Univerzita Karlova v Praze

Filozofická fakulta

Ústav anglického jazyka a didaktiky

Bakalářská práce

Lenka Tranová

Znělostní kontrast ve vietnamské angličtině

The Voicing Contrast in Vietnamese English

Praha, 2015

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

Prohlašuji, že jsem tuto bakalářskou práci vypracovala samostatně a výhradně s použitím citovaných pramenů, literatury a dalších odborných zdrojů.
Souhlasím se zapůjčením bakalářské práce ke studijním účelům.

V ….....dne.....… podpis.....

I declare that the following BA thesis is my own work for which I used only the sources and literature listed.
I have no objections to the BA thesis being borrowed and used for study purposes.

Firstly I would like to express many thanks to my supervisor doc. Radek Skarnitzl for his invaluable advice and great support with my thesis. My gratitude also goes to all the Vietnamese speakers who were so kind and willing to be recorded for my academic purposes. Last but not least, I am immensely grateful to my friend Alžběta Ambrožová for her friendship and support throughout.

Abstract

This thesis deals with the voicing contrast in Vietnamese-accented English. The theoretical part introduces the generally accepted phenomenon of voicing contrast and several theories aimed at generalization of the main tendencies in second language acquisition. The final part of the theoretical background addresses initial consonants in Vietnamese and Vietnamese English. The methodological section provides information about the informants, the recording, and data processing prior to the analysis itself.

I also present graphs and tables illustrating the statistical calculations – using ANOVA and Tukey's post-hoc tests – that identify the relations among the measured units. The results of this analysis show that in its initial stressed plosives, Vietnamese-accented English maintains a voicing contrast similar to that of a native English accent. The average Voice Onset Time values of lenis stops without prevoicing are slightly higher than those produced by American English (AmE) speakers, while the average values of voiced initial stops prove to be fairly close to their AmE equivalents. This affinity is attributed to the fact that prevoicing in Vietnamese exhibits strikingly similar values to AmE. The values for fortis initial plosives are shown to be higher in VE than AmE, due to the fact that in Vietnamese, the onset of voicing usually comes later than in AmE.

The thesis analyses how Vietnamese speakers of English deal with voicing contrast, as their mother tongue employs richer means of voice contrasting than English requires. The results we present show that Vietnamese speakers of English do not struggle to maintain the voicing contrast employed in native English production.

Keywords: *Voice Onset Time (VOT); Vietnamese, Vietnamese English, Voicing Contrast, Second Language Acquisition*

Abstrakt

Tato práce se zabývá znělostním kontrastem ve vietnamské angličtině. Teoretická část nabízí přehled týkající se znělostního kontrastu obecně, tak, jak se s ním setkáváme v angličtině. Následující kapitola představuje několik vybraných teorií, jejichž cílem je zobecnit hlavní tendence spojené s osvojováním cizího jazyka. Závěr teoretického úvodu se věnuje Vietnamštině a podstatě této práce – Vietnamské angličtině s ohledem na počáteční konsonanty. Samotné analýze předchází popis metody, který poskytuje informace o nahraných Vietnamských mluvčích angličtiny, o postupu nahrávání a zpracování dat.

Tabulky a grafy slouží k ilustraci statistických výpočtů provedených za použití ANOVA a Post-hoc testů, které rozpoznávají celkové a dílčí srovnání konkrétních vztahů. Výsledky analýzy ukazují, že angličtina s vietnamským přízvukem zachovává pro své počáteční přízvučné plozivy znělostní kontrast srovnatelný s rodilou angličtinou. Průměrné hodnoty doby nástupu znělosti pro znělé plozivy bez znělosti v okluzi byly naměřeny mírně vyšší, zatímco hodnoty pro znělé plozivy se znělostí v okluzi vykazují hodnoty téměř identické v porovnání s rodilými mluvčími americké angličtiny. Tuto shodu připisujeme průměrným hodnotám znělosti v artikulačním závěru, které má vietnamština taktéž velmi obdobné. Hodnoty pro počáteční neznělé plozivy jsou vyšší, což se zdá byt opět důsledkem vietnamštiny, kde je doba nástupu znělosti delší.

Hlavním cílem práce je analyzovat schopnost vietnamských mluvčích angličtiny vytvářet znělostní kontrast v závislosti na své mateřštině, která jej využívá ještě více než angličtina. Dosažené výsledky potvrzují výchozí předpoklad, že Vietnamci mluvící anglicky netrpí nedostatkem v užívání kontrastu znělosti, který je využíván rodilými mluvčími angličtiny.

Klíčová slova: *Nástup hlasivkového tónu (VOT); Vietnamština, Vietnamská angličtina, Znělostní kontrast, Osvojování druhého jazyka.*

Table of contents

List of Abbreviations

AmE American English

RP Received Pronunciation

VE Vietnamese English

L1 First Language

L2 Second Language

F_2 Second Formant

F_0 Fundamental Frequency

VOT Voice Onset Time

ACT After Closure Time

SA Superimposed Aspiration

SLA Second Language Speech Acquisition

MDH Markedness hypothesis

CEFR Common European Framework of Reference for Languages

1. Introduction

This thesis describes the voicing contrast in Vietnamese English and compares the acquired data with a native accent of English. A sizeable number of American researchers point to the issues surrounding Vietnamese learners of English, whose pronunciation appears to be one of the least comprehensible of all English Second language (L2) learners. This research is partly motivated by the following excerpt from a book by Zampini and Edwards in 2008:

> *Occasionally, large ESL programs offer courses tailored to a particular L1 group. In the early 1980s, for example, when substantial numbers of Vietnamese, Laotian and Cambodian refugees came as immigrants to North America, shared L1 pronunciation courses were established in many cities because the nature of these newcomers' pronunciation difficulties appeared to be both quantitatively and qualitatively distinct from those of other ESL students.* (Zampini, Edwards, 2008: 356)

My work may prove relevant and beneficial to L2 pedagogy, as recent studies are very much interested in the extent to which phonetic contrasts are present in non-native accents. Drawing inspiration from such studies, my thesis argues that First language (L1) strongly affects successful learning of such contrasts and explains variation found in production.

This study aims to explore the area of voicing contrast because Vietnamese speakers of English are generally considered to be unable to recognise the contrast. Furthermore, as proved in Cunningham's study, they tend to employ contrasting incorrectly (Cunningham, 2009). English spoken by native Vietnamese is generally considered to have low comprehensibility. One of the reasons is the elision of word-final consonant clusters that are in English grammatically and semantically loaded. Consonant clusters are problematic for Vietnamese to pronounce, especially those that occur word-finally. This study, however, focuses on the word-initial position, namely the initial plosives, and wanted to know if that is due to the lack of voicing contrast in the stressed initial plosives.

2. Theoretical Background

This section provides a theoretical basis for the main topic of this study, the voicing contrast in Vietnamese English. Since the native AmE accent will serve as the referential accent for this study, we first introduce English consonants, and the related phenomena of voicing and aspiration, which are quantified using VOT (Voice Onset Time). Following this, a section dealing with second language acquisition (L2) introduces several theories aimed at generalizing recurring phenomena in the acquisition of L2. Finally, there is a section looking into Vietnamese English with regard to syllable-initial consonants that are the main focus of this study.

The main topic of this work is voicing contrast in Vietnamese English, since it is the crucial phonetic feature aiding overall comprehension. Thus, we will primarily be looking at the delayed Voice Onset Time and prevoicing in Vietnamese English plosives in stressed positions. This is because fortis plosives are precisely where L1 English accents have a delayed VOT and lenis plosives prevoicing.

First of all, it should be noted that studies dealing with voicing contrast of obstruents are commonly based on the observation of stops. Furthermore, phonological systems in a variety of languages often utilize the measurement of aspiration as a means of differentiating and categorizing obstruents. (Skarnitzl, 2011: 64) The dimension of Voice Onset Time has been used as a measurable counterpart to two other features used to draw distinctions between consonants, which are more controversial and difficult to quantify. The first feature is the articulatory force, which distinguishes between fortis and lenis consonants and takes into account the state of the glottis and the amount of respiratory and muscular tension (Skarnitzl, 2011: 72). The second feature is the voiced–voiceless distinction describing vibration of the vocal cords. Lisker and Abramson (1964: 384 - 9) explains that “the preference for measuring aspiration stems from the fact that proportion of articulatory force continues to attain a dubious status” and the same may be said about voicing. This paper shall therefore employ the dimension of VOT to describe voicing contrast in Vietnamese English plosives as compared to native AmE plosives. Since English will be used as the referential language, the following section briefly comments on the English phonemic inventory.

2.1. English Consonants

English consonants are traditionally described with regard to their place and manner of articulation. The place of articulation depends on whether the contact occurs between the upper and the lower lip, or between the tongue (the active articulator) and the roof of the mouth (the passive articulator). Sometimes the constriction occurs in the glottis itself. In all English consonants except the bilabials and glottal ones, the occlusion is made with the tip, blade, or back of the tongue pressing against one of the areas on the roof of the mouth, namely the dental, alveolar, post-alveolar, palatal or velar area.

The English phonemic inventory contains twenty four consonants. With regard to the manner of articulation, the stop consonants include the following six: /p, b/, /t, d/, /k, g/; the glottal stop /ʔ/ can be included here as well. English also has nine fricatives /f, v/, /θ, ð/, /s, z/, /ʃ, ʒ/ and /h/, two affricates /tʃ, dʒ /, three nasal stops /m/, /n/, /ŋ/ and four approximants /w, l, r, j/.

The manner of articulation does not concern the parts of the mouth employed in the production of consonants, but rather the way the speaker produces them. In other words, the manner of articulation focuses on the acoustic qualities of the closure between the two articulators. To give an example, when we produce the approximants /l, r, w/, the articulators do not create turbulent airflow; they are closer to vowels, because of their relatively high sonority. The articulators are wide apart – not so wide as to create a vowel, and at the same time not enough to create friction. In the case of the fricatives, /s, z/, the articulators are very close to each other but not close enough to complete the closure; and the result is thus a high-pitched hissing sound. When the closure is complete as it is with stop consonants, e.g. /t, d/, the air cannot escape from the vocal tract. The name of stops is derived from this complete stopping of the air stream.

2.1.1. Plosives

As stop consonants make a so-called ‘plosion’ sound, they are also often labelled as plosives. The variety of names stems from the different points of view phoneticians assume when describing them. From the articulatory point of view, stop consonants, as has already been noted, completely impede the airflow in the oral cavity and they are therefore referred to as stops. The term ‘plosives’ focuses on their auditory quality. Plosives are produced in four phases: closing, compression, release, and post-release. The *closing* phase is when the articulators move to create a stricture behind which compressed air is accumulated, making the *compression* phase. The opening of the articulators is called *release* and this moment is accompanied by a perceptible plosion that gives the consonant one of its names. Right after

the release phase comes the *post-release* phase. The post-release phase is significant for our research, because a certain timing of articulatory gestures and voicing can cause the delay of voicing. The delay of voicing will be discussed to a great extent later.

Let us now briefly introduce voicing. English stops are found in pairs contrasting voiced and voiceless phonemes, also called lenis and fortis, such as the pairs/p, b/ /t, d/, /k, g/ where the left phoneme is always voiceless and often involves the delay of voicing.

2.2. Distribution

In relation to syllable positions, there is very little restriction in English; all six plosives occur initially, medially, as well as finally.

2.3. Voicing (Phonological)

Let us now briefly explain the occurrence of voicing as a phonemic distinction, i.e. voiced /b, d, g/ vs. voiceless /p, t, k/. First of all, voicing concerns a classification of speech sounds. In general, the voiceless set /p, t, k/ in English is produced with no vocal cord vibration, while /b, d, g/ are called voiced because in certain environment (for instance, a sequence of /d/ sounds produced in a row) the vocal cords vibrate when we pronounce them. However, looking at voicing as a phonetic concept, it is clear that voiced stops do not necessarily need to be voiced on the articulatory level.

2.4. Voicing (Phonetic)

Voicing can also refer to the articulatory process, for instance when a speaker pronounces /d/, in which the vocal folds vibrate. In general, though, we find very little voicing in the closing phase of the articulation of /b, d, g/; the voicing only starts before the release. The consonants /b, d, g/ may be fully voiced throughout the closure and compression phases if the speaker pronounces them slowly and carefully. In contrast, when produced in rapid speech, these sounds may have very little or no voicing at all. As Volín (2002: 73) points out, most of the English voiced stop consonants keep their voicing only when surrounded by vowels. Considering their distribution, voiced consonants appear to be voiceless or devoiced in most cases.

2.4.1. Devoicing

The loss of voicing in English lenis obstruents is typical for English and is recognized as devoicing. Devoicing generally appears with the plosives /b, d, g/ that are partially or fully devoiced in word-initial or word-final positions. English voiced consonants maintain their

voicing only in intervocalic position inside morphemes. In all other positions their voicing is partly or wholly lost.

In the words *pie* and *buy*, for instance, the bilabial plosives appear to be voiceless. Ladefoged (2010: 57) points out that most people perform very little voicing when the lips are pressed together in producing these words. Devoicing can also be exemplified in initial and final voiced plosives in the words *budget* [ˈb̥ʌdʒɪt] and *load* [ləʊd̥]. However, although both /b/ and /d/ lose their voicing, they still maintain their lenis character – for example, they never exhibit aspiration. Figure 2.1 below visualises a sound wave with no periodical pulsing[1] preceding the release (marked in pink) of the voiced obstruent /b/. Comparison with Fig. 2.2 shows the necessary contrast between voiced and devoiced plosives, because during the closure voiced stops contain the periodical pulsing that devoiced stops are missing. This pulsing takes place as a part of a process called prevoicing.

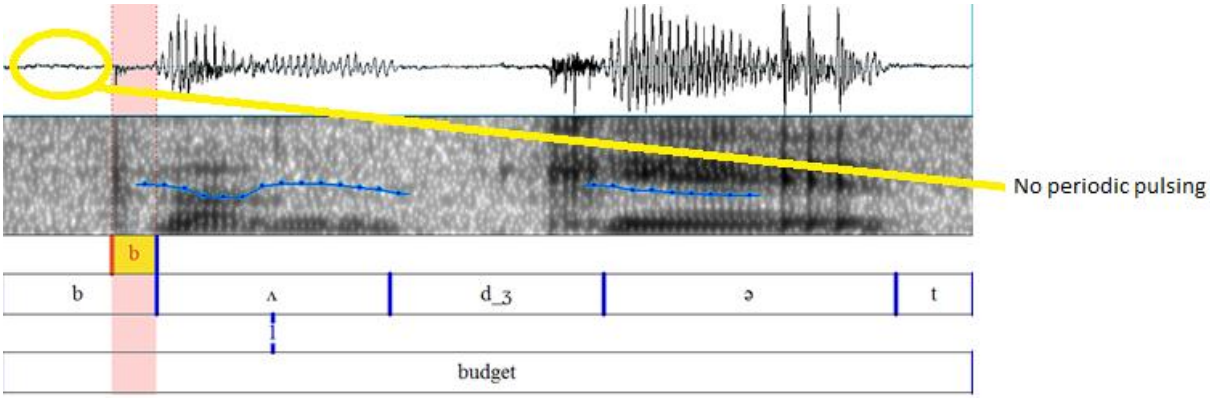

Fig. 2.1. – Devoicing[2]

2.4.2. Prevoicing

Another phenomenon, which occurs before the release of lenis plosives, is called *prevoicing*. In some languages, including English, initial voiced plosives are produced with prevoicing. As far as the four phases of stops are concerned, prevoicing occurs when voicing develops during the closure of the word-initial plosive before the release.

[1] Periodical pulsing is an instance of several complete sondvawes (cycles) following each other.
[2] Screenshot, PRAAT - from corpus of Vietnamese English, recorded by Lenka Tranová

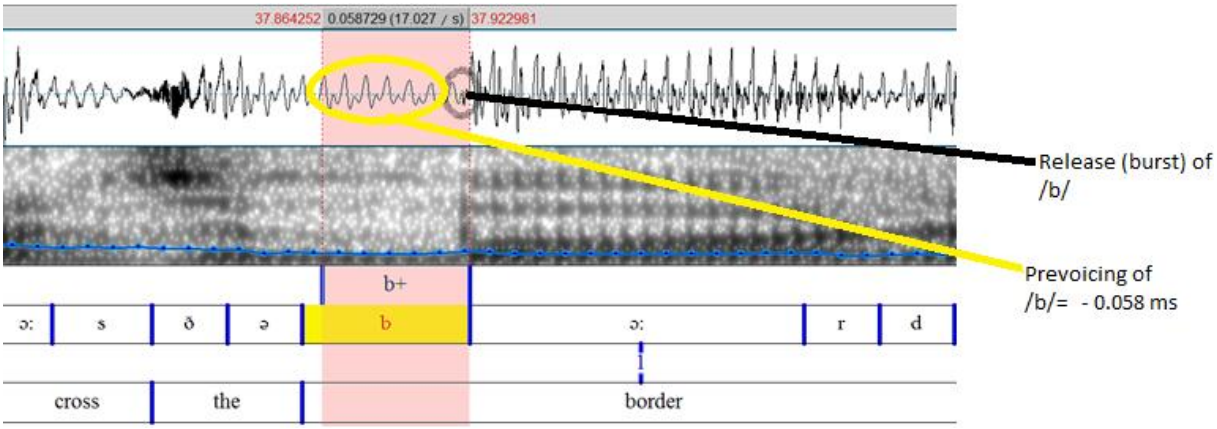

Fig. 2.2. – Prevoicing[3]

In that case, as the above Figure 2.2 displays, vocal cords vibrate during the closure of the initial plosive /b/. Prevoicing is characterized by negative Voice Onset Time values, which we measure from the moment of release to the left on the soundwave (Fig. 2.2). However, because the voiced plosives are often devoiced, the acoustic quality often recognised as the feature distinguishing between the voiceless /p, t, k/ from their voiced counterparts /b, g, k/ is labelled aspiration.

2.4.3. Aspiration

We will continue discussing obstruents from an acoustic point of view**.** As has been explained, obstruents are produced in four phases. The third phase, i.e. the release of voiceless stops, results in an audible plosion that can be characterized as a burst of noise. This is followed by a post-release phase, where the articulators are apart. However, the English voiceless stops /p, t, k/ do not immediately manifest an onset of voicing. In the moment after release, air typically escapes through the vocal cords, making a sound phonetically pronounced like [h] (Roach, 2009: 27) and this sound is called aspiration.

This audible h-like noise is demonstrated on a sound wave and a spectrogram in Figure 2.3 below. The acoustic feature of aspiration should thus be regarded as a feature automatically accompanying a large delay in voicing (Lisker, Abramson, 1964: 387). From an acoustic perspective, aspiration is characterized by turbulent glottal noise with delayed onset of vocalic tone (Ladefoged, 2010: 144). Looking at the pink section on the sound wave, we can notice that aspiration is displayed as a cluster of irregular pulsing. In this case the aspiration of /t/ takes 59 ms, which is – according to the general standard - enough to consider the delay of voicing as aspiration (Ladefoged, 2010: 151).

[3] Screenshot, PRAAT - from corpus of Vietnamese English, recorded by Lenka Tranová

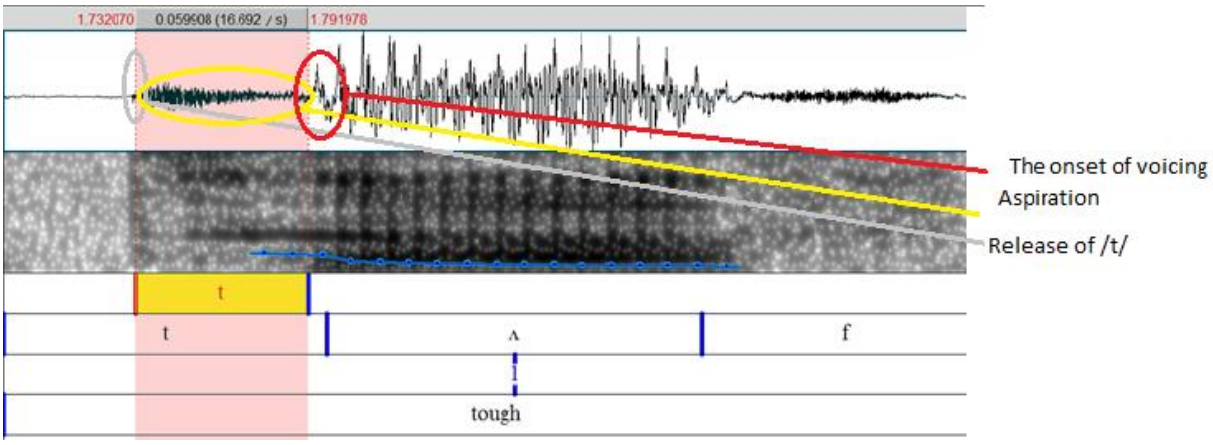

Fig. 2.3. – Aspiration[4]

In an initial consonant-vowel position, as in the word *tough* [t^{h}ʌf], aspiration occurs as a prominent acoustic feature that accompanies the release of /p, t, k/ and is marked in phonetics by the symbol [ʰ]. This irregular pulsing ceases with vocal cords coming together and generating the onset of voicing of the following vowel /ʌ/, which in contrast to aspiration displays periodical pulsing. From an articulatory perspective - right after the release of the stop - the vocal cords maintain an open position while the aspirated phonemes are pronounced. Alternatively, they are in a phonation posture when producing the unaspirated phonemes. Roach (2009: 87) explains that for the production of an aspirated sound, there needs to be an open glottis and a vocal tract without impediment.

To summarize the contrast between prevoicing and aspiration, the release of the voiced set of plosives is followed by a weak plosion (Roach, 2009: 22). This plosion happens simultaneously with the onset of voicing or even after it, which is why these stops are often pre-voiced. If we compare the Figures 2.2 and 2.3 we can clearly see that aspiration, in contrast to prevoicing, starts after the release of a stop and has a character of noise, while prevoicing starts often before the release and is periodical.

2.4.4. The Correlation between Voicing and Aspiration

In order to draw a complete picture of how the correlation between voicing and aspiration functions, we need to consider that voicing and aspiration depend on the timing of articulatory gestures–the vocal fold gesture and the oral articulatory gesture. The relation between the two is that the appearance of one in a spectrogram is conditioned by the absence of the other. Voicing takes the shape of periodic pulses at the frequency of the voice pitch while aspiration noise is in the frequency range of higher formants. It is worth noting that when a spectrogram shows the presence of voicing the aspiration noise is absent or obscured. The same thing

[4] Screenshot, PRAAT - from corpus of Vietnamese English, recorded by Lenka Tranová

happens when the aspiration noise is strongly present. In this case the periodic pulsing is usually imperceptible. The phonetic dimension of voicing – the primary glottal mechanism that controls it – is also responsible for producing parts of the features which are together taken to be the acoustic phenomena of aspiration and articulatory force (Lisker, Abramson 1964: 386-8).

2.5. VOT

Voice Onset Time can be called a unit of both aspiration and prevoicing as it marks the duration of time when they take place. Aspiration is measured by phoneticians with reference to time (the period of voicelessness) designated as Voice Onset Time (VOT). In the sixties, VOT was introduced as the first phonetic indicator able to distinguish voiced, voiceless unaspirated and voiceless aspirated plosives; (Skarnitzl, 2011: 66). VOT, as an indicator of the length of the period of voicelessness, is an acoustic manifestation of aspiration and prevoicing. In order to draw a clear distinction between aspiration and VOT we could say that aspiration is a phenomenon that is recognised by phoneticians based on the auditory perception, while VOT is mathematical quantity that they use to describe and measure aspiration. This study makes use of VOT to describe and asses the recorded material and VOT shall therefore be introduced and elaborated on more in depth in the following sections.

Voicing and aspiration both depend on the timing of articulatory gestures: the gesture of the vocal folds and the gesture of the oral articulator. Voice Onset Time is the time between the release of the fortis plosive and the beginning of voicing for the following sound. If we look at the release of a fortis plosive on a sound wave, to the right we can measure how many *ms* it takes until the first periodic pulsing of a following vowel appears. However, as we already demonstrated, prevoicing plosives have negative VOT values. Thus, if we look on a sound wave leftwards from the release of a lenis stop we can measure the time before the burst of the stop, which in case of voiced plosives with prevoicing brings negative VOT values (-*ms*).
Long and positive VOT, as has been mentioned, is characteristic for aspiration. However, consonants can also have short VOT or zero VOT as it is going to be explained in the following paragraphs.

Positive VOT may be also referred to as voicing lag, which is measurable in stressed fortis plosives occupying the onset of a syllable or medially before a stressed syllable: for instance, [tʰ] in *towel* [ˈtʰaʊəl] or [kʰ] in *become* [bɪkʰʌm]; voicing lag in English never appears in a coda. It also does not occur in voiceless stops when preceded by /s/, as in *skip* [skɪp]. In order

to obtain strong aspiration with longer positive VOT, /p, t, k/ need to occupy the onset of a stressed syllable. One example of a fortis aspirated plosive accompanied by weak aspiration (short VOT) is the word collide [kʰəˈlaɪd]. The short VOT is a result of the placement of the fortis stop within an unstressed syllable, as in the word competitor [kəmˈpʰetitər].

The case of zero VOT, also referred to as the instant of release, does not indicate that there is no aspiration; what happens is that there are two gestures – the release and the voicing – that occur at the same time (see 2.4.4). Such obstruents are voiceless but unaspirated. The voicing of the following vowel begins at or near to when the stop is released. The release is fully or nearly inaudible.

This contrast drawn between aspirated and unaspirated may give the impression that there are just two possible values. That is, however, not true. Aspirated and unaspirated are parts of a continuum of time measured in *ms*. One should therefore not think about the distinction strictly categorically, even though the intention is to categorise the two sets of stops.

2.5.1. Measuring VOT

To be able to compare aspiration in different languages, we need to look into how VOT is measured in *ms*. When the gauged figure is less than 20 *ms*, it is by phoneticians considered to be an unaspirated stop (Ladefoged, 2010: 151). The terms *voicing lag* and *voicing lead* have already been introduced, it remains to indicate their acoustic span/duration. Long-lag plosives are accompanied by long VOT duration ranging between 0 and 35 ms, while their counterparts, prevoiced stops, manifest voicing throughout the closure, see Fig. 2.2 and therefore display negative VOT on an oscillogram or spectrogram (Zampini, Edwards, 2008: 221).

Oscillograms, with their unambiguous indication of time, are considered to be more suitable for measurement of VOT than spectrograms; however, some phoneticians use spectrograms. There are researchers who, for instance, consider the start of aspiration to be the onset of vertical strokes starting from F_2 (2^{nd} formant) and higher, while others prefer to look at F_0. Although there are systematic differences between the two types of measurement, which range approximately between 5–10 ms, these are regarded as negligible and the two methods are considered to give highly correlated results. An example of a less common method is the measurement of duration of high frequency/pitch noise.

Although there are various approaches in measuring VOT, it is most often measured from the start of the first burst. This is useful to clarify, as voiceless obstruents such as /k/ are often accompanied by multiple bursts. In our measurements we are going to follow this most

common method and use an oscillogram to collect the data. As has been mentioned, aspiration ends with the onset of the first periodic pulsing of the following vowel. The same applies for VOT. The possibility to use spectrogram for our measurement would bring us undesirable complications since the difference between the onset of voicing (periodic oscillation of the vocal cords) on the one hand and the onset of F2 on the other can be up to 46 ms (Skarnitzl, 2011: 83-86).

2.5.2. Effects on VOT

It is important to note that syllable stress causes certain exaggeration of aspiration and the values of VOT are therefore higher. One of the reasons can be that the articulation of voiceless stops includes a greater glottal opening in stressed syllables than in unstressed ones (Ní Chasaide, 1987: 28-31).There are, however, other factors that have an impact on VOT and do not relate to a syllable's stressed or unstressed position within the word. Some relevant studies looking into these influencing factors are described below.

Miller, Green & Reeves (1986); Volaitis & Miller (1992) and Kessinger & Blumstein (1997) deal with the impact of speaking rate on the eventual value of VOT with /p, t, k/ consonants. The results state that the value of VOT decreases with faster speech. This, however, has not been conclusively shown in second language learners (Zampini, Edwards, 2008: 225).

Furthermore, place of articulation was also shown to have an impact on VOT, namely velar and uvular plosives were shown to have the highest values (Lisker, Abramson, 1964: 387). Among the factors possibly impacting the VOT are the general principles of aerodynamics. When producing a velar stop, the oral cavity has a lower volume of air compressed behind the closure. The lower the volume, the higher the pressure arising during the occlusion. The eventual release of this compressed air lasts longer than during the articulation of bilabial or alveolar plosives. From the opposite perspective, after a velar release there is a greater amount of air waiting behind the occlusion to be moved forward; this also lasts longer (Lisker, Abramson, 1964: 400).

In his study, Stevens (1998: 328) ascribes the importance to the place of contact of the passive and the active articulator during the closure, where the release of the closure between the two may be slowed down by the interaction between intraoral pressure and the mechanic pliability of the surface. Other influencing factors are mentioned by Cho and Ladefoged (1999: 209-212). They talk about the agility of the tongue, where the tip and blade are generally assumed to be faster and more flexible than the back of the tongue. Therefore, when the occlusion is created by the back of the tongue, we can assume the whole process to be slower and last

longer, thus increasing the VOT values. They also mention a slower closing of glottis when a velar sound is produced. We could again ascribe this to the intraoral pressure decreasing more slowly for velars than for plosives articulated at the front of oral cavity.

In brief, Cho and Ladefoged (1999), taking into account eighteen different languages, conclude that it is not possible to make a generally valid statement asserting that consonants articulated further back in the oral cavity have higher values of VOT. For this study it is essential to mention their conclusion that, although the VOT values stem from the aforementioned physiological and aerodynamic processes, different languages still have language-specific values of VOT.

2.5.3. VOT Cross linguistically

Before introducing the principles of voicing contrast in Vietnamese English, we should mention that different languages make different distinctions along the VOT continuum. English employs a two-way phonemic distinction /p, t, k/ vs. /b, d, g/. These, as we already know, differ in their phonetic realization as long-lag stops and short-lag stops. They are differentiated by their VOT. In some languages the situation may be different and even the voiced set of plosives may be characterized by a long VOT. This can be illustrated by the fact that in Hindi and Sanskrit, breathy-voiced consonants are often called voiced aspirated. Given such variation across different languages, it is worth looking at the various distributions of VOT.

Recent studies seem to be more and more interested in VOT because of its ability to distinguish between a native and non-native accent (Zampini, Edwards, 2008: 221). Collecting VOT data is beneficial for calculating a scale of VOT for different languages. Based on this data, programs have been developed that are able to classify foreign accents in spoken English according to the values of VOT. Such mechanism aiming at automatic accent detection was developed to certain extent by Hansen et.al in 2010. His experiment proved that we can successfully make an automatically working system that classifies accents using VOT in unvoiced stop.

2.5.4. Drawbacks to VOT

It should be taken into account that VOT is not suitable for formalizing the temporal relations involved in voiced aspirated phones. Further, differentiating voiced unaspirated and voiced aspirated plosives proves to be problematic when classifying plosives by VOT. Another issue

is that it is not possible to talk about Voice Onset Time for intervocalic plosives (Skarnitzl, 2011: 88).

Among the alternative approaches to plosives differentiation we should mention a study by Mikuteit and Reetz (2007). Mikuteit and Reetz address problems with VOT measurements and introduce an alternative method using lag times called After Closure Time (ACT) that they utilizes in describing East Bengali. East Bengali displays a four-way contrast of voiced/voiceless and aspirated/unaspirated oral stops and affricates in all word positions. The authors assert that VOT cannot sufficiently distinguish between the four categories. In the four-category languages VOT can only distinguish voiceless aspirated stops from voiceless unaspirated ones – according to the study, the negative VOT among voiced aspirated and unaspirated stops does not show such an evident distinction. Their research suggests that this new way of measuring can put an end to the additional concept of 'breathy voice', which we already know is typical for voiced aspirated stops. Therefore, they introduce an additional term called Superimposed Aspiration, abbreviated SA, which indicates the breathy part of the vowel that follows voiced aspirated stops, which is also known as breathy voice. The results that combine SA and ACT show that aspiration measured from the point of release is timed equally for voiced and voiceless obstruents. ACT in this study proves capable of distinguishing voiced aspirated and voiced unaspirated stops, and there is thus no need to search for the presence of breathy voice. Aspiration contrast in both voiced and voiceless consonants could be distinguished by ACT. Voiceless consonants have a long ACT and a short SA, while the reverse is true for voiced stops.

The usage of this method would certainly be interesting to apply to further research concerning VE accent. Since the conducted measurement did not need to differentiate voiced aspirated plosives the dimension of VOT was sufficient for this work. Furthermore the data could be compared to a larger amount of studies; the SA and ATC measurement are yet to be employed more frequently.

2.6. VOT in Foreign Accents

So far we have discussed aspiration in a linguistically universal sense. However, the purpose of this study is to focus on the language-specific processes which take part in determining the criteria by which a language distinguishes its stop consonants. Notably, there is a continuum of possible voice onset times, and various languages favour different points along this continuum to create oppositions among their stop consonants (Henton, Ladefoged,

Maddieson, 1993: 144). Moreover, recent studies have shown that VOT varies not only among different languages, but also among dialects or foreign accents (Hansen, Gray, Kim, 2010: 777-789). Interestingly, the onset of voicing has also been studied as a criterion differentiating individual speakers (Skarnitzl, 2011: 82). Therefore, suprasegmental properties of L2 speech such as VOT seem to be a reliable and accessible source for collecting data contrasting not only plosives. Given that the role of Ll in the perception of L2 has been observed on a prosodic level, this study intends to compare the data between American English and Vietnamese speakers of English.

2.7. Second Language Speech Acquisition

Since our target group of speakers are second language learners, the next section is dedicated to the influence of one's native language on their foreign language learning. The influence of foreign language learning is more obvious in pronunciation than, for instance, grammatical or lexical structure (Leather, James, 1996: 285). Therefore we are going to briefly comment on the influence we expect second language acquisition (SLA) to have on VOT values.

First language plays an influential role in the production of the language acquired later. The extent to which L1 of beginning learners influences their perception and production of L2 syllables and the extent to which this is an actual transfer from L1 to L2 has been the topic of a large number of studies. The aim of this study is to define, what tendencies towards mistakes are common to Vietnamese speakers. These are presumed to occur in onsets and codas of longer and more complex words (Zampini, Edwards, 2008: 5). There are perceptual differences between adults and children, as children have not yet developed all phonological categories of L1. This research will deal particularly with adult speakers and theoretical background regarding children's perception and production will therefore be omitted. Among the influencing factors that constrain L2 production in an elementary speaker both linguistic and non-linguistic conditions are included. Among these are:

> *voicing agreement, preceding linguistic environment, following linguistic environment, stress, intonation, coda length, and grammatical category, as well as non-linguistic/social factors such as gender, proficiency level, task, use of L2 at home, work, and socially, age of L2 learning, motivation, and length of stay affect L2 variation* (Zampini, Edwards, 2008: 269)

2.7.1. Compromise VOT values

Most of the studies in Zampini and Edwards (2008) claim that if second language acquisition takes place before the age of six the speakers are likely to produce native-like values of VOT,

while those who learn L2 later tend to produce compromise durations, somewhere between L1 and L2. The group of Vietnamese students that we recorded were all later learners; and we may therefore expect them to produce compromise VOT values.

2.7.2. VOT in Second Language Acquisition

A significant body of research reveals that in cross-language communication, VOT serves as an acoustic impulse sufficient to distinguish initial stop consonants in a variety of languages. Listeners who were tested on the differentiation of stop consonants tend to separate a continuum of stimuli varying in VOT into categories matching the system of consonants in their L1. They also present results that show adults with very little L2 proficiency to be the most sensitive to VOT differences (Leather, James, 1996: 274). Therefore, beside compromised VOT values, we can also expect that the informants of our study will produce plosives with VOT identical or close to L1 plosives. Assuming that Leather and James' conclusion is correct, VOT as a marker of voicing contrast is an important tool that even elementary learners of English can recognise and use.

2.7.3. The Theory of Equivalence Classification

This theory suggests that every human has a basic cognitive mechanism that helps them perceive and classify sounds. On the basis of the experience with different sound categories that the speakers have encountered since they were born, they create slots of sounds in their mind. Therefore, if there is a sound similar to the one they stored as, for example, /t/, they will classify it in the /t/ slot, even though the sound may not be completely the same. This process is called equating L1 and L2 sounds (Zampini, Edwards, 2008: 223).

This is particularly important in that we can expect the group of Vietnamese speakers whose speech we are examining to classify English plosives around the aspirated voiceless stop /t^h/ or prevoiced /d/ that is all present in Vietnamese phonemic inventory.

2.7.4. Foreign Accents

The above described research in the domain of accents has investigated the acoustic properties in the speech of L2 learners by comparing their production of certain sounds with those produced by a monolingual native speaker.

The investigated acoustic properties of plosives showed that even experienced speakers of L2 tend to compromise VOT values of stop consonants. The sounds produced by L2 speakers reveal that the range of VOT was neither that of L1 nor of L2. Those who manage to establish separate phonetic categories for L1 and L2 tend to be early learners under the age of six (Zampini, Edwards, 2008: 223). Later learners most probably inhabit equivalence

classification in the production of L2 plosives, or they tend to compromise their L1 VOT values with those of L2 (Zampini, Edwards, 2008: 42).

Importantly, considering the fact that aspiration is audible even to beginner learners of L2, it should be regarded as an important feature of accents of any kind. Hansen, Gray and Kim (2010) elaborate on this in their research of automatic detection of foreign accent, using VOT values in unvoiced stops.

2.7.5. Markedness Theory/Hypothesis

Markedness hypothesis (MDH) argues that not all differences between L1 and L2 will cause the same difficulty for an elementary speaker of L2. The theory of typological markedness holds a significant position in the research explaining the facts about L2 phonology (Zampini, Edwards, 2008: 112). We will first provide the necessary background by introducing the markedness hypothesis and the status of marked and unmarked. Afterwards we will give several examples to illustrate the hypothesis.

The markedness hypothesis and the principle of markedness were first introduced and elaborated on by the Prague School of Linguistics. Their pioneering figures were Nikolai Trubetzkoy and Roman Jakobson. The concept states that binary oppositions between certain linguistic representations, for instance voiced and voiceless obstruents, is not understood by linguists to be simply polar opposites. One element of the opposition is always considered in some way special, since it has a wider distribution both within a given language and across languages. This special status is characterised as markedness: the element with broader distribution is labelled as unmarked. Elements labelled as unmarked are also implied to be simpler, more basic or more natural than the marked member of the opposition. According to this classification, voiceless obstruents are considered unmarked relatives to voiced obstruents.

MDH suggests that markedness can anticipate and explain some of the difficulty an L2 speaker may face when learning individual L2 structures (Zampini, Edwards, 2008: 96) as illustrated by the following example:

One can assume that speakers of Vietnamese English may have difficulties when trying to produce aspirated sounds since these are marked, according to the theory. The degree of difficulty involved is predicted to correspond directly to the relative degree of markedness (Zampini, Edwards, 2008: 98).

It is worth noting that the transfer of L1 to L2 production is not the only thing that has a great impact on the speech of an L2 learner. Factors such as the ability to perceive foreign sounds, the particular differences or nuances together with the length of daily exposure to L2 and

experience in using a foreign language are also significant elements that determine the eventual quality of L2 speech. Markedness theory is one of the theories that are there to help us foresee difficulties a learner of a foreign language can encounter.

We can assume that Vietnamese speakers of English have a great advantage, because there is an aspirated plosive /t^h/ and pre-voiced /d/ in the mother tongue, Vietnamese. Nevertheless, we have to test not only if they are able to produce aspirated sounds but also if their distribution of VOT is similar to that of a native speaker of AmE.

2.7.6. Vietnamese speakers of English as L2

A study by Zampini and Edwards focused on L2 acquisition by Vietnamese speakers and describes the most common pronunciation problems that they have with English. The subjects struggled with consonant clusters in initial and final positions. Vietnamese speakers are known to struggle with consonant clusters, because there are no consonant clusters in their mother tongue. Zampini and Edwards' experiment tested the production of syllable-initial clusters and syllable-final clusters. For our study it is relevant to mention that the syllable-initial clusters produced by Vietnamese were found to correspond with the markedness hypothesis (Zampini, Edwards, 2008: 100). In our study we would like to test the syllable-initial position more and see if these speakers struggle even with combinations such as plosive plus a vowel.

2.8. Vietnamese and Vietnamese English

The above sections dealt with voicing and aspiration in general, together with second language acquisition. The following section will look into Vietnamese and Vietnamese English with regard to initial consonants and, most importantly, initial plosives, which this study compares to AmE plosives in terms of VOT.

2.8.1. Syllable-Initial Consonants

Stated earlier, this thesis is concerned with voicing contrast in Vietnamese English as it is an essential element without which there is a great deal of confusion between stops. Improper use of voicing contrast leads to misunderstandings and communication lapses. Considering the Markedness theory and the theory of Equivalence classification, speakers of Vietnamese English should struggle less in maintaining voicing contrast while speaking English, since the Vietnamese language contains both aspirated and pre-voiced sounds. However, although Vietnamese speakers of English are supposed to know these features of the English language from their L1, they still struggle to employ these markers of contrasting. In other words, their distribution of these markers differs from the distribution of a native English speaker. One such instance can be illustrated by contrasting the words *green* vs. *cream*. VOT matters

significantly since one typical feature of Vietnamese-accented English is the elision of consonants, especially in the syllable coda. This can be problematic, because English suffixes are semantically very important. The following example shows an instance where even a one-consonant coda can be deleted in Vietnamese English. The word *green*, when produced by Vietnamese speakers, was perceived and misinterpreted by listeners as *cream*. The problem rested on them producing a VOT of 112 ms after the release of /g/, making it sound like a voiceless aspirated stop (Cunningham, 2009). The fact that Vietnamese speakers are likely to eliminate consonant clusters at the end of words, together with aspirating plosives that are not aspirated in native English speech, causes great problems in keeping essential contrasts between similar—and not so similar—words.

Cunningham's study proved that Vietnamese speakers of English struggle even in word-initial positions. Moreover, they employ non-native-like distribution of positive VOT (aspiration) with the lenis plosive /g/. All in all, theoretical predispositions of L2 acquisition presage Vietnamese speakers of English to have little troubles whereas Cunningham's results prove the opposite. We will therefore examine the VOT values and draw conclusions based on our own measurements.

2.8.1.1. Distribution of Consonants

In contrast to English, which has very few restrictions in consonants placed in the word-initial position, Vietnamese is very specific in their distribution. Though not as strict as with word-final position in Vietnamese, the limited distribution of consonants should be mentioned and taken into account. If we compare the distribution in initial and final positions, we see that word initially can appear labial /ɓ, m, w/, labial dental /t, tʰ, n, l/, alveolar /ɗ, s, z/, palatal /tɕ, ɲ/, velar /k, ŋ, x, ɣ/, and glottal /ʔ, h/. In Vietnamese, initial /p, j, r/ appears very rarely and is considered to be primarily French in origin, appearing in loanwords. It is important to consider that many speakers realize /p/ as /b/ (Kirby, 2011). Among consonant clusters only /ʈ͡ʂ/ and /kw/ can appear word initially. With the word-final position the rules are stricter – there are no consonant clusters at all and phonemes occurring word finally are restricted to the set of /m, n, p, t, k, ŋ/, and / i̯, u̯/ semi vowels, sometimes transcribed as / j, w / (Hoang, 1965: 45).

2.8.1.2. Plosives

With regard to plosives, Singer (2012) considers glottal reinforcement and voiced implosives to be the most prominent characteristics of Vietnamese-accented English. One can easily deduce from the basic theories of second language acquisition, such as the aforementioned

theory of equivalence classification, the effect of L1 on L2. Vietnamese itself often includes glottal effects; therefore, glottal effects can be frequently observed in Vietnamese-accented English. Vietnamese speakers often realise voiced /b/ and /d/ as voiced pre-glottalized implosives [ʔɓ] and [ʔɗ]. Even if the sounds are produced as plosives, they will be, according to Singer, pre-glottalized, as in the noun phrase a bad man, which would be pronounced as [ə ˈʔɓæd mæn].

Another contrast to English is that Vietnamese does not have the /g/ phoneme, although it seems that a large number of speakers have it as an allophone. In the Vietnamese phonemic inventory we can instead find a voiced velar fricative [ɣ] that is most likely to be used by Vietnamese speakers of English to fill the slots of the English /g/. In Vietnamese English, word-final /g/ is extremely unlikely, due to the devoicing that happens word-finally in Received Pronunciation and General American English (Wells, 1982: 213). Therefore, at the ends of words the unvoiced [x] is more probable in VE and the Vietnamese speaker is likely to pronounce the word big [bɪg̊] as [bɪx].

Coronal consonants that are pronounced with the tip or the blade of the tongue are also specific to the Vietnamese accent. Singer (2012) points out that some consonants such as [t̪ʰ, l̪, s̪, z̪] are pronounced dentally, others laminally, using the blade of the tongue with the tip down behind the bottom teeth – these include [s̻, z̻, ç̻]. In the area of Ho Chi Minh City, located in the very south of Vietnam, two retroflex consonants can be found that Singer (2012) characterises as produced with a tongue tip curling back [ʂ, ʈʂ]. Vietnamese consonants exhibit a lot of complexity, so a great amount of attention should be paid to consonants pronounced by English-speaking Vietnamese, especially if we consider the difference between distribution of phonemes and allophones. In Vietnamese we can notice that some plosives appear as allophones while in English they are phonemes and vice versa.

2.8.1.3. Plosives and Aspiration

The following example shows a sound that is a phoneme in English while it is allophone in Vietnamese. Initial /t, tʰ/ are realized as apico-dental [t̪, t̪ʰ] or lamino-alveolar [t̻, t̻ʰ] (Kirby, 2011). Furthermore in Vietnamese the dental [t̪ʰ] is contrastive with [t⁼], which is unaspirated and alveolar (Singer, 2012). In other words dental aspirated [t̪ʰ] is a phoneme which means that Vietnamese speakers employ aspiration on phonological level whereas English speakers have aspirated sounds only as allophones.

Looking at the two remaining English fortis plosives, Singer (2012) asserts that aspirated [pʰ, kʰ] tend to be unaspirated when they are articulated by a Vietnamese speaker of English. This

is characteristic even for /p, k/ in initial stressed position. His conclusion leads us again to the fact that although Vietnamese speakers of English employ aspiration in their L1 on phonological level, they do not aspirate in English on phonetic level. Considering again the research of Cunningham (2009), who concludes that Vietnamese speakers aspirate even the lenis stop /g/, we can expect that the English speaking Vietnamese may struggle extensively with the aspirated sounds in English.

2.8.1.4. Voicing Contrast in Vietnamese

Unlike English, Vietnamese language employs a three-way voicing contrast between coronal stops: prevoiced *đ* /d/, voiceless unaspirated *t* /t/, and voiceless aspirated *th* /tʰ/. The prevoiced đ is distinguished by a long negative VOT. Slightly positive VOT characterizes voiceless *t*; in comparison, the aspirated *th* also has a positive VOT, but it is much longer than that of the plain voiceless *t*. English, in comparison, applies only a two-way voicing contrast in the initial position of words, between voiced or prevoiced /d/ and voiceless /t/ (Twist, Shamoo, Bauman, 2009: 4).

In conclusion to the theoretical introduction, which briefly compares English and Vietnamese phonological inventory and their means of contrasting, it is important to point out the following. Relatively recent studies (Singer, 2012), and (Cunningham, 2009) elaborating on Vietnamese-accented English inform us about facts that contradict the presupposition of L2 theories that Vietnamese speakers of English should struggle less in employing the voicing contrast in English. The following section describes methodology used to examine the voicing contrast in Vietnamese English.

3. Methodology

3.1. Informants / Speakers

In this kind of research, primary attention had to be paid to the selection of speakers. Their advanced command of spoken English was crucial. Further, in order to provide a representative group of speakers, it was important to test males and females in roughly equal proportion. For the purpose of this study fifteen recordings were obtained. Each of the fifteen informants produced a reading of a BBC News text. Each recording is about four minutes long and in total they comprise of one hour of recorded material. The following sections consist of more detailed information about the process.

3.1.1. Selection of Speakers

The following criteria were decisive for obtaining the spoken material. The command of spoken English of all the selected speakers was approximately B2 (according to CEFR). The participants in this research were university students pursuing a bachelor's or master's degree in the Czech Republic. As all but one of them have not been residents in the Czech Republic for more than four years, their command of Czech would not have been sufficient to enrol in a course taught in the Czech language. Therefore they opted for a university programme taught in English. They were asked to read a short coverage that appeared as a part of BBC news broadcasting. The reading of the text required a good knowledge of English, because we needed fluent and as-natural-as-possible connected speech, in order to be able to measure the VOT durations.

3.1.2. Gender

The gender makeup of the group studied was nine female students and six male students. The group intentionally consisted of close-to-equal numbers of both sexes so as not to collect a gender-biased sample.

3.1.3. Dialectal Background of the Vietnamese Speakers

It is worthwhile noting that compared to relative dialectal unity of the Czech Republic Vietnam has a rich diversity in vernaculars, which may result in inter-dialectal comprehension issues.

For the most part, the informants shared a dialect of the capital, Hanoi. This generally accepted standard can be found in the northern part of Vietnam. The only exceptions were speakers M6S and M7S (see Tab.3.1), who grew up in South Vietnam. It is known that over centuries, many highly specific dialects developed in the Vietnamese language. Vietnam has a series of integrating dialects. The heterogeneity of these dialects often cause that people from different regions hardly understand each other. In order to simplify, there are three mayor

dialects in Vietnam. Hanoi dialect is mostly considered to be the metropolitan standard and is spoken by people living in the North of Vietnam. Then there is a dialect of the imperial city Hué and southern capital Saigon. The dialects are typical for its pronunciation differences. Hanoi dialect is often described as a chopped speech compared to the Southerners, and its speech lacks retroflex consonants that we would find in the dialect of Hué (Thompson, 1987:18).

Since we did not want dialectal heterogeneity to affect our results, we recorded speakers coming predominantly from the North, in order for our corpus to maintain relative regional uniformity. Since none of the study participants, except for two, had been exposed to a native-English-speaking environment; their English second language acquisition took place in authentic Vietnamese-language teaching conditions. In light of the fact that all the speakers took language courses taught by Vietnamese teachers of English in Vietnam, we can say with certainty that their English is Vietnamese-accented English.

3.1.4. Age

The criteria of L2 acquisition and the level of English that was necessary for our research narrowed down the number of informant candidates considerably. Although there is a large community of Vietnamese living in the Czech Republic (CR), most of them do not speak English. The generation of Vietnamese that came to the CR after 1970 to work mostly speak in their mother tongue or in French. Their children forming the second generation of Vietnamese in the CR have a good command of English but their English is already Czech-accented. The group of Vietnamese that we recorded are all young students, aged mostly between 20–23, who were either visiting their relatives in the CR or came to study at Czech universities. The group is age-wise uniform. The average age of the recorded group is 22.4. The method that we used to find the participants for this study was – in simple terms – a friend-of-a-friend engine. The Table 3.1 shows the list of all the participants that we recorded and analysed.

Coding	***Gender***	***Age***	***Origin***	***Length of residence in the CR in years***	***Exposure to English in English-speaking environment***
F1	Female	21	Hanoi	3	yes
M1	Male	22	Bac Ninh	4	no
M2	Male	25	Vinh	2	yes
F2	Female	30	Halong Bay	1	no
F3	Female	26	Hanoi	2,5	no
M3	Male	31	Hai Phong	20	no

M4	Male	23	Hanoi	4	no
F4	Female	20	Hanoi	2	no
F5	Female	24	Hanoi	0,2	no
F6	Female	23	Hanoi	3	no
F7	Female	17	Hanoi	2	no
F8	Female	19	Hanoi	0,5	no
M5	Male	22	Hanoi	3	no
M6S	Female	30	South Vietnam	2	no
M7S	Male	21	South Vietnam	2	no

Table 3.1. – Table of informants

3.2. Corpus Recording

The following section is dedicated to the process of recording. In order to obtain the necessary material, we had to adjust the time and place of recording to the availability of the informants. Due to the workload of the participants all the recording sessions had to be obtained out of reach of the recording facilities of the Department of Phonetics at Charles University. Consequently, to make all the recordings in the front seats of a car turned out to be the most suitable solution.

Before the recording itself, the vehicle was parked in a calm and quiet area. In order to prevent the sound waves from deflecting off the dashboard and creating an undesirable echo, the dashboard was fully covered with a large piece of woollen material and the recorder itself was positioned on a similar echo-absorbing material. The informants were seated approximately 70 centimetres from the recorder. Each of the speakers had time to get acquainted with the text before the recording itself took place.

Once they were ready they were asked to read the first couple of sentences so that the input of the recorder could be adjusted to their voice volume. Afterwards they read text, approximately 500 words long, was composed of eight different BBS News extracts[5]. On average, it took the informant circa five minutes to read the text. The corpus was obtained using the portable recorder Edirol HR-09 with a 48 kHz sampling rate.

There were no significant gender differences noted during the execution of the study, only that some of the female speakers tended not to speak as loud as would be ideal for the purposes of recording.

[5] See attachments

3.2.1. Corpus / Data Processing

Having recorded all fifteen speakers, we proceeded to the processing of the obtained material. First we had to increase the volume. Although the input of the recorder was adjusted to the speaker's voice volume, a few of the respondents still produced material that required intensification of the volume. The volume was sometimes inconveniently low, because some of the speakers did not maintain consistent voice intensity during the reading. For this purpose we used Adobe Audition. This program also helped to remove the small amount of disturbance noise that could in some cases not be prevented during the recordings. However, these adjustments were minor and not at all detrimental to measuring Voice Onset Time.

Secondly, the recorded material was later processed using Praat, a phonetic computer program developed by Paul Boersma and David Weenink. Using this program, each of the recorded BBC news texts was cut into paragraphs for later convenience of phone segmentation.

3.2.2. Forced Alignment

The parsed material was then split into phones via the method of forced alignment using the P2FA tool (Yuan, J. & Liberman, M., 2008). The tool divided the individual recordings into phones, however, the programme is only designed to recognize either English or American native speech. Moreover, words that had not been incorporated into the programme's corpora were transcribed as SP (sound pause) or the aligner displayed all words that showed dubious segmentation as capitalized. For these reasons, manual segmentation was also necessary. The final and most important step was to manually correct the boundaries of the plosives we intended to examine.

3.2.3. Manual Selection of Plosives

For the sake of this study, only plosives in main stressed positions were selected. The selected phones were: occurrence of both voiced and voiceless stops immediately followed by a vowel.

Instances of stops that appear in stressed position but are followed by a consonant sound (e.g. [k̚]) were eliminated, as the plosives are already marked by devoicing and therefore they were not in the focus of this study. Furthermore, occurrences where /s/ precedes a stop were selected: although English phonotactics prefers no positive VOT for such stops in stressed position, Vietnamese English speakers may have acquired disparate tendencies. For this reason we included instances such as /sp-, st-, sk-/ to measure their VOT.

After the forced alignment divided the words into phonemes in the phone tier (Fig. 3.6), and we manually corrected their boundaries, the target plosives had to be bounded. The boundaries were marked, if we look at the sound wave, starting from the closure of the

obstruents and ending at the commencement of the first periodical sound wave of the following vowel.

Since a great number of stressed prevocalic obstruents are also word-initial, all the sound pauses, marked SP, preceding the stops, had to be checked. Owing to the fact that northern Vietnamese is specific for its so-called 'chopped' speech, the alignment marked a relatively large number of pauses, not only in front of the target obstruents. In cases where the stop was really preceded by a pause, the stop had to be removed due to the impossibility to clearly determine its closure phase.

Following this, an extra interval tier was added to delineate the target stressed postvocalic plosives. The extra tier was labelled VOT and unlike the phone tier consisted of the target stops only. These are similarly bounded, however, this time the demarcation started only from the burst, see Fig. 3.6, not from the closure of the stop. The end was, as usually, placed before the onset of voicing of the following vowel. When the last demarcation process was finished the data had to be extracted.

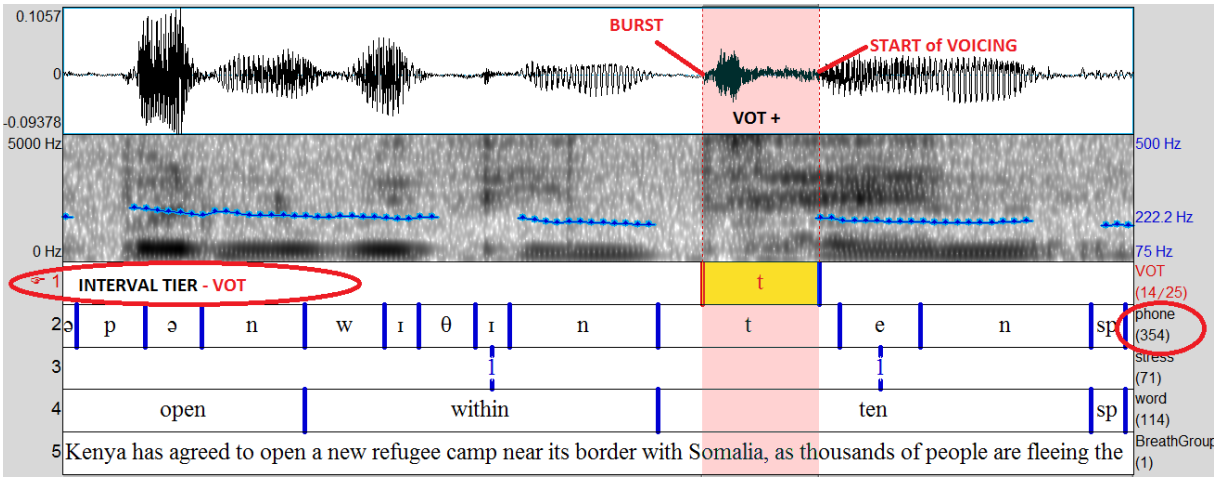

Fig. 3.6. – Bounded plosive /t/ in VOT interval tier

3.3. Data Extraction

The next step was to create a script that extracted all the necessary data from the VOT interval tier. The script generated VOT values based on the former demarcation of the target stressed plosives /p, t, k/, /b, d, g/ that follow a vowel, from both phone and VOT interval tier. The script compared the total time during which a certain plosive was articulated and the duration of either positive or negative VOT within this articulation. Finally, all minus and plus VOT values were extracted for all target stops. Later we transferred all the data to a MS Excel file where we ordered it in preparation for the analysis.

The Excel sheet contained six columns. Column number one contained all target obstruents – 776 in total. The second column contained information on the presence or absence of voicing,

and the third displayed the particular word in which the given stop occurred. In order to maintain the anonymity of the participants they were relabelled preserving only the information about their sex and origin. The two Vietnamese speakers that were originally from the South of Vietnam were coded with 'S' at the end. These measures were adopted due to possible phonetic deviations these speakers might exhibit from the speech of their Northern fellow-citizens, so they were placed at the very end of the table into a fourth column.

Moreover, it was necessary to label target obstruents that were preceded by /s/ in order to distinguish them from the rest. On that account one extra column was added. Finally, the sixth column consisted of the VOT values of each of the target obstruents.

In order to compare the data, we decided to use two statistical methods. The first one was ANOVA, thorough which we calculated the analysis of variance for aggregate relations. The second one was Tukey's HSD Test (Post-hoc test) that we used to identify the concrete relations among the measured units.

3.4. Hypotheses

As has been noted in the section on Vietnamese English, Vietnamese employs a three-way voicing contrast in its initial plosives. We can therefore hypothesise that:

H1: Vietnamese English stops in initial stressed position will exhibit very similar categories to those of English native speakers, as their L1 contains even higher voicing diversity in its word-initial stops (prevoiced plosives, voiceless unaspirated, and voiceless aspirated plosives) (Twist, Shamoo, Bauman, 2009).

This assumption is supported by Leather and James (1996: 274), as mentioned in 2.6.1, where we state that VOT proves to be most audible to elementary speakers. If even elementary speakers are sensitive to the changes of VOT we can assume that they are able to imitate and use it.

H2: It is assumed that there will be no significant differences between male and female speakers. There is a large number of studies that claim women to be quicker in terms of L2 acquisition (Farhady, 1982, Rod, E., 1994: 203). Some studies even assert that women master a foreign language and pronunciation faster, nevertheless, we did not notice any particular differences between the female and male respondents when listening to their speech.

H3: We may expect differences in VOT values produced by Southern Vietnamese speakers, since the Vietnamese language exhibits a rich regional variety. The northern, central and southern accents alone differ significantly. These regional dialects sometimes even cause difficulties in understanding. The regional variety could be transferred to L2.

H4: Vietnamese English will display significant shortening of VOT in combinations such as /s + p, t, k /. This assumption is made based on general knowledge of phonetics. Human speech often aims at being economic and tends to maintain the important contrast between words. When pronouncing the phoneme /s/, which is a high-frequency hissing sound, we already use up a relatively large amount of air held up in the lungs. As has been mentioned, aspiration also needs sufficient flow of exhaled air. Following this, combinations like /sp^{h}, st^{h}, sk^{h}/ are highly uneconomical and very difficult to pronounce and therefore we do not expect them to occur in Vietnamese-accented speech.

H5: As noted in the section 2.6 and 2.7, it can be anticipated that the VOT values produced by our study's informants shall not have native-like values as the speakers had not learnt L2 before the age of six. The informants are expected to produce compromised values of VOT.

4. Analysis

This chapter shows the results of several tests that were carried out in order to prove or disprove our formerly stated hypotheses. The results are illustrated in graphs and figures presented together with their analysis. Individual categories are organized according to the following variables: VOT, voicing; sex, origin, origin & sex, plosives preceded by /s/, and frequency of prevoicing occurrence. Following this, we compare our acquired VOT values for Vietnamese-accented English to the VOT data of four American English–speaking informants, published by Lisker and Abramson (1964).

As has been mentioned, prevoicing in voiced initial plosives is called voicing lag and is known for its negative VOT values. By contrast, aspiration, or in other words positive VOT, is expected to occur in voiceless plosives. Both voiced /b, d, g/ and voiceless /p, t, k/ plosives were compared within their class using Tukey's HSD test and ANOVA. The relations shown below were classified as significant according to these statistical analyses.

4.1. Ratio: VOT, Voicing

We first used ANOVA to analyse and compare the aggregate relation data, proceeding from the general to the specific. First, we looked into the effect of voicing on VOT. In the beginning of the analysis, we compared the voiced and voiceless plosives and their VOT values. Figure 4.1 below displays how VOT values (*ms*) change depending on whether a given plosive is voiced or voiceless; the difference is significant ($p < 0.001$) based on the equation: $F(1,757) = 393{,}37$.

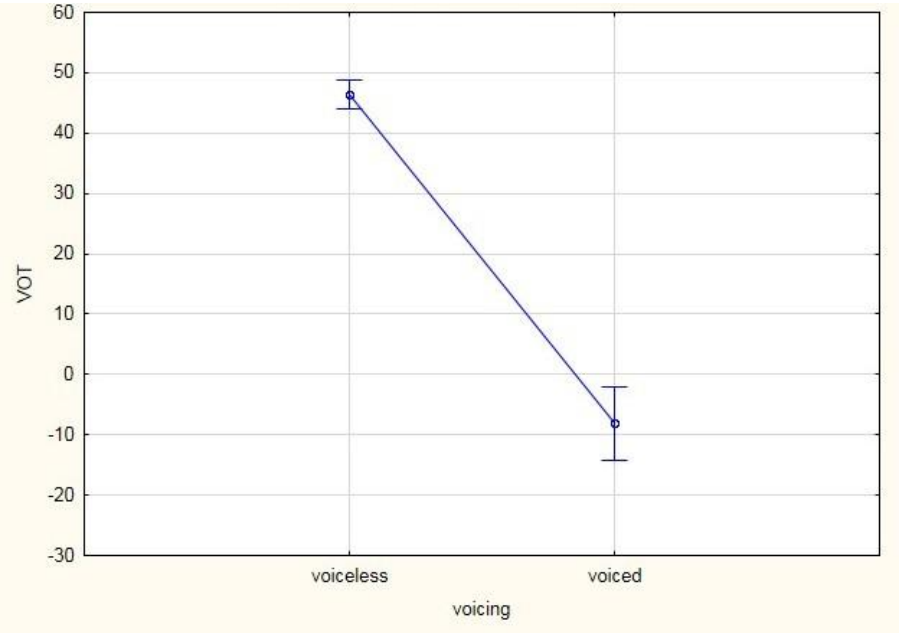

Fig. 4.1. – Ration: VOT is lower for voiced plosives
(The error bars in all figures denote the 95% confidence intervals)

Vietnamese English maintains the two-way plosive contrast based on VOT values similarly to English spoken by native speakers. Voiceless stops are characterized by higher VOT values, while their voiced counterparts have lower values, as the figure shows. The level of significance here is $p < 0.001$ based on $F(5,753) = 97,057$.

The individual voiced plosives (see Fig 4.2) are dispersed either in the VOT plus or minus area. We start our observation with plosives that are characterised by negative VOT values. We can observe that the phonemes /b/ and /d/ are pre-voiced and their VOT ranges between -15 and -10 *ms*, whilst the phoneme /g/ displays positive VOT values of ± 10 *ms*.

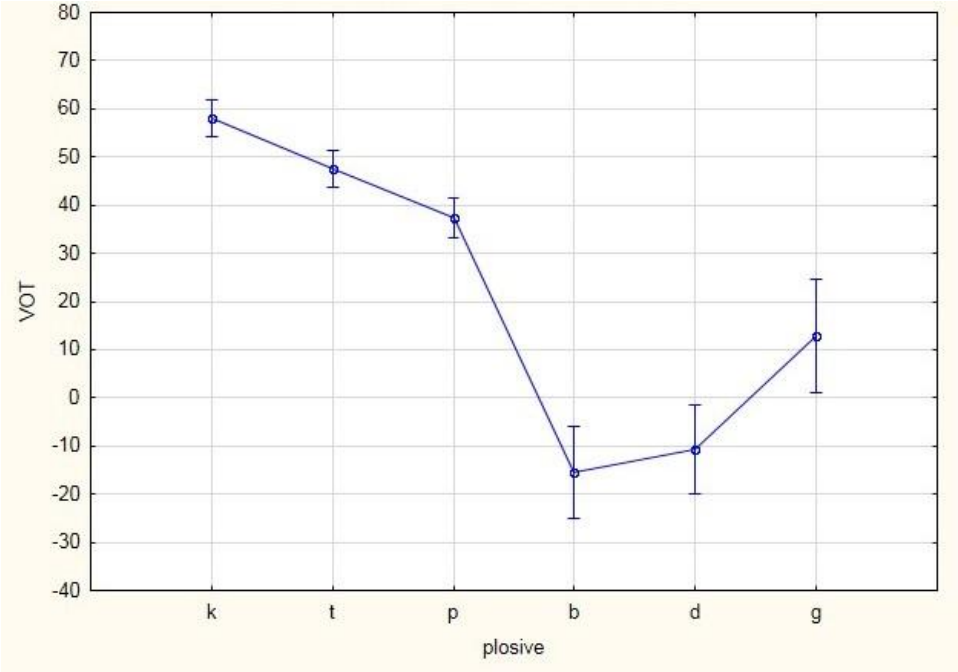

Fig. 4.2. – Plosives and VOT in Vietnamese English

As has been stated, plosives produced further back in the oral cavity are characterized by higher VOT values. Vietnamese English displays the same tendency. Considering the voiced set of plosives, we can classify /g/ as unaspirated stop because the value is less than 20 *ms*.

Similarly, the voiceless velar plosive /k/ has the highest VOT value from the voiceless set, because it shares the velar place of articulation. Correspondingly, dental plosives – voiced or voiceless – have higher values of VOT than the bilabial plosives pronounced at the very front of the mouth. Referring to the above, the Vietnamese English plosives, in terms of VOT, do not deviate from the way native English uses the amount of VOT do differentiate voiced and voiceless plosives. Positive VOT longer than 20 *ms* is likewise found with fortis stops, while lenis stops are bellow 20 *ms* and have negative VOT.

4.2. Plosives & Sex

Languages can display differences that are gender-specific and therefore we also tested whether voicing contrast in Vietnamese English evinces such distinctions. Figure 4.3 below displays the target set of fortis and lenis plosives produced by male and female informants. It can be observed that proportionally there are no major particularities given by the sex of the speaker. The only exception seems to be the voiced stop /b/ where the difference in VOT values is statistically significant according to Tukey's HSD Test: $p < 0.05$ based on F (5,747) = 3, 2276.

The VOT values produced by the male informants are lower than those produced by females. Among males the VOT values for /b/ reached up to - 40 *ms*, while among females the values approximated 0 *ms*.

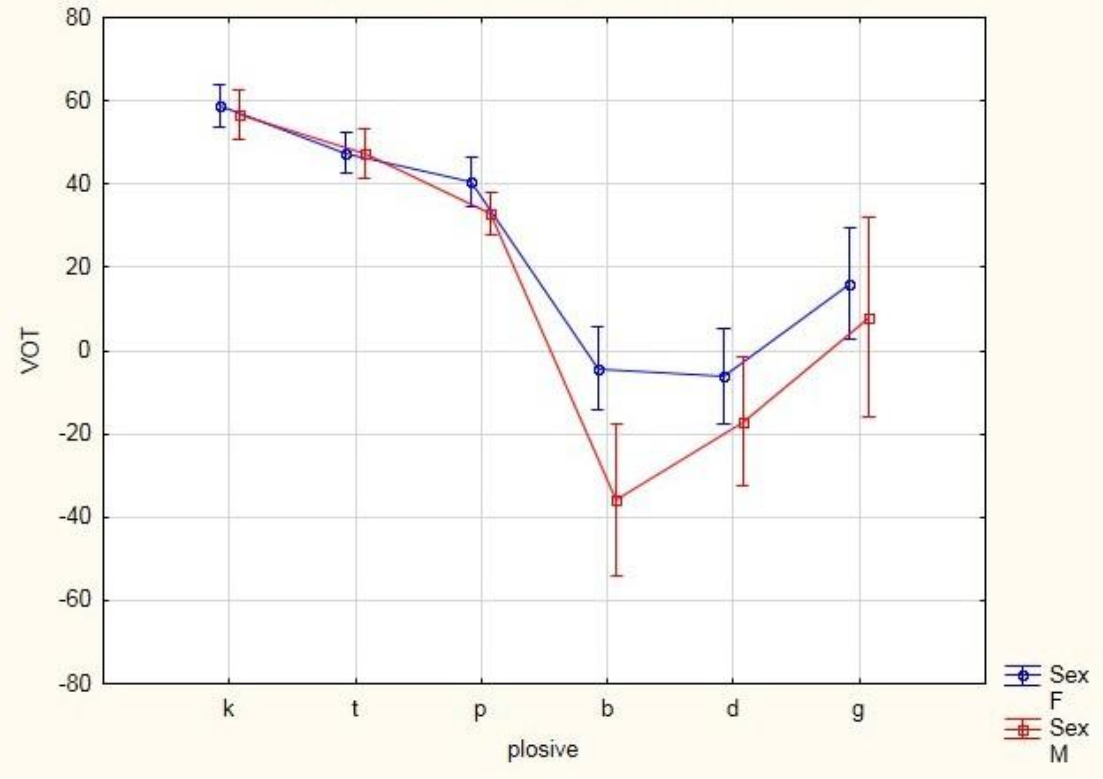

Fig. 4.3. – The influence of sex on VOT

In short, significance is reached only with the /b/ plosive – the error bars do not overlap (see Fig 4.3). Male Vietnamese seemed to produce more instances of pre-voiced /b/ than females. Men are expected to have larger vocal tract size and it is thus generally assumed to be easier for them to produce prevoicing. Van de Weijer and Van der Torre (2007:102) in their study point to the fact that the occurrence of prevoicing with males is 20% higher than with females. We can therefore assume that prevoicing is influenced by sex. Moreover, the place of articulation, which determines the extent to which the vocal tract can be expanded, may also

results in higher chance of prevoicing. The smaller the volume, the faster the supraglottal pressure increases and the more difficult it is to produce prevoicing (Van Alpen & Smits, 2004). The authors assert that bilabial plosives tend to be prevoiced more often because when pronouncing them, it is possible to maximally expand the vocal tract behind the point of constriction. Bilabial plosives can make use of the expansion of: pharyngeal walls, soft palate, oral cavity plus tongue surface and parts of the cheeks, which cause larger vocal tract and easier prevoicing. There are physiological and aerodynamic influences that are bound to anatomical gender differences and together they affect the occurrence of prevoicing.

4.3. Plosive & Origin

The following test investigates whether the acquired corpus evinces any differences for the two informants that were originally from the south of Vietnam. In respect to the Vietnamese language that shows rich regional variety we may expect that regional specificities were transferred from L1 to Vietnamese English. We decided to compare voiceless and voiced plosives separately. Figure 4.4 below shows the VOT of voiceless plosives for all the fifteen recorded speakers. The two Southern Vietnamese are marked with 'S' at the end.

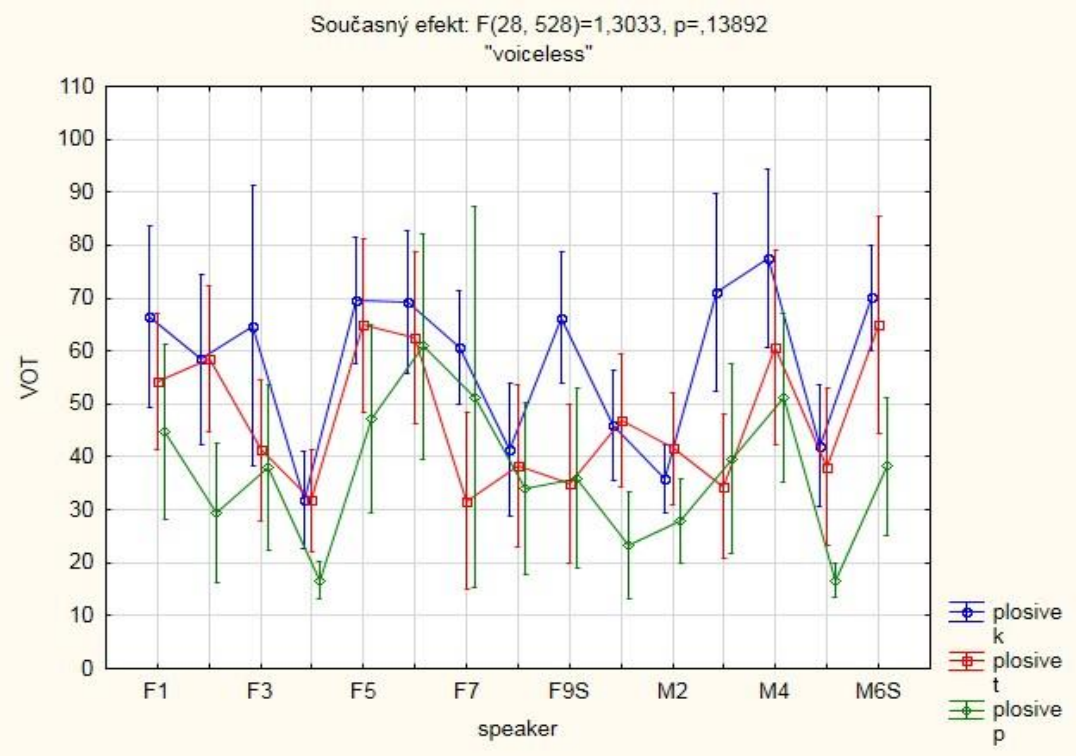

Fig. 4.4. – VOT of individual Vietnamese speakers, voiceless plosives

Compared to the majority of the North Vietnamese speakers, the Southerners do not show significant differences in the production of plosives. The comparison of the constituents using

post hoc tests yields insignificant results and the same can be said about the voiced plosives displayed in Figure 4.5.

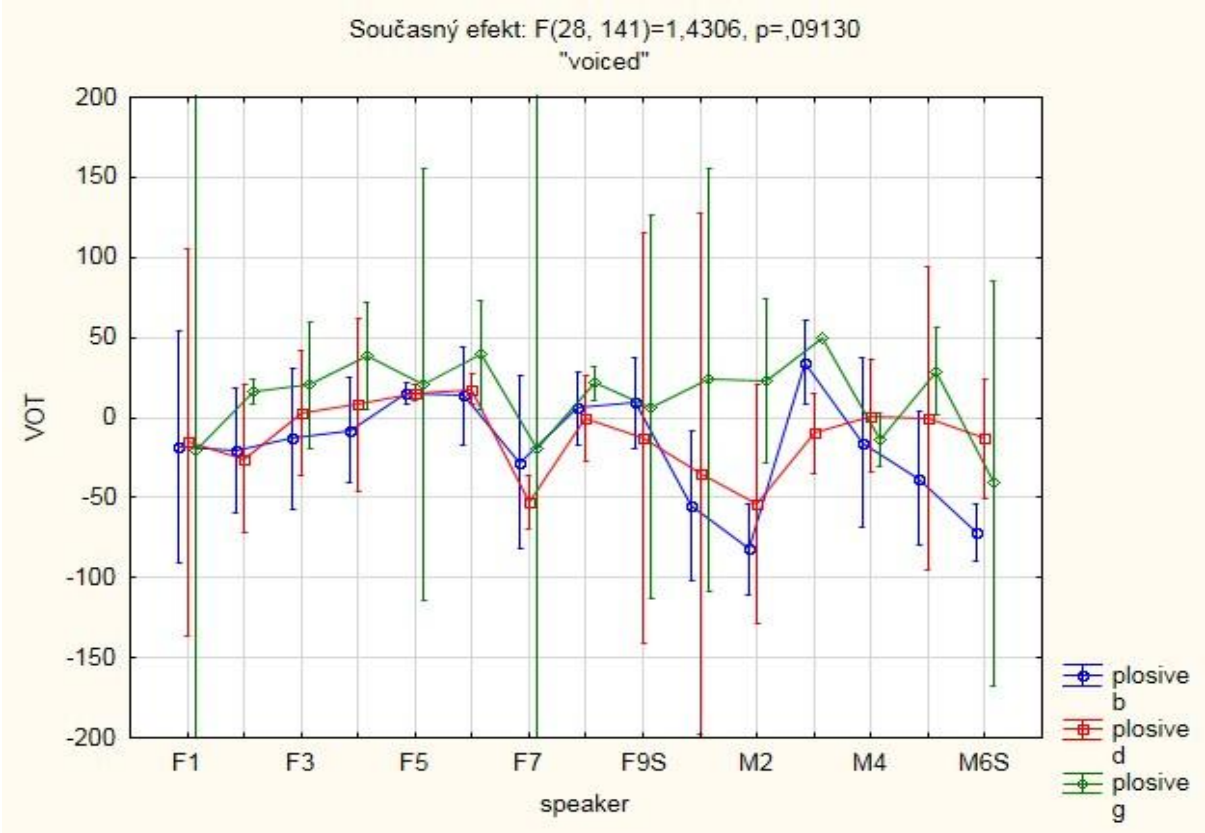

Fig. 4.5. – VOT of individual Vietnamese speakers, voiced plosives

4.3.1. Plosive & Origin, Sex

Next, we decided to compare sexes within the category of the Southern Vietnamese speakers. This test similarly did not reveal any significant differences. We can therefore assume that there are no fundamental differences between sexes in producing stressed voiced and voiceless stops in Vietnamese English (see Figures 4.4 and 4.5). Although the two above experiments did not reveal any regional differences, we should point out the small sample consisting of only two speakers tested. For a more reliable result a bigger sample is necessary.

4.4. Plosives Preceded by /s/

The general tendency in native English realization of initial stressed stops is to differentiate VOT for fortis plosives preceded by /s/ and without /s/. The tendency is that VOT is lower for plosives with /s/ and higher for plosives without /s/. In order to compare the Vietnamese English VOT behaviour, we measured it in all /s/+/p, t, k/ combinations. When we had all the data from the fifteen recordings, we examined whether the combinations with /s/ differ from the stops without /s/.

If we look at the data in Fig. 4.6, they overall show that Vietnamese speakers of English also have this difference in VOT. The equation F (2,567) = 6,6830 resulted in p =.001 which shows us a highly significant results that correspond to the native like tendency to have the VOT for /p, t, k/ higher than for /sp, st/. The graph displays solid error bar dispersal with mainly lower VOT values for the voiceless stops preceded by /s/. The only deviation the speakers exhibited was that they did not maintain the opposition /s/ versus /sk/. The VOT values for /k/ seemed not to keep the difference between these two because of its relatively high VOT for /sk/ combination - see Fig. 4.6.

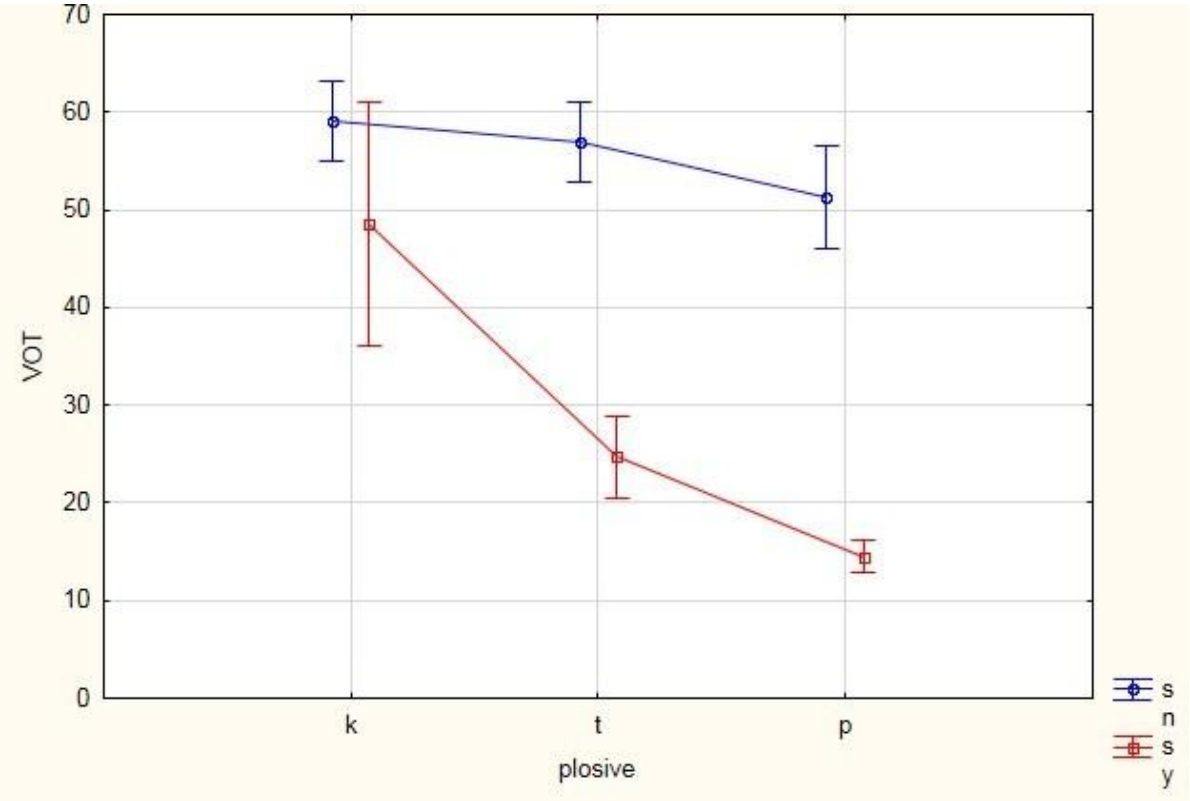

Fig. 4.6. VOT in fortis plosives preceded by /s/ yes x no

Considering the recorded words, it turned out that this contrast was caused by the repetitive occurrence of /sk/ throughout only one word ‘rediscovered’, therefore the results of higher VOT for /k/ preceded by /s/ may not be significant (Tukey: p>0.5).

Although the consonant /k/ had long VOT, the result of the statistical analysis including the deviation still yields highly significant difference for fortis plosives with /s/ and without /s/. We can therefore assert that the informants maintain difference in VOT that correspond to native-like behaviour.

4.5. Frequency of Prevoicing Occurrence

Finally, we looked at the set of voiced stops where we measured the occurrence of prevoicing, in other words, negative VOT. The highest occurrence of prevoicing was measured with the lenis plosive /g/, while /b/ and /d/ showed similar probabilities of being accompanied by prevoicing. Below in Table 4.5 we can see the exact numeral representation of each of the particular voiced plosives.

	negative	***positive***
/b/	37	51
/d/	28	32
/g/	9	29

Tab. 4.5. – Number of occurrences of positive or negative VOT with particular lenis plosives

Only the high percentage of prevoicing with /g/ (see Figure 4.7) turned out to be significantly different from the other two, proving to be statistically marginally significant, with probability $p = 0.064$. We used the chi-squared test with the calculation $\chi 2\ (2;\ n = 186) = 5.49$.

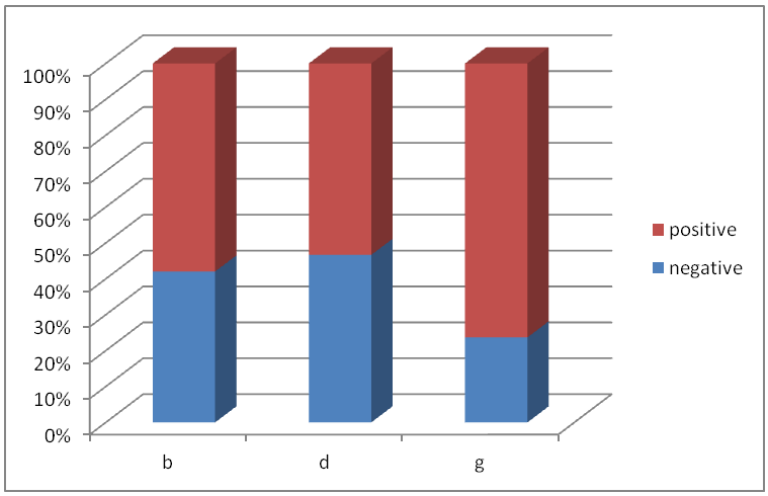

Fig. 4.7. – VOT Comparing the percentage of occurrence

All in all, the lenis plosives exhibited similar distributions of prevoicing and positive VOT. The deviation measured with the voiced velar plosive /g/ may be explained by the general tendencies for velar stops to be produced with longer VOT.

The following part closes the analysis section, making comparison with the average Vietnamese English VOT values and the average VOT values of American English speakers. For the reference average values for native English word-initial stops, we used the data published by Lisker and Abramson (1964), see Tab. 4.6 below. We comment on the compromised values of VOT with reference to those of native AmE speakers and we explain the particularities of Vietnamese-accented English.

	p	t	k	b	d	g	b+	d+	g+
Av./ms	28	39	43	7	9	17	-65	-56	-45
Group Av. / ms	36.6			16.5			-55.3		
				-19.4					

Tab. 4.6. – Average VOT values for initial stops in AmE[6] (+ indicates for all tables the presence of prevoicing)

	t (plain)	th (aspirated)	đ (prevoiced)
Av./ ms	15	90	-58

Tab. 4.7. - Average VOT values for Vietnamese plosives[7]

	p	t	k	b	d	g	b+	d+	g+
Av./ ms	37.27	47.44	57.73	18.98	18.31	30.39	-65.36	-43.59	-47.97
Group Av./ ms	46.27			21.7			-55		
				-9.63					

Tab. 4.8. - Average VOT values for VE plosives

As can be seen in Table 4.8, the average VOT values for VE prevoiced lenis stops, which we recorder, are very similar to those of native AmE speakers. The reason for such accuracy may be that the AmE values are close to the values of prevoicing in Vietnamese, compare Tab. 4.6 and 4.7 The VOT values for the Vietnamese English voiced plosives that do not undergo prevoicing are slightly higher than those produced by AmE speakers.

The voiceless Vietnamese English stops produced by our informants exhibit higher positive VOT values in comparison with the set of fortis obstruents produced by AmE speakers. This set of VE stops seems to exhibit compromised values of VOT, which means that the average VOT value for VE fortis plosives is somewhere in between the VE and AmE. The compromised values are most probably a result of the influence of L1 (Vietnamese). VE exhibits positive VOT values that are higher than those produced by the AmE speakers investigated by Lisker and Abramson.

In conclusion, the voicing contrast in Vietnamese English exhibits similar categories as native English. Our data indicate that Vietnamese-accented English maintains the two-way contrast that is typical in native English speech. In relation to VOT, the values for lenis prevoiced plosives are nearly identical with the values of AmE speakers, while lenis stops lacking prevoicing show slightly higher values. Fortis stops tend to have a higher positive VOT, probably due to the higher VOT typical for L1 Vietnamese fortis plosives.

---

6 Lisker, L. and Abramson, A. S., 1964, p. 410

7 Twist, A., Shamoo, J., Bauman,J. 2009, p. 5

5. Discussion

It is worth noting that, although the use of VOT is very popular due to its relative ease of measuring and variability across languages, there are quite a few pitfalls that arise when describing plosives. Mikuteit and Reetz (2007) in their research address problems of VOT measurements that stem from the inability to differentiate voiced aspirated plosives. However, considering the fact that Vietnamese maintains only three-way contrasting of its plosives and English is a language that employs even fewer distinctions among its plosives – that is, a two-way contrast – the measurement with the use of VOT is sufficient for this study, which does not face the problem of differentiating voiced aspirated plosives. Voiced aspirated plosives, however, occur in VE. Cunningham (2009) in his study deals with these plosives in Vietnamese-accented English. Nevertheless, the Vietnamese we recorded did not pronounce anything similar to them.

The results of the analysis show that Vietnamese-speaking learners of English maintain a two-way contrast for English plosives. The voicing contrast in Vietnamese English is therefore shown to follow native-like tendencies. Firstly, the analysis compared VOT with respect to voicing, which proved to follow native-like standards and with respect to place of articulation, as the results were higher for plosives produced further back in the mouth – VOT was shown to be longest in the case of velars and shortest in the case of bilabials.

Secondly, we looked into the influence of gender on the characteristics of plosives, which proved to be significant with the consonant /b/, whose VOT was much lower for male speakers. This, however, may be ascribed to the anatomical differences that occur not only between the sexes but also among speakers in general so in order to draw a more general conclusion, further analysis is required. Moreover, the gender-based differences were also compared in terms of southern or northern origin, which, however, did not yield any significant differences between the sexes, nor between the Southern and Northern accent. We should point out the fact that the investigated sample of Sothern Vietnamese consisted of only two informants, so further research is necessary to assure these results.

Thirdly, the analysis looked at the VOT for plosives preceded by /s/, which follows the native English standards of omitting aspiration. VE has the contrast between fortis plosives preceded by /s/ and without /s/. This contrast is significant although there was deviation displayed by /k/, which had a considerably higher positive VOT. The higher VOT values could have been caused by /k/ that was repeatedly occurring in only one word, and therefore further analysis to confirm this tendency is necessary.

Further, we compared the presence of positive and negative Voice Onset Times in lenis plosives. The result showed similar percentages in the distribution of positive and negative VOTs for lenis plosives except for the velar /g/, where prevoicing occurred only in 15% of instances in the recorded corpus.

Finally, we compared the average values of VOT in our analysis of Vietnamese-accented English with average values for native speakers of American English, published by Lisker and Abramson (1964). The result showed values higher for fortis stops, while prevoiced plosives displayed average VOT values very similar to native AmE. This resemblance can be ascribed to the fact that the average values of prevoicing in Vietnamese correspond to the average values for Vietnamese-accented English produced by our informants. Our speakers may have classified prevoicing in English similarly to prevoicing in the L1 – Vietnamese đ bears a negative VOT very similar to that of the English voiced alveolar plosive /d/. Considering the fact that all of the informants were late learners of L2 – they did not acquire English before the age of six – they do not have separate phonetic categories for L1 and L2 VOTs.

6. Conclusion

This thesis was aimed at describing the voicing contrast in Vietnamese English. The theoretical part contains a brief account on the theory that is necessary to understand the concepts of voicing, aspiration and measurement based on the values of VOT. It also comments on second language acquisition and the way pronunciation features can be transmitted to L2 – in our case, English. Furthermore, Vietnamese and Vietnamese English are described with regard to their initial consonants so that the later comparison of our study's results can draw upon this theory. We developed several hypotheses based on the fact that Vietnamese is a language with a three-way plosive contrast while English has a two-way contrast, as well as on several second language acquisition theories.

The Analysis section showed that Vietnamese English follows the native English standards of forming the voicing contrast. In the conclusion of that section we compared the average values of VOTs produced by our informants with those produced by native speakers of American English. The results show specific VOT values for Vietnamese-accented English.

Since our informants have been shown not to struggle with maintaining the contrast between initial English stops, we can assume that when L1 contains a richer voicing contrast in plosives, the L2 learner will not struggle with the acquisition of voicing contrast in initial consonants. Nevertheless, as has been examined in Cunningham (2009) it may not always be the case. His study presents us with results showing that Vietnamese speakers of English may have positive VOT even with lenis plosives. It is therefore necessary to first analyse the foreign speech and accordingly prepare a teaching method. Such method can be based on repetitive listening. Learners of foreign language struggling with a certain feature of voicing contrast (such as aspiration), should record their speech and later re-listen to it. Since positive VOT is recognizable even to elementary speakers, practicing it is essential in order to achieve comprehensibility. Since the VE speakers we worked with do not struggle with aspiration, they should be a subject to different L2 teaching method than for instance Czech students of English. Czech accented English in most of the cases lacks aspiration.

7. Bibliography

Sources Quoted

Cunningham, U. (2009). "Phonetic Correlates of Unintelligibility in Vietnamese-accented English". *Proceedings, Fonetik 2009*. University of Stockholm. 1 - 4.

Cho, T. & Ladefoged, P (1999). "Variation and Universals in VOT: evidence from 18 languages". *Journal of Phonetics* 27, 207 - 229.

Farhady, H. (1980). *Justification, Development, and Validation of Functional Language Testing*. Ph.D.dissertation, University of California.

Rod, E. (1994). *The Study of Second Language Acquisition*. HK: Oxford University Press. 200 - 203.

Hansen, J. H. L., Gray, S .S. & Kim, W. (2010). "Automatic Voice Onset Time Detection for Unvoiced Stops (/p/, /t/, /k/) with Application Accent Classification". *Speech Communication* 52. 777 - 789.

Henton, C., Ladefoged, P., & Maddieson, I. (1992). Stops in the World's Languages. *Phonetica* 49 (2), 65 - 101.

Hoang Thi Quynh Hoa. (1965). *A Phonological Contrastive Study of Vietnamese and English*. A Thesis in English, Graduate at Faculty of Texas Technological College.

Jessen, M. (1998). "Phonetics and Phonology of Tense and Lax Obstruents in German". Amsterdam: John Benjamin's publishing Company.

Kessinger, R. H. & Blumstein, S. (1997). "Effects of Speaking Rate on Voice-Onset Time in Thai, French, and English". *Journal of Phonetics*, 25. 143 - 168.

Kirby, J. P. (2011). "Vietnamese (Hanoi Vietnamese)". *Journal of the International Phonetic Association 41(3)*. 381 - 392.

Ladefoged, P. (2010). *A Course in Phonetics*, (3rd ed.). Fort Worth: Harcourt Brace College Publishers. 141 - 147.

Leather J. H., & James A. R. (1991). "The Acquisition of Second-Language Speech". *Studies in Second Language Acquisition* 13. 269 - 269.

Lisker, L., & Abramson, A. S. (1964). "A Cross-Language Study of Voicing in Initial Stops: Acoustical Measurements". *Word* 20. 384 - 422.

Mikuteit, S., & Reetz, H. (2007). "Caught in the ACT: The Timing of Aspiration and Voicing in East Belgali". *Language and Speech* 50 (2). Kingston Press Ltd.

Miller, J. L., Green, K. P., & Reeves, A. (1986). "Speaking Rate and Segments: A look at the relation between speech production and speech perception for the voicing contrast". *Phonetica* 43. 106 - 115.

Ní Chasaide, A (1987). "Glottal Control of Aspiration and Voicelessness". *Proceedings of the 11th International Congress of the Phonetic Sciences,* 6, 28 - 31. Tallin: ISPhS.

Roach, P. (2009). *English Phonetics and Phonology, A practical course* (4th ed.) Edinburg: Cambridge University Press.

Singer, E. (2012). "Vietnamese English". University of York. Canada. Retrieved from http://www.yorku.ca/earmstro/asia/Vietnamese_accent.pdf (accessed: 1 June 2015)

Skarnitzl, R. (2011). *Znělostní Kontrast Nejen v Češtině*. Praha: Nakladatelství Epocha.

Stevens, K. N. (1998). *Acoustic Phonetics*. Cambridge, Massachusetts: MIT Press

Thompson, L. (1987). *Vietnamese Reference Grammar*. Honolulu: University of Hawaii Press

Twist, A. Shamoo, J., & Bauman, J. (2009). "Pulling out all the Stops in Vietnamese: A delineation between native and non-native Vietnamese speech for voice onset time". University of Maryland. 4 - 7.

Volaitis, L., & Miller, J. L. (1992). "Phonetic Prototypes: Influence of place of articulation and speaking rate on the internal structure of voicing categories". *Journal of the Acoustical Society of America* 92. 723 - 735.

Volín, J. (2002). *IPA-Based Transcription for Czech Students of English.* Charles University in Prague: Karolinum.

Van de Weijer, J., & Van der Torre, E. J. (2007) *Voicing in Dutch: (De)voicing phonology, phonetics, and psycholinguistics.* Amsterdam: John Benjamin's Publishing Company.

Wells, John C. (1982), *Accents of English*. Cambridge: Cambridge University Press.

Yuan, J., & Liberman, M. (2008). "Speaker identification on the SCOTUS corpus". *Proceedings of Acoustics '*08.

Zampini, M, & Edwards, H. J. (2008). *Phonology and Second Language Acquisition*. Amsterdam: John Benjamin's Publishing Company.

Sources Consulted

Darwin, C.J. & Pearson, M. (1982). "What Tells Us When Voicing Started?". *Speech Communication* 1/1982. 29 - 44.

Gonet, W., Rózanska, K. (2003) "Voice Onset Time in Word Initial Lenis Plosives in the Speech of Four BBC Presenters". Uniwersytet Marii Curie-Skłodowskiej.

Hwa-Froelich, D., Hodson, B., & Edwards, W., etc. (2002). "Characteristics of Vietnamese Phonology", *American Journal of Speech- Language Pathology* 11. 264 - 273.

Kohler, K. J. (1982). "F0 in the Production of Lenis and Fortis Plosives". *Phonetica* 39. 199 - 218.

Kováčová, Ch., & Romport M. (1976). *Konfrontace Zvukové Stránky Vietnamštiny a Češtiny a Její Pedagogická Aplikace.* Kandidátská disertační práce, Katedra Lingvistiky a Fonetiky, Univerzita Karlova v Praze, Filozofická Fakulta.

Leather, J. & James, A. R. (1996). "Second Language Speech." *Handbook of Second Language Acquisition*. San Diego: Academic Press. 269 – 316.

MacNeilage (Ed.), "The Production of Speech". New York: Springer-Verlag. 11-38

Maddieson, I. (1984). *Patterns of Sounds*. Cambridge: Cambridge University Press.

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

Sawashima, M. & Hirose, H. (1983). "Laryngeal Gestures in Speech Production". *Speech Production and Speech Modelling NATO ASI Series* 55. 289 – 322.

Slavická, B. (2011). *Praktická Fonetika Vietnamštiny*. Praha: Karolinum.

Smith, C. L. (1997). "The Devoicing of /z/ in American English: Effects of local and prosodic context." *Journal of Phonetics*, 25. 471 - 500.

Stevens, K. N., Blumstein, S. E., & Glicksman, L., etc. (1992). "Acoustic and Perceptual Characteristics of Voicing in Fricatives and Fricative Clusters". *Journal of the Acoustical Society of America* 91. 2979 - 3000.

Vaissière, J. (1997). "Phonological Use of the Larynx". *Larynx* 97. 115-126.

8. Appendices

- BBC news with Kyle Andrews.

- The Italian Senate has passed a tough austerity budget, including cuts of forty billion euros over the period of three years. The lower house must also adopt the measures in a vote on Friday. Italy has one of the largest debts in the eurozone, and the latest efforts aim to avoid any need for a bail-out. Both the government and the opposition are well aware of the fact that Italy is under close scrutiny due to its large debts. At the end of last week, the IMF urged Italy to ensure decisive implementation of spending cuts.

- Kenya has agreed to open a new refugee camp near its border with Somalia, as thousands of people are fleeing the region's worst drought in sixty years. The Prime Minister of Kenya announced that a camp which can fit up to 80,000 people would open within ten days. Some ministers had feared opening the camp would encourage more Somalis to cross the border, but the Prime Minister said that turning away the refugees was not an option. In his opinion, that would amount to ethnic cleansing.

- Buranda Airways is expected to hold an emergency meeting of its board of directors. The airlines have been experiencing problems over fulfilling admission criteria of the Star Alliance, a global airline network. The directors were angered by the latest offer and accused the Alliance of pursuing their own interests in the negotiations. The board, acting on the instructions of the shareholders, were ordered to arrive at a decision by Friday.

- You're listening to the news from the BBC in London.

- A colourful, spindly-legged toad that was believed to be extinct has been rediscovered in the forests of Borneo. Scientists from the university in Sarawak found three of the missing animals up a tree during a night-time search. The team had spent eight months searching for the toad. According to Dr Robin Moore of Conservation International, it is good to know that nature can surprise us when we are close to giving up hope, especially amidst our planet's escalating extinction crisis.

- A rare Jane Austen manuscript has sold for almost a million pounds in London, three times more than its estimated price. Auction house Sotheby's had originally valued the unfinished novel, entitled The Watsons, at about a quarter of a million pounds. The manuscript, originally owned privately, was purchased by an independent institution. It is assumed that Austen wrote the tale, about a young woman who returns to her father's household after being brought up by her aunt, in 1804. Austen published six complete novels and died in 1817 at the age of 41.

- BBC news.