottal air pressure) as a straightforward identification of the two sets of plosives. According to his results, only about 5 percent of all the stops could be identified as belonging to either fortis or lenis group based on the air pressure measured. In other words, most of the plosives could not be identified due to great overlap among the values. It is true, though, that Lisker only recorded one speaker, who may not be representative of the whole population of American speakers, not to mention possible variances among American English and English as spoken in other regions. However, the results seem extreme enough to suggest that fortis/lenis distinctions is neither sufficient

not reliable enough. Furthermore, a degree of stress on a particular syllable seems to have an important role when measuring the air pressure. Therefore, a different measure may serve better to distinguish the two sets of English plosives.

2.2 Voice Onset Time

2.2.1 Definition

Another possible way to differentiate plosives was suggested by Lisker and Abramson in 1964. In order to find "some single best measure by which to separate the two phoneme categories" (Lisker and Abramson 1964, 385) they propose to measure the time between the release of the stricture and the onset of voicing, which allows for differentiating among the plosive categories generated by the voiced/voiceless and aspirated/unaspirated oppositions. The authors define "the amount or degree of voicing of a plosive as the duration of the time interval by which the onset of periodic pulsing either precedes or follows release" (386). In other words, they propose to see aspiration as a delay in voice onset, and to differentiate stops based on the voice onset time (VOT) measured as the time between the release of occlusion and voice onset. In their study they prove that differences in voice onset time are sufficient to distinguish different types of plosives in eleven languages, even though the boundaries of individual languages differ. Generally, VOT for voiced unaspirated plosives is negative, because the vocal cords vibrate during the explosion. For voiceless unaspirated sounds the VOT is low, up to 30 milliseconds. Higher values are usually found for voiceless aspirated plosives. However, VOT times for voiced breathy and voiced unaspirated plosives overlap, and thus it is impossible to separate these two categories based on VOT measured. The authors explain this fact by stating that what distinguishes them is rather the kind of voicing than its duration (Lisker and Abramson 1964, 419).

2.2.2 Differences in VOT

As has already been mentioned at the end of previous section, boundaries between plosive categories based on VOT are different in different languages. Skarnitzl (in print) emphasizes the amount of attention paid to voice onset time since the pivotal article by Lisker and Abramson. Therefore, many aspects have been examined since, such as further differences among VOT within one language based on factors like position in a word, place of articulation or nature of the following vowel. Also, socio-linguistic aspects affecting individual speakers have been considered. In the following paragraphs, the most important findings will be reviewed. For a much more extensive overview see for example Skarnitzl (in print).

Segmental and suprasegmental features

Lisker and Abramson already in 1964 noted that mean VOT changes with the place of articulation within one language. Their study shows that velar plosives like $/k, g/$ normally have larger VOT than bilabial or alveolar plosives. As Cho and Ladefoged (1999) state in their article, speakers do not produce these different values deliberately. Rather, "observed VOT is just the inevitable consequence of the physiological movements and the aerodynamic forces" (225) of the articulation of individual plosives. However, in 1967, Lisker and Abramson extended their original research to focus on the effect of various contexts on the values of VOT. The first important conclusion is that the distinction clearly visible when only individual words were considered is somewhat obscured when longer stretches of speech are analyzed. In the more complex setting, it "is no longer possible to completely separate contrasting homorganic categories solely on the basis of the VOT measure" (Lisker and Abramson 1967, 9).

Another factor significant for VOT emphasized by Lisker and Abramson (1967) especially in languages like English is word stress. As aspiration is directly linked to the values of VOT, and as it is generally believed that the

amount of aspiration varies in stressed and unstressed syllables, it follows that such a dependence should be observed. The authors draw two tentative conclusions from their experiments. First, stressed $/p, t, k/$ tend to be produced with longer delays in voice onset than the unstressed ones. Second, $/b, d, g/$ show a contrary trend - those that are stressed tend to have slightly smaller values of VOT. Thus, the distinction between the two contrasting categories is more clearly defined for stressed than the unstressed syllables (Lisker and Abramson 1967, 16). Again, the differences in the values of VOT are smaller in connected speech than in isolated words as can be seen from table 2.1.

	Stressed	Unstressed
Words	70	46
Sentences	41	35

Table 2.1: Mean voice onset time for voiceless plosives in English in different contexts.
Source: Lisker and Abramson (1967, 17).

In connection to the empirical part of this thesis, it should be also noted that voiceless plosives preceded by an $/s/$ phoneme even in stressed syllables are pronounced as unaspirated, and thus with VOT values below 30 ms. The authors also examine the effect of sentence stress on the values of VOT. No significant differences were found with one notable exception. In cases where a plosive-initial word occurs in a sentence-final position which coincides with the sentence stress, the VOT is significantly larger than the mean for plosives in stressed syllables generally. However, for the unstressed syllables in the same position within a sentence, no significant difference was found.

In addition, the rate of speech was examined to determine whether it affects VOT values. Kessinger and Blumstein (1997) showed that higher speech rate does affect VOT in some positions. According to their results, VOT for voiceless aspirated plosives increases significantly with lower speech rate. However, VOT of voiced unaspirated sounds remains relatively stable

across various speaking rates. Therefore, at faster speech rates, there is larger overlap between the two categories.

Another aspect that has been considered as affecting VOT is the vowel following a particular plosive. Lisker and Abramson (1967) did not find any proof of the VOT values being dependent on the following vowel. Also, Kessinger and Blumstein (1997, 155) assert that vowel context seems to have no apparent effect of VOT. However, Higgins *et al.* (1998) suggest longer VOTs for high than low vowels.

Socio-linguistic features

Apart from differences caused by the language itself, further differences may arise because of the various sociological features of the speakers. Allen *et al.* (2003) noted that there are stable differences among individual speakers in VOT duration. In this study, the authors carried out an experiment that proved talker-specific differences in VOT in monosyllabic words with plosives in initial position, even after statistically controlling for contextual differences such as speaking rate. In a related study, Theodore *et al.* (2009) elaborated the topic by focusing on the effect of two contextual features (speaking rate and place of articulation) on individual speakers' VOTs. Their results suggest that VOT changes differently for individual speakers as an effect of a different speaking rate, which implies a systematic variability in speech. However, this change proved to be quite stable across the three places of articulation, suggesting that "the effect of speaking rate on VOT for a given talker is the same for labial and velar voiceless stops" (Theodore *et al.* 2009, 3979).

Other researchers tried to establish the effect of gender on VOT values. Morris *et al.* (2008), for example, after controlling for the phonetic environment and speech tempo, report no male/female differences in VOT observed among either the voiced or voiceless syllables (314). Therefore, the physical gender differences that cause aerodynamic and articulatory differences were found to have no significant effect on the VOT measured. However, as the

study focused on individual syllables, a difference may be found when phrases or sentences are considered. Robb *et al.* (2005) support this view by examining short phrases to encourage naturalistic productions. Their results show that although the durations of individual syllables were similar for men and women, VOTs for voiceless plosives were longer for women than for men, although the same effect was not visible for voiced plosives.

Another feature that has been examined is the effect of age on VOT. The most important finding is the increased variability of VOT values for both children and old people (Sweeting and Baken 1982).

Given the variability of VOT in speech of different individuals, it has been suggested that VOT values may be used for identification of a particular speaker. What is most interesting from the point of view of this thesis is the fact that VOT may be also used to detect dialects or foreign accents (Hansen et al 2010). The authors use VOT in unvoiced stops of Indian, Chinese and American speakers of English to show that "average VOTs are different among three different language groups, hence making VOT a good feature for accent classification" (788). In other words, based on VOT measured, it may be possible to differentiate among different accents.

2.2.3 Measurement of VOT

As Lisker and Abramson point out in their article (1967), unlike other speech sounds, it is quite easy to define measures of voicing for plosives, because the end of closure for a stop is fairly well marked by a burst. Therefore, the onset of this burst can be used as a "convenient reference point for making measurements of voicing duration" (2). However, it is also important to note that different plosives in different contexts may show bursts of varying prominence.

While the beginning of the voice onset time is usually easily found, the other boundary is subject to much larger degree of disagreement. Originally, Lisker and Abramson (1967) suggest to consider the onset of pulsing in spec-

trograms as the end of the VOT. Nowadays, however, the sound wave is usually used rather than spectrograms and the beginning of voicing is thus identified with the beginning of the first complete period after the burst (Skarnitzl, in print). For further details on the approach used in this thesis and examples, see section 3.1.3.

2.3 Second language acquisition

As the present thesis deals with English as spoken by Czech students, it seems appropriate to consider, at least briefly, also the question of second language (L2) acquisition with specific focus on English pronunciation. In the following paragraphs, we will describe the main aspects of foreign language learning, what features determine the degree of foreign accent, and how instruction can affect foreign pronunciation.

2.3.1 Learning pronunciation

According to Setter and Jenkings (2005), pronunciation as a part of second-language teaching is given different significance in different parts of the world, but overall tends to be largely marginalized (1). Cenoz and Lecumberri note in their article that "as a result of historical and social events, the English language shows frequent lack of correspondence between sound and spelling [..] [which] contributes to making pronunciation one of the most difficult areas in the acquisition of English as a foreign language" (Cenoz and Lecumberri 1999, 4). Despite that, though, and despite the great importance of pronunciation for clear communication, it is often not paid enough attention by students of a foreign language themselves and their teachers as well, apart from the most basic acquaintance with unknown phonemes. However, errors in pronunciation of individual phonemes are not likely to cause misunderstanding, while errors at higher levels may significantly hinder communication (Cenoz and Lecumberri 1999, 4). In English, for example, suprasegmental features

such as word-stress, rhythm and intonation are very important and have a key role in acquiring a more native-like pronunciation. However, even when learners might wish to acquire a native pronunciation, they may find it difficult to overcome the deep-rooted features adopted for their native tongue as it operates partly at subconscious level.

As there are many different regional varieties of English, another complication concerning pronunciation is which one to choose as a model for language learning. The two basic models that are used most often are British and American English. European countries as well as former British colonies usually use British English, while American English is used in countries like Japan and south American countries (Setter and Jenkings 2005, 2). Within these two broad types, though, there is a large number of different dialects. In their pronunciation, language learners usually target those dialects that are perceived as 'proper,' while other regional varieties are considered not good enough (Setter and Jenkings 2005, 2). Thus, Received Pronunciation (RP) and General American (GA) dialects are most often chosen as a model for students' pronunciation (Setter and Jenkings 2005, 2). However, with the increasing use of English as an international language, many other non-native varieties are being used as the goal for students of English (Cenoz and Lecumberri 1999, 4).

2.3.2 Degree of foreign accent

In human communication, perception of speech sounds has an important role. Because of that, perception of foreign accent in speech of language learners have been investigated extensively. Although the focus and method of those studies differ greatly (for a review see for example Piske *et al.* 2001), there are several factors that are often identified as having an impact on the degree of perceived foreign accent in speech of L2 students. While most of the studies focus on immigrants who are in everyday contact with the L2 they are learning, many of the conclusions apply to other students of L2 as well.

The factor that is mentioned most often is the age of L2 learning. There are theories stating that "there is a maturational limit (usually set around puberty) beyond which it is simply impossible to acquire an L2 (or certain aspects thereof) to native levels" (Singleton 2001, 79). In addition, Scovel (1988 in Piske *et al.* 2001), when talking about accent, asserts that those who begin exposure to L2 after the age of 12 cannot ever pass themselves off as native speakers in their pronunciation. In other words, there is the so-called "critical period" (CP) for speech learning, and if learning begins after the end of such a period, it is not possible to ever achieve a complete native-like proficiency of a second language. However, researches disagree in the exact value of the CP. Furthermore, there may be multiple critical periods, each applying to a different aspect of language learning. As Piske *et al.* (2001) assert, "the first ability to be lost would be the one needed to develop a native-like pronunciation of an L2" (195).

On the other hand, several other studies seem to indicate that even those who begin later may achieve native-like proficiency. For example, Bongaerts *et al.*(1995) assert in their experimental study that there are late-learners who passed for native speakers phonologically (43). Some researchers even suggest a better performance by late learners than by the early ones. However, these conclusions usually apply to the initial situation, not a long-term pronunciation acquisition. Initially, adult learners may have advantage over young children, but it is only temporary. In a longer horizon, it is the younger learners who achieve best results (Piske *et al.* 2001, 196).

The second most often examined factor is the amount and nature of L2 learning. Obviously, the degree of proficiency is closely related to the amount of time spent in the foreign country or in the company of native speakers. The findings so far suggest that in early phases of L2 learning, an additional experience has a far more significant impact on the degree of foreign accent than in the later phases for more experienced students (Piske *et al.* 2001, 199).

The fact that most of the studies focus on immigrants may also help explain some of the findings. In case of immigrants, children often enter the school system in the foreign country, where they are subject to an explicit second-language instruction. This, however, is mostly unavailable for adults. Thus, children may generally achieve better results in L2 pronunciation also because they are more likely to undergo a formal education in foreign language.

There is also a question whether children and adult learners employ the same strategies in L2 learning. It has been suggested that while adult learners transfer their first-language processing strategies into their second language, children are able to acquire the same strategies that are employed by the native speakers (Singleton 2001, 80). Thus, especially for later learners of L2, their native tongue plays an important role, as it is often found that phonological features of L1 are transferred into L2 pronunciation (Wong 2008, 43).

Another important feature to consider is thus the native tongue. In the 1960s, many researchers used the so-called Contrastive Analysis (CA), which suggests that transfer between L1 and L2 "accounted for and predicted all errors" (Major 2008, 64). However, later the empirical evidence disproved this approach, or at least the predictive ability of CA. Also, other types of errors were found that could not be accounted for by L1 transfer. These mistakes could sometimes be linked to the mistakes made during L1 acquisition (Major 2008, 65).

When considering the effect of L1 on L2 learning, one should consider not only a comparison of the set of phonemes in the two languages, but also other aspects, such as their position in syllables. There are several theories on how the interaction of the two languages works, and how it affects language acquisition (see for example Flege *et al.* 1995). As Major (2008) points out, "a great deal of research has demonstrated that similar sounds tend to be more difficult than dissimilar sounds. The reason seems to be that the

larger the differences are, the more easily they tend to be noticed; therefore, learning is more likely to take place" (71). In contrast, where the differences are small, they remain unnoticed, and thus transfer prevails. Furthermore, it should be noted that the effect of native language is closely related to age and experience of L2 learner. Some authors argue that L1 interferes mainly at the earlier stages of language learning, but at later stages, developmental effects more like those present in L1 learning take its place (Major 1987 in Setter and Jenkins 2005, 4).

Apart from these, one may also consider psychological aspects, such as strength of motivation. There are many variables that may be used to measure motivation, usually a subjective importance of language learning for students is used. In a study by Flege *et al.* (1995), a "concern for L2 pronunciation" and "integrative motivation" were found to be significant predictors of the degree of foreign accent, although they accounted for only 3 % of the variance (Flege *et al.* 1995). Apart from problems with quantification of motivation resulting from subjective evaluation, it should also be pointed out that while motivation may have a positive effect on pronunciation of L2, it is rarely strong enough to enable late learners to achieve an accent-free pronunciation (Piske *et al.* 2001, 202). In connection to motivation, some people may also have larger aptitude for mimicking foreign sounds. While correlation of this feature with musical ability has not yet been experimentally identified, it has been proven to have a significant effect on L2 pronunciation (Piske *et al.* 2001, 202).

2.3.3 Effect of instruction

Given the topic of this thesis, the most important question to ask is whether it is possible to improve pronunciation of second-language learners by targeted instruction. The experimental results so far are not very encouraging for teachers, as the effect is found to be either very limited, or not significant at all. Flege and Fletcher (1992) report the number of years of English-language

instruction to account for only 5 % of the variance in the foreign accent ratings obtained for native Spanish speakers learning English (375).

One of the reasons for the small effect, that has already been mentioned, may be the little amount of attention given to practicing pronunciation. Studies mentioned so far usually do not focus on examining the effect of a targeted phonetic and phonological instruction, they mostly take into account English instruction generally. In addition, as they often focus on immigrants who attend schools in the foreign country, pronunciation is probably not given much space. Therefore, it might be interesting to focus on this aspect when considering the effect on L2 pronunciation.

In a study carried out by Wong (2008), the impact of an eight week course in English phonology on pronunciation of Chinese speakers was examined. The course focused on four areas to improve students' pronunciation. First, their awareness of the errors they made was raised by comparing the English and Cantonese phonologies; second, the rules governing the pronunciation of the problematic sounds were explained; third, awareness of sense groups was increased by analyzing the linguistic features of text types, explaining the meaning of the texts, and physically marking the sense groups on paper before reading; and fourth, learners were asked to imitate how native-speakers read. The students participating in the study were undergraduate students of English. According to her results, the students showed improvement in many aspects of their pronunciation, such as the pronunciation of plural and past-tense morphemes, and the pronunciation of /ɬ/. However, in other areas of pronunciation the course seemed to have no effect, such as realization of the front open vowel (/æ/) in words like "bad". As the author points out, the reason for this finding may be that English phonemes with similar native-tongue counterparts may be more resistant to change (45). Despite that, though, consciousness-raising of problems seems to significantly help improve second-language pronunciation.

2.3.4 Czech students of English

For Czech speakers of English, there are several aspects of English pronunciation that cause significant problems due to the phonological differences between English and Czech. First, on the level of individual phonemes it is the realization of the dental fricatives /θ, ð/, which are often pronounced as alveolar instead of dental. Another feature that often marks Czech English as foreign is the pronunciation of [ŋk] instead of /ŋ/ for 'ng' spelling (Volín and Skarnitzl 2010, 272).

Another problem, the one that this thesis focuses on, is the realization of plosives. In Czech, plosives can be distinguished based on the voiced/voiceless opposition. In some positions, plosives may be re-categorized, so that voiced plosives are pronounced as voiceless and vice versa, but the distinction is always valid. As has been already mentioned, in Czech the phonological and phonetic voicing correspond. There is no need for further differentiation by aspiration like in English. Therefore, Czech students speaking English often do not aspirate voiceless plosives in stressed syllables, and pronounce voiced plosives as fully voiced even at the word-initial positions. On the other hand, in cases where students do aspirate voiceless plosives, they over-generalize and aspirate even plosives preceded by /s/. In addition, as the rules of co-articulation and assimilation are different in these two languages, this can also lead to pronunciation of different phonemes.

On suprasegmental level, students are often not aware of the reduced pronunciation of common grammatical words, which results in a completely different rhythm and thus a heavy foreign accent. Another feature that adds to the foreign accent is the absence of linking in Czech English. In Czech, a word-initial vowel is usually preceded by a glottal stop, while in English there are several ways of how to link the vowel to the preceding word based on the phoneme that precedes it. Also, as Czech is a syllable-timed language, while English is a stress-timed language, another important difference is in the rhythm of speech.

Apart from problems resulting from L1 interference, there are other aspects that may influence the pronunciation of Czech students of English. In the experimental part of the thesis, pronunciation of university students of English studies is examined. As has already been mentioned, contact with native speakers as well as motivation play important roles in the degree of foreign accent. It may be argued that university students are more likely to have higher motivation for learning English and English pronunciation in particular. Also, during their studies they are in contact with native speakers at the department.

In the following part of this thesis, we will examine pronunciation of Czech students of English and American studies at the Faculty of Arts, Charles University in Prague in two sets of recordings with particular focus on plosives. The first recording was taken at the beginning of the students' studies, and the second after the first semester in which they all attended a weekly lectures and seminars in phonetics. We will examine VOT values in both stressed and unstressed positions, and try to determine whether there was a significant improvement in the pronunciation.

3 Empirical Part

In the previous part, a theoretical overview of the VOT and second language learning was presented. In this part, the theoretical background will be used to carry out an experimental analysis of the recordings of Czech students of English. There are several questions this part will try to answer. The most important one is whether the speech of the speakers improved in terms of aspiration after one semester of a phonetic course, and, if so, whether this improvement is statistically significant. In addition, we will examine further differences based on gender and place of articulation.

The first section describes the participants, recordings and software used, methodology, and also proposes the most important research questions. Furthermore, it also provides a brief description of the content of the phonetic course. In the second part, the statistical analysis itself is carried out. This includes not only figures based on summary statistics, but also tests that were run to verify the proposed hypotheses. As this thesis relies heavily on the results of these tests, their basic properties are also described. The last section concludes with the discussion of the results.

3.1 Material and Method

3.1.1 Participants

For the analysis, recordings of 20 students of English and American Studies at the Faculty of Arts, Charles University in Prague were used. The recordings were made for a seminar work required during the course. Each student was

asked to read a short text, adopted from the BBC, at the beginning of his or her first semester, and then again after one semester of the mandatory course in Phonetics and Phonology. All students read the same text on both occasions, which allows for a direct comparison. From now on, the two sets of recordings will be referred to as winter and summer semester recordings.

The recordings for the analysis were selected so that there would be 10 males and 10 females. The individual speakers were given a code consisting of four letters based on the name of the speaker, two digits identifying the year in which the student took the course, two or three letters standing for one of the 10 different texts used and a code for whether it is the winter or summer semester recording. The break down of the speakers and the texts read for any of the two semesters can be seen in appendix A.

From the table it is clear that the distribution of texts among students is not purely coincidental, which may cause a slight bias. However, as the number of male students is significantly lower than the number of females in each year, it was impossible to select a more randomly distributed sample.

3.1.2 Methodology

In each of the ten texts available, all instances in which a plosive is followed by a vowel were found. These include plosives in both stressed and unstressed position, and also plosives that are preceded by an /*s*/ phoneme. Most of the foreign names were excluded, but place names (like states) and common names of firms or organizations were included. Then, all actually pronounced plosives in those positions were labeled in Praat[1]. A table with all the words in which a plosive could occur was made, in which each token was coded in terms of the type of plosive (/*p*, *t*, *k*/), status of whether it can be used or not and its position in a word.

There were five possible positions in which a plosive could occur. The first was a word-initial stressed position (S1), such as in words *part* or *car*.

[1]For more details on the VOT measurement see section 3.1.3

The second was a stressed-syllable position, but other than word-initial (S2), such as *attempt*. Unstressed positions, in words like *criticized* were marked by U. All grammatical words belong to this category. The last two positions, SX1 and SX2, refer to those instances in which a plosive is preceded by /*s*/ in a stressed syllable that is either word-initial or not, such as *space* or *experience*. The reason for distinguishing between individual positions is that each is expected to have different VOT values[2]. The position assigned to each plosive was based on how the word was actually pronounced by a speaker, not how it should be pronounced by a native speaker. The resulting table in Excel for each speaker separately can be seen on the accompanying CD with all the data. Out of almost six thousand possible occurrences, 4988 could be used for further analysis. A more detailed breakdown of the numbers can be seen in tables 3.2 and 3.3.

Status	**Number**
Fine	4988
Excluded	33
Non-usable	699
Total	**5720**

Table 3.2: Number of occurrences based on Status. See text below for details.

Position	**Number**	**p**	**t**	**k**
S1	1589	503	391	695
S2	305	162	90	53
U	2727	408	1810	509
SX1	298	79	212	7
SX2	69	30	33	6
Total	**4988**	**1182**	**2536**	**1270**

Table 3.3: Number of occurrences based on Position and Plosive. See text above for details.

The "non-usable" status indicates those instances in which a plosive was not pronounced at all, it was pronounced as some other phoneme or there

[2]For details see theoretical sections 2.2.2 and 2.3.4

was no following vowel. It also includes instances in which a plosive was pronounced as voiced, typically /*t*/ pronounced as alveolar flap. Sometimes, even though a plosive was pronounced as it should, it was impossible to distinguish the exact boundaries of the VOT. In those cases, the plosive was given the "excluded" status. Mostly, it was due to no visible plosion in either the wave or spectrogram. Also, in /*st*/ sequences, the boundary between the two phonemes was sometimes impossible to find. In table 3.3 we can see the number of occurrences based on position first for all the data together, and then separately based on the place of articulation.

3.1.3 VOT boundaries

As already mentioned in the theoretical part, the VOT boundaries are not considered difficult to find. The approach used in this thesis was based on Machač and Skarnitzl (2009) and Skarnitzl (in print).

First, the plosives were identified based on the absence of formant structure. This rule was always applicable for the identification of the end of the plosive, as we were interested only in those cases where a vowel followed. Vowels are characterized by the presence of full formant structure, and thus the boundary could be easily found (Machač and Skarnitzl 2009, 30). In cases where this rule could not be applied (for example the /*st*/ sequences), additional rules, such as the presence of fricative noise were applied.

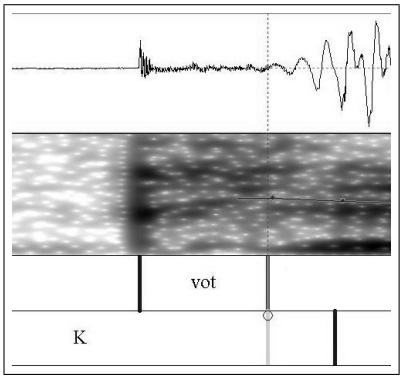


Figure 3.2: VOT boundaries.

Second, the beginning of the VOT was identified with the beginning of a plosion. The voiceless plosives we were interested in mostly have relatively prominent plosions. The basic example can be seen in figure 3.2. Cases when the explosion could not be identified in either the waveform or the spectrogram (mostly for /p/) were excluded. Sometimes, not one but multiple explosions were visible. In those cases, usually the first plosion was identified as the beginning of the VOT, as can be seen in figure 3.3. However, two exceptions were made to this general rule. If the first explosion was very close to the beginning of the plosive, so that almost none closure would be present, the second explosion was marked as the beginning of the VOT. In addition, if the first plosion was very small (as seen in the waveform) and not prominent in the spectrogram, the second plosion was used. These cases can be seen in figure 3.4.

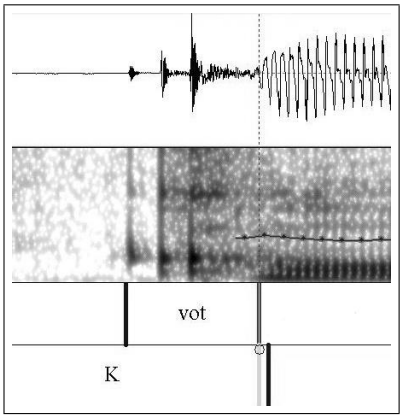
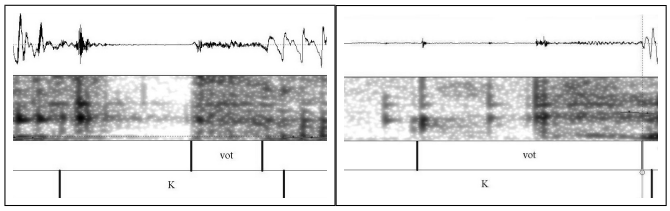


Figure 3.3: VOT boundaries for multiple explosions.

Third, the beginning of voicing, and thus the end of VOT was identified with the first prominent period in the waveform. This could be seen in figures 3.2, 3.3 and 3.4.

3.1.4 Phonetic instruction

The first recording was taken at the beginning of the first semester of a two-semester mandatory course in Phonetics and Phonology. The second



(a) First exception (b) Second exception

Figure 3.4: Two possible exceptions to the rule on multiple explosions.

recording was taken at the beginning of the second semester. During the first semester of the course, students are introduced to the terminology and the system of English phonemes from the articulatory and perception point of view. In addition, the students learn the International Phonetic Alphabet and are expected to complete a seminar work, which consists in the phonetic prediction for an RP speaker and phonetic transcription of part of their own recording. A particular emphasis is put on the differences in pronunciation of Czech and English phonemes, including vowel differences and dental fricatives. Aspiration is also covered in the first semester as an important aspect of these differences, and it is also a part of the requirements for the allophonic transcription practiced throughout the semester.

3.1.5 Research Questions

There are several interesting questions the thesis attempts to answer.

1. Do the VOT values in various positions in both winter and summer semester recordings correspond to the values of native speakers?
2. Does the phonetic instruction in the first semester alter the VOT values?
3. Are the VOT values in stressed syllables significantly higher for the summer semester recordings?
4. Are there any differences between values for males and females?
5. Are there any differences observed for different plosives?

These question will be dealt with in section 3.2. In analyzing the data, we expect to find that phonetic instruction does have a significant effect on the VOT values in stressed syllables; the values are expected to increase as a result of the increased awareness. However, in cases where a plosive is preceded by /s/, the values are expected to remain the same or decrease in case the plosives were aspirated in the first recording. In addition, we expect to find no differences between the values of males and females. As the theoretical part suggests, differences in values based on place of articulation should be observed, with the values increasing as the place of articulation moves backwards in the vocal tract.

3.1.6 Software used

The recordings were made in a sound-treated recording studio of the Institute of Phonetics at the Faculty of Arts. They were labeled in Praat version 5.2.30[3]. The VOT was extracted using a Praat script. The obtained data was organised and filtered for various uses in Microsoft Excel and in Gretl version 1.9.5.[4] Descriptive statistics, such as mean, median, standard error and boxplot graphs were found using Gretl. Bar charts were done in Excel. All tests, such as the t-tests, their non-parametric variants and tests for normality and equality of variances were done in *R*, version 2.13.1[5].

3.1.7 Statistical tests

In the present thesis, several statistical tests were carried out to determine whether changes in the data are significant or not. To analyze differences between winter and summer semester recordings, two tests were used. Both tests work with pairs of data, so we only included those values that occurred in the appropriate position in both recordings. This makes the number of

[3]Home page: http://www.fon.hum.uva.nl/praat
[4]Home page: http://gretl.sourceforge.net
[5]Home page: http://cran.r-project.org

observances slightly smaller than what is worked with in the figures presented in the following sections.

The most important test that was used is a matched-pairs t-test. This test is appropriate because there is a connection between the two sets of recordings, as each pair was recorded by the same person in two different settings. The null hypothesis is that the means of the VOT values in the winter and summer semester recordings are equal. We can re-formulate the null hypothesis to state that the random sample $d = [d_1, d_2, \ldots, d_n]$ comes from a distribution with the mean equal to zero, where $d_i = x_{1i} - x_{2i}$ for $i = 1, 2, \ldots n$ is the difference between the pairs of values. Thus, we test the hypothesis

$$H_0 : \mu_d = \mu_1 - \mu_2 = 0$$

against the alternative: $H_A : \mu_d = \mu_1 - \mu_2 \neq 0$.

The value of t-statistics for this test is given by the following formula:

$$t = \frac{\overline{d}}{s_d}\sqrt{n}$$

The standard error is given by:

$$s_d = \sqrt{\frac{\sum_{i=1}^{n}(d_i - \overline{d})^2}{n-1}}$$

where $\overline{d} = \frac{\sum d_i}{n}$ is the average of the differences between the pairs of values. The number of degrees of freedom is given by $(n-1)$. It should be noted that n is not the number of individual measurements, but the number or pairs. The critical region for a two-tailed test is given by $\{t : |t| \geq t_{1-\frac{\alpha}{2}}(n-1)\}$, where α is the chosen level of significance. If the value of the test-statistics is in the critical region, we reject the null hypothesis.

There are two assumptions that should be tested to verify that the use of this test is appropriate. The first one is normality of distribution for each of the two sets of data we are comparing. We check this using the Shapiro-

Wilk test. The second one is homogeneity of variances, which could be tested by a basic F-test. However, as this test assumes normal distribution of the data and this was mostly violated, a more robust Lavene's test was used in most cases. The null hypotheses of both the Lavene's test and Shapiro-Wilk tests are that the respective assumptions are fulfilled. When one or both of these assumptions are violated, the t-test may give biased results. For large number of observances, it was shown that the requirement on normality can be violated without serious consequences (Volín 2007, 113). However, when both assumptions are violated, Volín (2007, 112) recommends to use a non-parametric variant of the t-test, Wilcoxon signed-rank test.

For this test we define $z_i = x_{1i} - x_{2i}$ for $i = 1, 2, \ldots, n$ with median Θ. The null hypothesis tested and the alternative are:

$$H_0 : \Theta = 0 \qquad H_A : \Theta \neq 0$$

The test statistics for a two-tailed test is given by:

$$w = min\{w_+, w_-\}$$

$$w_+ = \sum_{i=1}^{n} \varphi_i R_i \qquad and \qquad w_- = \sum_{i=1}^{n} \phi_i R_i$$

where R_i is the rank of each non-zero $|z_i|$ ordered in ascending sequence, φ_i and ϕ_i are indicator functions $\varphi_i = 1$ for $z_i > 0$ and $\varphi_i = 0$ otherwise and $\phi_i = 1$ for $z_i < 0$ and $\phi_i = 0$ otherwise. In other words, we first sum the rank of the positive and negative differences separately, and then find the smaller value. In our analysis, in most cases both tests were run for comparison.

In addition, two-sample t-test and its non-parametric variant, Mann-Whitney test were used in two instances. More details on these tests and their use in phonetics can be found for example in Volín (2007).

3.2 Statistical Analysis and Results

3.2.1 Overall results

In table 3.4, we can see the mean, median and standard error of the VOT values for both winter and summer semester recordings in all the positions. Throughout this section, the values are always given in milliseconds. It is clear that there is an increase in both mean and median values for all positions. However, even in the summer semester recordings, when a plosive is preceded by /s/, the values are still below the boundary 30 ms. The table also shows us that even before the course, the average values in positions S1 and S2 were well above this boundary, so that on average students aspirated their plosives in stressed syllables.

	Mean		**Median**		**St.err.**		**N**	
Position	ws	ss	ws	ss	ws	ss	ws	ss
S1	40.691	48.379	35.2	44.6	24.715	27.695	790	799
S2	41.527	45.746	37.9	42.75	23.221	26.25	135	170
U	34.114	37.407	30.4	33.3	17.652	19.141	1364	1363
SX1	23.2	24.433	21.4	23,3	11.886	12.25	150	148
SX2	25.539	29.625	22.5	27.1	14.164	14.052	33	36
Total:							**4988**	

Table 3.4: Mean, median and standard error of the VOT values separately based on position in both sets of recordings, and a number of ocurrences in each semester. In this and all the following tables, the VOT values are in miliseconds.

First, we can compare the obtained values to values of native speakers as shown in table 2.1 in the theoretical part. Obviously, both stressed and unstressed values are significantly higher for native speakers. To compare individual plosives, we can use data as reported by Klatt (1975). His data and data on comparable positions from our analysis are provided in table 3.5.

However, when comparing the data, we should bear in mind that there are some differences between Klatt's approach and the approach adopted in this thesis. First, only stressed word-initial clusters were examined by Klatt. For

Plosive/ cluster	Mean VOT in Klatt	Number	Ws mean VOT	Ws number	Ss mean VOT	Ss number
p	47	15	28.784	248	38.641	255
t	65	15	45.753	191	51.292	200
k	70	15	46.35	351	53.903	344
sp	12	15	13.854	39	13.91	40
st	23	15	26.528	107	28.236	105
sk	30	15	25.3	4	31.633	3

Table 3.5: VOT values for word-initial plosives and different consonant clusters in stressed positions as reported in Klatt for native speakers (1975, 19) and as derived from our analysis.

that reason, we also include only data on positions S1 and SX1. the second difference is the approach to determining VOT boundaries. Therefore, minor differences in values could be expected due to this. Obviously, the values for word-initial plosives are lower than they are for native speakers, even in the summer semester recordings. On the other hand, values for clusters starting with /*s*/ are slightly higher than they should be, although still below the 30 ms boundary.

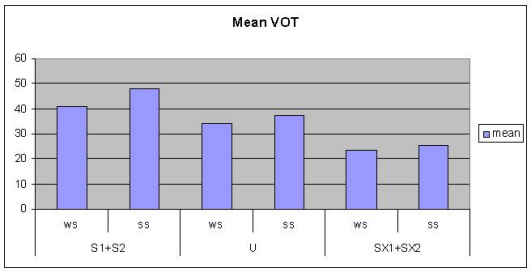


Figure 3.5: Differences in mean VOT in various positions for the winter and summer semester recordings separately. In this and all the following figures, the VOT values are in milliseconds. Number of occurrences: $n_{S,ws} = 925$, $n_{S,ss} = 969$, $n_{U,ws} = 1364$, $n_{U,ss} = 1363$, $n_{SX,ws} = 183$, $n_{SX,ss} = 184$.

Let's now take a closer look at the data obtained in our analysis. Figure 3.5 shows the stressed and unstressed positions in comparison. An increase between the winter and summer semester values is clearly visible. However, because of the high standard deviations, mean values may be misleading.

Therefore, in figure 3.6 we can see more about the distributions of the values for stressed positions (S1 and S2) alone.

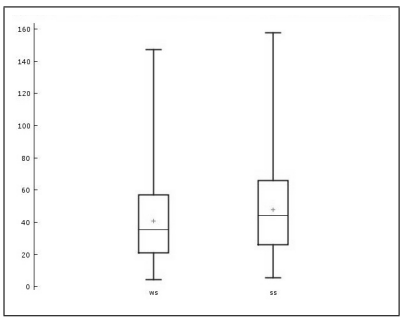


Figure 3.6: Boxplot of the VOT values in stressed positions for the two sets of recordings. The central box encloses the middle 50 percent of the data, the "whiskers" extend to the minimum and maximum values. A line is drawn across the box at the median and a "+" sign identifies the mean. $n_{ws} = 925$, $n_{ss} = 969$.

In the boxplots we can see that the mean is higher than median because of a relatively few values that are very large. For that reason, in further tables and figures, median will be used instead of mean as it is less vulnerable to extreme values.

From the two figures, it seems as though the difference between winter and summer semester recordings is quite large. A t-test was carried out to test whether this difference is significant. Before computing the test statistics, the assumptions of the test have to be checked. The first assumption is that the distributions of the two data sets are normal. The resulting p-values are both < 0.001, so we have to reject the first assumption. The second implicit assumption is that the variances of the two data sets are equal. The resulting p-value of the Levene's test is 0.001937, which means that we also have to reject the null hypothesis at the significance level $\alpha = 0.05$. Because neither of the assumptions is fulfilled, the results of the t-test may be biased. Therefore, a non-parametric Wilcoxon test will be carried out as well. However, the results of the t-test will be included for comparison, as the large number of occurrences may help avoid the bias. The resulting test statistics and p-values

are:

$$t\,(851) = -8,502 \qquad p - value < 0.001$$
$$w = 120333.5 \qquad p - value < 0.001$$

As the p-values are very small, we do not get the exact numbers, and thus we cannot compare them to each other. However, both tests give the same result - we reject the null hypothesis on $\alpha = 0.05$. In other words, there is a significant change in the VOT values in stressed positions between the winter and summer recordings. From the previous figures we can also see the direction of the change; the mean VOT values increased.

We can also check whether the same significant increase can be observed for the VOT values in unstressed position. In figure 3.5 the difference is visible, but it is smaller than for the stressed position. The p-value of the Levene's test is 0.1152 which means we cannot reject the equality of variances. The p-values of the normality test are both < 0.001, so we reject the normality assumption.

The resulting test statistics for unstressed position are $t\,(1159) = -5.8792$, $p - value < 0.001$ and $w = 265602.5$ with $p - value < 0.001$, which means that we reject the null hypothesis. This result suggests that while it is not necessary to aspirate plosives in unstressed positions, it may be easier for the students to increase VOT values for plosives in both stressed and unstressed positions than in the stressed ones only.

The third possible group of positions in which a plosive could occur is when a plosive is preceded by /s/ in a stressed position (SX1 and SX2). The number of plosives in these two positions was not very large, only 168. The p-value of the Levene's test was 0.6563, so that we cannot reject the null hypothesis about equality of variances. The normality assumption, however, is not fulfilled in either case.

The test statistics are $t\,(167) = -1.5852$ ($p - value = 0.1148$) and $w = 5620$ (0.01928). In this case, for $\alpha = 0.05$, the two tests yield different results.

While based on the t-test we cannot reject the null hypothesis about the homogeneity of variances, the opposite is true based on the Wilcoxon test. As only the assumption on normality was violated, which may be offset by the still relatively high number of observances, it is not easy to decide which test should be given priority. However, given the very small values of the tests on normality and the much lower number of observances than in other cases, Wilcoxon test seems more reliable as it is more robust. Thus, we could say that the null hypothesis can be rejected on $\alpha = 0.05$. This result would confirm the conclusion from the previous paragraph that it is easier for the students to increase the VOT values indiscriminately in all possible positions. However, as both the values of winter and summer semester recordings are well below 30 ms, students still do not aspirate plosives that are preceded by /s/.

3.2.2 Individual speakers

As could be expected, there were differences among individual speakers observed in the data. In figure 3.7 we can see the difference in median VOT in stressed positions (S1 and S2) for each of the speakers separately. While some of the speakers' VOT values were well above the 30 ms boundary already in the winter semester recording, others did not improve enough even after one semester. In addition, even similar changes may be of different importance in that changes from below-30 to above-30 ms are the most important, as the border between aspirated and unaspirated plosives is crossed. More than half of the students can be said to have improved visibly based on the figure.

We could try to test whether these improvements are significant or not. Given the fact that the average number of observances for individual speaker is only 42, in case the p-value of the Shapiro-Wilk normality test is less than 0.05 for any of the two semesters, only the Wilcoxon test was used. The resulting p-values can be seen in the figure as well. Overall, out of the 20 speakers, nine can be said to have improved significantly ($p < 0.001$ or

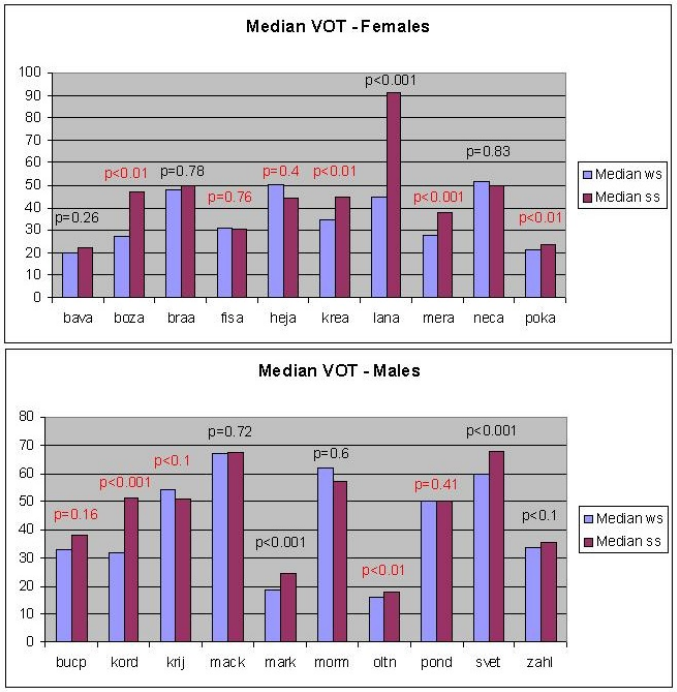


Figure 3.7: Comparison of the VOT median in stressed positions for individual speakers. Above each pair there is a p-value of either the t-test (black) or the Wilcoxon test (red) based on whether the assumption on normality was fulfilled. Number of occurrences ranged from 28 to 49.

$p < 0.01$). VOT values of another two students changed marginally, as their p-value was still < 0.1. In the case of "krij", the change seems to be negative. However, for this speaker, both winter and summer semester values were well above the 30 ms boundary. In the values of the remaining students, no significant change was found.

3.2.3 Gender differences

An interesting question raised in the theoretical part is whether there are any differences between the data for males and females. We can see the median values for different positions in figure 3.8. It seems as though for the unstressed positions the differences are very small. Also, for the plosives following /*s*/ the differences do not seem to be large, and the small difference present in the winter semester recordings is almost evened in the second

semester recordings. The main difference seems to be in stressed positions, where the values for males are higher in both semesters. For this position, we can check whether the differences between winter and summer semester recordings are significant for males and females separately.

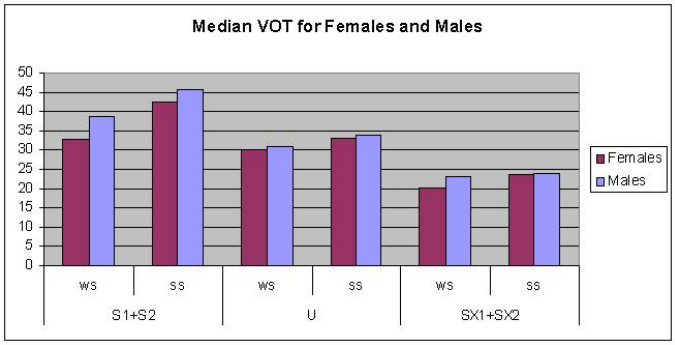


Figure 3.8: Median VOT values for males and females in various positions. Number of occurrences: $f_{S,ws} = 458$, $f_{U,ws} = 677$, $f_{SX,ws} = 98$, $f_{S,ss} = 482$, $f_{U,ss} = 680$, $f_{SX,ss} = 101$, $m_{S,ws} = 467$, $m_{U,ws} = 687$, $m_{SX,ws} = 85$, $m_{S,ss} = 487$, $m_{U,ss} = 683$, $m_{SX,ss} = 83$.

When checking the assumptions of the test for females, we again have to reject the null hypotheses about normality of distribution (both < 0.001) and equality of variances (< 0.001). For males, we also reject normality of distributions (again both < 0.001). However, the equality of variances cannot be rejected on $\alpha = 0.05$ as the $p-value = 0.9248$. The resulting test statistics are:

$$t_M\,(432) = -4.4083\,(< 0.001) \qquad w_M = 33718.5\,(< 0.001)$$
$$t_F\,(418) = -7.3686\,(< 0.001) \qquad w_F = 26786.5\,(< 0.001)$$

Obviously, in both cases we reject the null hypothesis about the equality of means, so that there is a significant change in the VOT means for both males and females.

In the theoretical part, a question of whether there are any differences between males and females was also raised. We can test that as well. For the test, the data on stressed and unstressed positions were used only, excluding

data on plosives preceded by /s/. Unsurprisingly, assumption on normality had to be rejected for all data samples, so both the two-sample t-test and Mann-Whitney test were carried out. A basic F-test was run to verify equality of variances, and thus to determine the exact specification of the t-test.In both winter and summer semester recordings, the homogeneity of variances had to be rejected. For winter semester recordings, the results are:

$$t-test: \qquad t(2282) = -3.2459 \qquad p-value = 0.0012$$
$$M-Wtest: \qquad w = 604949.5 \qquad p-value = 0.0016$$

We can see that in the winter semester, the difference between males and females was very significant, with VOT for males higher than for females. The results for the summer semester are:

$$t-test: \qquad t(2282) = -3.2459 \qquad p-value = 0.3869$$
$$M-Wtest: \qquad w = 604949.5 \qquad p-value = 0.7364$$

Based on p-values, we cannot reject the null hypothesis that the VOT values are the same for males and females in the summer semester recordings. Therefore, based on the results of all the tests performed, it seems as though females improved more than males, as their VOT values were significantly lower than VOT of males in winter, but not in summer semester recordings.

3.2.4 Individual plosives

In the theoretical part it was suggested that there should be a relationship between the VOT values and the place of articulation. In figure 3.9 we can see the median VOT values separately for /p/, /t/ and /k/. We can partly support the conclusions that the VOT values increase as the place of articulation moves backwards. There seems to be a large difference between the values of /p/ and the values of /t, k/. The difference between the latter two

phonemes, though, is not as prominent.

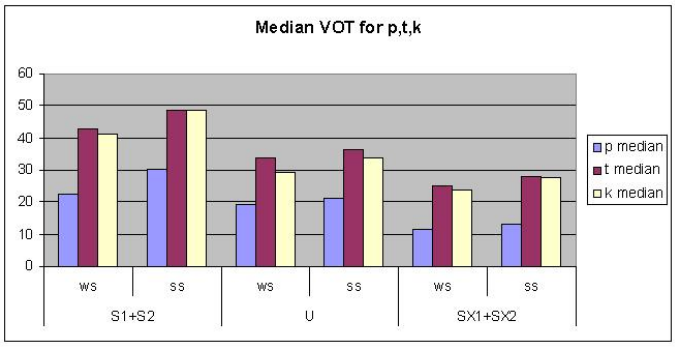


Figure 3.9: Median of the VOT for individual plosives for different positions in a word. Number of occurrences: $p_{S,ws} = 319$, $p_{U,ws} = 214$, $p_{SX,ws} = 53$, $p_{S,ss} = 346$, $p_{U,ss} = 194$, $p_{SX,ss} = 56$, $t_{S,ws} = 233$, $t_{U,ws} = 895$, $t_{SX,ws} = 123$, $t_{S,ss} = 248$, $t_{U,ss} = 915$, $t_{SX,ss} = 122$, $k_{S,ws} = 373$, $k_{U,ws} = 255$, $k_{SX,ws} = 7$, $k_{S,ss} = 375$, $k_{U,ss} = 254$, $k_{SX,ss} = 6$.

These conclusions can be tested by a two sample t-test. We test whether there are differences between the plosives in winter and in summer semester recordings separately. As in the previous section, we only include the stressed (S1 and S2) and unstressed (U) positions of each plosive. As we can only test two plosives at a time, three tests for each of the semesters had to be run. First, however, the assumptions have to be tested. The null hypothesis on the normality of distribution has to be rejected for all the six separate sets of values (3 plosives in two semesters) as all the corresponding p-values are < 0.001. The results of the test on homogeneity of variances and the Mann-Whitney test for the winter semester are[6]:

$$F_{pt}(532, 1127) = 0.9677\,(0.67) \qquad w_{pt} = 188982\,(< 0.001)$$
$$F_{tk}(1127, 627) = 0.9386\,(0.37) \qquad w_{tk} = 331675\,(0.02704)$$
$$F_{pk}(532, 627) = 0.9083\,(0.25) \qquad w_{pk} = 96163\,(< 0.001)$$

Obviously, the null hypothesis about the homogeneity of variances cannot be rejected for any of the three possible pairs of plosives. The p-values of

[6]Again, there were no significant differences between the t-test and Mann-Whitney test for $\alpha = 0.05$

the test comparing /p/ and /k/, and /p/ and /t/ were < 0.001 suggesting a significant difference between the two sets of values. However, for the difference between /t/ and /k/, we cannot reject the null hypothesis that the two sets are the same.

For the summer semester recordings, the results are:

$$F_{pt}(539, 1161) = 1.5042\,(< 0.001) \qquad w_{pt} = 229246\,(< 0.001)$$
$$F_{tk}(1161, 629) = 0.7387\,(< 0.001) \qquad w_{tk} = 325435\,(< 0.001)$$
$$F_{pk}(539, 629) = 1.111\,(0.2) \qquad w_{pk} = 110095\,(< 0.001)$$

In this case, the assumption on the equality of variances had to be rejected for /p/ and /t/, and /t/ and /k/ pairs of values. However, for the last pair /p/ and /k/ it could not be rejected on significance level $\alpha = 0.05$. For all the three pairs of plosives, the p-values suggest that the differences were highly significant for the summer semester recordings.

The next question is again whether there are significant differences between the winter and summer semester recordings. The results are given in table 3.6.

Plosive	df	t-statistic	w-statistic	Normality	Variance
p	296	-6.8639 (< 0.001)	11866.5 (< 0.001)	< 0.001 < 0.001	< 0.001
t	213	-2.7747 (0.006016)	8804.5 (0.002935)	< 0.001 < 0.001	0.3275
k	340	-4.8533 (< 0.001)	20412 (< 0.001)	< 0.001 < 0.001	0.167

Table 3.6: T-test and Wilcoxon test results and tests of assumptions for $/p, t, k/$ separately.

The numbers in the "normality" and "variance" columns are p-values of the corresponding tests; in the "normality" column the two values correspond to the two sets of recordings. The p-values of the t-test and the Wilcoxon test are given below the test statistics in parentheses. Obviously, in all the

cases we have to reject the assumption on normality of distribution. However, for /t/ and /k/, the test for the equality of variances gives relatively high p-values, so that the null hypothesis cannot be rejected for $\alpha = 0.05$. Most importantly, we reject the null hypothesis of both the t-tests and Wilcoxon tests, which means that in all three cases there was a statistically significant increase in the VOT values.

3.3 Discussion of Results

From our analysis, several important conclusions can be reached. First, based on the values obtained it can be shown that the students generally do aspirate plosives in stressed positions when not preceded by /s/. However, the VOT values are significantly lower than native speakers' values. This is not surprising given that in the students' mother tongue, Czech, voiceless plosives are not aspirated. Where preceded by /s/, the students did not aspirate plosives in any of the two semesters, even though the values increased in the summer semester recordings just as they did for other positions. As already mentioned, this suggests that it might be easier for students to increase their VOT values indiscriminately for all positions rather than for the stressed position only.

Second, the most interesting question was whether, generally, there is a significant improvement in the VOT values between the two sets of recordings. The performed tests proved that there is, not only for all the speakers together, but also for 9 out of 20 individual speakers. In other words, the test suggest that phonetic course helped improve the VOT values to a more native-like levels. This is quite a high number given the fact that the second recording was taken after only one semester of a two-semester course, so that further improvement may be expected.

Third, it was shown that there is a difference of the VOT values based on the place of articulation. In the winter semester, however, the prediction from the theoretical part could not be fulfilled. The difference was limited to

/p/ with relatively small values, and /t, k/ with higher ones, as also the tests of significance confirm. In the summer semester recordings, the difference between /t/ and /k/ became significant, while the previous differences were kept as well. In addition, the tests confirmed that VOT values of all the plosives increased significantly. Therefore, the VOT values for /k/ must have increased more than those for /t/. The fact that there are differences between the values based on the place of articulation supports the suggestion presented in the theoretical part that there might be articulatory factors that make VOT values for velar sound higher than for alveolar of bilabial ones.

Fourth, both males and females' VOT values increased when examined separately. In the winter semester, a very significant differences were found between the values of males and females. These differences became insignificant in the summer semester, suggesting that females improved more. This could be partly due to the fact that the values for females were lower, and thus even after their higher increase in VOT values, they only managed to "narrow the gap". However, the relatively small number of male students when compared to female students, and the fact that most of males read the same text[7] may at least partly explain the reason why male speakers had higher VOT values on both occasions.

However, there are several aspects that should be taken into consideration when analyzing the results. The most obvious problem is the fact that in about half of the cases, none of the two important assumptions of the test were fulfilled. The normality assumption may not cause problems as the Wilcoxon test may be able to overcome that. In addition, the comparison of the t-test and Wilcoxon test statistics suggests that the large number of observances was usually sufficient to ensure the correct results. However, the violation of the homogeneity of variances may still cause some bias in the results.

Furthermore, even if we can accept the results as they are, there are several

[7]See appendix A

important points that should be considered. First, there was no control group for our analysis, so that we cannot be sure that the phonetic course was the sole cause of the improvement. It is possible that the time difference alone may be responsible for the improvement. In addition, the students were exposed to pronunciation of native speakers during the first semester of their studies, which could also account for part of the improvement. Further studies should try to separate those various influences to determine what effect does the phonetic course has in itself.

Another point is the fact that we did not distinguish between changes at low-level values (below 30 ms) and high-level values (above 30 ms). A change that shifts the VOT value from below-30 to above-30 level is obviously significant in that the boundary between unaspirated and aspirated plosives is crossed. A change at already high levels does not add to the perception difference as much as at lower levels.

Another aspect for further consideration would be the value of VOT with respect to not only word-stress level, but also phrase and sentence stress. In speech of a native speaker, the overall stress should result in a certain level of VOT, and there should be differences even within plosives that are all labeled as aspirated, as was also mentioned in the theoretical part.

In addition, the large differences between individual speakers may not be only a result of different pronunciation abilities, but also, as the theoretical part suggests, a result of different speech rates. Obviously, higher speech rate results in shorter time of all phonemes, and thus also lower VOT values. Therefore, speech rate should be statistically controlled for when comparing individual speakers to each other. However, when considering differences between winter and summer semester VOT values, speech rate should not influence them as much as the absolute VOT values.

Therefore, further studies should consider the question why some individuals improved significantly, while the others did not improve so much, or not at all. Several factors could be suggested that could affect individual's

susceptibility to instruction and ability to incorporate theoretical knowledge into his or her own speech. Also, the nature of previous study of English language should be taken into account when assessing the results.

4 Conclusion

The aim of the present thesis was twofold: first, to provide a theoretical introduction to voice onset time and second language acquisition in connection to Czech speakers of English. Second, it attempted to determine the effect of targeted phonetic instruction on the pronunciation of non-native speakers with specific focus on aspiration of plosives.

From our analysis, several important conclusions can be drawn. First, on average students did aspirate their plosives in both stressed and unstressed positions even in the first semester recordings, just like a native speaker would. However, in comparison to native speakers, the students' VOT values were significantly lower. On the other hand, in positions when a plosive followed /*s*/, the VOT values were slightly higher for students than for native speakers, but still below the 30 ms, and thus the plosives were not aspirated in either of the two semester recordings.

The most important finding is that there was a highly significant increase in the VOT values for both stressed and unstressed positions between the winter and summer semester recordings. In addition, when examined separately, nine speakers out of 20 could be said to have improved significantly based on data in stressed positions. In positions following /*s*/, the results are marginally significant. Furthermore, when males and females were examined separately, both groups were found to have improved significantly as well.

The study also supports the findings of previous research in several key points. First, the place of articulation proved to have a significant effect on VOT values in both semesters, although the differences are fully present only

in the summer semester recordings. VOT values for each plosive separately have also increased significantly between the two semester recordings.

While the results seem quite straightforward, several other aspects should be considered for further study to gain better understanding of the influence of instruction on non-native pronunciation. For example, not all factors that influence VOT values, such as sentence stress or speech rate, were taken into account. In addition, factors outside the language, such as the length and nature of English studies, could also help illuminate the differences between individual speakers.

The last thing that should be considered is that the second recording was taken after only one semester of the two-semester course. While the theory on aspiration was already covered during the first semester, it may take the students more time to incorporate theoretical knowledge into their speech. Therefore, it would have been necessary to take a third recording after the completion of the whole course to see whether any further improvements occurred.

In conclusion, while the analysis carried out in this thesis came up with several interesting conclusions, it was by far not exhaustive. There are many aspects that should be further developed to provide a much clearer picture about the factors that determine the VOT values in pronunciation of non-native speakers of English.

Bibliography

[1] Allen, J. S.; Miller, J. and D. DeSteno. (2003) "Individual talker differences in voice-onset-time." *Journal of the Acoustical Society of America* 113, 544-552.

[2] Bongaerts, Theo; Planken, Brigitte and Erik Schils. (1995) "Can late starters attain a native accent in a foreign language: A test of the critical period hypothesis." In Singleton, David and Zsolt Lengyel (Eds.). *The age factor in second language acquisition.* Clevedon: Multilingual Matters, 30-50.

[3] Cenoz, Jasone and Luisa Garcia Lecumberri. (1999) "The acquisition of English pronunciation: learner's views." *International Journal of Applied Linguistics.* Vol. 9, No. 1, 3-15.

[4] Cho, Taehong and Peter Ladefoged. (1999) "Variation and universals in VOT: evidence from 18 languages". *Journal of Phonetics* 27, 207-229.

[5] Flege, James E. and Kathryn L. Fletcher. (1992) "Talker and listener effects on degree of perceived foreign accent." *Journal of the Accoustical Society of America* 91, 370-389.

[6] Flege, James E.; Munro, Murray J. and Ian R.A. MacKay. (1995) "Factors affecting strength of perceived foreign accent in a second language." *Journal of the Accoustical Society of America* 97, 3125-3134.

[7] Hansen, John H.L.; Gray, Sharmistha S. and Wooil Kim. (2010) "Automatic voice onset time detection for unvoiced stops (/p/,/t/,/k/) with application to accent classification." *Speech Communication* 52, 777-789.

[8] Higgins, Maureen B.; Netsell, Ronald and Laura Schulte. (1998) "Vowel-related differences in laryngeal articulatory and phonatory function." *Journal of Speech, Language and Hearing Research* 41, No. 4.

[9] Kessinger, Rachel H. and Sheila E. Blumstein. (1997) "Effects of speaking rate on voice-onset time in Thai, French, and English." *Journal of Phonetics* 25, 143-168.

[10] Klatt, D.H. (1975) "Voice Onset Time, Frication, and Aspiration in Word-Initial Consonant Clusters." *Journal of speech and hearing research* 18, 686-706.

[11] Ladefoged, Peter. (1971) *Preliminaries to Linguistic Phonetics*. Chicago: The University of Chicago Press. Midway Reprint 1981.

[12] Lisker, Leigh. (1970) "Supraglottal Air Pressure in the Production of English Stops." *Language and Speech*. Vol. 13, Part 4, 215-230.

[13] Lisker, Leigh and Arthur S. Abramson. (1964) "A Cross-Language Study of Voicing in Initial Stops: Acoustical Measurements." *Word* 20, No.3, 384-422.

[14] Lisker, Leigh and Arthur S. Abramson. (1967) "Some Effects of Context on Voice Onset Time in English Stops." *Language and Speech* 10, 1-28.

[15] Machač, Pavel and Radek Skarnitzl. (2009) *Principles of Phonetic Segmentation*. Praha: Epocha.

[16] Major, Roy C. (2008) "Transfer in second language phonology." in Edwards *et al.* (2008). *Phonology and Second Language Acquisition*. John Benjamins Publishing Company.

[17] Morris, Richard J. (2008) "Voice onset time differences between adult males and females: Isolated syllables." *Journal of Phonetics* 36, 308-317.

[18] Piske, Thorsten; MacKay, Ian R.A. and James E. Flege. (2001) "Factors affecting degree of foreign accent in an L2: a review." *Journal of Phonetics* 29, 191-215.

[19] Roach, Peter. (2000) *English Phonetics and Phonology.* Cambridge: Cambridge University Press. Third edition.

[20] Robb, Michael; Gilbert, Harvey and Jay Lerman. (2005) "Influence of Gender and Environmental Setting on Voice Onset Time." *Folia Phoniatrica et Logopaedica* 57, 125-133.

[21] Setter, Jane and Jennifer Jenkins. (2005) "State-of-the-Art Review Article: Pronunciation." *Language Teaching* 38, 1-17.

[22] Singleton, David. (2001) "Age and Second Language Acquisition." *Annual Review of Applied Linguistics* 21, 77-91.

[23] Skarnitzl, Radek. (In print) *Znělostní kontrast nejen v češtině.* Praha: Epocha. Chapter 5.

[24] Sweeting, Patricia M. and Ronald J. Baken. (1982) "Voice onset time in a normal-aged population." *Journal of Speech and Hearing Research* 25, 129-134.

[25] Theodore, Rachel M.; Miller, Joanne L. and David DeSteno. (2009) "Individual talker differences in voice-onset-time: Contextual influences." *Journal of the Acoustical Society of America* 125, No. 6, 3974-3982.

[26] Volín, Jan. (2007) *Statistické metody ve fonetickém výzkumu.* Praha: Epocha.

[27] Volín, Jan and Radek Skarnitzl. (2010). "Suprasegmental Acoustic Cues of Foreignness in Czech English." *The Speaker and Language Recognition Workshop* Brno, Czech Republic, 271-278.

[28] Wong, Matilda. (2008) "Can Consciousness-Raising and Imitation Improve Pronunciation?" *The International Journal of Learning* 15, No. 6, 43-47.

Shrnutí

Hlavním cílem této práce je zjistit, zda díky prvnímu semestru kurzu Fonetiky a fonologie angličtiny došlo ke statisticky významnému zlepšení hodnot nástupu hlasivkového tónu (voice onset time, VOT) v nahrávkách studentů Anglistiky a Amerikanistiky. Práce je rozdělena do dvou částí: první část představuje teoretické pozadí, zatímco v druhé části jsou analyzovány nahrávky 20 studentů z hlediska hodnot VOT a jejich změn.

Voice onset time je definován jako doba trvání mezi uvolněním závěru a nástupem znělosti při artikulaci explosiv (Lisker and Abramson 1964, 387). Na základě jejich definice je pak možno nahlížet na aspiraci jako na hodnoty VOT vyšší než 30 milisekund. U rodilých mluvčích by v angličtině měly být aspirovány explozivy $/p, t, k/$ v přízvučných pozicích, pokud nejsou předcházeny hláskou $/s/$. V nepřízvučných pozicích tyto explozivy aspirované být nemusí, i když měření ukázala, že i tyto hodnoty většinou bývají nad hranicí 30 ms. U znělých exploziv $/b, d, g/$ pak bývají hodnoty VOT záporné, protože hlasivky kmitají v průběhu celé artikulace.

Mnoho výzkumů bylo zaměřeno na zkoumání faktorů, které ovlivňují hodnoty VOT, a to jak lingvistické, tak faktory mimo jazykové. Jako hlavní se už u Liskera a Abramsona (1964) ukázalo místo artikulace. Čím zadnější místo artikulace, tím vyšší hodnoty VOT byly pozorovány, takže velární $/k/$ mívá vyšší hodnoty než alveolární $/t/$ a bilabiální $/p/$. Dalším často zkoumaným faktorem je rozdíl mezi izolovanými slovy a spojitým projevem. Výzkum naznačuje, že rozdíl mezi hodnotami v přízvučných a nepřízvučných slabikách se ve spojitém projevu poněkud stírá.

Dalším faktorem, který zkoumaly např. Kessinger and Blumstein (1997), je tempo řeči. Podle jejich výsledků se v přízvučných slabikách snižují hodnoty VOT při vyšším tempu. Autorky také zkoumaly vliv typu následující samohlásky na hodnoty VOT, ale nezjistily žádný významný efekt. Naproti tomu Higgins *et al.* (1998) objevil vyšší hodnoty pro vysoké samohlásky. Mezi mimo jazykové faktory, které byly zkoumány, patří především vliv pohlaví. Morris *et al.* (2008) nepřišli na žádnou závislost, zatímco Robb *et al.* (2005) zjistil vyšší hodnoty pro ženy.

Důležitým faktorem výslovnosti je také proces osvojování cizího jazyka. Setter a Jenkins (2005) zdůrazňují, že výslovnost bývá při výuce cizího jazyka do značné míry marginalizována a to i přes to, že chybná výslovnost může výrazně ztížit komunikaci i s mluvčím s jinak výbornými znalostmi daného jazyka. Další překázkou při osvojování výslovnosti jsou zakořeněné zvyky z rodného jazyka, kterých si mluvčí často nejsou vědomi.

Kvůli významu výslovnosti pro komunikaci byly často zkoumány faktory ovlivňující míru cizího přízvuku. Hlavním uváděným faktorem je věk studentů cizího jayzka. Podle jedné z teorií existuje takzvané "kritické období," po jehož uplynutí je nemožné dosáhnout úrovně rodilého mluvčího. Tyto hodnoty se mohou pro jednotlivé oblasti daného jazyka lišit a mezi autory nepanuje shoda o přesných hodnotách, nicméně výslovnost je často uváděna jako oblast s nejnižší hodnotou, kdy schopnost dosáhnout úrovně rodilého mluvčího se ztrácí již kolem 12 let.

Vzhledem k zaměření praktické části práce je zajímavý výzkum, který se zabývá vlivem cílené výuky fonetiky na výslovnost. Dosavadní empirické výsledky nejsou jednoznačné; zatímco někteří autoři vliv zamítají, nebo alespoň nepovažují za důležitý, jiní autoři přichází s výrazně povzbudivějšími výsledky. Je ovšem nutno poznamenat, že přístup jednotlivých studií není vždy srovnatelný. Studie, které se zaměřují specificky na fonetickou výuku, bývají optimističtější než autoři, kteří zkoumají vliv výuky angličtiny všeobecně.

Druhá, empirická část práce se zabývá analýzou nahrávek studenů Anglistiky a Amerikanistiky. Dvě nahrávky 20 studentů prvního ročníku, 10 žen a 10 muzů, byly porovnávány z hlediska aspirace exploziv v různých pozicích v rámci slova. První nahrávky byly pořízeny na začátku prvního semestru povinného dvousemestrálního kurzu Fonetiky a fonologie. Druhé nahrávky pak byly pořízeny na začátku druhého semestru. V obou případech byli studenti požádáni o přečtení krátkého textu. Každý student četl jeden z deseti textů v obou semestrech, je tedy možné přímé srovnání. V každém textu byly nalezeny všechny výskyty exploziv předcházející samohlásku v přízvučných i nepřízvučných slabikách a také v případech, kdy exploziva následovala po /*s*/. Jednotlivé nahrávky byly v Praatu popsány a byly označeny jak hranice hledaných slov a exploziv, tak hranice VOT. Za začátek VOT byl označen začátek exploze, konec byl určen tam, kde začíná první kompletní perioda následující po šumové části signálu (Skarnitzl, in print). Z téměř šesti tisíc výskytů jich 4988 mohlo být použito pro další analýzu.

Kromě grafických reprezentací hodnot VOT bylo použito i několik statistických testů k ověření významnosti pozorovaných rozdílů. Hlavními testy byly párový t-test a jeho neparametrická varianta Wilcoxonův test. Kromě nich byly použity testy na ověření předpokladů a dvouvýběrový t-test.

Analýza naměřených hodnot VOT ukázala, že studenti aspirují explozivy v přízvučných i nepřízvučných pozicích, ale neaspirují explozivy v přízvučných pozicích, pokud je předchází hlaska /*s*/. To odpovídá způsobu, jakým by explozivy měly být vyslovovány rodilým mluvčím. Nicméně porovnáme-li hodnoty s údaji pro rodilé mluvčí (Liskerem a Abramson 1967, Klatt 1975), hodnoty českých mluvčí jsou výrazně nižší a to pro přízvučné i nepřízvučné pozice. Naproti tomu jsou hodnoty pro pozice, kdy exploziva následuje po /*s*/, o něco vyšší než u rodilých mluvčí, i když stále pod hranicí aspirace 30 milisekund. Ze získaných dat je patrné, že došlo ke zvýšení hodnot VOT mezi nahrávkami z prvního a druhého semestru ve všech třech pozicích, přízvučných, nepřízvučných i následujících po /*s*/. Statistiskými testy bylo

zkoumáno, jestli je toto zvýšení pro jednotlivé pozice signifikantní či nikoli. Z výsledků lze vyvodit, že zatímco pro přízvučné a nepřízvučné pozice toto zvýšení signifikantní bylo (p-hodnoty odpovídající příslušnému t-testu byly < 0.001), pro explozivy následující /s/ byly výsledky na hranici signifikance. Výsledky tak naznačují, že pro studenty může být jednodušší nejprve zvýšit VOT hodnoty pro všechny pozice najednou, a teprve později začít případně mezi jednotlivými pozicemi více rozlišovat. Opakováním testů pro jednotlivé mluvčí bylo dále zjištěno, že u devíti z dvaceti mluvčích došlo k významnému zvýšení hodnot VOT v přízvučných pozicích.

V závislosti na faktorech zmiňovaných v teoretické části byly také zkoumány rozdíly mezi hodnotami VOT samostatně pro muže a ženy. Pro obě skupiny došlo podle výsledků testů k významné změně hodnot VOT mezi prvními a druhými nahrávkami; na základě grafů můžeme dále říci, že došlo k jejich zvýšení. Další zajímavou otázkou bylo, zda existují rozdíly mezi hodnotami VOT pro muže a ženy. V tomto případě byly porovnávány hodnoty VOT v přízvučných i nepřízvučných pozicích dohromady. V nahrávkách z prvního semestru byl zjištěn významný rozdíl, přičemž hodnoty pro muže byly výrazně vyšší než pro ženy. V nahrávkách z druhého semestru byl ale tento rozdíl už nevýznamný. Z těchto výsledků lze usoudit, že došlo k výraznějšímu zlepšení u žen než u mužů.

Posledním zkoumaným faktorem byl vliv místa artikulace na hodnoty VOT. Podle závěrů předchozích studií shrnutých v teoretické části se dalo očekávat, že explozivy se zadnějším místem artikulace budou vykazovat vyšší hodnoty VOT. Tento závěr byl potvrzen v nahrávkách z druhého semestru, zatímco v prvním semestru se ukázal jako signifikantní jen rozdíl mezi /p/ na jedné a /t, k/ na druhé straně. Z toho vyplývá, že i když testy prokázaly významné zvýšení hodnot VOT pro všechny tři explozivy zvlášť, toto zlepšení bylo výraznější pro /k/ než pro /t/.

Závěrem lze říci, že výsledky analýzy potvrzují z velké části předpoklady z teoretické části. I když předpoklady jednotlivých testů nebyly vždy splněny,

hodnoty testových statistik byly ve většině případů natolik jednoznačné, že můžeme říci, že cílená výuka fonetiky má významný vliv na hodnoty VOT v nahrávkách nerodilých mluvčí angličtiny. Nesmíme ale zapomínat, že toto zlepšení nemusí být jen důsledkem tohoto kurzu, ale může plynout i z faktu, že studenti byli v období mezi nahrávkami v kontaktu s rodilými mluvčími. Proto by se další analýzy měly zaměřit na oddělení jednotlivých možných vlivů a jejich vlastní přínos pro zlepšení výslovnosti, stejně jako na analýzu dalších aspektů výslovnosti pro komplexnější představu o vlivu cílené výuky fonetiky.

Appendix A

Table of speakers and texts read

Gender	Speaker	Text	Number
Female	MERA09	AMA	1
	BOZA09	DSA	1
	POKA09	EMA	1
	BAVA09	JLA	1
	FISA09	MSA	3
	HEJA09		
	LANA09		
	BRAA09	RWA	2
	KREA08		
	NECA09	ZAA	1
Male	BUCP08	AMA	1
	POND08	DSA	1
	KORD09	ED	5
	KRIJ09		
	MACK09		
	SVET09		
	ZAHL09		
	OLTN08	MMA	1
	MARK09	PJ	2
	MORM09		
Total	20	10	20

Table A.1: Breakdown of the speakers and texts read for any of the two sets of recordings