Univerzita Karlova v Praze

Filozofická fakulta

Bakalářská práce

2015 Lenka Tranová

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/diplomovou práci vypracoval(a) 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 express many thanks to my supervisor doc. Radek Skarnitz 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.

Abstract

This thesis deals with the voicing contrast in Vietnamese accented English. The theoretical part introduces the generally accepted phenomena of voicing contrast, and several theories aiming at generalizing the main tendencies in acquiring a second language. The final part of the theoretical background is devoted to Vietnamese and Vietnamese English where we addressed the initial consonants. The methodological section provides information about the informants, recording, and data processing prior to the analysis itself.

Furthermore, graphs and tables illustrate the statistical calculations using ANOVA and Tukey's post-hoc tests that identify the aggregate and concrete relations among the measured units. The results of the analysis show that Vietnamese-accented English maintains comparable voicing contrast in its initial stressed plosives as a native English accent does. The average Voice Onset Times values of the lenis stops without prevoicing shows to be slightly higher, while the average values of voiced initial stops prove to be similar or close to similar those produced by American English (AmE) speakers, which we assign to the fact that pre-voicing in Vietnamese exhibits strikingly similar values. The values for fortis initial plosives showed to be higher due to such quality typical for Vietnamese language, where the onset of voicing takes longer.

The thesis analyses how Vietnamese speakers of English have developed the notion of voicing contrast, as their mother tongue employs richer means of voice contrasting than English requires. The results we achieved prove that Vietnamese speakers of English do not struggle maintaining 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. V následující kapitole jsme také představili několik vybraných teorií, jejichž cílem je zobecnit hlavní tendence probíhající při osvojování cizího jazyka. Na závěr teoretického úvodu se dostaneme k Vietnamštině a k podstatě této práce - Vietnamské angličtině s ohledem k počátečním konsonantům. Samotné analýze předchází popis metody, který poskytuje informace o námi nahraných Vietnamcích mluvících anglicky, o postupu nahrávání a zpracování dat.

Tabulky a grafy ilustrují statistické výpočty provedené 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 srovnatelný znělostní kontrast jako rodilá angličtina. Průměrné hodnoty doby nástupu znělosti pro znělé plozivy bez znělosti v závěru byly naměřeny mírně vyšší, zatímco hodnoty pro znělé plozivy s přítomností znělosti v závěru vykazují hodnoty téměř identické v porovnání s rodilými mluvčími americké angličtiny. Tuto shodu připisujeme průměrným hodnotám znělosti v závěru, které má vietnamština taktéž obdobné. Hodnoty pro počáteční neznělé plozivy se ukázaly být vyšší, což se zdá byt opět důsledkem vietnamštiny, kde je doba nástupu znělosti delší.

Hlavním smyslem práce je analyzovat, jak si vietnamští mluvčí angličtiny rozvinuli schopnost vytvářet znělostní kontrast v závislosti na své mateřštině, která znělostní kontrast využívá ještě bohatěji než angličtina. Dosažené výsledky potvrzují náš předpoklad, že Vietnamci mluvící anglicky netrpí nedostatkem v užívání kontrastu znělosti, který je využit v řeči rodilých mluvčích angličtiny.

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

Table of Contents

List of Abbreviations 9
1 Introduction 10
2 Theoretical Background 11
2.1 English Consonants 11
2.1.1 Plosives 12
2.2 Distribution 13
2.3 Voicing (Phonological) 13
2.4 Voicing (Phonetic) 13
2.4.1 Devoicing 13
2.4.2 Prevoicing 14
2.4.3 Aspiration 15
2.5 VOT 17
2.5.1 Measuring VOT 18
2.5.2 Effects on VOT 19
2.5.3 The Recognition of VOT 20
2.5.4 Voicing Aspiration Correlate 20
2.5.5 VOT Drawbacks 21
2.5.6 VOT in Foreign Accent 22
2.6 Second Language Speech Acquisition 22
2.6.1 VOT in Second Language Acquisition 23
2.6.2 The Theory of Equivalence Classification 23
2.6.3 Foreign Accent 24
2.6.4 Markedness Theory/ Hypothesis 24
2.6.5 Vietnamese speakers of English as L2 25
2.7 Vietnamese and Vietnamese English 26
2.7.1 Syllable Initial Consonants 26
3 Methodology 29
3.1 Informants / Speakers 29
3.1.1 Selection of Speakers 29
3.1.2 Gender 29
3.1.3 Origin Language Acquisition 29
3.1.4 Age 30
3.2 Corpus Recording 30

3.2.1 Corpus / Data Processing 31
3.2.2 Forced Alignment 31
3.2.3 Manual Selection of Obstruents 32
3.3 Data Extraction 33
3.4 Hypotheses 33
4 Analysis 35
4.1 Ratio: VOT, Voicing 35
4.2 Plosives & Sex 37
4.3 Plosive & Origin 38
4.3.1 Plosive & Origin, Sex 40
4.4 Plosives Preceded by /s/ 40
4.5 Frequency of Prevoicing Occurrence 41
5 Discussion 44
6 Conclusion 46
Bibliography 47

List of Abbreviations

AmE American 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 intends to describe the voicing contrast in Vietnamese English and eventually compare the acquired data with a native accent of English. Sizeable numbers of American researchers point to the issues surrounding Vietnamese learners of English whose pronunciation appears to be one of the least comprehensible, as they have vocabulary and grammar that cannot be orally communicated in an efficient manner. This excerpt that appears in a book published by Zampini in 2008 party motivated this research

> *Occasionally, large ESL programs offer courses tailored to a particular L1group. 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 2008: 356)

This work may be of partial relevance and benefit to L2 pedagogy, as recent studies seem to be very much interested in the way non-native phonetic contrasts are presented. This thesis was initiated having been inspired by such studies in mind. The following paper argues that L1 strongly affects successful learning of such contrasts and explains variation found in production.

2 Theoretical Background

This theoretical background is composed to provide a base for the main preoccupation of this study, which is voicing contrast in Vietnamese English. Firstly, the native AmE accent will serve as a referential accent for this study therefore this chapter introduces English consonants, and the relative phenomena of voicing, aspiration, and mainly Voice onset time (VOT). Following this, a section dealing with second language acquisition (L2), introduces several theories that aim at generalizing the recurring phenomena in the acquisition of L2. Finally, a section looking into Vietnamese English with regard to the syllable-initial consonants that are, for the sake of this study, the main focus of our interest.

The main intent of this paper is voicing contrast in Vietnamese English since it is a crucial phonetic feature that aids overall comprehension. Therefore, the delay of onset of voicing will mainly serve as a means of describing Vietnamese English plosives in stressed position.

Before all it should be noted that the studies dealing with voicing contrast of obstruents are commonly based on the observation of stops. Also, phonological systems often utilize the measurement of aspiration as a means of differentiating and categorizing obstruents in a variety of languages. (Skarnitzl, R. 2011: 64) The dimension of the onset of voicing has been used as a measurable counterpart to the two more controversial and difficult to quantify discrimination elements. These are firstly the dimension of articulatory force, which looks into fortis and lenis division and takes into account the state of the glottis, and the amount of respiratory and muscular tension (Skarnitzl, 2011: 72). The second element deals with voiced and voiceless division describing the vibration of the vocal cords. (Lisker, 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. Therefore this paper shall employ the dimension of VOT (voice onset time) to describe the quality of Vietnamese English consonants.

2.1 English Consonants

English consonants are traditionally described with regard to their place and manner of articulation. The place of articulation means the contact between the upper and lower lip, between the active articulator tongue and the passive articulator that is the roof of the mouth or with glottis itself. The occlusion - exclusive of the bilabial and glottal - is made with the

tip, blade, or back of the tongue pressing against one of the areas on the roof of the mouth - namely dental, alveolar, post-alveolar, palatal or velar area.

English phonetic inventory employs twenty four consonants. Stop consonants with regard to the manner of articulation include the six following: /p, b/, /t, d/, /k, g/, we can also include here a glottal stop /ʔ/. Nine are fricative /f, v/, /θ, ð/, /s, z/, /ʃ, ʒ/ and /h/, two affricates /tʃ, dʒ / three nasal sounds /m/, /n/, /ŋ/ and four approximants /w, l, r, j/.

The manner of articulation does not consider parts of the mouth employed in the production of consonants, but how the speaker produces them. In other words the manner of articulation focuses on the acoustic qualities of the closure between the two articulators. To illustrate it in an example, when we produce approximants, e.g. /l, r, w/, the articulators do not create such a turbulent airflow; they are closer to vowels, because of their relatively high sonority. The articulators are wider apart, however, not too wide to create a vowel; and, at the same time not close 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 therefore the result is a high pitched hissing sound. With a stop consonant /t, d/, the closure is complete and the air cannot escape from the vocal tract.

2.1.1 Plosives

As English stop consonants make a so-called ‘plosion’ sound they are also often labelled as obstruents or plosives. The variety of denominations stems from the separate point of views 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; therefore mark them as stops or obstruents. The term ‘plosives’ brings into focus another quality and that is their auditory quality. English plosives are produced in four phases: closing, compression, release and post-release phase. The closing phase is when the articulators move to create a stricture behind which is accumulated compressed air waiting to be released. The opening of the articulators is called *release* and this moment is accompanied by a perceptible plosion that is the reason for this labelling. Right after the release phase follows the post-release phase, which indicates what happens right after release. English obstruents are found in pairs comprising of voiced and voiceless phonemes, also called lenis and fortis. Such as the pair /t, d/, where the left phoneme is always voiceless.

2.2 Distribution

In relation to syllable positions, in English there is little restriction, all six plosives occur initially, medially and also finally.

2.3 Voicing (Phonological)

Having come upon the occurrence of voicing in the phonemic categories, i.e. voiced /b, d, g/ and voiceless /p, t, k/, which identifies these two sets of obstruents, the phenomenon as such will be briefly explained. Firstly voicing can be referred to as a classification of speech sounds. In general the set of voiceless /p, t, k/ in English are produced with no vocal cords vibration, therefore they are classified as voiceless, while /b, d, g/ as voiced, because the vocal cords vibrate when we pronounce, for instance, a sequence of /d/ sounds produced in a row. However, as phonetic voicing indicates the voiced stops do not need to be necessarily voiced on the articulatory level.

2.4 Voicing (Phonetic)

Secondly voicing can also refer to the articulatory process, for instance when a speaker pronounces /d/, in which vocal folds vibrate. Nevertheless if we look at the closing phase /b, d, g/, in general, it manifests very little voicing during the closure; the voicing only starts before the release. The possibility for /b, d, g/ to be fully voiced even during the entire closure and compression phase is yet possible, if the speaker pronounces them slowly and carefully. In contrast, if produced in rapid speech, they may have very little or no voicing at all.

As (Volin, 2002: 73) points out, most of the English voiced stop consonants keep their voicing only when occupying an intervocalic position.

2.4.1 Devoicing

Hand in hand with voicing for English is typical another acoustic phenomenon called devoicing - typically in English lenis obstruents. Plosives /b, d, g/ are in word initial or final positions partially or fully devoiced. English voiced consonants maintain their voicing only when they occur in intervocalic position inside morphemes. In all the other positions their voicing is partly or wholly lost.

Considering the words pie or buy both bilabial obstruents appear to be voiceless. (Ladefoged, 1991: 57) points out that most people perform very little voicing when lips are pressed together while producing these words. Therefore the major distinction may be asserted to rest on the amount of VOT.

We can exemplify devoicing in initial and final voiced plosives in words: *budget* [ˈb̥ʌdʒɪt] and *load* [ləʊd̥]. However, although these initial plosives lose their voicing, they still maintain their lenis character as for instance they can never exhibit aspiration. The Figure 2.1 bellow shows no presence of periodical pulsing before the release of the voiced obstruent /b/.

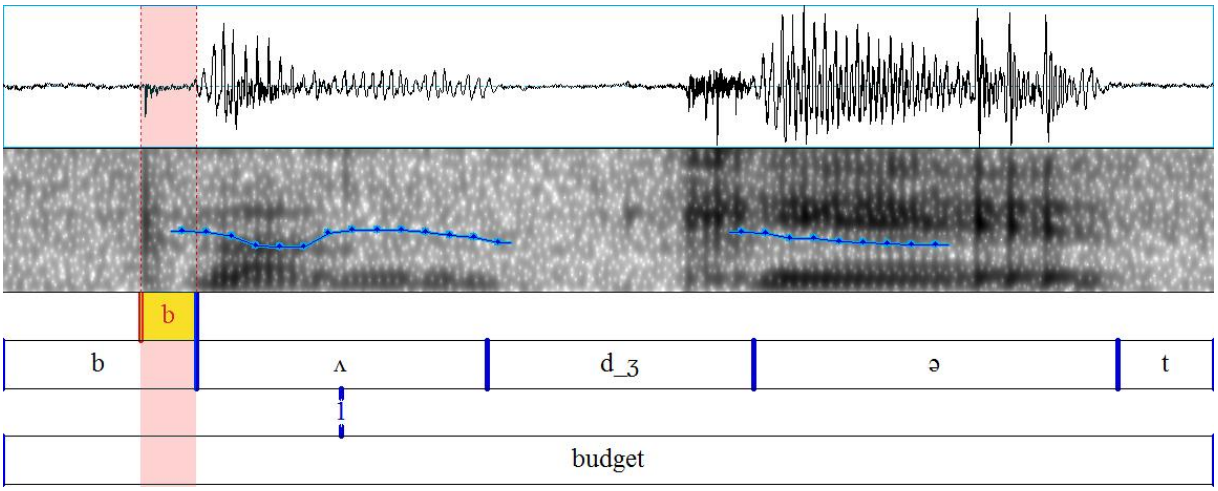

Fig. 2.1. – Devoicing

2.4.2 Prevoicing

One more phenomenon called *prevoicing* shall be introduced. Prevoicing takes place before the release of the fortis obstruent; therefore it is dealt with before the discussion of aspiration. In some languages, as in English, initial voiced plosives are produced with prevoicing. Referring to the four phases, which generate obstruents, the occurrence of prevoicing presents itself when the voicing develops during the closure of the initial plosive before the release.

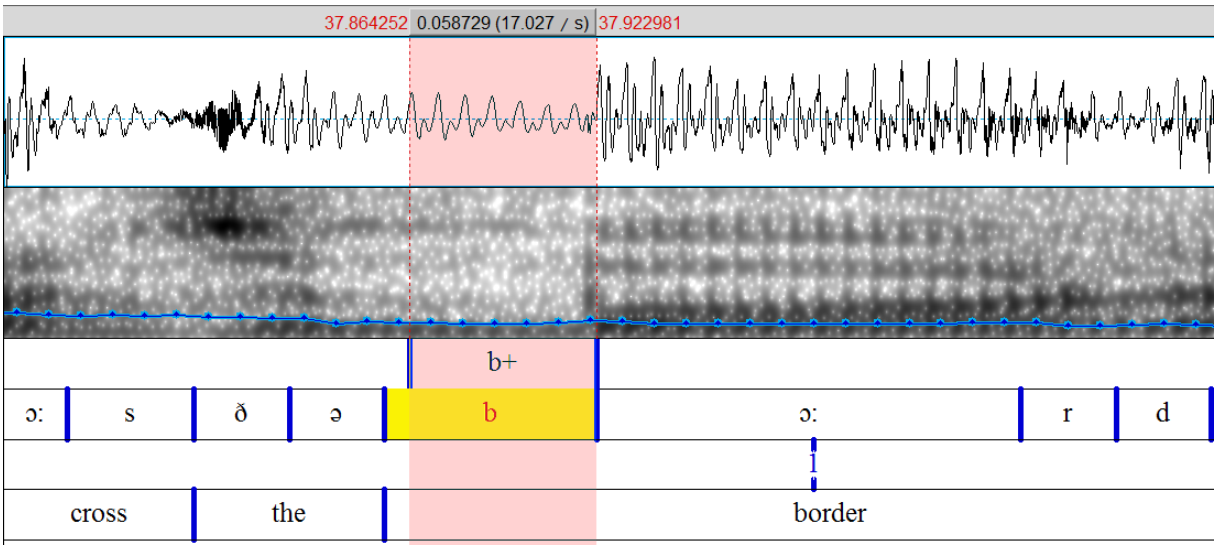

Fig. 2.2. – Prevoicing

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 - voicing lead that is acquired by the fully voiced /b, g, k/. The acoustic quality that distinguishes the voiceless counterparts /p, t, k/ to this set of stops is labelled as aspiration. The Voice onset time may be called a unit of *aspiration* as it is the duration of the period of time when the aspiration takes place. Both phenomena shall be introduced and elaborated on in the following passage.

2.4.3 Aspiration

The topic of obstruents will be carried forward together using the acoustic point of view**.** As has already been explained obstruents are produced in four phases. The third phase the release of these voiceless stops results in an audible plosion that is characterized as a burst of noise. Then there is a post release phase, when the articulators are apart, however, English voiceless stops /p, t, k/ manifest no immediate onset of voicing. For this moment is typical that the air escapes through the vocal cords, making a sound of a phonetically pronounced [h] (Roach, 2009: 27). This audible noise is called aspiration (below viz. Fig 2.3) demonstrated on a sound wave and a spectrogram.

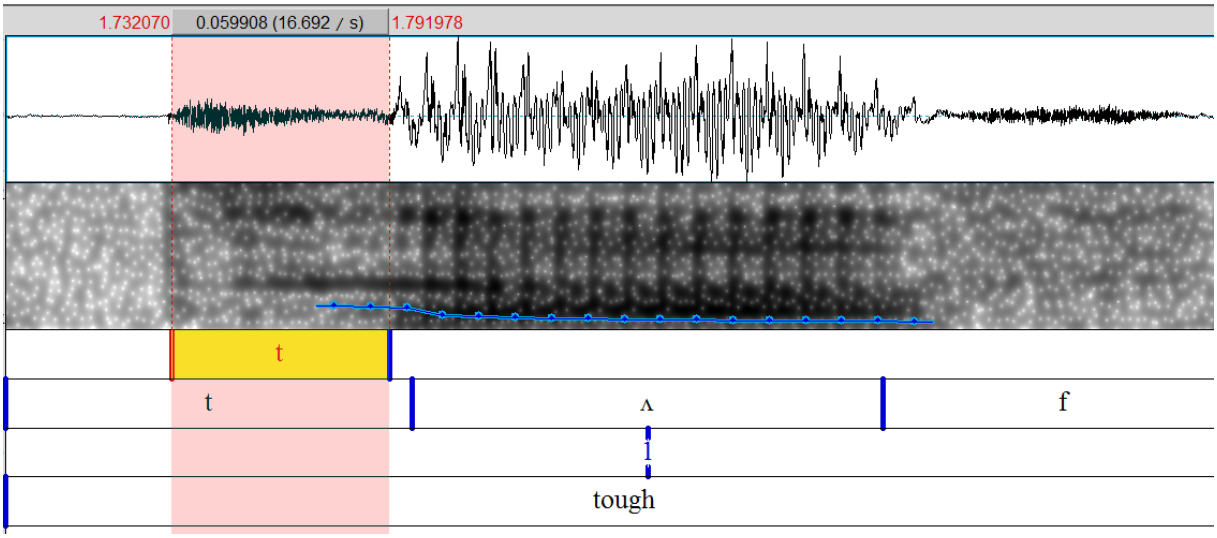

Fig. 2.3. – Aspiration

After the post-release phase, which is visible on the sound wave as periodic pulsing, vocal cords come together and the voicing of the following sound starts.

Considering the acoustics, aspiration is characterized by turbulent glottal noise with delayed onset of vocalic tone. (Ladefoged, 1991: 9) the noise feature of aspiration should be regarded then as an automatic accompanying feature to the large delay of voicing (Lisker, 1964: 387). In the 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/, in phonetics symbolized as [ʰ].

In contrast the release of the voiced set of plosives is followed by weak plosion as (Roach, 2009: 22) explains. According to Roach, it happens at the same time or even after the onset of voicing. For our purpose, the most noticeable acoustic difference for the set of voiced and voiceless obstruents is then the aspiration of /p, t, k/.

Considering the articulatory gestures, the vocal cords maintain an open position right after the release of the stop while producing the aspirated phonemes, while they are in a phonation posture when producing the unaspirated phonemes. (Dixit, 1997: 87) explains that for the production of aspirated sound there needs to be an open glottis and vocal tract without impediment.

2.5 VOT

To begin with, aspiration, when considered with reference to time (period of voicelessness) is designated by phoneticians as VOT. In the sixties the first phonetic indicator was introduced that was able to distinguish voiced, voiceless unaspirated and voiceless aspirated plosives implemented as Voice Onset Time (VOT) (Skarnitzl, 2011: 66). VOT, when concerned with reference to time (period of voicelessness), is an acoustic manifestation of Aspiration, Devoicing, and Prevoicing. The aforementioned voicing and aspiration both depend on the timing of articulatory gestures: the vocal folds gesture and the oral articulator gesture. Voice onset time is enclosed between the release of the fortis obstruent and the beginning of voicing for the following sound.

We shall focus now on voicing that starts after the release. The voicing that starts before the stop is released has already been discussed. For English voiceless obstruents, in initial stressed position, is typical the already introduced aspiration. Long positive VOT values are characteristic for aspiration. However, there is also short VOT; or zero VOT.

Positive VOT may be also referred to as voicing lag, which is measurable in stressed lenis obstruents occupying the onset of a syllable or medially before a stressed syllable, as for instance /t/ in *towel*, /k/ in *become* [bɪkʰʌm]*;* it never appears in a coda. This, however, excludes the position when preceded by the /s/ phoneme) as in *skip* [skɪp].

In order to obtain strong aspiration measured as longer positive VOT, /p, t, k / needs to occupy the onset of a stressed syllable and needs to be followed by a stressed vowel. As vowels are the core of the syllabic stress they are one of the essential circumstances. The instance when a lenis aspirated obstruent is accompanied by weak aspiration (short VOT) can be illustrated on a word collide [kʰəlaɪd].

The short VOT is a result of the placement of the lenis stop within an unstressed syllable as in the word competitor [kəm'pʰhetitə]. It is important to notice that the stress causes a certain exaggeration of aspiration.

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 appear at the same time, 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 inaudible and is close to zero.

The labelling of 'aspirated' and 'unaspirated' may give the impression that there are just two extremes, aspirated and unaspirated that is, however, not true. Aspirated and unaspirated is a continuum, it is a smooth scale between the two categories. The scale is the launch of VOT therefore one should not think strictly categorically, although the intention is to categorise the two set of stops.

2.5.1 Measuring VOT

To be able to specify the exact aspiration definition, we need to look into how VOT is measured in *ms*. When the gauged figure is less than 20 ms, it is generally considered to be an unaspirated stop (Ladefoged, 2010: 151). As the terms *voicing lag* and *voicing lead* have already been introduced it remains to indicate their acoustic span/duration. Long lag obstruents are accompanied by long VOT duration ranging between (0-35 ms.), while their counterpart, pre-voiced stops manifest voicing throughout the closure and therefore they display (if we regard oscillogram or spectrogram) negative VOT (Zampini, 2008: 221).

Oscillogram with its unambiguous time indication is considered to be more suitable for the measurement of VOT than spectrogram; however, some phoneticians proceed from spectrograms. The spectrogram analysis might, for instance, look at the start of aspiration as the outset of vertical strokes starting from F_2 (2^{nd} formant) and higher, other prefer the onset of F_0. Although we can note that there are systematic differences between the two types of measurement, which range approximately between 5-10 ms, those are, however, regarded as negligible and the two methods are considered as maintaining a very high correlate.

Among the less common methods can be listed, as an example, the measurement of duration of high frequency/pitch noise.

Although there are various approaches in measuring VOT, most commonly it is measured with the start of the first burst. This is useful to clarify especially when a voiceless obstruent such as /k/ is often accompanied by multiple bursts. The end of aspiration is considered to be the onset of voicing that is defined to start with the commencement of the first complete period following right after the noise output - burst or aspiration.

The positive figures of VOT can, however, be two periods shorter than the values identifying the end of aspiration. The difference between the onset of voicing – periodic oscillation of vocal cords – on one the hand and the onset of F2 on the other hand can comprise of 46 ms difference (Skarnitzl, 2011: 83-86).

2.5.2 Effects on VOT

It is important to notice that syllable stress causes a certain exaggeration of aspiration and the values of VOT are therefore higher. There are however other factors that have an impact on the values of VOT that do not relate to stressed or unstressed position within a word. The most recent ones are mentioned in the following studies.

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

Furthermore the place of articulation also proved to have an impact on VOT as the figure of VOT changes together with the place of articulation; namely, velar and uvular explosives proved to have the highest figures (Lisker, 1964: 387).

One of the possible factors impacting the VOT values are the general principles of aerodynamics. When producing the velar stop the oral cavity has a lower volume of air compressed behind the closure if we are looking in the direction of the exhaled air. The result is the lower the spatial volume the higher the pressure that raises during the occlusion. The eventual release of this compressed air lasts longer than the release of bilabial or alveolar plosives. From the opposite point of view after the velar release there is a greater amount of air waiting behind the occlusion to be moved forward that also lasts longer.

As the second influencing factor (Cho and Ladefoged, 1999: 209-212) subsume the agility of the tongue, where tip and blade are generally assumed to be the fastest and most flexible in contrast to the back of the tongue. Therefore if the back creates the occlusion we can assume the whole process to be slower and take longer, which increases the VOT values.

Among other influencing factors we can also mention the place of contact of the passive and active articulator during the closure, where the release slowing down of the closure between these two may be caused by the interaction between intraoral pressure and the mechanic

pliability of the surface (Stevens, 1998: 328). Cho and Ladefoged also mention a slower closing of glottis when the velar sound is produced. They explain this as a result of slower decreasing of intraoral pressure as plosives articulated at the front of oral cavity, which is closer to the teeth, are accompanied with intraoral pressure that is decreased much faster.

In brief, (Cho and Ladefoged, 1999) after the observation of eighteen different languages conclude that a generally valid statement asserting higher values of VOT for consonants with the articulation placed further back in the oral cavity is not possible to make. For this study it is essential to mention their conclusion stating that although the VOT values stem from the aforementioned physiological and aerodynamic processes, mentioned above, different languages still employ language specific values of VOT.

2.5.3 The Recognition of VOT

Starting from the broad perspective, before introducing the principles of (voicing contrast in) Vietnamese English, we need to mention that stop consonants in different languages employ distinct voice onset time durations. English employs two way phonemic distinction /p, t, k/ vs. /b, d, g/ these, as we already know, differ by the phonetic realization such as long-lag stops and short-lag stops. The duration of VOT differentiates them. With some languages the situation may be different and even the voiced set of plosives may be characterized by a long VOT. To illustrate, in Hindi and Sanskrit breathy-voiced consonants are often called voiced aspirated. Owing to such a variation across different languages it is worth looking at the various distributions of VOT that help to test the native and non-native figures/duration (Zampini 2008: 221).

2.5.4 Voicing Aspiration Correlate

As stated above, aspiration/VOT is enclosed between the release of a stop and the voicing of the following sound. Therefore, to draw a complete picture of how voicing 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 in the case when the spectrogram shows a presence of voicing the aspiration noise is absent or obscured. The same happens when the aspiration noise is strongly present. In this case the periodic pulsing is usually not

distinguishable. The noise feature of aspiration is regarded as automatic concomitant of a large delay in voice onset. The phonetic dimension of voicing, the primary glottal mechanism that controls it is also responsible for producing part of the features that are considered to be acoustics phenomena of aspiration and articulatory force (Lisker, 1964: 386-8). Moreover, voiceless stops manifest greater glottal opening in stressed syllables than in unstressed ones. (Chasaide, 1987: 28-31)

2.5.5 VOT Drawbacks

It should be taken into account that VOT is not suitable for formulizing the temporal relations in regard to voiced aspirated phones. Also the differentiations of voiced unaspirated and voiced aspirated obstruents prove to be a problematic aspect of the classification of obstruents when measured by VOT. Among others it can be mentioned for instance the fact that it is not possible to talk about voice onset time when the obstruent is in intervocalic position (Skarnitzl, 2011: 88).

Mikuteit (2007) in his research addresses problems of VOT measurements and introduces alternative measurement of lag times called *After Closure Time* (ACT) that he utilizes in describing four-category language East Bengali. East Bengali displays a four-way contrast of voiced/ voiceless and aspirated/ unaspirated oral stops and affricates in all word positions. Mikuteit asserts that VOT cannot sufficiently distinguish between the four categories, in languages that display for-way contrasting. In these languages only the voiceless aspirated stops can be distinguished from the voiceless unaspirated ones – negative VOT of the voiced aspirated and unaspirated stops in his view do not show such an evident distinction. According to his research this new way of measuring can put the end to the extra notion of ‘breathy voice’, which we already know is typical for voiced aspirated stops. Therefore he introduces an additional term called SA – *Superimposed Aspiration* that is a breathy part of the vowel that follows voiced aspirated stop also known as breathy voice.

The results that combine SA and ATC prove that aspiration measured from the point of release is timed equally for voiced and voiceless obstruents. ACT in this study proves to be able to distinguish those voiced aspirated and unaspirated stops; therefore there is no need to search for the presence of breathy voice anymore. Aspiration contrast in both the voiced and voiceless category could be distinguished by ACT. For voiceless consonants the ACT part appears to be long and the SA part short while the reverse proves to be the instance of the voiced stops.

This research argues that the reason for aspiration, not being identical across languages, is due to aspiration being the manifestation of various underlying phonological functions; however, this claim needs to undergo more research in order to be confirmed.

2.5.6 VOT in Foreign Accent

So far language universal processes have been discussed in regard to aspiration; however, the purpose of this study is to focus on language specific processes, which take part in forming special criteria for its stop consonant discrimination. Noticeably, there is a continuum of possible voice onset times and various languages favour different points along this continuum in order to create oppositions among their stop consonants (Ladefoged, 1993: 144). Moreover, latest studies have proven that VOT not only shows differences among diverse languages but also among dialects or foreign accents (Hansen, 2010: 777-789). Interestingly, the onset of voicing has been also studied as a criterion differentiating individual speakers (Skarnitzl, 2011:82). Therefore suprasegmental properties of L2 speech, such as Voice Onset Time (VOT) seem to be a reliable and accessible source for collecting data contrasting not only obstruents. Considering the fact that the role of Ll in the perception of L2 has been observed on prosodic level this study intends to compare the data between English standard and Vietnamese speakers of English.

2.6 Second Language Speech Acquisition

Since our target group of speakers are second language learners, the next section is devoted to the influence of native language on foreign language. The influence of foreign language learning is most obvious in speech in comparison with, for instance, grammatical or lexical structure (James, 1996: 285). Therefore we are going to shortly comment on SLA as we expect this influence to be projected also in the VOT values.

L1 plays an influential role in production of the newly acquired language. To what extent L1 of beginning learners influences their perception and production of L2 syllables and how much of those is actually transferred from L1 to L2 has been enquired into and elaborated on in a large number of studies. What challenges the researches is the possibility to find universal constraints on syllable structure. These are presupposed to occur in longer and more complex words, in onsets and codas. (Zampini 2008: 5) There are perceptual differences between adults and children, as children do not have all the phonological categories of L1 developed. This research will deal particularly with adult speakers; therefore any theoretical background

regarding perception of children will be omitted. Among the influencing factors that constrain the L2 production in an elementary speaker we include not only linguistic but also non - linguistic aspects. Among those are namely:

> *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, 2008: 269)

Most of the studies in (Zampini 2008: 270) claim that if the second language acquisition takes place before the age of six the speakers are likely to produce native-like values of VOT, while those who learnt L2 later tend to produce compromised durations somewhere between L1 and L2.

2.6.1 VOT in Second Language Acquisition

A significant body of research findings reveal that the cross-language perception and production of VOT serve as a sufficient acoustic impulse to distinguish initial stop consonants in a variety of languages. Listeners who were tested on differentiating stop consonants tend to separate a continuum of VOT varying stimuli into categories matching the system of consonants of their L1. The results also showed that adults with very little proficiency of L2 have been found to be most sensitive to VOT differences. (James, 1996: 274).

2.6.2 The Theory of Equivalence Classification

This theory suggests that every human acquires a basic cognitive mechanism that helps him/her to perceive sounds and classify them. Owing to the physical exemplars of a certain category that the listener has had chance to hear since s/he was born the listener creates slots of sounds in his/her consciousness. Therefore if there is a sound similar to the one s/he stored as for instance /t/ s/he will classify it in a slot of /t/ although the sound may not be completely the same. This is called equating of the L1 and L2 sounds (Zampini 2008: 223)

This is namely important, as it can be expected that the group of Vietnamese speakers, whose speech we are examining, will classify English aspirated plosives around their aspirated voiceless stop /t^h/ that is present in Vietnamese.

2.6.3 Foreign Accent

The above brings us to point out that research in the domain of accents has been investigating the acoustic properties in the speech of L2 learners comparing their production of certain sounds with the one produced by a monolingual native speaker.

To summarize, stop consonants showing VOT values produced by L2 speakers reveals that even experienced speakers of L2 tend to compromise VOT values. In other words the range of VOT was neither of L1 nor of L2 property. Those who manage to establish separate phonetic categories for L1 and L2 tend to be only early learners until they reach the age of six (Zampini 2008, 223). Later learners most probably inhabit equivalence classification; they tend to transfer their L1 VOT values in the production of L2 obstruents (Zampini 2008, 42).

Importantly, considering the fact that aspiration is audible even to the beginner learners of L2 it should be regarded as an important feature constituting accent of any kind. Hansen (2010) elaborates on this in his research of automatic detection of foreign accent using VOT values in unvoiced stops. Our study, will compare VOT values of Vietnamese accented English in order to possibly reveal Vietnamese accented specialities in English.

2.6.4 Markedness Theory/ Hypothesis

The markedness theory argues that not all the differences that are held between L1 and L2 will cause the same difficulty for an elementary speaker of L2. Typological markedness holds a significant position in the research explaining the facts about L2 phonology (Zampini 2008: 112). Firstly the Markedness hypothesis and the status of marked and unmarked will be introduced to provide the necessary background. Afterwards it will be illustrated in several examples in order to clarify the hypothesis.

The markedness hypothesis and the principle of markedness was first introduced and elaborated on by the Prague School of Linguistics. Their pioneering figures were Nikolai Trubetzkoy and Roman Jakobson. The concept is that binary oppositions between certain linguistic representations, as for instance voiced and voiceless obstruents, had not been understood to be simply polar opposites. One element of the opposition was always made believe to be in a way special, since it had a wider distribution as within a given language and also across languages. In order to characterize this special status the markedness value had been imposed on this opposition, since the element that was more distributed than the other of

the opposition was labelled as unmarked. The label unmarked also included a characteristic of simpler, more basic or more natural - than the marked member of the opposition. According to this classification the voiceless obstruents can be called – relative to voiced obstruents - unmarked. Markedness hypothesis (MDH) insinuates that markedness can anticipate some degree of difficulty an L2 speaker may face when learning individual L2 structures. Therefore, in other words typological markedness can, to some extent, explain the difficulty that the learners are confronted with. (Zampini 2008: 96)

One can assume that our speakers may suffer difficulties when trying to produce aspirated sounds since these are according to the theory marked. The constructions related to other representations by markedness principles are predicted to cause learning problems. The degree of difficulty involved is predicted to correspond directly to the relative degree of markedness (Zampini 2008: 98).

It is worth noting that not only the transfer of L1 to L2 production has a great impact on the accented speech of L2 learner. Factors such as ability to perceive foreign sounds, the particular differences/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. A large number of studies have been preoccupied with L1 - L2 nuances in quality and pronunciation that occur in consonant clusters in onsets and codas.

To conclude, judging from the above it appears that some speakers encounter greater troubles since they have to learn more marked structures than the others. However, since our informants are acquainted with aspiration in their L1, they do not need to learn a new marked structure, the production of aspiration should, therefore, not cause any issues.

2.6.5 Vietnamese speakers of English as L2

To illustrate what kind of speech production mistakes a Vietnamese speaker may make in English, we can mention one of the studies Zampini looked into which focused on L2 acquisition by the Vietnamese. The Vietnamese speaking English struggled with consonant clusters in initial and final positions. The experiment tested the production of syllable-initial clusters and syllable-final clusters, while the former was found to correspond with the markedness hypothesis the latter- syllable-final clusters - resulted in ceiling effects. The study subscribes this deviation to the high level of English that the tested speakers exhibited (Zampini 2008, 100).

2.7 Vietnamese and Vietnamese English

The above sections dealt with voicing and aspiration in general; together with the second language acquisition. The following section is going to look into Vietnamese English with regard to initial consonants as those are the major topic for the L2 group.

2.7.1 Syllable Initial Consonants

As has been stated this thesis is concerned with voicing contrast in Vietnamese English as it is an essential element without which there is a high occurrence of confusion between stops. Such mixing up of voicing contrast leads to misunderstanding and communication lapses.
One such instance can be illustrated by contrasting two words: (green vs. cream). The amount of VOT matters significantly since the Vietnamese accented English is typical for 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 one consonant coda can be deleted in Vietnamese English. The word green produced by Vietnamese speakers was perceived and misinterpreted by listeners as cream. The problem rested on them producing 112 ms VOT after the release of /g/ that was pronounced as voiceless aspirated stop (Cunningham, 2009). Considering the fact that Vietnamese speakers are likely to eliminate consonant clusters at the end of words together with the misleading amount of VOT causes great problems in keeping essential contrast not only between similar words.

2.7.1.1 Distribution of Consonants

In contrast to English that has very little restrictions in consonants placed in the initial position, Vietnamese is very specific in distribution of its consonants. Although it is not as strict as with regard to word final position, the limited distribution of consonants should be mentioned and taken into account. If we should compare the distribution of consonant in the initial and final positions we can notice a significant difference. The initial position, as it is with English, can occupy any consonant of the Vietnamese set of consonants. Those are labial / ɓ, m, w/, labio dental /t, tʰ, n, l/, alveolar / ɗ, s, z/, palatal /tɕ, ɲ/, velar /k, ŋ, x, ɣ /, glottal /ʔ, h/. In Vietnamese the initial /p, j, r/ appears very infrequently and is considered to be primarily French of origin - appearing in loanwords. It is important to consider that many speakers realize /p/ as /b/ (Kirby, 2011).

With regard to consonant clusters only /ʈʂ/, /kw/, can appear in initial position. With the final position the rules are stricter - there are no consonant clusters at all and the phonemes occupying the position are restricted to a set of /m, n, p, t, k, ŋ/, / i̯, u̯/ semi vowels, sometimes transcribed as /w, j/ (Hoang Thi Quynh Hoa, 1965: 45)

2.7.1.2 Plosives

Speaking of 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 the second language acquisition, as for instance, 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/, /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 sounded as [ə ‘ʔɓæd mæn].

Further contrasting indicates that Vietnamese does not have the /g/ phoneme, although it seems that a large number of speakers have it as an allophone. In Vietnamese phonetic inventory we can find a voiced velar fricative [ɣ] instead that is most likely to be used by Vietnamese speaking English to fill the slots of the English /g/. In Vietnamese English word final /g/ is extremely unlikely, because the devoicing that in English happens at the end of words the unvoiced [x] is more probable to occur. This can be illustrated on the example of *big* [bɪg̊] x [bɪx].

Coronal consonants that are pronounced with the tip or the blade of the tongue are also Vietnamese accent specific. (Singer, 2012) points out that some consonants such as: [t̪ʰ, l̪, s̪, z̪] are shaped dentally, others laminally, using the blade of the tongue with the tip down behind the bottom teeth that can be demonstrated like so: [s̻,z̻,ɕ̻]. In the area of Ho Chi Minh City, located in the very south of Vietnam two retroflex consonants can be found that (Singer, 2012) characterizes as produced with a tongue-tip curling back [ʂ, ʈʂ]. Vietnamese consonants coronal consonants seem to exhibit a lot of complexity, therefore a great amount of attention should be paid to the consonants pronounced by each English speaking Vietnamese.

2.7.1.3 Plosives and Aspiration

Following the Vietnamese phonotactics stressed aspirated [pʰ, kʰ] produced by native speaker of English tend to be unaspirated when produced by a Vietnamese speaking English. This is characteristic even for /p, k/ in the initial stressed position. There is one allophone that may be similar to English aspirated [tʰ] and that is Vietnamese aspirated and dental [t̪ʰ]. Moreover [t̪ʰ] is contrastive to English [t̄] that is unaspirated alveolar (Singer, 2012). The voiced stops are often produced as implosives as for instance the initial /t, tʰ/ are realized as apico - dental [t̪, t̪ʰ], lamino alveolar /t̻ t̻ʰ/ (Kirby, 2011).

2.7.1.4 Voicing Contrast in Vietnamese (Aspiration, Devoicing, Prevoicing)

Unlike English, Vietnamese language employs a three way voicing contrast between coronal stops: pre-voiced *đ*, voiceless unaspirated *t*, and voiceless aspirated *th*. The pre-voiced *đ* is distinguished by a long negative VOT. Slightly positive VOT is 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 pre-voiced /d/ and voiceless /t/.(Twist, Shamoo, Bauman, 2009: 4).

3 Methodology

3.1 Informants / Speakers

In this kind of research, the primary deal of attention had to be paid to the selection of speakers. Their advanced command of spoken English was crucial. Secondly, in order to provide a representative group of speakers it was favourable to acquire material that represents in proportion both males and females. For the purpose of this study 15 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 provide us with one hour of recorded material.

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 B2 (of CEFR) and higher. The participants of this research were university students following either bachelor or master degree in the Czech Republic. As all of them, except for one, have not been resident in the Czech Republic longer 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 pathway taught in the English language.

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 number of both sexes as not to obtain gender biased sample.

3.1.3 Origin Language Acquisition

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

The informants were of majorly unique vernacular origin, i.e. Hanoi dialect located in the Northern part of Vietnam. The only exception concerned speakers M6S and M7S, who grew up in South Vietnam. Following this the corpus might have benefited from its proportional regional specificity, which is the Northern Vietnamese regional language.

As for the nature of the command of English, the informants' second English language acquisition took place in authentic in Vietnamese language teaching environment. All the speakers took language courses instructed by Vietnamese teachers of English. Moreover, none of the study participants, except for two, had been exposed to a native English speaking environment.

3.1.4 Age

We could assert the sample is age-wise proportional since the participants were aged mostly between 20 - 23, and the average age was of 22.4. The participants were chosen mostly based on their English proficiency and willingness to participate in the research, most of the indulgence to participate in the research came from the university students.

Signage	***Gender***	***Age***	***Origin***	***Length of the residence in CR/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	2	no

Table 3.1. – Table of informants

3.2 Corpus Recording

Due to the workload of the informants all the recordings had to be accommodated to the informants' time and location availability. Therefore the regular recording room facilities at

the Department of Charles University could not be used. Consequently, the most plausible solution occurred to be a passenger car, where all the recordings were made in the front seats. The vehicle was parked in a calm and quiet area. In order to prevent the sound waves from deflecting on the dashboard and creating an undesirable echo, the dashboard was fully covered with a large piece of woollen material, so the recorder itself was positioned on a similar echo absorbing material. The informants were asked to read BBC news extracts - each text was approximately 500 words long.

The corpus was obtained using the portable recorder Edirol HR-09, with 48 kHz sampling rate. 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. After they had finished they were asked to read the first couple of sentences so that the input of the recorder could be adjusted according to their voice volume.

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

3.2.1 Corpus / Data Processing

Although the input of the recorder was adjusted to the speaker's voice volume, a few of the respondents still produced material that needed intensification of the volume. For this purpose we used Adobe Audition programme. This programme also helped to clear some of the recordings of the disturbance noise that scarcely yet still occurred during the obtaining of the material. These adjustments were not at all detrimental for the measuring of voicing contrast in this study.

The recorded material was later processed using PRAAT, where each of the recorded BBC news text was cut into paragraphs for the 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 resulting material divided the individual recordings into phones, however, the programme is designated to recognize only native speech corpus either English or American, therefore a manual segmentation was also necessary. Moreover words that had not been incorporated in the programme's corpora were transcribed as SP (sound pause). The aligner seemed to display all the words that showed

dubious segmentation as capitalized, therefore a script that could remedy this shortage had to be produced. The final step was to manually correct the obstruent boundaries.

3.2.3 Manual Selection of Obstruents

For the sake of this study, only obstruents in the main stressed position were selected. The criteria were both voiced and voiceless stops immediately followed by a vowel and not preceded by sound pause (SP). Instances of obstruents, though in stressed position but followed by a consonant sound e.g. [k̥l] were eliminated, as the identification of such obstruents is already marked by devoicing and therefore were not in the focus of this study. Furthermore, occurrences where /s/ sound precedes an obstruent were also selected, although English phonotactics prefers no positive VOT for its stops in stressed position, Vietnamese English speakers may have acquired disparate specificities.

After the capital upper case letters were substituted by lower case letters, using the prepared script, and the forced alignment generated phones to the third interval tier the target obstruents had to be bordered. The boundaries were marked, if we look at the sound wave, starting from the closure of the obstruents on the left and ending before the first periodical sound wave of the following vowel commenced, on the right see Fig. 2.4.3 Aspiration. Since a great number of the stressed prevocalic obstruents are also word initial all the sound pauses, marked SP, preceding the stops had to be checked. Northern Vietnamese is specific for its so-called ‘chopped’ speech; therefore the alignment marked a relatively large number of pauses, not only in front of the target obstruents. In cases when the stop was really preceded by a sound pause it had to be removed due to impossibility to clearly determine the closure phase of the obstruent.

Following this, an extra interval tier was added to delineate the target stressed postvocalic obstruents. The extra tier was labelled VOT and unlike the third tier consisted of the target stops only. These are similarly bordered with the difference that this time the demarcation started only from the burst. The end was placed in the same position as in the third interval tier and that was before the onset of voicing of the following vowel. After the last demarcation proceeding was finished the data had to be extracted.

3.3 Data Extraction

The next step was to create a script that extracted all the necessary data from the fourth interval tier. The script generated VOT values based on the former demarcation of the target obstruents from both the third and fourth interval tier. An MS Excel file was created where all the data was transferred in preparation for analysis. The Excel sheet contained five columns. Column number one contained all the target obstruents 776 in total. The second column contained information of 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 only two Vietnamese speakers that were originally from the South of Vietnam were coded with an additional remark on their regional origin and placed at the very end of the table. These measures were adopted due to possible deviations these speakers might exhibit in comparison to their Northern fellow-citizens.

Moreover, it was necessary to label those target obstruents that were preceded by /s/ in order to discriminate them from the rest. One extra column was added that indicated the presence of /s/ for each of the words. Finally, the fifth column consists of the VOT values of each of the target obstruent.

Finally ANOVA analysis of variants was used for aggregate relations and Tukey's HSD Test (Post-hoc test), were used to identify the concrete relations among the measured units.

3.4 Hypotheses

As it has been noted in the section on Vietnamese English, Vietnamese is a three category language that employs a three way voicing contrasting of its initial obstruents. Therefore we can assume that:

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

This presumption can be even be supported by (Leather, James, 1996: 274) as mentioned in 2.6.1, where we mention that VOT proves to be most audible to elementary speakers.

H2: It is assumed that as there should be no significant differences between male and female speakers.

H3: We may expect differences in VOT values produced by Southern Vietnamese speakers, since Vietnamese language exhibit a rich regional variety that could have been transferred to L2.

H4: Vietnamese English will display significant shortening of VOT in combinations such are: /s + p, t, k /

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

4 Analysis

This chapter performs several tests that were made 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: ratio, sex, origin, origin & sex, plosives preceded by /s/, and frequency of prevoicing occurrence.

Prevoicing in voiced initial plosives called, as we have already elaborated on, voicing lag. Voicing lag is known for its negative VOT values. The occurrence of positive VOT is expected to be measured with voiceless aspirated plosives that display positive VOT values.

The following relations were classified as significant according to both Tukey's HSD test and ANOVA.

At the end of this chapter we compare our acquired of VOT of Vietnamese accented English to the VOT data of four American English speaking informants published by Lisker and Abramson, in 1964.

4.1 Ratio: VOT, Voicing

Firstly we used ANOVA to analyse and compare the aggregate relation data and proceeded from the general concentration to the more specific. Firstly the effect of voicing on the level of VOT was looked into. At the onset of the analysis, we compared the voiced and voiceless plosives and their VOT values. The figure 4.1 below displays how VOT values (ms) change according to the move between the voiced and voiceless continuum is significant $p < 0.001$ based on equation: $F(1,757) = 393,37$

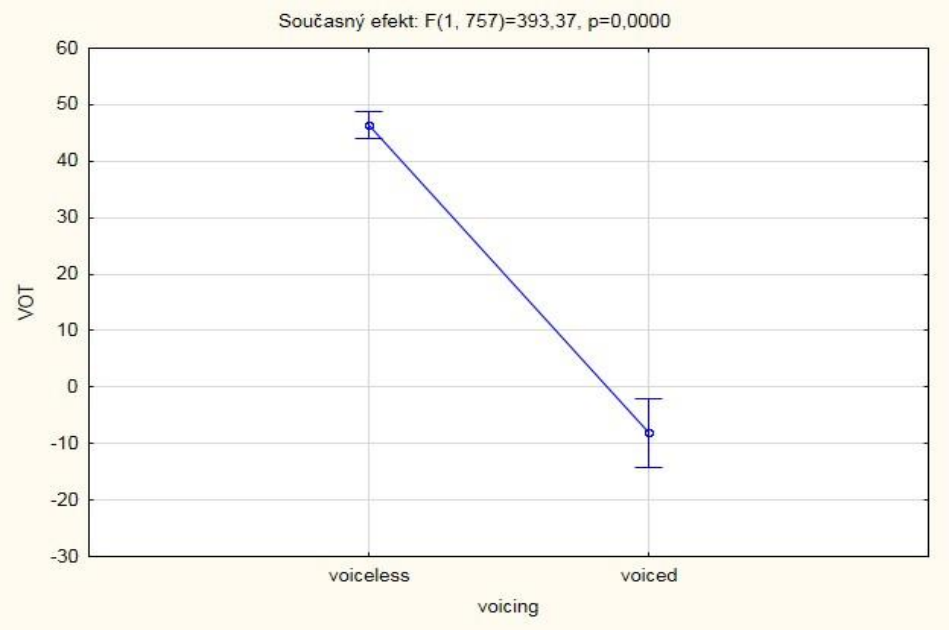

Fig. 4.1. – Ratio: decreasing VOT towards the voiced continuum. The error bars in all figures denote the 95% confidence intervals.

Vietnamese English maintains the two ways plosive contrasting based on VOT values similarly as native speakers of English. Voiceless obstruents are characterized by the higher VOT values, while the voiced counterpart manifests the lower values, as the figure displays. The set of voiced plosives see Fig 4.2. are dispersed in the VOT plosive area, each set occupying the space of either higher VOT or lower VOT values.

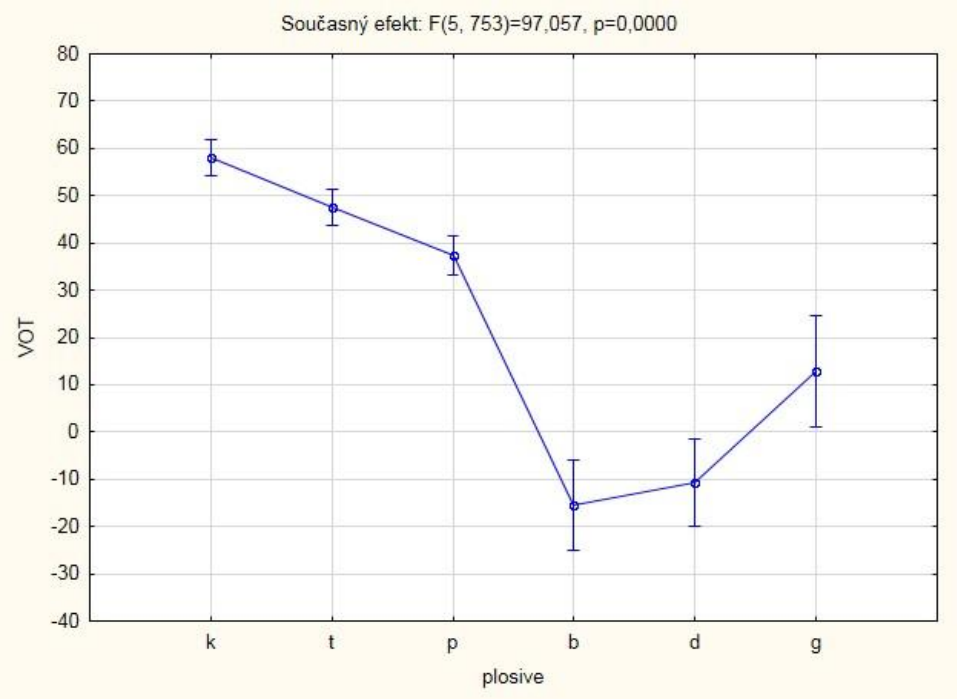

Fig. 4.2. – Plosives and VOT in Vietnamese English

As has been stated obstruents produced further back in the oral cavity are characterized as obtaining higher VOT values. Vietnamese English displays the same tendency. Velar voiceless plosive /k/ evinces the highest VOT value from the voiceless set. Similarly the highest value for the voiced set of stops is represented by velar /g/. Correspondingly the dental plosives either voiced or voiceless acquire higher values of VOT than bilabial plosives pronounced at the very front of the mouth. Based on $F(5,753) = 97,057$, $p < 0.001$.

Referring to the above Vietnamese English sample examined in our study does not deviate from the general characteristics of native English.

4.2 Plosives & Sex

Languages can display differences that are gender specific; therefore we also tested if voicing contrast in Vietnamese English evinces similar deviations. The below attached figure 4.3. displays the target set of fortis and lenis obstruents produced by males and females informants. It can be observed that proportionally there are no major deviations in sexes. The only exception seems to be represented by /b/ where the difference in VOT values is according to Tukey's HSD Test statistically significant $p < 0.05$ based on $F(5,747) = 3,2276$.

The VOT values produced by the male informants, compared to those produced by females, are lower. In the males the VOT values for /b/ approaches up to - 40 ms while in females the values approximated 0 ms.

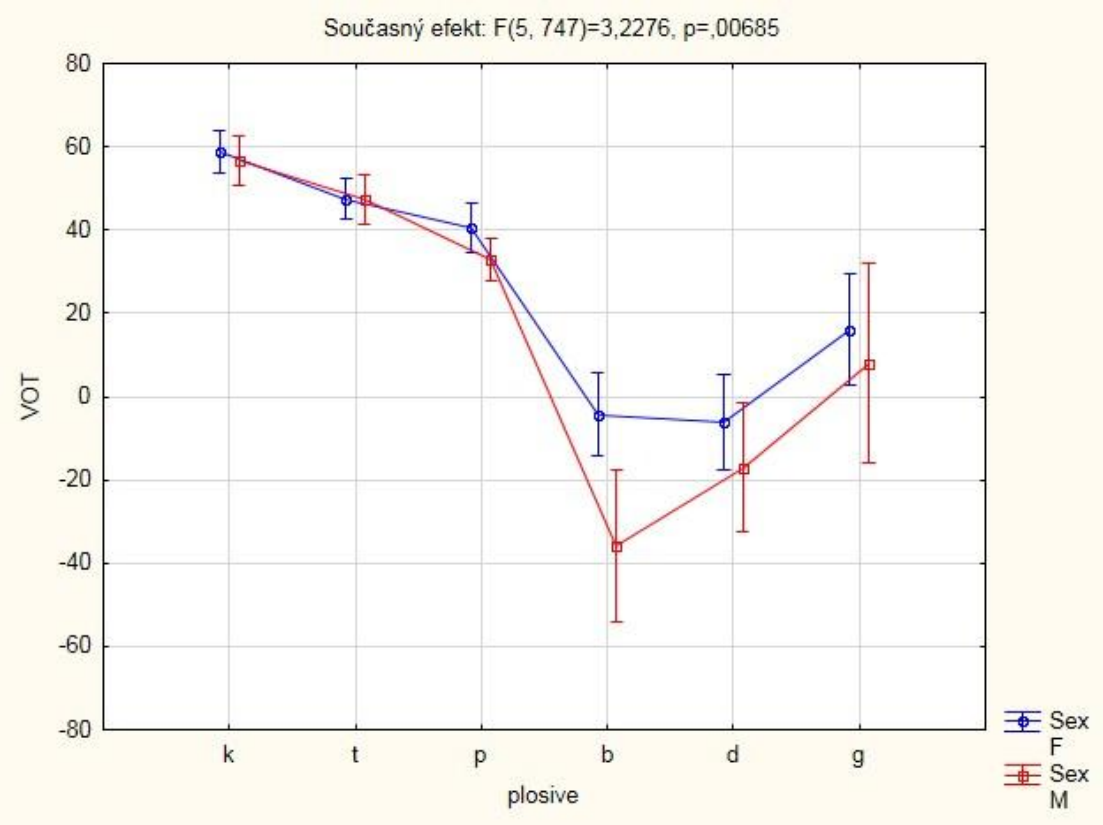

Fig. 4.3. – The influence of sex on VOT

In short, the significance is caused only by the /b/ plosive, see Fig 4.3 where the error bars do not overlap.

4.3 Plosive & Origin

The following test investigates if the acquired corpus evinces any differences for informants that were originally from the South of Vietnam. As has been noted, Vietnamese language shows a rich regional variety; therefore we may expect that regional specificities were transferred from L1 to Vietnamese English. We decided to compare voiceless and voiced plosives separately. The following figure 4.4. shows VOT of voiceless plosives for all the recorded speakers. The Southern Vietnamese are coded with ‘S’ at the end.

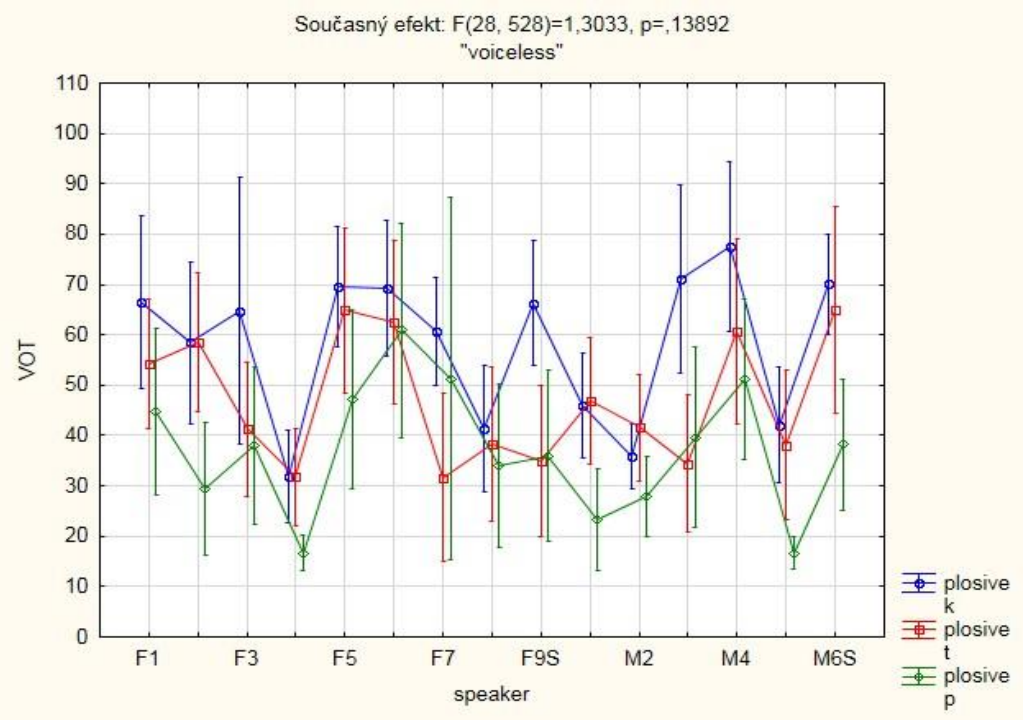

Fig. 4.4. – VOT of individual Vietnamese speakers, voiceless plosives

Firstly we can see that the equation F (28,528) = 1.3033 showing the probability p > 0.1 do not assess the voiceless plosives produced by F9S and M6S as significant. The Southerners compared to the majority of the North Vietnamese speakers do not show significant differences in the production of plosives. The same can be said about the voiced plosives displayed in figure 4.5. below where F (28.141) = 1.4306, p > 0.05.

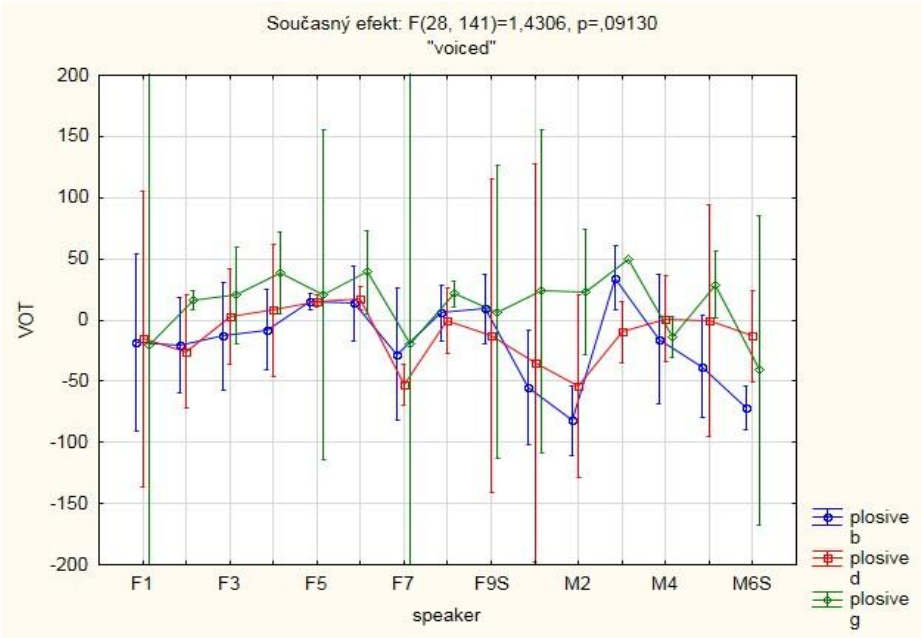

Fig. 4.5. – VOT of individual Vietnamese speakers, voiced plosives

4.3.1 Plosive & Origin, Sex

Secondly 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.

4.4 Plosives Preceded by /s/

In order to thoroughly compare all the values of VOT in a syllable stressed position, also the instances of fortis stops preceded by /s/ were selected from the group of the target obstruents and their values of VOT were compared to the general tendencies of native English speakers. Overall the data in Fig. 4.6. yields native like disposition, showing solid error bars dispersal with lower VOT values with the voiceless stops with /s/ preceding them.

The only instance of deviation the speakers exhibited was with the voiced obstruent /k/, where the error bars overlap. The VOT values for /k/ did not follow the English native standard, since the values for voiceless plosives preceded by /s/ proved to have relatively higher VOT values than the English native standard would require the equation $F(2,567) = 6.6830$, resulted in $p = .001$ which is classified as highly significant.

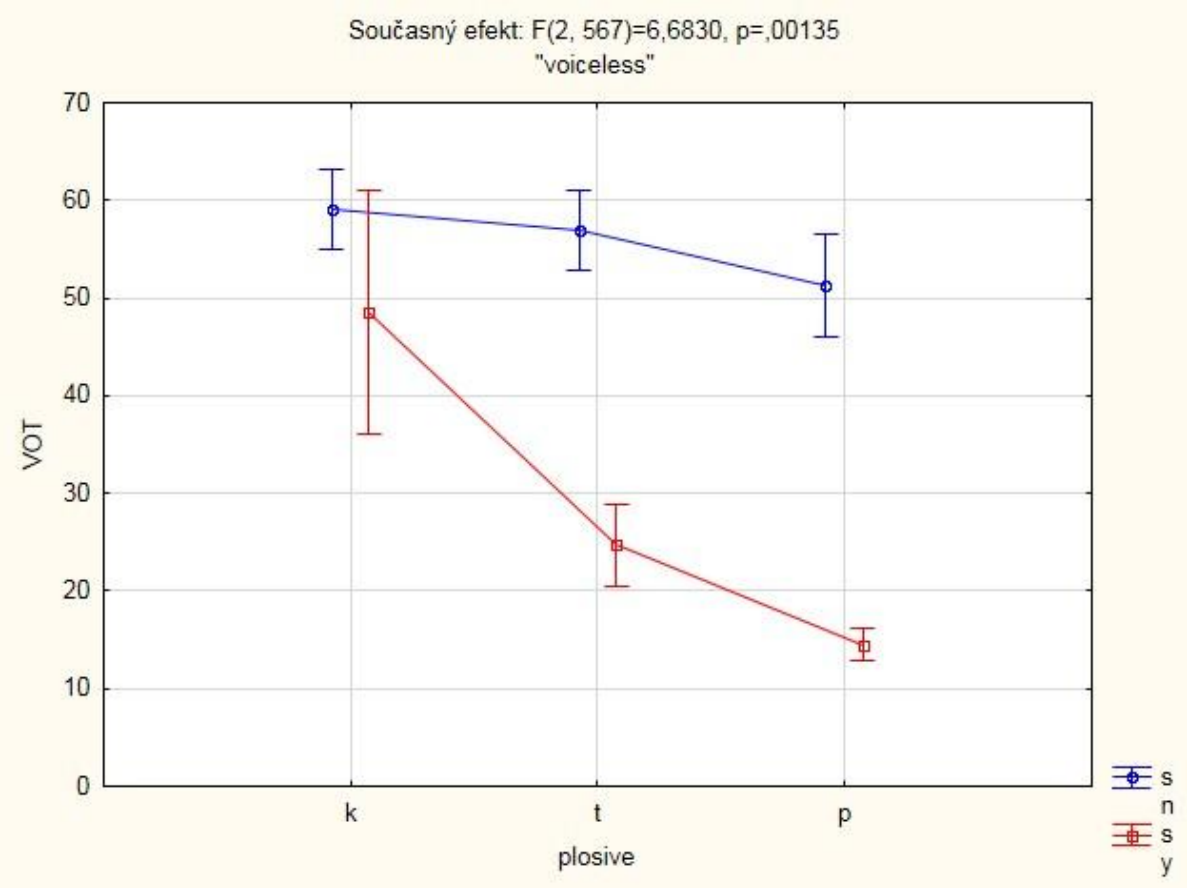

Fig. 4.6. – VOT in fortis obstruents preceded by /s/ yes x no

However, further analysis revealed that this contrast was caused by the repetitive occurrence of /k/ throughout only one word 'rediscovered'. Therefore the results of higher VOT for /k/ preceded by /s/ cannot be accepted as significant (Tukey: $p > 0.5$) In conclusion the informants adhere not to aspirate in voiceless plosives preceded by /s/.

4.5 Frequency of Prevoicing Occurrence

Finally the attention was turned to the set of voiced stops where we measured the occurrence of prevoicing, in other words, negative VOT. Firstly we compared the percentage of prevoicing occurrence. The Fig. 4.5 denotes that the highest occurrence of prevoicing was measured with the lenis obstruent /g/, while /b/ and /d/ showed similar percentage probability to be accompanied by prevoicing. Only the high percentage of prevoicing with /g/ occurred to be significantly different from the other two, proving to be statistically marginally significant with probability $p = 0.064$, we used chi-squared test with the calculation $\chi 2$ (2; n = 186) = 5.49.

Below in table 4.5 we can see the exact numeral representation of each of the particular voiced obstruent.

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

Tab. 4.5. – Numbers of occurrence of positive or negative VOT with particular lenis plovies.

All in all, the lenis plosives exhibited similar distribution of prevoicing in comparison to positive VOT. The only difference was measured with the voiced velar plosive /g/ whose distribution seems to favour more positive Voice Onset Time see figure 4.7. This deviation may be explained by the general tendencies for velar stops, which in their production tend to be accompanied with longer VOT.

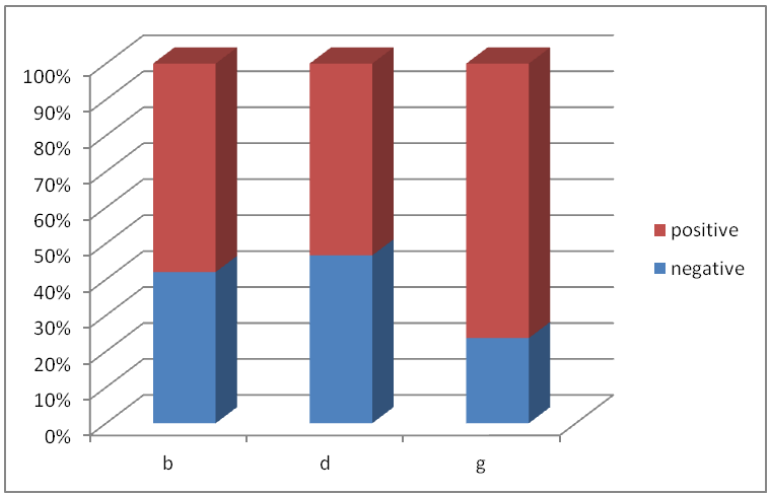

Fig. 4.7. – VOT comparing the percentage of occurrence

In conclusion, in the analysis section we compared the average Vietnamese English VOT values with the average VOT values of American English speakers in order to evaluate if there are compromised values of VOT in comparison to native English speakers and any specialities of Vietnamese accented English in relation to 2.6.3. The reference average values for native English word initial stops used referential VOT values published by Lisker and Abramson (1964) see Tab. 4.6 bellow.

	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 American English[1] (+ Shows for all tables the presence of prevoicing)

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

Tab. 4.7. – VOT values for Vietnamese plosives[2]

	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. – VOT values for Vietnamese English plosives

As can be seen in table 4.8, the average values for our target group of Vietnamese-English speakers for the pre-voiced lenis stops shows that compared to speakers of AmE they produce VOT values very similar to native English speakers. The reason for such accuracy may rest in the fact that the numbers are close to the values of prevoicing in Vietnamese, see Tab. 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.

In comparison, the voiceless stops produced by our informants exhibit higher positive VOT values than the set of fortis obstruents produced by native English speakers. This set of stops seems to exhibit compromised values of VOT, since Vietnamese as L1 shows VOT values higher than those produced by AmE speakers investigated by Lisker and Abramson.

In conclusion, the voicing contrast in Vietnamese English exhibits similar categories as native English. Our data prove that Vietnamese accented English maintain two-way contrasting that is typical in native English speech. In relation to VOT, the values for lenis pre-voiced plosives are bordering on being identical with the values of AmE speakers whilst, lenis stops lacking prevoicing showed values slightly higher. With reference to fortis obstruents tend to have positive VOT higher, which seems to be due to higher VOT in L1 typical for Vietnamese fortis plosives.

[1] Lisker, L. a Abramson, A. S., 1964, pp. 410
[2] Twist, A. Shamoo, J., Bauman,J. 2009, pp. 5

5 Discussion

The results of the analysis show that Vietnamese-speaking learners of English maintain two-way contrasting for English plosives; the voicing contrast in Vietnamese English showed, therefore, to be native English like in plosive contrast production.

Firstly, the analysis compared VOT in respect to voicing, which proved to follow the native like standards also in respect to the place of articulation, as the results were higher for plosives produced further back in the mouth - VOT has proved to be the longest in the case of velars and the 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 one consonant /b/, whose VOT was majorly lower in male speakers. This, however, may be subscribed to the anatomical differences that in general occurred not only between sexes but also among individual speakers; therefore in order to draw more general description further analysis is required. Moreover the gender-based differences were also compared across the southern or northern origin, which however did not yield any significant differences between sexes not even between the Southern and Northern accent.

Thirdly the analysis consulted the VOT for plosives proceeded by /s/ which seem to follow the native English standards of omitting aspiration before this consonant. The only deviation was displayed by /k/ that seemed to acquire significantly higher positive VOT, nevertheless this deviation seemed to appear only throughout only one word, and therefore further analysis to justify this tendency is necessary.

Fourthly, the analysis was directed into comparing the occurrence of lenis plosives and their either positive or negative Voice Onset Times. The result showed similar percentage in distribution of positive and negative VOTs for lenis plosives except for the velar /g/, where prevoicing occurred only in 15% of the instances throughout 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 lenis stops missing prevoicing, while pre-voiced plosives displayed average VOT values very similar to native AmE. This resemblance can be subscribed to the average values of prevoicing in Vietnamese that respond

to the average values for Vietnamese accented English produced by our informants. Prevoicing for English might have been by our speakers /d/ classified as prevoicing in their L1 Vietnamese *đ* that bears negative VOT very similar to English voiced alveolar plosive. Considering the fact that all of the informants were late learners of L2 – they did not acquire English before the age of six- they are not likely to have separate phonetic categories for L1 and L2 VOTs.

It is worthwhile noting that although the measurement based of VOT is very popular due to its relative feasibility of measuring and variability across different languages there are quite a few pitfalls that arise when describing the plosives. Mikuteit (2007) in his research addresses 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 only two-way contrasting the measuring with the use of VOT- which does not face the problem of differentiating voiced aspirated plosives – is sufficient for this study . Nevertheless, research using the methods suggested by Mikuteit and Henning may yield interesting results for Vietnamese accented English.

6 Conclusion

This thesis intended to describe the voicing contrast in Vietnamese English. The theoretical part brings a brief account on the theory that is necessary to understand the concept of voicing, aspiration and the measurement based on the values of VOT. It also comments on the second language acquisition and the way pronunciation features can be transmitted to L2 - in our case English. Furthermore Vietnamese and Vietnamese English is described, with regard to its initial consonants so that the later comparison of the results can draw upon this particular theory its conclusions. We developed several hypotheses based on the knowledge of Vietnamese being a language with three-way contrasting, English being a language with two-way contrasting together with regard to second language acquisition theories.

The Analysis section proved Vietnamese English to follow the native English standards of forming the voicing contrast. In conclusion of this section we compared the average values of VOTs produced by our informants with those produced by native speakers of American English. The results prove to have specific VOT values for Vietnamese accented English that differentiate its accent from the native accent of English. 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.

Despite the thorough work this study performed examining the differences in phonetic utterance of Vietnamese English; it is without doubt that the topic could be looked into further to a greater extent. Therefore the paper does not aim to be prescriptive in how to approach the topic. Nevertheless, we could assert the following:

Since our informants have proven not to struggle with maintaining the contrast between the initial English stops we can assume that when L1 contains richer variety of voicing contrast in plosives the L2 learner may not struggle with this type of acquisition (voicing contrast in initial consonants acquisition). Therefore, as phonetics regards, it would be advisable L2 learners majorly focus on features not common in L1 as to master L2.

This also follows that SLA (Second Language Acquisition) researchers and tutors should not omit to focus on phonetic features that are not present in the mother tongue, they work with, so that they learn and can pass on how achieve an L2 oral comprehensibility. Finally, language textbooks should regard the absence of features in the native language in order to guide the learners through the phonetic pitfalls of the L2.

Bibliography

Sources Quoted:

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

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

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

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

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

Ladefoged, P. (2010). *A Course in Phonetics, Third Edition*. Fort Worth: Harcourt Brace College Publishers, pp. 141 – 147.

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

Ní Chasaide, A (1987). "Glottal Control of Aspiration and Voicelessness". In: *Proceedings of the 11th International Congress of the Phonetic Sciences*, 6, p. 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. Available on-line from http://www.yorku.ca/earmstro/asia/Vietnamese_accent.pdf (accessed: 1 June 2001)

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

Stevens, K.N. (1998). *Acoustic Phonetics*. Cambridge, Massachusetts: MIT 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, pp.4-7.

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

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

Zampini, M, & Hansen Edwards, 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, pp. 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.

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

Hwa-Froelich, D., Hodson, B., Edwards, W., etc. (2002). "Characteristics of Vietnamese Phonology", American Journal of Speech- Language Pathology, vol. 11, August 2002, 264-273… skrtnout autory a dat etc.

Kohler, K. J. (1982). F0 in the production of lenis and fortis plosives. Phonetica, 39, str. 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, Filosofická Fakulta..

Leather, J. & James, A. R. (1996). "Second Language Speech." In: Ritchie, W. C. & Bhatia, T. K. (eds.), Handbook of Second Language Acquisition (269-316). San Diego: Academic Press.

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

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

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

Sawashima, M. a Hirose, H. (1983). "Laryngeal Gestures in Speech Production". Speech Production and Speech Modelling NATO ASI Series Vol. 55 1990. 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, str. 471-500.

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

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