Filosofická fakulta Univerzity Palackého
Katedra anglistiky a amerikanistiky

Subconscious Imitation of Perceived Phonetic Properties and its Relationship to Phonology

(Bachelor thesis)

Author: Tomáš Sedláček
Supervisor: Mgr. Václav Jonáš Podlipský, Ph.D
Olomouc 2014

I declare that I elaborated this bachelor thesis on my own, using solely the sources listed in the references. The pilot study was conducted as part of the Acoustic Phonetics class, opened by the Department of English and American Studies of Palacky University in Olomouc, taught by Václav Jonáš Podlipský, Ph.D. and supervised by Šárka Šimáčková, Ph.D in the Spring Semester 2013, in collaboration with Kateřina Braunová, Pavel Hrdlička, Anna Morales, Veronika Neřoldová, Adam Petrásek and Jitka Voříšková

In Olomouc on 27th of June 2014 Tomáš Sedláček

Acknowledgement

I would like to thank my supervisor Mgr. Václav Jonáš Podlipský, Ph.D. for his guidance and useful advice throughout the creation of this thesis. My thanks go also to Kateřina Braunová for providing me with useful literature and helpful advice.

Table of Contents

1. INTRODUCTION

The current study is concerned with the comparison of Czech and English VOT. More specifically, it aims to replicate the study conducted by Nielsen (2011), who investigated phonetic imitation of extended or truncated VOT of English speech.

VOT refers to the time elapsed between the burst of the plosive and the onset of voicing for the following segment. Phonologically speaking there is "evidence in favour of there being only three [modal] values" (voiced; voiceless unaspirated and voiceless aspirated plosives), strongly suggested by the fact that no language contrasts more than three homorganic plosives (Cho and Ladefoged 1999, 226). Nevertheless, VOT is a continuous measure and languages may arbitrarily choose the values for the contrast, therefore the actual phonetic contrast may not concord with the phonological one (ibid 226). In case the voicing starts before the release of the plosive, it is described as *prevoiced*, with the VOT going to the negative value. On the other hand, if the voicing starts immediately or shortly after the release of the plosive, we describe it as *short-lag*. The plosive is described as *long-lag,* if the voicing for the succeeding vowel is postponed by a period of voiceless noise.

The contrast between English voiced and voiceless plosives is in many cases not based on the voicing per se but rather on other phonetic characteristics, which serve contrastively. One of the main acoustic measures, also in the focus of the current study, is the VOT. Ladefoged (2011, 198) argues that English contrasts voiceless plosives /p, t, k/, which are usually produced with long VOT, compared to the voiced ones /b, d, g/, most often produced with partial or no voicing at all. The voiced plosives appear to be distinguished from the voiceless ones by zero or very short VOT.

The pilot study is based on Nielsen (2011), who exposed listeners to English word-initial /p/ with extended and reduced VOT. "The results revealed that participants produced significantly longer VOTs after being exposed to target speech with extended VOTs," moreover, the author also found a generalization "to new instances of the target phoneme /p/ and the new phoneme /k/" (Nielsen 2011, 132). The new instances of the target phoneme /p/ and the new phoneme /k/

represent /p/ and /k/ words that were not used in the listening material. Moreover, the study found that after being exposed to tokens with reduced VOT, the participants did not imitate the reduction. The author speculated that the imitation was asymmetrical, that is VOT extension *was* imitated but VOT reduction was *not*. The reason is most likely due to the fact that reducing VOT could result in neutralization of the voicing contrast between the English phonologically voiceless plosives that are aspirated when implemented phonetically and the phonologically voiced stops that are phonetically short-lag voiceless plosives.

The aim of the present thesis is to test Nielsen's hypothesis. A pilot study is reported, which was conducted by students of Acoustic Phonetics class, opened by the Department of English and American Studies of Palacky University in Olomouc. The course was taught in the Spring Semester 2013 by Václav Jonáš Podlipský, Ph.D. and supervised by Šárka Šimáčková, Ph.D. This pilot study was based on Nielsen (2011), a study presented in more detail in Chapter 4.4.1, that is to say, it was also focused on the voicing contrast between homorganic plosives but unlike Nielsen, it was concerned with Czech plosives.

Skarnitzl (2011, 71) argues that the Czech plosives "transcribed as /p, t, k/ are produced with no voicing or aspiration, whereas /b, d, g/ typically retain full voicing."[1] Word-initially the Czech voiced plosives /b, d, g/ are said to be prevoiced, with negative VOT.

Firstly, it is necessary to define some central phonetic and phonological terms and present the properties of Czech and English plosives, which is the aim of Chapter 2. This chapter also deals with the difference between phonetic and phonological voicing and phonetic cues that may help the listener to categorize a plosive as either voiced or voiceless. Chapter 3 presents the factors that have to be accounted for when measuring VOT. Previous studies of speech imitation are to be discussed in Chapter 4. Chapter 5 presents the present study, which aims to test the hypothesis that was derived from Nielsen's speculation, it also provides suggestions for further research. Chapter 6 summarizes the findings of the study.

[1] The source is originally in Czech, the citation and any further citations of Skarnitzl (2011) were translated by the author of this thesis.

2. PHONETIC BACKGROUND

The aim of this section, divided into 3 subsections, is to provide phonetic background for the present study.

The first concern will be the general categorization of consonants.

Secondly, I will focus on comparing the system of Czech and English plosives.

The last subsection aims to distinguish the phonological from the phonetic voicing. In this subchapter some crucial terms will be defined. Furthermore, the main cues that may help the listener categorize a certain plosive to either the voiced or voiceless category, will be presented.

2.1 Consonantal Categorisation

Vocal tract comprises of articulators which make the realization of speech sounds possible. The upper part of the oral and pharyngeal cavity comprises of passive articulators due to their static nature. They form the ridge on which the lower parts of the vocal tract, the active articulators, can make various kinds of constrictions and hence shape the vocal tract to create a range of distinct sounds. The upper parts of the vocal tract were given names for an easier classification based on the place of articulation, presented in Figure 1.

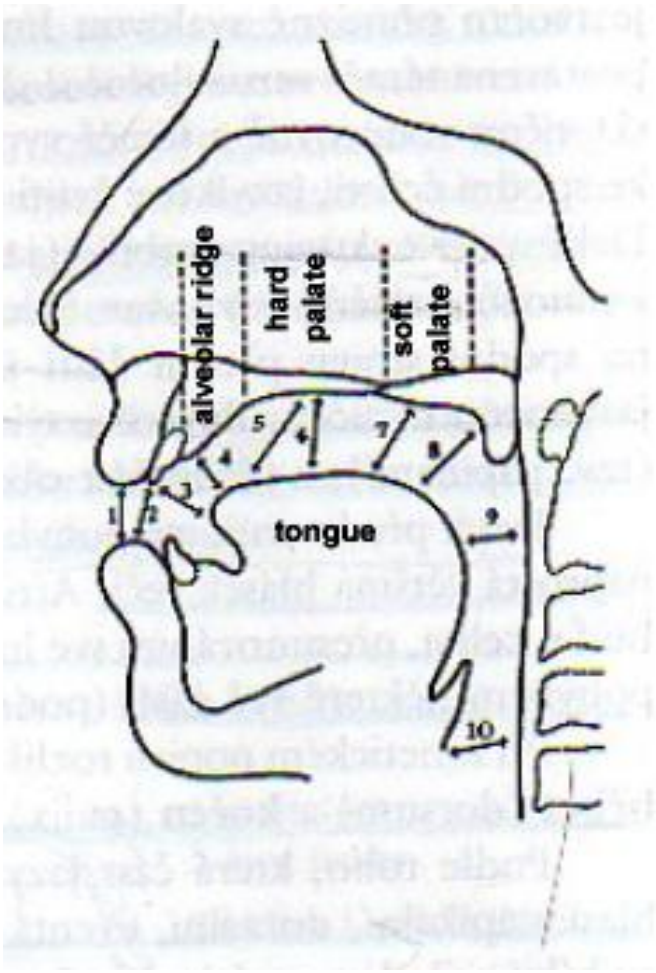

Figure 1 represents the main regions of articulatory movements to produce speech sounds. The numbers represent the places for the following sounds (adapted from Palková 1994, 66):

1) bilabial
2) labio-dental
3) dental
4) alveolar
5) retroflex
6) palatal
7) velar
8) uvular
9) pharyngeal
10) laryngeal

Place of articulation is not the only criterion by which speech sounds can be contrasted. Another way of classifying consonants is based on the manner of articulation. With respect to the manner of articulation, Ladefoged (2011, 14) states that "the articulators may close off the oral tract for an instant or a relatively long period; they may narrow the space considerably; or they may simply modify the shape of the tract by approaching each other."

Another classification offered is based on the grade of loudness. The consonants are distinguished according to their loudness, i.e. sonority. A consonant in which there is a complete closure or a constriction causing friction are called obstruents, exhibiting less sonority. This variety sometimes includes a contrast between voiced and voiceless taxonomy. The other category in which "there is only partial closure or an unimpeded oral or nasal escape of air; such articulations, typically voiced, and frequently frictionless, i.e. without a noise component, may share many phonetic characteristics with vowels" and are called sonorants, exhibiting more sonority than obstruents (Gimson 2001, 149).

Demarcation of those wide categories is motivated not only with respect to articulation and perception but also in terms of phonological systems and phonotactic properties of world languages (Duběda 2005, 69).

2.2 The Characteristics of Plosives

Plosives are typically dissimilar from other types of consonants owing to their realization. Three stages are recognized throughout their production.

The first stage constitutes the coming-together (closing) of the articulators to create a closure somewhere in the vocal tract. "In this stage, there is often […] a transition audible in the preceding sound segment and visible in an acoustic analysis as a characteristic curve of the formants of a preceding sound" (Gimson 2001, 150). During this stage, there is also another process involved within the vocal cavity, namely the soft palate is raised to obstruct the airflow to the nasal cavity, so that the air cannot escape through the nose. Considering the spectrogramic representation of this stage, it causes some difficulties to locate the closing precisely, because "the blocking of the airstream is not a sudden event but occurs more gradually as the articulators come together" (Hayward 2000 177).

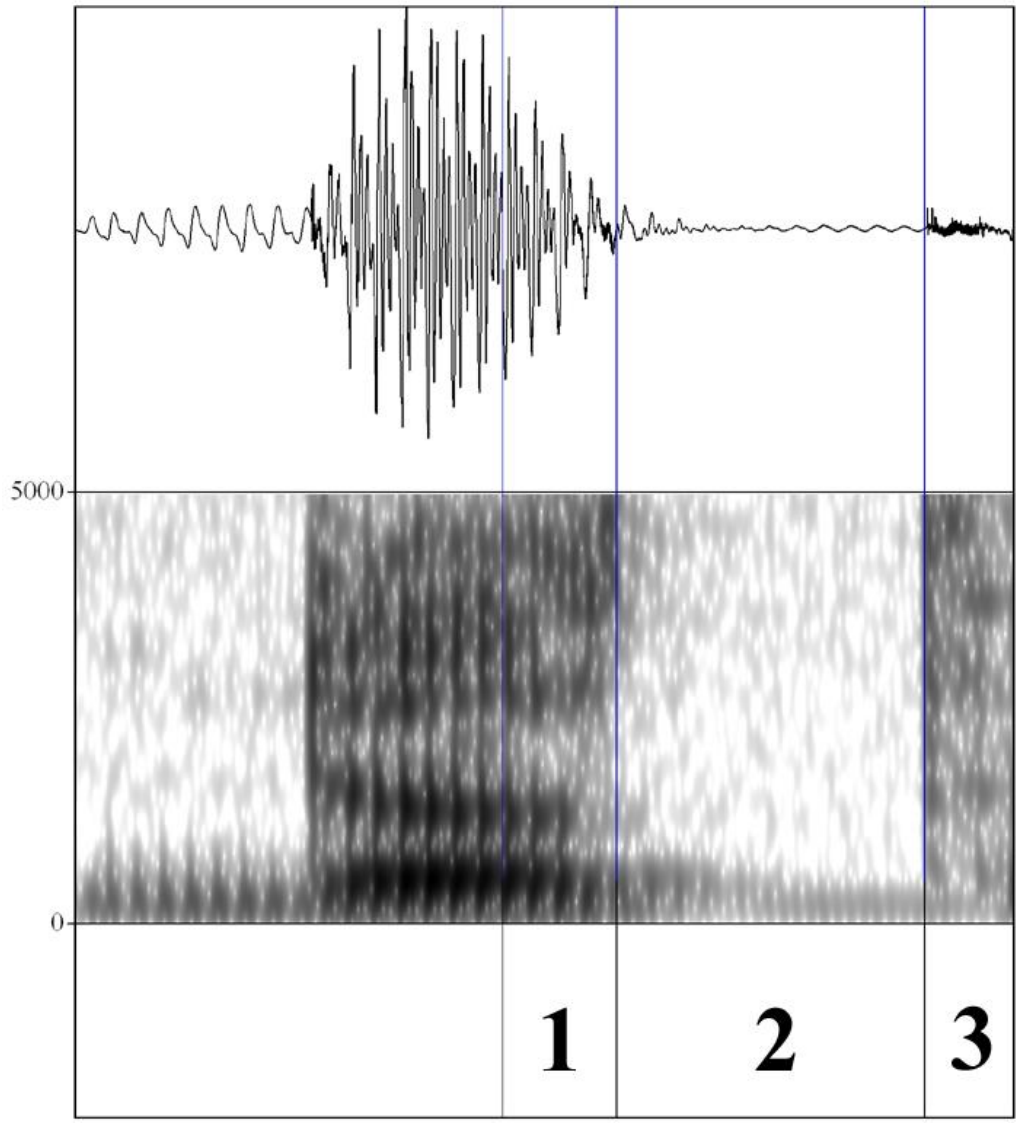

Figure 2 represents /dot/ of the Czech word **dotek** "touch", as produced by a male speaker of native Czech. The three phases of the plosive are marked by the numbers. The number 1 represents the closing phase, 2 marks the closure phase and the left line of 3 shows the beginning of the release phase.

In the Figure 2 we can see such closing of the two articulators into the constriction of the vocal tract. This closing phase is marked by the number 1. In the waveform, we can notice a regular decrease of the amplitude throughout each of the succeeding periods of the vowel occurring prior to the plosive.

The second stage, by definition the closure (compression) represented by the number 2 in the Figure 2, involves the accumulation of air in the vocal cavities, caused by the constriction made in the vocal tract and the action of the

lungs, pushing the air in the direction of the mouth. Within this stage the vibration of the vocal folds may be present. However "if the vocal folds are not vibrating, no sound will be produced at all, resulting in a period of silence" (Hayward 2000, 177).

Concerning articulatory characteristics of individual plosives, it is necessary to note that the duration of the closure period universally decrease in this particular order /p/ > /t/ > /k/. Pressure proportions in the vocal tract can account for this phenomenon. The more the space behind the closure, the longer the increase in pressure persists to reach the necessary threshold for the realization of the burst. (Duběda 2005, 86)

In the final stage, called the release (burst), the beginning of which is marked by the left line of the number 3 in the Figure 2, the two articulators making the obstruction, become rapidly detached from each other, therefore an abrupt outflow of air can be observed. We can say that the air escapes with an explosion, for this particularly property, the stop consonants are called plosives. "This is generally easy to locate on a spectrogram, because it corresponds to a brief vertical spike, though it is weaker for labials than for other places of articulation" (Hayward 2000, 177). If the voicing is retained throughout the second stage, the vibration of the vocal folds will persist for the duration of the release stage if a vowel follows. On the contrary, if the second stage is voiceless, there may also be a period of silence throughout the release of the plosive, before the onset of the voice of the following vowel.[2]

2.2.1 Czech Plosives

Czech distinguishes, with respect to the place of articulation, 4 contrastive voiced and voiceless pairs of plosives. The bilabial /p, b/; alveolar /t, d/; palatal /c, ɟ/ and velar /k, g/, represented by the Table 1. For the purposes of the research, presented in Chapter 5 of this thesis, I will focus more closely on bilabial and alveolar plosives.

[2]Also known as "aspiration". Chapter 2.3.6 will be concerned more closely with this phenomenon.

	Bilabial	Labiodental	Alveolar	Postalveolar	Palatal	Velar	Glottal
Plosive	p b		t d		c ɟ	k g	

Table 1 shows the phonemic inventory of Czech plosives, when a pair is present, the right phoneme is always voiced, the left one voiceless respectively (adapted from Dankovičová 1999, 70)

Bilabial plosives are produced by pressing the lips together against each other, the muscular activity is greater for the production of /p/ than for /b/ (Palková 1994, 223). During the production of those bilabials, the tongue is located on the base of the oral cavity, without participating in the production of this speech sound. The soft palate is raised to block the air from flowing through the nasal cavity. One of the cues for the place of articulation is considered the concept of locus, which according to Delattre et al. (1955, 769) is "a place on the frequency scale at which a transition begins or to which it may be assumed to 'point'". According to Palková (1994, 223), the F2 locus for the bilabial plosives is as low as 600 Hz.

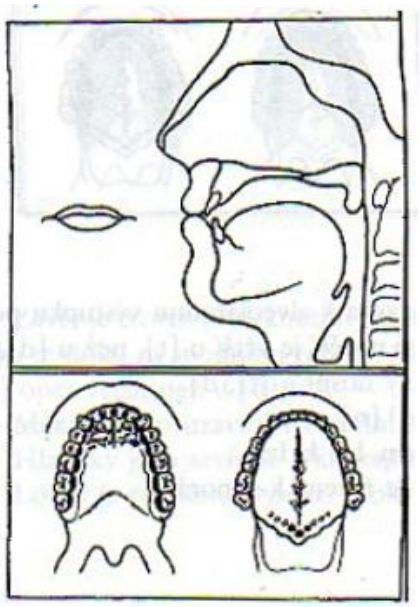

Figure 3 represents the articulation of the Czech voiceless /**p**/ and voiced /**b**/ bilabial plosives (adapted from Palková 1994, 223).

Another pair of plosives, called the alveolars, is the one that is produced by the tip of the tongue pressing on the alveolar ridge, alongside the upper teeth. The muscular activity is again greater for the articulation of voiceless /t/, compared to the voiced /d/. The soft palate again obstructs the nasal cavity and the F2 locus is in the mid zone, around 1600 Hz (Palková 1994, 224).

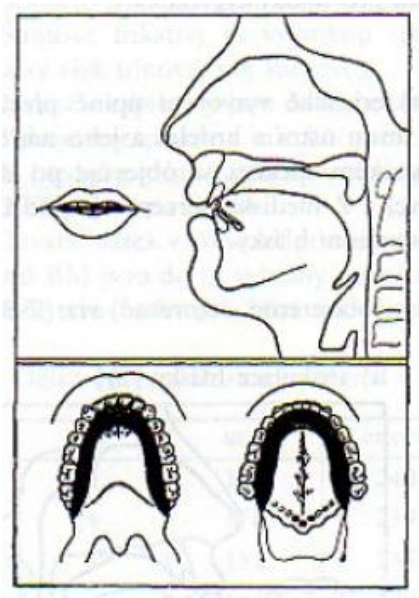

Figure 4 represents the articulation of the Czech voiceless /**t**/ and voiced /**d**/ alveolar plosives (adapted from Palková 1994, 224).

2.2.2 English Plosives

Contrary to the system of Czech plosives, English contrasts only 6 pairs of phonemes. The bilabial /p, b/; alveolar /t, d/ and velar /k, g/. Owing to the fact that the crucial research of Nielsen (2011) used as stimuli words beginning with /p/ and /k/, I will only present the details of bilabial and velar plosives.

	bilabial	labio-dental	dental	alveolar	palato-alveolar	palatal	velar
plosive	p b			t d			k g

Table 2 shows the phonemic inventory of English plosives, when a pair is present, the right one is always voiced, the left one voiceless respectively (Adapted from Ladefoged 2011, 43).

The plosives, classified as bilabial, are primarily obstructed at the lips which come together to create the barrier to the airflow, nasal cavity is obstructed by raising the soft palate (Gimson 2001, 161). The F2 locus seems to vary from about 700 to 800 Hz (Hayward 2000, 186).

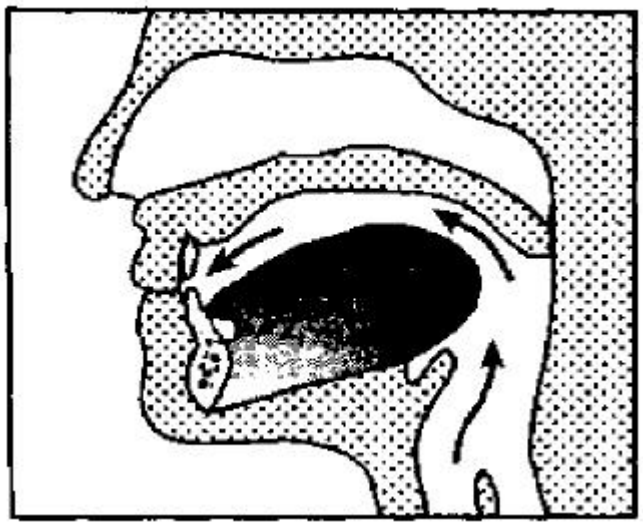

Figure 5 represents the articulation of the English voiceless /**p**/ and /**b**/ bilabial plosives (adapted from Gimson 2001, 161).

Another category that I would like to present comprises of velar plosives. The closure for the production of velar stops is created with the back of the tongue pressing against the velum and the soft palate blocking the air from entering the nasal cavity (Gimson 2001, 167). According to Gimson (2001, 167), the velar plosives are highly dependent on the vowels found in their immediate adjacency.

The author argues that when "a front vowel follows[,] […] the contact will be made on the most forward part of the soft palate and may even overlap on to the hard palate" (ibid 167). On the other hand when the velar is followed by a back vowel, "the contact on the soft palate will be correspondingly retracted" (ibid 167). We can notice such difference in the Figure 6.

.

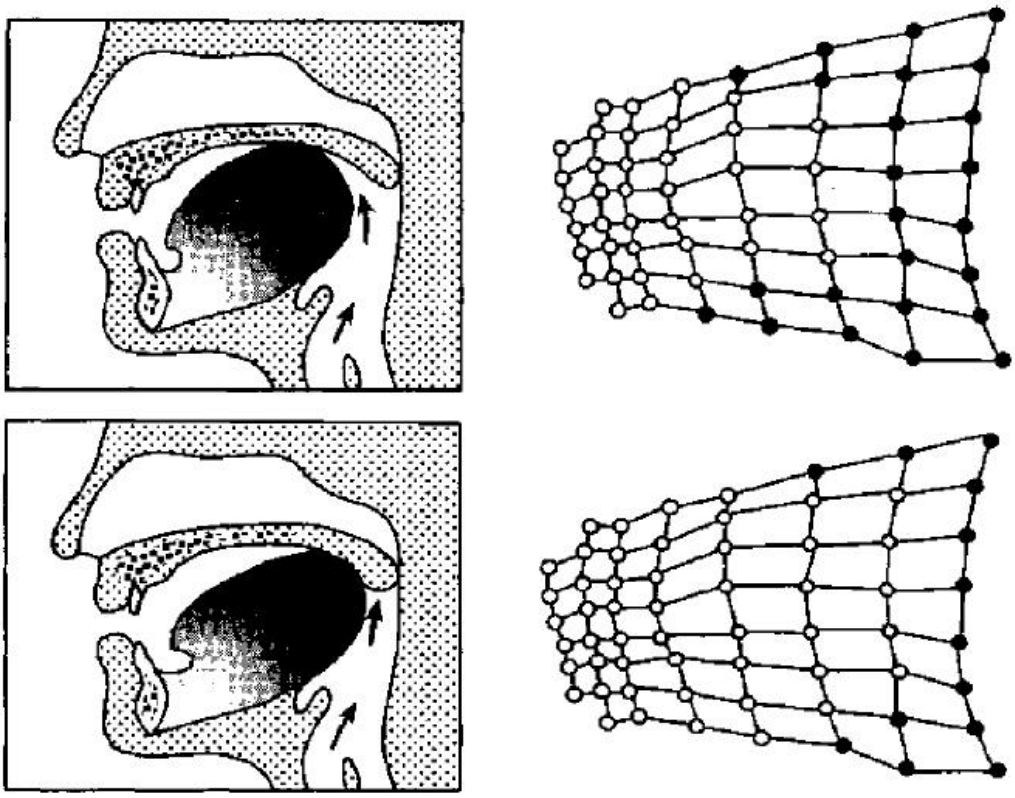

Figure 6 represents the production of the English velar plosives. The upper figure shows the production of the velar plosive, followed by a front vowel, in this case the high front vowel /i/. In contrast, the lower figure shows the production of velar plosive, followed by a back vowel, in this case the open back vowel /ɑ/ (adapted from Gimson 2001, 167).

2.3 Phonetic and Phonological Voicing

Focusing on voicing contrast of plosives, we have to note that not all languages contrast both voiced and voiceless plosives[3]. However, in languages that contrast plosive sounds with respect to voicing, we must make a distinction between different concepts of voicing because using this term may lead to confusion.

[3] A fact that will be discussed in more detail in Chapter 2.3.1

There are two distinct ways of using the terms *voiced* and *voiceless*. Considering phonetics, they can be used "to designate presence or absence of vocal fold vibration during the consonantal interval" (Hayward 2000, 178). On the contrary, if we focus on phonology, the terms "may be used to refer to opposed members of a phonological contrast, which might or might not be pronounced as their label suggest" (ibid 178). We can hence say that phonological voicing does not always correlate with phonetic voicing.

Due to the fact that phonological and phonetic voicing do not always correlate in every language[4], it was suggested to use the contrast based on tenseness, that is the consonants are referred to as either fortis or lenis (tense or lax) (Kohler and Künzel 1978). This distinction is usually concerned with articulatory strength and closely correlated with voicing (Duběda 2005, 81). In many languages the voiceless obstruents are usually more tense and therefore constitute the fortis group, whereas the voiced obstruents are less tense and are thus called lenis (ibid 81).

Arguably, not every single phonetic property, distinguishing the voiced and voiceless plosives is relevant with respect to every position within the word (Skarnitzl 2011, 113). For instance plosives in their word-final position have different properties, with respect to voicing and aspiration. Maintaining voicing and distinguishing aspirated plosives from unaspirated may prove difficult, especially before a pause, since the vocal folds are anticipating being spread for the intake of breath. This may lead, in many languages, to neutralization of the voicing contrast and for that reason also the aspirated plosives manifest themselves as unaspirated (ibid 114-115). This leads us to the fact that some languages with both voiced and voiceless stops do not contrast them in every position. One such example is Czech, which exhibits word-final devoicing, therefore there is no contrast of voicing in this position.

When plosives occur word-initially, the contrast is based for example on the phonetic voicing, aspiration, voicing during closure, the nature of F0 and the onset frequency and the extent of F1 transition. It is evidently not possible to use the duration of the closure or the duration of the preceding vowel. The following subsections aim to present some of the main phonetic cues to categorize a certain

[4] Such as in the case of Germanic languages, the phonetic voicing is typically not a very reliable cue to distinguish the sound as voiced or voiceless. To use fortis/lenis distinction is much more appropriate (Duběda 2005, 81).

sound as either voiced or voiceless. Repp (1979, 173) argues that there are "trading relations between different cues" that serve to distinguish "the same phonetic contrast." The author speculates that making an alteration to one cue may be compensated for by manipulating another cue in an opposing way without changing the final "phonetic percept" (ibid 173). According to the author, such interchangeability of relations can be accounted for the wide variety of cues. Nevertheless, "for a given phonetic distinction there may be a natural hierarchy of cues" (ibid 174). This is consistent with Francis et al. (2008, 1234), who argue the multiplicity of acoustic properties which "have the potential to function as perceptual cues to categorization under appropriate circumstances," even though, the cues are not "weighted equally in a given contrast."

2.3.1 Voicing during Closure

Universally, we can say that voiceless plosives are found much more frequently in languages. Duběda (2005, 85) investigated sample of world's languages and found 92% of them to have voiceless plosives without further modifications, whereas only 67% did also have voiced ones. The author speculates more frequent occurrence of voiceless over voiced plosives, and generally all obstruents as well, is caused most likely by physiological factors, that is the constriction causes the supraglottal pressure to rise and thus exacerbates aerodynamic conditions for phonation (ibid 85). We can thus say that the higher the supraglottal pressure the more difficult to produce a voiced plosive. This is also connected with a quasi-universal property, that is the voiceless plosives tend to be longer than the voiced ones.

The fact that voiced plosives are less common and much harder to produce suggests that voicing during closure may function as a lesser significant voicing cue, especially in languages such as English in which phonological voicing very often does not concord with phonetic voicing. For example Raphael (1972, 1301) argues that "the presence of voicing during the closure period […] does have some cue value, although it is minor." Winn and Chatterjee (2012) are also in agreement with such suggestion. According to the authors, "voicing during closure is not thought to be essential for the perception of the voicing feature" (ibid, 1473).

2.3.2 The Durational Influence of the Preceding Vowel

Even though this thesis is concerned with plosives in the syllable-initial position and the perceptual cues that relate to those plosives, it is important to mention one of the crucial cues for plosives that occur word-finally and are preceded by a vowel.

Warren and Marslen (1988, 21) suggest that "the presence or absence of voicing in word-final stops can be marked by a number of cues, which include vowel and closure duration, [...] [and] the properties of the burst." The authors also suggest the distinction between the spectral cues and durational cues. The former "directly signals the place of articulation of the final stop," that is after hearing the last period of the vowel, the listener can determine the sound (ibid 21). The durational cues, however, are somewhat different. Namely, it is not concerned with the specific quality of the acoustic spectrum but the perceived duration "relative to a given set of phonologically determined criteria for the vowel lengths associated with the voicing contrast" (ibid 22).

One of the researchers concerned with this phenomenon in American English was Raphael (1972). He argues that the "duration of vowels before final consonants is both a powerful and sufficient acoustic cue […] for differentiating between the classes of those cognate pairs" (ibid 1296). According to the author, when we neutralize voicing cues in word-final plosives and make the vowel longer than 200 ms, the consonant following the vowel seems to be recognized as voiced. On the other hand, modifying the vowel to be shorter than 200 ms leads the perceptual judges to categorize the consonant as voiceless (ibid 1296). We can thus say that "the perception of 'voicing' of the final consonant increases as the ratio of the duration of final consonant to preceding vowel decreases" (ibid 1296). The finding of Raphael (1972) is consistent with that of Warren and Marslen-Wilson (1988), who found similar correspondence. More specifically, Warren and Marslen-Wilson's findings "show especially clearly the operation of the durational criterion," the voiceless responses are predominant when the listeners hear around 130 ms, the voiced ones, however, begin to be prominent only when at least 150 ms of the vowel is heard (Warren and Marslen-Wilson 1988, 24).

2.3.3 Duration of the Closure

One of the cues for voicing differences of plosives is the duration of the closure. Intervocalic plosives /p, b/ in American English were investigated by Lisker (1957, 47) with a result of consistent difference in the closure duration, which is thus considered "a major cue to the voiced vs. voiceless distinction" (Lisker 1957, 47). According to the author the closure duration constitutes "the time interval between termination of the vowel-formant transition preceding the stop and onset of the transition to the following vowel" (ibid 43). "Spoken at moderate conversational speed" /p/ was found to have an average of 120ms whereas /b/ averaged 75ms (ibid 43). Brunschweiler (1997, 354) also concerned with "the temporal patterns of the voicing contrast in medial position." The findings suggest, similarly to Lisker (1957), that the plosive voicing contrast is reflected in dissimilar values for the closure duration (Brunschweiler 1997, 372).

2.3.4 The Nature of F0

Another cue relevant to the voicing contrast is the fundamental frequency (F0), even though, according to Raphael (2005, 191), "the role of F0 […] cueing the voiced-voiceless distinction is generally considered secondary, its effects are measurable in experimental conditions." Abramson and Lisker (1985, 26) also suggest that there is a connection between the "voicing state of an initial consonant and the F0 height and movement at the beginning of that contour." Following a voiced plosive, F0 tends to be lower and consequently rises upward. For the voiceless plosive, on the other hand, the F0 will go from higher value to lower ones (ibid 26). This distinctive feature is normally referred to as the "F0 perturbation and has been found in a wide variety of languages" (Whalen et al. 1992, 2152).

Abramson and Lisker (1985, 32) studied the power of this particular cue in relation with VOT and found only a "modest effect of fundamental frequency shifts on judgements of consonant voicing." More specifically, testing tokens with ambiguous VOT values with flat F0, and subsequently manipulating the F0, were found to help categorize the plosive as part of one "category or the other […] in a forced-choice test" (ibid 32). However, when the VOT sets a word securely into voiced or voiceless category, F0 is unable to effect this categorization (ibid 32).

Whalen et al. (1992, 2153) concerned with the "influence of the F0 perturbation on identifying VOT continua." In the experiment onset F0 value was adjusted to go against the standardly observed perturbation phenomenon, thus "the F0 information [did not] match that of VOT" (ibid 2153). Voicing judgements with speeded and unspeeded responses were assessed. The authors speculated that if the participants would show a slower response for the tokens with adjusted F0, compared to tokens with unmodified F0 values, it will serve as an evidence of the fact that "all acoustic consequences of a speech gesture contribute to the perception of speech, even if the labelling fails to show it" (ibid 2153). The authors found that "high F0s slowed responses to short VOTs" and "low F0s slowed responses to long VOTs" even in unambiguous VOT tokens (ibid 2158). Thus contrary to Abramson and Lisker, this research indicates that listeners hearing a speech sound tend to make "use of all information" they have available, even if such information is phonetically redundant (ibid 2159).

2.3.5 Onset Frequency and the Extent of F1 Transition

Another possible correlate of voicing to be considered is the behaviour of the first formant. In Lisker's (1975, 1547) words "significant effects [were found] to depend on the behavior of the first formant (F1) frequency immediately following voice onset."

Similarly to Lisker, Benkí (2001, 2) argues that "the timing relation between oral release and laryngeal vibration produces […] acoustic cues, primarily manifested in differences in formant transitions." The author also suggests that during the production of voiced plosives in Consonant-Vowel (CV) patterns, the presence of the fundamental frequency right after the release of the plosive makes the rise of the F1 from the consonant to "the vowel steady-state frequency" audible (ibid 2-3). The F1 for the voicing onset of the CV patterns, in which the consonant is a voiceless plosive, is much higher than that of the word with a voiced plosives, "[s]ince voicing onset occurs much later," the F1 is excited when "the CV transition […] is close to the vowel steady-state configuration" (ibid 3). This particular delay is referred to as "F1 cutback" (ibid 3).

Lisker (1975, 1547) suggests utilizing a complement to VOT, which was proposed by Stevens and Klatt (1974). Lisker calls this complement "voiced

transition duration" ("VTD"), and according to the author it constitutes "transition duration minus VOT" (Lisker 1974, 1547). Few difficulties are, however, listed by Lisker, namely that F1 cutback "is simply very hard to measure [and] it is neither all that easy to fix the time at which F1 […] reaches full amplitude," also the F1 amplitude stability is questionable (ibid 1548). The author's findings suggest that VTD is a factor highly influence by context and thus "F1 transition is not a requirement for stops to be heard as /b, d, g/" (ibid 1550).

It has been noted that there may sometimes be overlap between F1 and F2 patterns, which primarily arise from different sizes of vocal tracts, which is prominent especially when we compare the vocal tracts of children and adults (Nearey 1989, 2088). The author speculates that by "reformulating absolute formant frequencies as proportions of a speaker's formant frequency range" may solve such problem. (ibid 2088)

2.3.6 Aspiration and VOT

Discussing the differences between voiced and voiceless contrastive pairs of plosives, we have to make a reference to the term "aspiration". Acoustically speaking, aspiration manifests itself by the presence of turbulent glottal noise and delayed voice onset time (Skarnitzl 2011, 65). According to Ladefoged (1971, 9), the feature of aspirated-unaspirated plosives is concerned with "the state of the vocal cords during and immediately after the release of an articulatory stricture."

During unaspirated sounds, if the sound in question is voiced, the vocal folds vibrate throughout the closure and the release period. On the other hand, if the sound is voiceless, the vibration of the vocal folds starts at the release point of the constriction (ibid 9).

Focusing on the duration of aspirated sounds, the vocal folds are in a voiceless position during the period of closure and the release. In addition, there is a time lag between the release of the consonant and the onset of voicing for the succeeding vowel (ibid 9).

According to Dixit (1987), aspiration is conditioned by three essential requirements: open glottis, unimpeded supra-glottal vocal tract and stronger air flow rate through the glottis.

Voice onset time (VOT) is a more precise term, which is arguably able to adequately differentiate the phones, such as /p, t, k/ from /b, d, g/ and can

effectively correlate with the distinction of fortis/lenis phones. The term is useful mainly when plosives occur word-initially. VOT became popular in 1960s, even though it had been coined in the late 19th century. During the 60s, phoneticians became more concerned with the search of a measurable property.

Lisker and Abramson were one of the most important phoneticians who argued that it would be "feasible to relate the degree of voicing of a stop to the time relation between the burst and the onset of pulsing" (Lisker and Abramson 1967, 2). They investigated both citation forms of words and also words in connected speech. In the former they found that "this measure of voice onset time completely separates the two phonemic categories" (1). Whereas in the latter case "the separation is less sharp; there is some overlap along the dimension of voice onset time" (ibid 1).

Phonologically speaking, there is "evidence in favour of there being only three [modal] values" (voiced; voiceless unaspirated and voiceless aspirated plosives), strongly suggested by the fact that no language contrasts more than three homorganic plosives (Cho and Ladefoged 1999, 226). Nevertheless, VOT is a continuous measure and languages may arbitrarily choose the values for the contrast, therefore the actual phonetic contrast may not concord with the phonological one (ibid 226). In case the voicing starts before the release of the plosive, the plosive is described as *prevoiced*, with the VOT going to the negative values. On the other hand, if the voicing starts immediately or shortly after the release of the plosive, we describe it as *short-lag*. The plosive is described as *long-lag,* if the voicing for the succeeding vowel is postponed by a period of voiceless noise.

Languages that correlate the phonological voicing with the phonetic one base their contrast on phonetically voiced and voiceless plosives. Throughout the duration of those voiced plosives there is a noticeable voice bar within the closure period. In this case the plosive is said to be prevoiced with negative VOT (ibid 64-65) The contrasting voiceless plosive, on the other hand, occurs with little or no discernable aspiration, but following Van Alphen and Smits[5] (2004, 456), the VOT is still considered positive. The contrast mentioned in this paragraph is illustrated by Figure 7.

[5] This study was concerned with the voicing contrast of Dutch initial plosives. Dutch contrasts phonetically voiced (prevoiced) and voiceless plosives.

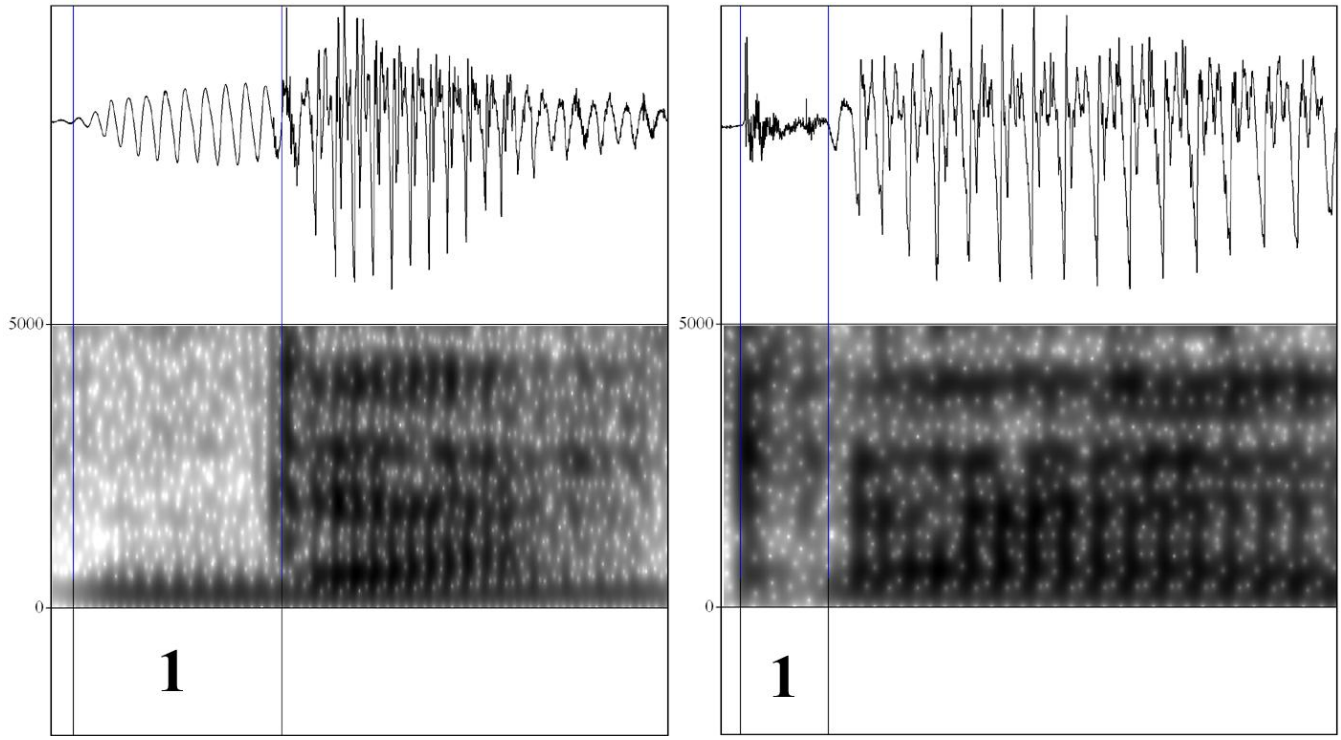

Figure 7 represents the voicing contrast of most of the languages in which the phonological contrast correlates with the phonetic one. The picture shows waveforms and spectrograms of the first syllables /da/ of the Czech words **datel** "woodpecker" (on the left) and the first syllable /ta/ of the Czech word **tady** "here" (on the right), as pronounced by a native female speaker of Czech. Focusing on the left figure, the left line of the number 1 marks the beginning of the prevoicing and the right line marks the release of the plosive. Focusing on the right figure, the left line of the number 1 marks the release of the plosive and the right line the onset of the following vowel.

In case of no such phonological-phonetic correlation, the contrast may be based on phonologically voiced but phonetically voiceless plosives and phonologically voiceless plosives, which manifest themselves with a long period of turbulent noise producing a delay to the onset of the following vowel. Figure 8 illustrates such contrast.

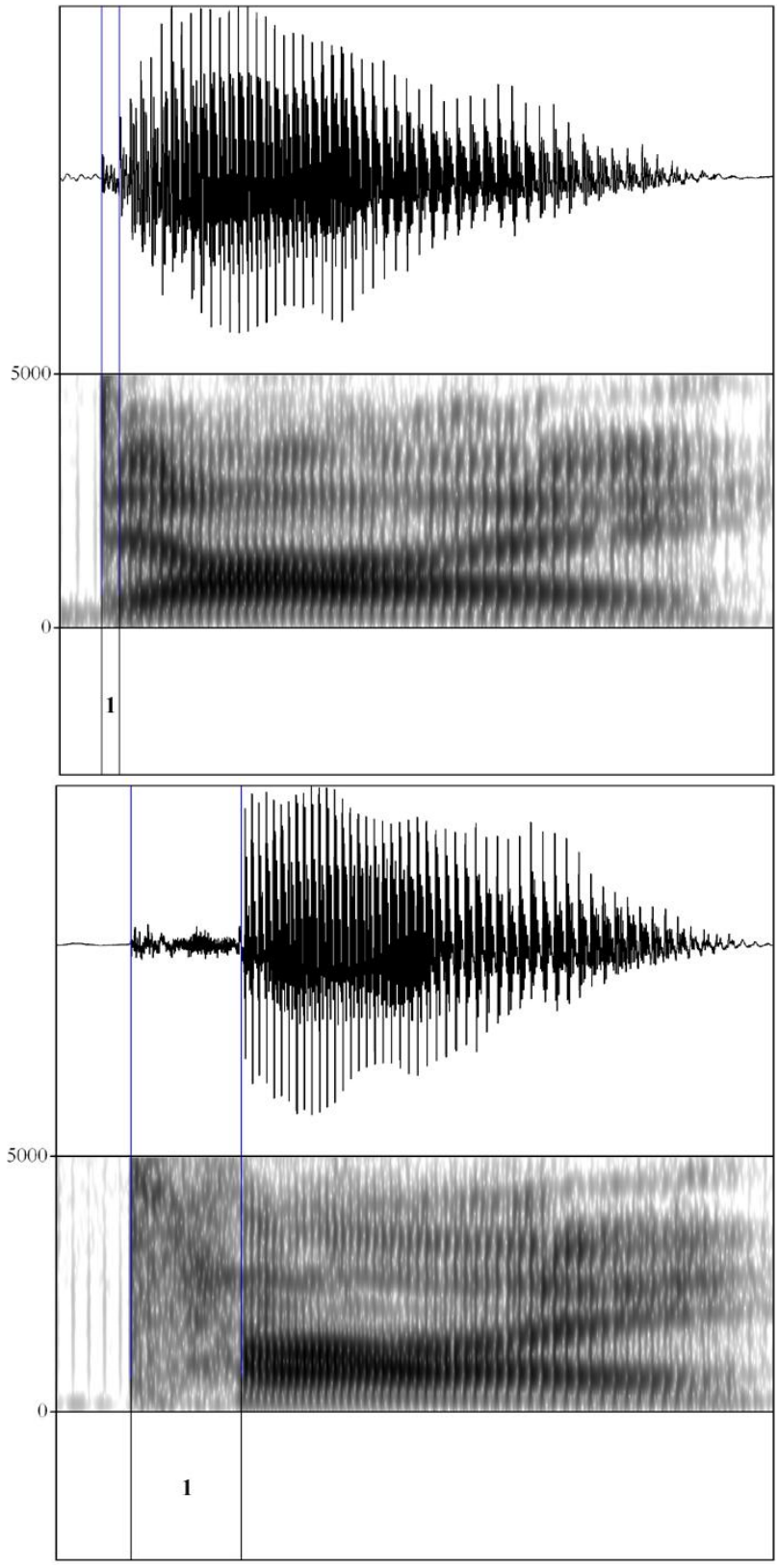

Figure 8 shows the waveform and spectrogram of English words **die** /daɪ/ (top) and **tie** /taɪ/ (bottom), as pronounced by a male native speaker of British English. In both cases the left line of the number 1 marks the release and the right line the onset of voicing.

2.3.7 The Intensity of the Burst Noise / Aspiration

One of the cues, possible to signal the voicing distinction could be the relative amplitude of aspiration noise. Repp (1979) argued that the cues considered as "primary," such as duration and frequency of aspiration, receive much more attention than those such as relative amplitude and bandwidth. However these latter "secondary" cues "are likely to play an important role in perception" (Repp 1979, 174).

The author was concerned with "the amplitude relationship between the aspiration noise" and the following periodic portion of the vowel (ibid 174). Repp tried to establish whether this relationship formed "a cue for the perception of the voicing distinction in syllable-initial stop consonants" (ibid 174). Focusing on English, the author suggested that due to the contrast between the phonologically /voiced/ and /voiceless/ plosives in this particular language, it is possible that "the primary perceptual cue is not the abstract temporal property of delay in voicing contrast […] but the presence and amount of aspiration during that delay" (ibid 174). This suggests that increasing aspiration amplitude in contrast with the intensity of the succeeding vowel would lead to an easier recognition of a sound as a part of the voiceless category (ibid 175). The experiment revealed the perceptual significance of the complete noise portion, not only of the aspiration itself, but also of the intensity of the burst. According to the author this may be explained by the laxness of the burst observed within the production of voiced stops (ibid 187).

3. FACTORS AFFECTING VOT

This chapter aims to list the possible factors that were found to have an influence on VOT. As we will be concerned with measuring VOT we have to be aware of the systematic variation it undergoes and the factors that cause it to vary. The main factors are presented in the following subchapters, i.e. the effect of place of articulation, gender influence, the context of the plosive and the second language experience.

3.1 The Effect of Place of Articulation

As I mentioned previously, the duration of the closure decreases in this order, according to the place of articulation, /p/ > /t/ > /k/. VOT, on the other hand, decreases in the order of /k/ > /t/ > /p/ and therefore works contrary to the duration of closure (Duběda 2005, 86). Such an influence is also observed by Cho and Ladefoged (1999, 208), who also found that "the further back the constriction, the longer the VOT". Duběda (2005) makes 3 arguments as to why the duration VOT, with respect to a particular place of articulation, works in this manner.

Firstly, he mentions the aerodynamic processes. The onset of voicing needs a certain level of trans-glottal pressure and for the constrictions formed more in the back of the vocal tract, the air cumulates behind this constriction and consequently the more the back the plosive is, the more time needed to achieve the aforementioned level of pressure.

Another principle concerns with the bio-mechanic properties of jaws. According to this principle, the articulators that are located closer to the front of the vocal tract, and thus further from the axis of rotation, come apart faster and therefore the air accumulated during their closure is also released faster, compared to the plosives formed more back in the vocal tract.

The author also lists the principle of the extent of the articulatory contact area. The parts of the main active articulator, namely the tongue, that are at the front, such as the tip or blade of the tongue, do not form as extensive constriction as the back part of the tongue (ibid 86). The parts of the tongue are presented in Figure 9.

.

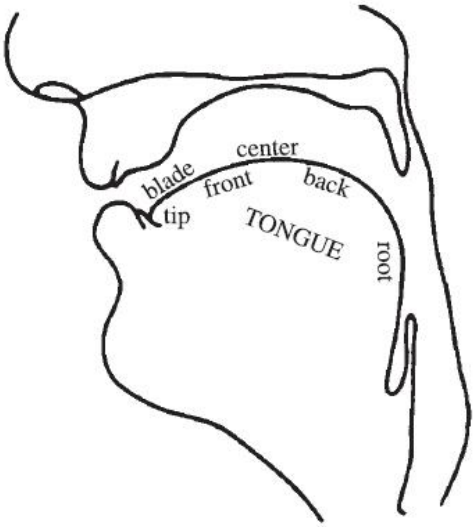

Figure 9 represents the parts of the tongue (adapted from Ladefoged 2011, 10).

	CLOSURE [ms]	positive VOT [ms]	DURATION OF THE PLOSIVE [ms]	PERCENTAGE OF CLOSURE [%]	PERCENTAGE OF VOT [%]
p	92,4	16,5	108,9	84,5	15,5
b	70,2	10,6	80,8	86,9	13,1
t	75,2	20,0	95,2	78,5	21,5
d	47,2	10,0	57,2	81,5	18,5
c	64,7	46,1	110,9	58,4	41,6
ɟ	49,1	25,4	74,6	66,0	34,0
k	75,1	32,1	107,2	69,9	30,1
g	59,8	19,5	79,3	75,3	24,7

Table 3 represents the duration of Czech plosives (adapted from Machač, 2006, 144).

The findings of Duběda (2005), and Cho and Ladefoged (1999) are consistent with Machač (2006), whose research was concerned with temporal and spectral structure of Czech plosives. Table 3 indicates that the percentage of the closure duration is the highest for the bilabial plosives and decreases when the constriction is made more back in the vocal tract. Contrary to the closure, the burst is the longest for the velar plosives and decreases the more front the constriction is. There is, however, an exception, namely the Czech palatal plosives /c, ɟ/ go against this universal tendency. Machač explains this exception by stating that the palatal stops have a larger contact area than other plosives. The author also found that "voiceless bilabial plosive /p/ and velar /k/ are about 1/3 longer

than their voiced counterparts /b, g/, voiceless palatal plosive /c/ is approximately 1/2 longer than voiced palatal plosive /ɟ/ and voiceless alveolar /t/ about 2/3 longer than the voiced one /d/" (ibid 143).

Van Alphen and Smits (2004, 486) found that among the speakers the production of prevoicing was more frequent for labial plosives than for alveolar ones. The results of the study suggested that "the listeners rely more strongly on prevoicing for labials than for alveolars," since the error rate of voicing identification was higher for the labial tokens produced without prevoicing than for the alveolar ones (ibid 486). The authors propose that the voicing distinction for alveolars is communicated more in the burst which is stronger and longer than for the labials. It was also suggested that the place of articulation of /t/ and /d/ was slightly different.

3.2 Gender Influence

Morris (2008, 309) argues that one of the factors that can "affect speech acoustics, including VOT, is speaker sex." However, the logic for the gender influence is related to the previous section, namely the author speculates that the difference between genders could be due to the size of the vocal tract. Van Alphen and Smits (2004, 464-465) found female speakers to produce less prevoiced tokens than males. The authors account this to the vocal tract size as "men tend to have larger vocal tracts than women […] and therefore the supraglottal pressure rises less quickly in the former," which facilitates the production of prevoicing (ibid 464-465). The difference in the duration of prevoicing was not significant (ibid 465).

Considering aerodynamics, the result of the smaller size of the vocal tract would manifest "in longer VOTs because of the additional time required to achieve the transglottal pressure difference as occurs for the smaller area of velar plosives" (Morris 2008, 309). The author also argues that there is a greater precision of articulatory movements found in the speech of women which "could be related to longer VOTs for voiceless plosives or the wider differences between voiced and voiceless plosives" among women (ibid 309). The author suggests that it is difficult to consider the gender difference as a factor, which could influence the VOT, since previous studies conducted in this field vary considerably. Studies, such as Koenig (2000) found that females produced longer VOTs than males for voiceless consonants. This study was supported by Ryals et al. (1997), who found,

besides the longer VOT of female speakers in voiceless plosives, shorter negative VOT for voiced plosives in male speech. On the other hand Smith (1978) found male/female differences for voiceless plosives. Moreover, he also found longer positive VOTs among males for voiced plosives. This is consistent with Peng et al.(2014, 74)[6], who found male speech to consist longer VOTs in unaspirated plosives, furthermore female speech was found to consist of longer VOTs in aspirated plosives. This tendency was also found in Robb et al. (2005, 130-131). On the other hand Morris (2008, 315) found no significant influence of gender on VOT. He argued that any differences found in his study "may be more closely related to individual differences than sex-based ones" (ibid 315). According to the author, it is more necessary to have vowel environment, place of production and speaking rate under control (ibid 316).

To sum up, the investigation made into the field of gender influence on VOT offers contradictory results and it is thus difficult to consider the gender difference as a factor, which could influence the VOT.

3.3 Contextually-Driven Influence

One of the significantly influential factors is whether or not the plosive appears in a stressed or unstressed syllable. According to Lisker and Abramson (1967, 15) the stress is reflected in the VOT measure. In English, aspiration is always present when a voiceless plosive is word-initial and in a stressed syllable, whereas in an unstressed syllable the aspiration is not present (ibid 15). Results of Lisker and Abramson's research present a couple of generalizations. Firstly they suggest that "stressed /ptk/ tend to be produced with longer delays in voice onset than do unstressed /ptk/" (ibid 16). Furthermore the voiced plosives /bdg/ without the phonetic voicing go against the inclination of /ptk/, "they tend to have greater VOT values in unstressed than in stressed position" (ibid 16). In addition, /bdg/ are less likely to be produced with unbroken voicing. Consequently the demarcation between the contrastive voiced-voiceless categories is a bit fussy in unstressed occurrences (ibid 16). The author also found that the mean voice onset time for /ptk/ "is about four times greater in isolated words than in sentences"

[6] The focus of this study was to investigate voice onset time of initial stops in Mandarin and Hakka, in particular the effect of gender differences. The said languages contrast phonetically voiceless unaspirated and aspirated phonemes.

(ibid 17). We can account this to the fact that stressed plosives are already very much reduced in sentences (connected speech), which as a result seems to limit the effects of the stress (ibid 17).

Another significant factor, with the possibility to influence the VOT values is the height of the succeeding vowel. Yavas (2007, 493) was concerned with this influential factor. He suggested that it is possible to observe longer VOT when plosives are followed by more high vowels compared to the plosives found in low vowel context. The reason that the author suggests for this phenomenon is that the effect "is related to abruptness of the pressure drop […] [as] [h]igh vowels have more obstruct cavity than low vowels" (ibid 493). That being said, if the tongue anticipates the position for the following high vowel, the air pressure will arguably lower more subtly, whereas for the plosives followed by a low vowel, we can expect the release of the air pressure to be more abrupt (ibid 493). The author confirmed the hypothesis only for bilabial and velar plosives, yet for the alveolar plosives it was not confirmed (ibid 496-497). The tendency of a longer VOT after high vowels is consistent with Wildermuth (2006, 258), who found that "VOT increases significantly as the subsequent segment changes from a low vowel to a high vowel".

3.4 Potential influences of second language experience

Due to the arguments discussed in this subchapter, the choice of the participants in the current study (Chapter 5) was set to include only speakers with minimal exposure to L2.

It has been also well established that L2 may have an influence on L1, in the field of phonology. For instance Pavlenko (2000, 179) suggests that "the human perceptual system remains somewhat flexible throughout the life course and carries out modifications in response to changes in sensory input." Leather and James (1996, 279) also suggested the "restructuring of the acoustic-phonetic space." This means that the parametric values for L1 may differ for monolinguals compared to speakers who acquire second language, a movement towards L2 values may be observed (Pavlenko 2000, 179). Such phenomenon may ultimately lead to the attrition of L1, that is the loss of one's native language. The suggestion made in the previous sentence is supported by Major (1992, 204), who found that "one's language is not a fixed and stable system but rather a fluid and changeable

one that is highly subject to the influence of a well-developed second system." A study, concerning with the shift of language among children belonging to a ethnic minority in the U.S., added that the younger the children were, when they started learning L2 (English), the greater the effect of L2 on the patterns of language use in their homes (Wong and Fillmore 1991, 341). The author thus, similarly to Major (1992), suggested a partial or complete loss of native language when the immigrant children learn English (Wong and Fillmore 1991, 341).

Even though VOT is only one feature that characterizes the production and perception of speech, Pavlenko (2000, 180) argues that VOT values play a crucial role as correlates of "overall judgements on the 'nativeness' or 'accentedness' of one's speech". One of the studies that found a possible influence of L2 on L1, as far as VOT is concerned, was Van Alphen and Smits (2004) who examined the production and perception of Dutch word-initial stops. The findings suggested asymmetric perception of voicing, namely "the presence of prevoicing alone provides enough evidence for a listener to be sure that the plosive is voiced," on the other hand, when prevoicing was missing it did "not provide enough evidence for a listener to categorize the plosive as voiceless" (Van Alphen and Smits 2004, 486). The study also presented somewhat contradicting results since "prevoicing was absent in 25% of voiced plosive productions across 10 speakers" (ibid 486). Owing to other perceptual cues, for some of these plosives without prevoicing, the misperception didn't occur. The authors speculated that there might have been a certain influence of English which could cause this absence of prevoicing. The subjects, who were to identify the plosives, had received 6 years of formal English education, as well as having been exposed to English in the television and radio (Van Alphen and Smits 2004, 487). Such influence was also noted by Major (1992, 201), who also found native speakers of American English who immigrated to Brazil in their adulthood to change their VOTs "in the direction of native Portuguese". The shift of the participants' VOT was also evident for the participants whose proficiency in Portuguese was far from perfect (ibid 201).

4. SPEECH IMITATION

It has been shown that when two interlocutors talk to each other, their speech tend to become more similar over time. However in cases where one of the

interlocutors dislikes or wants to define himself against the person with whom he is conversing, it may happen that his speech becomes more dissimilar. The tendency to become more similar with the ambient speech was noted by Sancier and Fowler (1997, 422), who concerned with the influence of L1 on L2 or vice versa. Consequently, we may say that articulatory gestures tend to alter according to the ambient language or dialect.

When we talk about imitation, we have to mention the fact that there is variability in persons' speech patterns, nevertheless, the interlocutors are able to surmount such differences and classify the sounds in the same phonemic category as was intended by one of the interlocutors (Pardo 2006, 2382). More importantly, the differences are also to be found in various productions of a word, pronounced by one speaker, the productions will not match exactly at acoustic-phonetic level (ibid 2382). This may lead us to the fact that "phonetic level imitation is likelier to be graded and inexact, rather than perfectly compliant with the acoustic, articulatory and phonetic detail of the spoken model for imitation" (2382). This suggestion is supported by Krauss and Pardo (2004). In addition, they suggest that "identity may work for descriptions of relatively abstract syntactic and lexical [...] representation, but not for representations at the phonologically level [...] [since] repeated phonetic elements (even within the same talker) are not physically identical" (Krauss and Pardo 2004, 203).

4.1 The Automaticity of Imitation

To proceed any further, we have to first note the difference between speech imitation and convergence. General imitation is said to be a process during which a plain mimicry can be observed, one person simply "copies the actions of others" (Garrels 2006, 49). This is supported by the automatic theories of speech imitation, which suggest that there is no space for any factors to influence speech imitation, production is driven automatically by perception and leads to inevitable imitation (Pardo 2006, 2382).

Contrary to speech imitation, speech convergence is said to be less automatic, subject to many factors "including talker physiology, dialect, affect, and social/situational" (Pardo 2012, 754). Whether such social factor comes into play and to what point its influence reaches, was investigated by Babel (2009, 137). To account for the possible social condition of the imitation, participants

were questioned with regard to the degree of attractedness and the bias against the Afro-Americans or Caucasians, moreover they were presented with a photo of the person after whom they were to repeat the stimuli (ibid 138-139). The author found evidence against the automaticity of imitation, both the attractiveness and the racial bias had a slight influence on the degree of imitation. (ibid 138) The author, however, speculated that only divergence may be with certainty classified as a conscious process, because it is a clear social tool "used by talkers to manipulate the social distance between themselves and interlocutors" (ibid 126).

Such divergence from the talker's speech pattern was also discussed by Krauss and Pardo (2004, 204). According to them, throughout the verbal interaction the degree of convergence varies with respect to different participants (ibid 204). That is to say when talkers interact with each other, they may converge to a certain degree in some features whereas they may not show convergence with their interlocutor in other features or may even diverge from them instead (ibid 204). Nevertheless, such departures from convergence are unlikely to be random. More precisely, they "frequently [...] reflect social processes that are fundamental to the interlocutors' interpersonal relationship and the ways they define the interaction situation" (ibid 204). Those departures may be accounted to ethnic identity and accentedness, causing to diverge in speech when the talker is of a different ethnic background. Furthermore the convergence was also found to be connected with the social status of the listener. A symmetry in pitch convergence was found "between a talk-show host and his guests [depending] on the guest's status relative to that of the host [...] higher-status guests changed less than their lower status counterparts" (ibid 204). This is consistent with Pardo (2006, 2384), who suggests that "dialect formation and change [...] are influenced by social relationships between interacting talkers."

To sum up, it seems that phonetic imitation is to be to some degree automatic, but certain factors, such as dialectal, affectional or social come into play.

4.2 Theories in Speech Imitation

According to theories of speech perception, there seems to be a perceptual problem of speech variability, the listeners have to surmount this variability in order to understand (Goldinger 1998, 251). Such differences can be accounted to

the distinctions in vocal tracts, glottal waves, speaking rate and dialectal differences of the speaker (ibid 251-252). Speech is normalized at perception, as phonetically redundant, surface information is filtered out. According to Halle (1985, 101), when people learn new words, they practically "never remember most of the salient acoustic properties that must have been present in the signal that struck our ears". This seems to support the theory of abstract representation of speech sounds. It was, however, suggested that even though normalization is said to occur very early in the process of processing a speech sound, the voice information remains, more precisely, Green et al. (1991, 533) suggests that "neutralization of talker differences for the purposes of phonetic categorization does not result in a loss of detailed information about the talker."

Alternative to normalization, Goldinger (1998), argues for the episodic models of speech representation, according to which words should match to episodic memory traces directly, without the need to normalize the speech sound (Goldinger 1998, 264). This was also found by Jacoby (1983), who suggested a non-analytic behaviour of word perception, being compared to prior episodes. Due to the fact that Goldinger (1998) aimed to "assess the benefits of an episodic view," he used a non-hybrid episodic model of Hintzman (1986), namely MINERVA 2 (Goldinger 1998, 254). In this particular model, the stored episodes are taken to a logical extreme "assuming that all experiences create independent memory traces that store all perceptual and contextual details" (ibid 254). The episodic memory system is, however, criticized by Pardo (2006), who suggests that this system "fails to explain the effect of a transient social factor on phonetic convergence" (Pardo 2006, 2390).

There is a continual debate between researchers concerning speech perception as having "acoustic, auditory, articulatory, gestural, or more abstract phonological parameters" (Pardo 2006, 2382). To determine whether speech perception covertly provides phonetic parameters for its realization, a close connection of perception and production was assumed (ibid 2382). Such link clearly evokes the motoric theory of Liberman and Mattingly (1985).

Another theory closely related to the motoric theory is direct realism. This theory is based on Gibson (1966), namely on his view concerning visual event perception. Gibson distinguishes the events and the informational media through which the event could be perceived. The direct realism differs from motor theory

mainly in allowing the idiosyncrasies of the speaker to play their roles, for instance "the speaker's vocal tract […] affect and other variables also structure acoustic speech signals distinctively" (Sheffert and Fowler 1995, 682).

4.3 L1 Changes after the Critical Period

It has been established that children are, after a certain period of life, not able to fully acquire phonological rules of a dialect which is spoken in a different area than to which they were exposed (Chambers 1992, 690). That being said, one might think that after a certain age, the production of one's speech might remain static, as was once acquired from the ambient of person's early stage of life. Nevertheless, the inability to acquire a completely new accent does not mean that speech is not subject to change, due to the influence of the new dialectal environment (Babel 2010, 437).

Having moved to California from Minnesota as an adult, Babel (2009, 1-2) noticed that her speech production "has lost many of its Minnesotan features." Speakers of native Californian dialect did not, however, deem her speech the same as their own (ibid 2). The research of Munro et al. (1999, 401) backs the aforementioned proposition of Babel (2009), as their findings suggest that adults are indeed able to acquire some of the phonetic aspects of a second dialect, if not phonological ones (Munro et al. 1999, 385). More specifically, the speech of native Canadians living in Alabama was evaluated by native speakers of both Alabama and Canada, with both groups judging the speakers as having "more Canadian-sounding than American-sounding" accent (ibid 401). Nevertheless, compared to native Canadians that did not move to Alabama, they were rated to have "an intermediate degree of American accent" by the two groups (ibid 401).

An effect of long term dialectal phonetic convergence was also found among "university students from the north of England," who studied in southern England for the period of 2 years (Evans and Iverson 2007, 3814). In particular, the authors concerned with phonetic convergence within vowels. After 2 years, the participants were found to be "able to change their spoken accent" (ibid 3824). However, the results revealed that "the changes in production were not accompanied by changes in perception" (ibid 3824). Most of the investigated 5 pairs of male roommates, interacting during the academic year, were found to converge in their speech "after approximately 1,5 months of cohabitation" (Pardo

et al. 2012, 196). What is more important, after spending about 3,5 months living together, all of them were found to converge (ibid 196). The investigated roommates were not previously acquainted. The data were collected at four intervals throughout the academic year and included item duration and vowel spectra. Those two assessments were compared to the test of perceptual similarity, conducted by a separate set of listeners. Parallel to Evans and Iverson (2007), Pardo et al. (2012) also investigated the possible shifts of vowel formants. Nevertheless, the results went against the ones of Evans and Iverson (2007). The author explained this dissimilarity by "the fact that most of [the students] were from New York state area and had already been attending Columbia University for at least one year" (Pardo et al. 2012). It is also necessary to mention that similarly to Babel (2009), the participants of the research of Pardo et al. (2012) were questioned about their perceived closeness.

Research made in this area suggests that we may be able to relate the fact that people may accommodate some features of their native dialect to those of the dialect which is being spoken around them, to the possible influence of L2 on L1.

One of the phonetic features was found to alter for a Brazilian Portuguese native, already behind the critical period for language acquisition, after spending 4 months in the USA (Sancier and Fowler 1997, 433). More specifically, the subject imitated a feature of native English (L2), namely the lengthening of VOT. Surprisingly, the findings showed that there was also a transfer of the lengthened VOT to the subject's L1 (Brazilian Portuguese). Brazilian Portuguese, which contrary to English, contrasts voiceless plosives with short VOT and voiced plosives with negative VOT (ibid 432). The results of this research seem to indicate that the process of phonetic convergence may occur on the level super-ordinate to that of a specific language.

To sum up, even though there is a certain period, variable among L2 learners, after which it is not possible to completely learn a foreign dialect or language for that matter, L2 seems to be able to cause L1 phonemes to shift in the direction of L2 phonemes.

4.4 Short Term Convergence

As it was mentioned in the previous subsection, subjects were found to converge within a long time span. However, for the research I'm going to present in the

following chapter, it is more important to mention convergence found to occur within a short term span, "both when individuals interact in conversation and when listeners rapidly repeat words over headphones" (a phenomenon called shadowing) (Pardo 2013, 2).

One of such prominent studies was made by Goldinger (1998), who proposed, based on the result of his study, that spoken words are perceived and produced on the basis of storage of specific episodes or exemplars of encountered words including all idiosyncratic details, which may later facilitate perception. Since speakers spontaneously imitated the phonetic properties of the perceived words, Goldinger concluded that the collection of exemplars represented in the speaker's memory serves as the basis for production and that recently experienced instances have a greater weight especially for words with lower frequency of occurrence of which speakers store fewer exemplars. The frequency of a word was also found to have an effect when the focus of the study was the imitation of vowels (Babel 2010). However, according to Pardo (2013, 2), despite the fact that this "procedure might be more likely to evoke phonetic convergence in shadowing studies, its use impedes the generalizability of the findings to more natural settings of language use," the low-frequency word criterion is not applicable in conversational setting.

Similar to Goldinger (1998), a significant post-exposure imitation effect was found by Goldinger and Akuma (2004, 719). However, this research differed greatly, because rather than using shadowing, they extended Goldinger (1998) by investigating printed-word naming. Contrary to Goldinger (1998) study, the participants did not immediately shadow the responses but there was a gap of 5 days between the exposure to the stimuli and the post-exposure recording. The imitation was observable in the results, even though such crucial change in the methodology was presented, which suggests that "reading aloud involves more than [just] print-to-sound conversion; it also taps memory for prior perceptual episodes" (ibid 719). This research thus argues for the exemplar theory, suggested by Goldinger (1998).

A study, conducted by Shockley et al. (2004), also found a convergence of the participants, shadowing the model speech. The authors replicated Goldinger (1998) but used different stimuli. The results, equally to Goldinger (1998), show better imitation after shadowing the stimuli (Shockley et al. 2004, 425). In spite of

the similarities, a couple of discrepancies also arose. In particular, Shockley et al. "did not find evidence that the number of prior exposures to the token words […] significantly influenced the reports of better imitation" (Shockley et al. 2004, 425). The authors speculated that the differences in methodology might have had an impact on the results. Unlike Goldinger (1998) who used 10 different model speakers, just one was included in the listening part of Shockley et al. (2004). In addition to the number of speakers, the stimuli were "more uniform in their phonetic property than were Goldinger's, in that all of [their] words began with voiceless stops" (ibid 425). Surprisingly, the arrangement of the presented tokens, while testing the AXB similarity, had a significant effect on the judgements imitations (ibid 425). That is, the perceptual judges marked the shadowed recording (B) as a better imitation, only when this token was presented last, after the stimuli (X) (ibid 425). Whereas, "when shadowed tokens were presented first (A) and baseline tokens […] last (B), the shadowed tokens were not identified as better imitations of the target tokens more often than chance" (ibid 424).

An interesting phenomenon was noted by Bourhis and Giles (1977), who found Welsh adults to diverge[7], by adapting a Welsh-accented dialect when being exposed to a speaker of Received Pronunciation. It is important to mention that the RP interviewer was concerned with Welsh in the modern era, more specifically, he called the vitality and function of the Welsh language into question (ibid). However, a second group, who attended Welsh language classes that were aimed to improve their further carrier, was found to converge with the RP speaker (ibid). The findings thus suggest that the convergence towards the interlocutor's dialect depends on whether or not the participants agree with the model speaker. A replication of this research was made by Babel (2010), who found all participants to accommodate, to various extent, towards the model speech (Babel 2010, 452).

The research of Babel (2010) concerned with phonetic convergence of the New Zealand English in response to Australian English. Contrary to Bourhis and Giles (1977), whose research may have had the perceptual judgments of convergence or divergence based on lexical items, syntactic structures or phonetic features, Babel (2010) tried to determine, whether the speakers were "using a slightly different pronunciation to socially distance themselves" (Babel 2010,

[7] This phenomenon refers to the increasing dissimilarity of the participant and the model speech.

441). Similarly to Bourhis and Giles, the research also included the positive and negative conditioning, that is the participants were randomly assigned to either the group that was flattered or insulted by the speaker of AuE. In addition, the author also added a task to evaluate the participants' biases towards New Zealand or Australia, by performing the "Implicit Association Task," (hereafter IAT) (ibid 442). Contrary to Bourhis and Giles, the feelings that the task stimulated, did not seem to alter the convergence, the participants were found to converge, regardless of whether they were assigned to the Positive or Negative Condition (ibid 452). Nevertheless, the biases towards a language, as examined by the IAT, found to have an effect. In particular, with increasing positive bias towards Australia the convergence was more prominent (ibid 452). This seems to suggest that convergence is to a certain degree automatic but some social factors may also come into play, such as the participants' pre-existing sentiments towards the speaker.

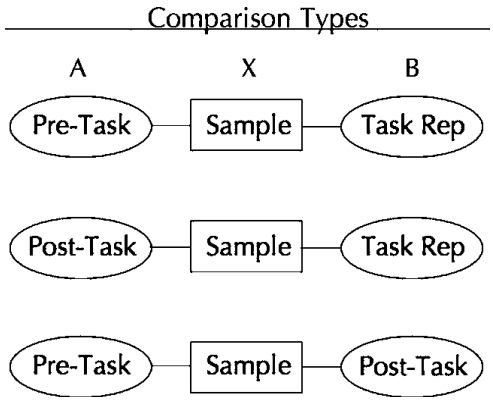

Figure 10 represents the trial structure for perceptual judges in the study of Pardo (2006). Each trial consisted of 3 repetitions of the same item. The first item and the last item were compared for similarity with the middle one. The sample item was either of the giver's or receiver's task session production. Task repetition featured the partner's repetition of the same phrase (Adapted from Pardo 2006, 2386).

Pardo (2006, 2382) focused on the conversational interaction and the degree to which they increased phonetic similarity. The author used a set of listeners who judged the recorded tokens according to Figure 10. She found robust evidence of phonetic convergence between the conversational partners. Evidence

for phonetic convergence was found already in the early period of the conversation and increased with the course of time, being prominent even after the interaction ceased (ibid 2388). The author suggested that female talkers may be more sensitive to the indexical features of the talker, which may lead to a more detailed recognition of the partner's phonetic forms (ibid 2388). It is somewhat surprising that the effect of gender and conversational role did not follow the expected predictions. More specifically, males were found to converge more often than females and givers were found to converge more often than receivers (ibid 2388). The fact that males converged more often thus suggests that "attention rather than absolute perceptual sensitivity is moderating these effects" (ibid 2388). Pardo et al. (2010) based their study on Pardo (2006) with the exception of adding the conscious imitation goal, measures of the talker's articulation rates and the pre-task, post-task vowel formants investigation to the focus of the study. They found imitation to occur only when receivers were instructed to imitate, while instructing givers did not result in imitation (with the exception of male givers, whose convergence was more or less the same as in Pardo (2006)) (Pardo et al. 2010, 2261). The findings for both the articulatory rate and the vowel formants did not suggest a faithful imitation (ibid 2261).

4.4.1 Nielsen's Crucial Research (2011)

The main research upon which the present study is built, is the one conducted by Nielsen (2011). The research has a certain base in Shockley et al. (2004), namely the manipulated feature being VOT, owing to the simplicity for constructing the stimuli. The author also uses the proposition of the exemplar based on Goldinger (1998), that is the fact that stronger specificity is expected for the words with lower frequency of occurrence. Additionally, the exemplar view also suggests that with increased immediate repetition, the imitation tends to be more salient, thus target words to which the subjects were exposed before, as opposed to the novel words, would receive a stronger specificity effect. The AXB perceptual assessment, utilized by both Goldinger (1998) and Pardo (2006) was not employed. According to the author "overall perceptual assessments integrate multiple acoustic phonetic dimensions and thus provide more holistic" and for that matter more powerful measure of imitation (Nielsen 2011, 133). However, the author is more concerned with what is being imitated rather than possible

general tendency of imitation. The generalizability of the VOT imitation was tested, that is whether "imitation of positive VOT can be generalized to a different word or phoneme" (ibid 133). If so, it would indicate that the incoming speech is coded on the level of phoneme or feature, suggesting that the "unit of phonological representation [...] responsible for the imitation effect has to be smaller than a lexical word" (ibid 133). Two experiments were conducted, the first of which was concerned with the imitation of extended VOT and the second one with the imitation of truncated VOT. Due to the fact that truncating the VOT could neutralize the voicing contrast between the English /voiceless/ aspirated and the /voiced/ short-VOT unaspirated plosives, it was predicted that only the extended VOT would be imitated.

The stimuli for the production consisted of 150 words, in particular 120 test words and 30 fillers. Out of the test words, 100 began with /p/, out of which the participants were exposed to 80 of them (40 high and 40 low frequency words), 20 novel p-words and 20 non-exposure target words beginning with /k/. The threshold for the low-frequency words was set to be below 5 tokens/million and for high-frequency words to be above 50 tokens/million and was determined from both Kučera and Francis (1967) and CELEX2 (1995) (Nielsen 2011, 134).

Both of the experiments consisted of four blocks. The first block consisted of warm-up reading, in which the participants were presented the words on monitor and were asked to read the words silently, without pronouncing them. In the second block the participants were instructed to read the words aloud, producing the baseline recording of the same words, they read in the warm-up reading block. In the third block, participants were exposed to the model speech. In the last block, participants were instructed to produce the post-exposure recording, which consisted of the same words that appeared in the baseline recording and warm-up reading.

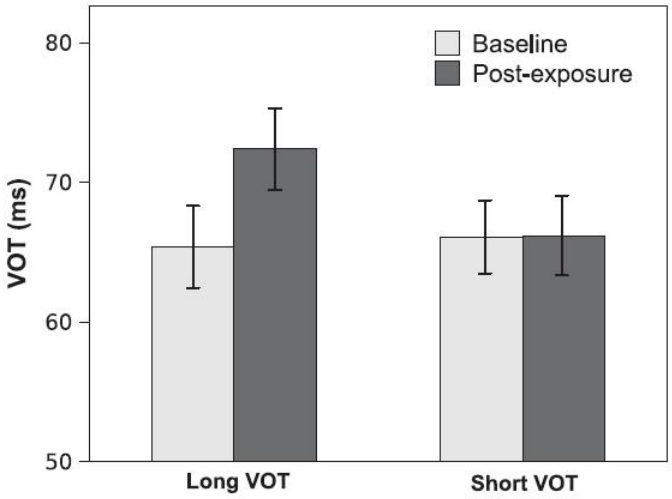

Figure 11 compares the baseline and post-exposure productions of both of the two experiments.

As it was predicted, we can see in the Figure 11 that the post-exposure recordings showed significant effect of imitation after the participants were exposed to the words with extended VOT. In contrast, words with truncated VOT showed no such effect. Judging from the Figure 12, the results also revealed that the "imitation of extended VOT was generalized at a sub-lexical level," as the VOTs increased in the post-exposure block even for the novel /p/-words and non-exposure /k/-words (ibid 136). Also the "magnitude of generalized imitation was larger for" novel /p/-words than for non-exposure /k/-words (ibid 136). We can thus conclude that the "extended VOT in modelled speech was coded at *both* phonemic *and* feature levels, and subsequently affected the two levels […] independently" (ibid 136). The effect of word-frequency did not seem to have an effect and thus both target p-high and target p-low words were combined into one group (ibid 135).

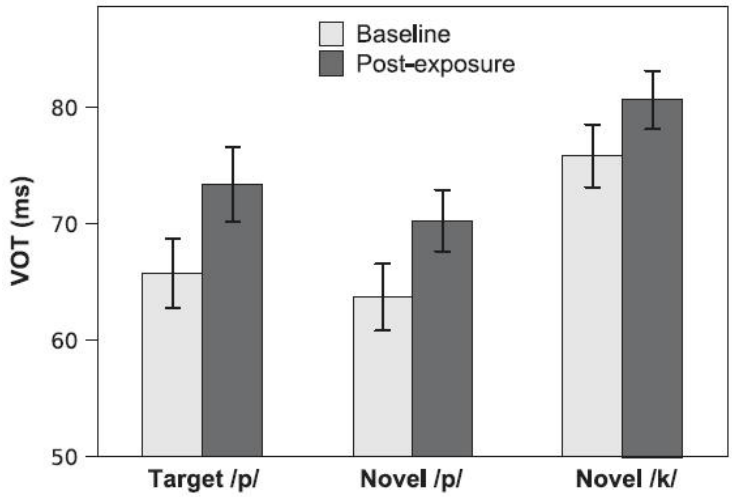

Figure 12 represents the baseline and post-exposure recordings of target /p/-words, novel /p/-words and non-exposure /k/-words.

5. THE PRESENT STUDY: DOES IMITATION INDICATE PHONOLOGICAL SELECTIVITY? THE CASE OF PREVOICING IMITATION IN CZECH

Even though, there seems to be extensive amount of studies, concerning with imitation in English (mainly on Accents of English, for instance Babel (2010) on NZE and Australian; Bourhis and Giles (1977) on Welsh dialect of English and RP; and convergence towards tokens with extended VOT, such as Goldinger (1998) or Nielsen (2011)), there seems to be a gap with respect to Czech.

As mentioned in previous chapter 2.3.6, phonological voicing contrast correlates with phonetic contrast in Czech. To be more explicit, following this established fact and focusing on voicing contrast of Czech plosives, the VOT in Czech is negative (the case of pre-voicing) in phonetically voiced plosives and short (positive) in case of phonetically voiceless plosives.

The English voicing contrast works, however, quite differently from that of Czech one. In case of English, the phonological contrast does not concord with the phonetic one. English thus contrasts phonetically voiced plosives (with zero or short VOT) and phonetically voiceless aspirated plosives (with long VOT). The aforementioned research of Nielsen (2011), found participants to converge only

when the stimuli (voiceless plosives) had an extended VOT, in the case of stimuli with reduced VOT, the convergence did not occur. Such phenomenon could be accounted to the preservation of voicing contrast. This seems to suggest that imitation may be phonologically selective.

Analogically to Nielsen (2011), our research was concerned with imitation of artificially extended and reduced VOT. Following Nielsen's logic, a prediction was made. In particular, we suggested that after being exposed to the artificially lengthened VOT (prevoicing), the participants would imitate this modification. The reduction, on the other hand, would not be imitated, due to the fact that it could weaken the voicing contrast between the phonetically voiced plosives (with negative VOT) and phonetically voiceless plosives (with zero or short-lag VOT). Analogically to Nielsen (2011), the research was divided into two experiments.

5.1 General Methods

The pilot study used only two participants, listed in Appendix 1. Only males were chosen analogically to Pardo (2006) and van Alphen (2004). According to Pardo (2006), phonetic convergence seems to be more likely when both the model speaker and the shadower are of the same sex. Van Alphen (2004) also found male speakers to produce pre-voicing more frequently. This was contrary to Nielsen (2011), who used both male and female participants.

In order to avoid influences from L2, the choice of the participants was limited to those who had minimal exposure to a non-native language.

Both of the participated were used in the Experiment 1 and 2, having a break of 20 minutes between the two experiments. The order of the two experiments was fixed for both of the participants, the Experiment 1 occurred before the Experiment 2. In the Experiment 1, participants were exposed to the stimuli with extended VOTs, whereas in the Experiment 2, they listened to the stimuli with truncated VOTs.

The stimuli, analogically to Pardo et al. (2012) who used native speakers of English without any hearing or speech impairment, were produced by native speaker of Czech.

The used material differed from Nielsen (2011), because the more front the constriction, the longer the pre-voicing. On the other hand, Nielson (2011) used /p/-initial words as exposure, novel /p/- and non-exposure /k/-words as target

material, owing to the fact that aspiration increases with the backness of the closure, that is, aspiration is shorter in /p/ and longer in /k/. Pre-voicing is very short in /g/ and therefore truncating such pre-voicing would not make a sound that is different from the commonly encountered sound. Therefore, we chose 28 /d/-words (out of which 10 were used in the production as the target words) as exposure, 10 novel /d/- and 10 non-exposure /b/-words as target words. Also 18 fillers were used as the exposure material (out of which 5 were used as target words) and 15 non-exposure fillers were used as the target material, as represented by the Table 4.

The stimuli for the exposure and production targets were constructed from Czech disyllabic words, the tested segment was followed by short or long vowel, high vowels were excluded. The stimuli were selected analogically to Goldinger (1998), who suggested that imitation of low-frequency words is more likely, compared to high-frequency words. Czech National Corpus (Syn 2010) was used and the word-frequency was set to 18 tokens per million. The concordance was created only of the words in their specific forms, not all of their forms were included. The full list of the used stimuli is presented in Appendix 2.

	EXPOSURE	TARGET	
		OLD	NEW
D short V	18	5	6
D long V	10	5	4
B short V	-	-	5
B long V	-	-	5
Fillers short V	-	-	9
Fillers long V	18	5	6

Table 4 represents the number of words that the participants heard (exposure) in the listening part and those that they produced (target) in the baseline and post-exposure recording. Target-Old are words that the participants heard in the exposure block and then produced in the baseline and post-exposure recording. Target-New constitutes novel /d/-words and non exposure /b/-words, not used in the listening part.

The procedure was based on Nielsen (2011), and consisted of four blocks, as presented by the Figure 13. In the warm-up reading, the participants read the randomized production list presented to them on a piece of paper. The second block consisted of baseline recording of word-initial target /d/, novel /d/ and non-exposure /b/-words. In the third block, the participants were exposed to the /d/-words with artificially modified VOTs, all the exposure stimuli were heard 5 times in random order. The last block consisted of post-exposure recording of target /d/-words, novel /d/-words and non-exposure /b/-words. In both the baseline and post-exposure recording, the participants produced 3 tokens of each word in random order.

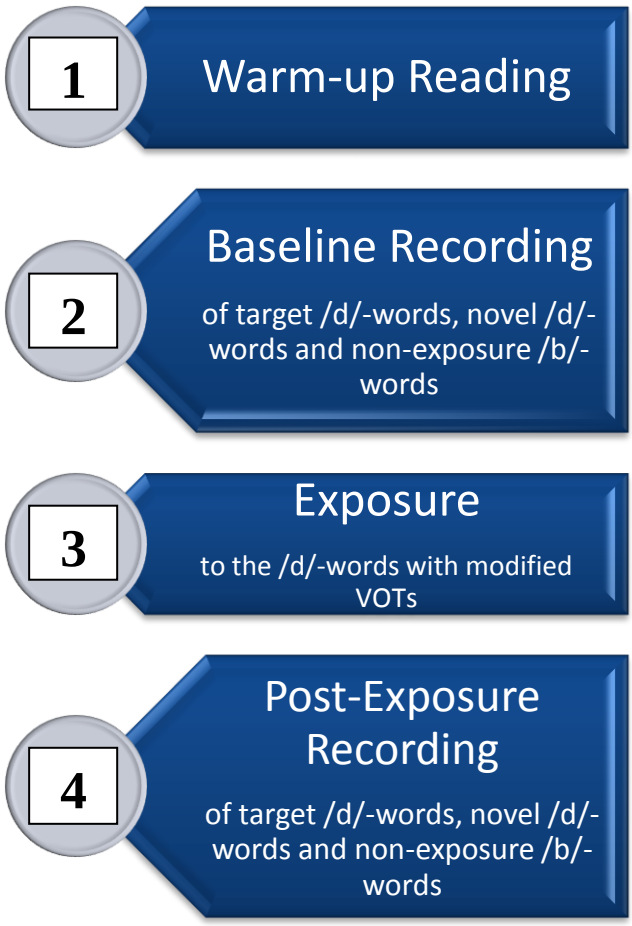

Figure 13 represents the procedure used in both the Experiment 1 and Experiment 2.

Following the methods of Van Alphen (2004), the data were annotated and measured in Praat by the students of Acoustic Phonetics class, adhering to strict rules: the beginning of the pre-voicing was set at the onset of noise after the

release of the plosive and the end of the pre-voicing was set at the beginning of the vocal fold vibration. Out of the 3 produced tokens, the most clearly articulated token was chosen by the students of Acoustic phonetics class.

5.2 Experiment 1 – Imitation of Extended VOT

The first experiment investigated, if the target speech with extended VOT of the employed /d/-initial word, would be imitated. It also aimed to find out if the extended VOT would be transferred to the new instances of /d/-words, not heard in the exposure block. Moreover we also focus on the possible transfer of the extended VOT to the non-exposure /b/-words.

5.2.1 Methodology

The used methods were already presented in 5.1 General Methods, in this case the exposure block consisted of the stimuli with artificially extended VOTs, according to the following form: **original*1.35 + 25 msec**.

5.2.2 Results

Factorial ANOVA was used to analyze the data, with VOT as the dependent variable. The independent variable constituted target /d/-words, the novel /d/-words and non-exposure /b/-words; and also the baseline and extended condition.
Factorial ANOVA revealed no significant effect of extended condition for neither the first participant [F (1, 54) = .49314, p = .48555], presented in the Figure 14, nor the second participant [F (1, 54) = 1.5862, p = .21329], presented in the Figure 15.

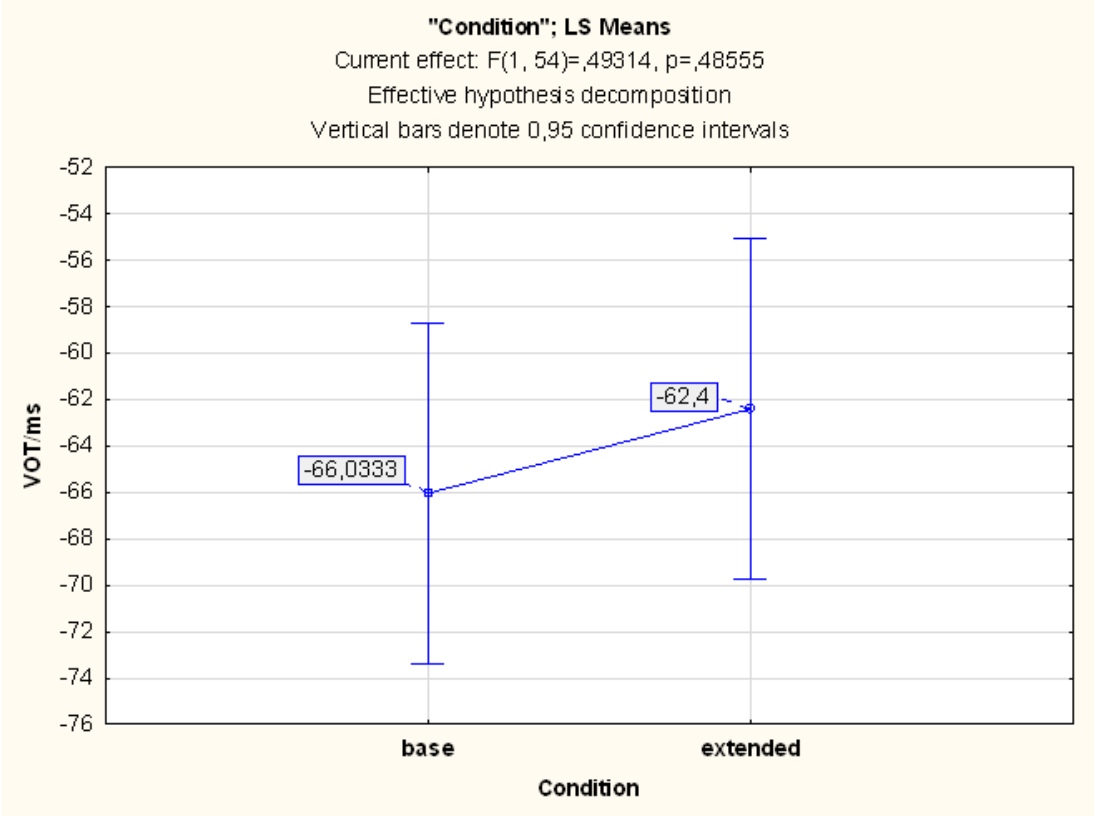

Figure 14 compares the VOT means of the extended condition and baseline recording of the first participant.

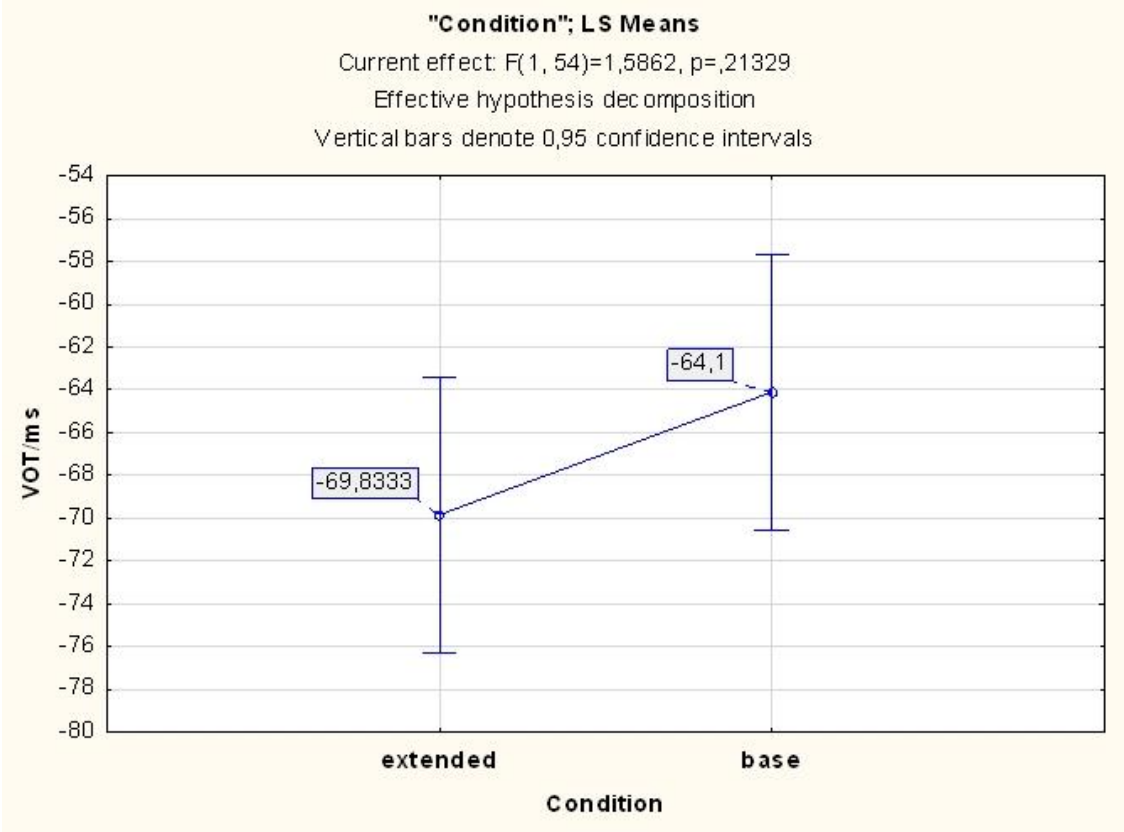

Figure 15 compares the VOT means of the extended condition and baseline recording of the second participant.

However Factorial ANOVA did reveal a slightly significant effect of exposure word for the first participant [F (2, 54) = 2.9872, p = .05882], the /b/-words were produced with longer VOT than found in /d/-words. The second participant showed even more significant effect of the exposure word [F (2, 54) = 6.3787, p = .00326]. The Figure 16 and Figure 17 represent those results.

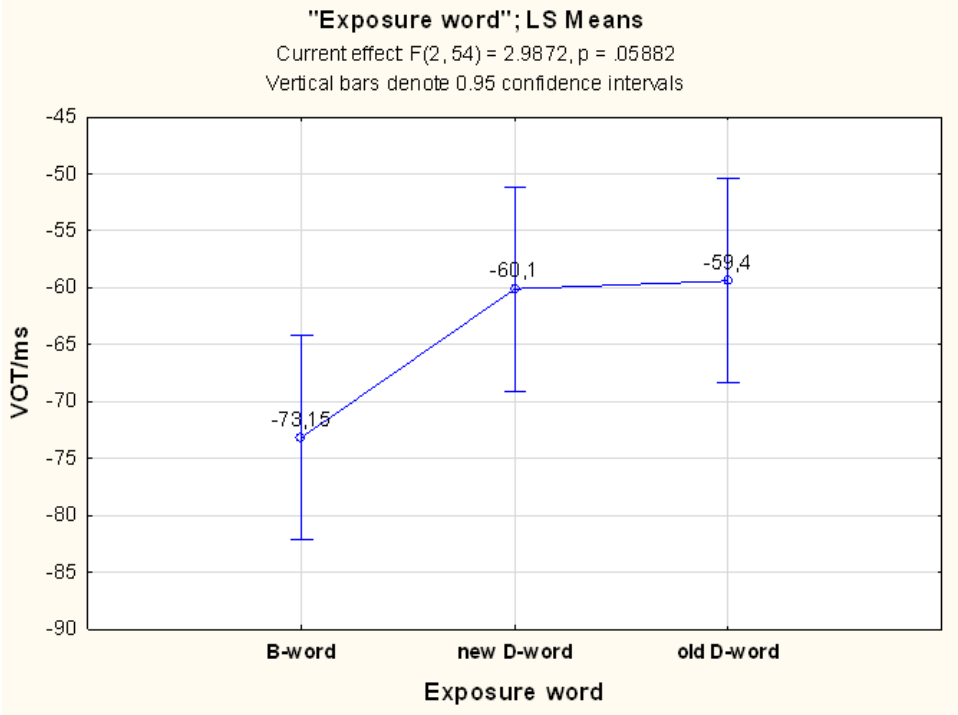

Figure 16 represents the effect of exposure word for the first participant. Old /d/-words were heard in the exposure block, new /d/-words and non-exposure /b/-words were not.

Figure 17 represents the effect of exposure word for the second participant. Old /d/-words were heard in the exposure block, new /d/-words and non-exposure /b/-words were not.

5.2.3 Discussion

Contrary to Nielsen (2011), the results of the current study did not reveal any significant tendency to imitate the extended VOT (prevoicing), the hypothesis was thus not proven. We might account this to the number of participants, the current study consisted of only 2 participants, which is not enough to make good statistical conclusions.

However one important finding was revealed. After being exposed to the extended condition a slightly significant effect of exposure word was found for the first participant and even more significant tendency for the second participant, namely the VOT was found to be longest in non-exposure /b/-words. Nevertheless a less significant tendency for /b/-words to be produced with longer prevoicing was found not only in extended condition, but also in the baseline recording, meaning it is more likely due to the universal characteristics of bilabial plosives to have a longer closure than the alveolar ones, mentioned in chapter 3.1.

5.3 Experiment 2 – Imitation of Truncated VOT

In this experiment, we aimed to investigate the linguistic selectivity of phonetic imitation. It was hypothesised that truncated VOTs would not be imitated, due to the fact that it could weaken the phonological contrast between voiced and voiceless plosives.

5.3.1 Methodology

The used methods were the same as presented in 5.1 General Methods, in this case the exposure block consisted of the stimuli with artificially truncated VOTs according to the following form: **original*0.15**.

5.3.2 Results

Factorial ANOVA was also used to analyze the data, with VOT as the dependent variable. The independent variable constituted target /d/-words, the novel /d/-words and non-exposure /b/-words; and also the base, extended condition.

The factorial ANOVA did not reveal a significant effect of truncated condition for the first participant [F (1, 53) = 1.2516, p = .26829], presented in the Figure 18. A surprisingly significant effect was found for the second participant [F (1, 54) = 10.958, p = .00166], presented in the Figure 19, which suggested that

the second participant extended his VOTs, after being exposed to stimuli with truncated VOTs.

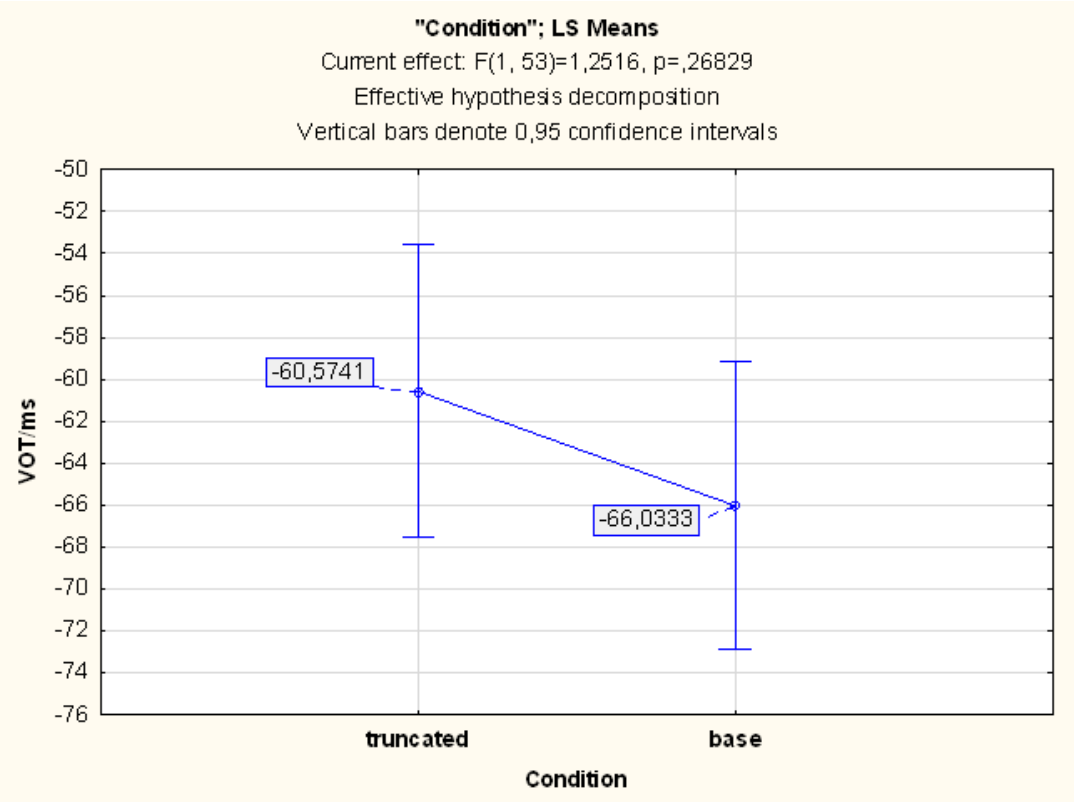

Figure 18 compares the VOT means of the truncated condition and baseline recording of the first participant.

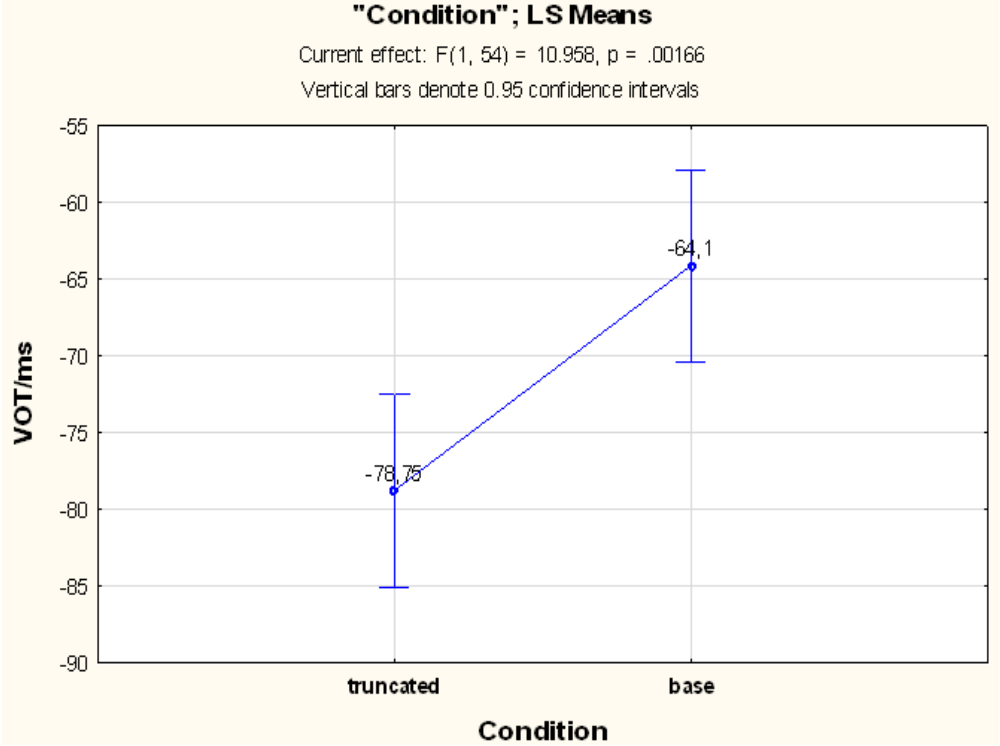

Figure 19 compares the VOT means of the truncated condition and baseline recording of the second participant.

5.3.3 Discussion

As we predicted, the results found no significant tendency to imitate the truncated VOT, owing to the fact that it could endanger the voicing contrast.

However, the post-exposure tokens, produced by the second participant are somewhat harder to explain. Surprisingly, this participant extended VOT after being exposed to the truncated condition. One of the possible reasons could be the fact that the participant might have either consciously or unconsciously perceived the tokens as having too short VOT and may have consciously or unconsciously tried to normalize his production. Another possible reason could be the articulatory rate. In the baseline recording, the articulatory rate of participant's speech might have been faster than that of the exposure. With faster articulatory rate, the time spent on production of individual speech sounds decreases. This suggests that the VOT is shorter in faster speech. Consequently, if the participant imitated the slower tempo, he would have to slow down and the prevoicing would thus become longer, and as such it could explain the results found in this experiment, exemplified by the Figure 19.

5.4 General Discussion

Unlike previous studies, which found participants to converge towards their model speakers, such as (e.g. Babel 2010, Nielsen 2011, Shockley et al. 2004), the present study does not indicate such tendency.

Basing our research on Nielsen (2011), we hypothesized that the extended VOT would be imitated, whereas the truncated VOT would not be imitated, due to the preservation of the phonological voicing. However, we found no significant effect of imitation in the expected direction. Arguably, one of the reasons for no such effect to be observed, could be the number of the subject, 2 participants are not enough to make good statistical conclusions. Another possible reason that might have influenced the present study could be the fact that both of the experiments were conducted on one day with only a pause of 20 minutes in between. The order of experiments was fixed, Experiment 1 was conducted first. Consequently the exposure material from the Experiment 1 could have still been present in participants' memory and might have interfered with the succeeding Experiment 2.

5.5 Further Research

The pilot study was conducted to investigate the imitability of VOT in Czech. However, it did not help to prove the hypothesis of imitating extended VOT and not imitating truncated VOT. Therefore there is still a lot to be done in this field.

In order to make better statistical conclusions, further research on the imitability of VOT should include more participants.

Also it would be interesting to use different participants for each of the experiments to avoid any interference from the preceding experiment, or at least have the order of the two experiments varied across the participants.

The articulatory rate could have crucially influenced the results in the Experiment 2, further research should also measure the articulatory rate, if such anomaly arises.

The current research used only male participants of native Czech. Nevertheless, due to the fact that studies concerning with gender influence on VOT offer no coherent findings, further research may also include female participants. The comparison of male and female VOT values would surely offer interesting data.

6. SUMMARY

Two aims are in the focus of this bachelor thesis. First focus is to summarize previous studies, concerning speech imitation. Secondly, it aims to compare the VOT of Czech and English. More specifically, it focuses on testing the hypothesis of VOT imitability among Czech native speakers, based on Nielsen (2011).

First part of the thesis offers general phonetic background, concerning consonantal categorization and the comparison of characteristics of Czech and English plosives. Moreover it also focuses on presenting the difference between phonetic and phonological voicing. In addition it offers some of the main phonetic cues, which may help categorizing the plosives.

The following part is concerned with VOT and the factors which might cause VOT to vary, such as the effect of the place of articulation, the influence based on gender, influence driven by the context or the second language experience.

Chapter 4 summarizes the studies in the field of speech imitation, it offers some basic theories and also presents the fact that L2 may influence L1 to a certain degree even after the critical period, sometimes it may even cause an L1 attrition, that is the complete loss of one's native language.

The last content chapter (chapter 5) focuses on the current study. It introduces the study and offers general methods for the two experiments. Both of the conducted experiments are presented with respect to their specific methodology, the findings they offer and the discussion of the results. Additionally, the general discussion is presented, followed by the suggestions for further study.

Based on Nielsen (2011), the main research question was focused on whether or not native male speakers of Czech would imitate the extended and/or truncated VOT. The hypothesis was made, namely if imitation were to take place, it would most likely occur after the exposure of the participants to the extended condition. The truncated condition should not have an effect, due to the fact that it would weaken the voicing contrast. Unfortunately, the research did not find

significant data for imitation in the proposed direction, neither in the Experiment 1, nor in the Experiment 2.

7. APPENDICES

7.1 Appendix 1

Subject	Gender	Age	Exposure to L2
1	M	26	Only on Vacations
2	M	31	Spent a few weeks in Germany and Romania

7.2 Appendix 2

	EXPOSURE	TARGET	
		OLD	NEW
D SHORT	18	5	6
D LONG	10	5	4
B SHORT	-	-	5
B LONG	-	-	5
FILLERS SHORT	-	-	9
FILLERS LONG	18	5	6

	EXPOSURE	TARGET	
		OLD	NEW
D SHORT	**18** daněk datle dacan daman datum danost defekt despekt	**5** datum dezert defekt dotek dozor	**6** datel davy dekolt dehet dovoz dostih

	delikt dezert dekor derviš dorty dotek dorost doklad dozor dotaz		
D LONG	**<u>10</u>** Dánsko dárce dárky dáma dávit démant déčko délka dómy dóže	**<u>5</u>** dáma dávit déčko démant dóže	**<u>4</u>** dáseň dávka démon dóza
B SHORT	-	-	**<u>5</u>** bacil basa bečet beton

			bodec
B LONG	-	-	**5** bázeň bádat bérec béčko bába
FILLERS SHORT	-	-	**9** salaš larva harfa řemen celer žezlo losos norma mosaz
FILLERS LONG	**18** sádra šátek záhyb fádně žába sádlo chátra žábry zápor léčka cévy céčka létat rébus lóže móda	**5** sádra šátek cévy céčka móda	**6** zákal cákat fábor léčit šéfka sólo

	nóta zóna		

8. SHRNUTÍ

Tato bakalářská práce si stanovuje dva hlavní cíle, a to za prvé shrnout předešlé výzkumy provedené v oblasti imitace řeči. Druhým předmětem práce je porovnání českého a anglického VOT (doba nástupu hlasivkového tónu). Přesněji řečeno, zaměřuje se na otestování hypotézy imitace VOT mezi českými rodilými mluvčími. Tato hypotéza je založena na výzkumu Nielsen (2011).

V první části se práce snaží představit základní fonetické znalosti, jako například na jakém základě jsou konsonanty zařazeny do kategorií. Součástí této kapitoly je také porovnání vlastností explozív češtiny a angličtiny. Také je zde zmíněn rozdíl mezi fonetickou a fonologickou znělostí. Zahrnut je i seznam hlavních fonetických podnětů, které mohou pomoci zařadit explozívu mezi znělé či neznělé.

Následující část se zabývá VOT a faktory, které mohou způsobit změny ve VOT. Mezi tyto faktory patří například místo artikulace, vliv pohlaví, vliv kontextu nebo znalost cizího jazyku.

Čtvrtá kapitola shrnuje výzkum v oblasti imitace řeči a představuje základní teorie v této oblasti. Zabývá se také vlivem cizího jazyku na jazyk mateřský, obzvláště po tzn. kritickém období. Je také známo, že cizí jazyk může ovlivnit jazyk mateřský do takové míry, že způsobí jeho úplnou ztrátu, tj. jazykovou atrici.

Kapitola 5, jakožto poslední obsahová kapitola, se zaměřuje na současný výzkum, seznamuje čtenáře s úvodem a představuje obecné metody použité v obou experimentech. Oba experimenty jsou pak představeny na základě jejich specifické metodologie, výsledků, které z těchto výzkumů vyplývají a následující diskuze o těchto výsledcích. Kapitolu uzavírá obecná diskuze k experimentům a návrhy pro případný další výzkum.

Výzkumná otázka této práce je založena na studii Nielsen (2011), tj. výzkum se snažil zjistit, jestli rodilí mluvčí češtiny mužského pohlaví budou imitovat prodloužené nebo zkrácené VOT. Hypotéza byla založena na principu

zachování znělostního kontrastu, tj. předpokládala se imitace pouze prodlouženého VOT, protože imitace zkráceného VOT by mohla tento znělostní kontrast narušit. Bohužel výsledky provedeného výzkumu platnost hypotézy neprokázaly. Imitace nebyla prokázána ani v jednom z experimentů.

9. WORKS CITED

Abramson, Arthur S., and Leigh Lisker. 1985. "Relative power of cues: F0 shift versus voice timing." *Phonetic linguistics: Essays in honor of Peter Ladefoged*: 25-33.

Baayen, R.H., R. Piepenbrock, and L. Gulikers. 1995. The CELEX Lexical Database (Release 2) [CD-ROM]. Philadelphia, PA: Linguistic Data Consortium, University of Pennsylvania.

Babel, Molly. 2009. "Phonetic and social selectivity in speech accommodation." Berkeley: University of California, dissertation.

Babel, Molly. 2010. "Dialect divergence and convergence in New Zealand English." *Language in Society*, 39, 437–456.

Babel, Molly. 2012. "Evidence for phonetic and social selectivity in spontaneous phonetic imitation." *Journal of Phonetics* 40, no. 1: 177-189.

Benkí, J.R. 2001. "Place of articulation and first formant transition pattern both affect perception of voicing in English." *Journal of Phonetics* 29, no. 1: 1-22.

Bohn, Ocke-Schwen, and James Emil Flege. 1992. "The production of new and similar vowels by adult German learners of English." *Studies in Second Language Acquisition* 14.02: 131-158.

Braunschweiler, N. 1997. "Integrated Cues of Voicing and Vowel Length in German: A Production Study." *Language and Speech* 40, no. 4: 353-376.

Chambers, Craig. 1992. "The design and implementation of the self compiler, an optimizing compiler for object-oriented programming languages." Stanford: Stanford University, dissertation.

Cho, T., and P. Ladefoged. 1999. "Variation and universals in VOT: Evidence from 18 languages." *Journal of Phonetics* 27, no. 2: 207-229.

Dankovičová , Jana. 1999. "Czech." *Handbook of the International Phonetic Association: A Guide to the Use of the International Phonetic Alphabet*. Cambridge: Cambridge University Press: 70-73.

Delattre, Pierre C., Alvin M. Liberman, and Franklin S. Cooper. 1955. “Acoustic Loci and Transitional Cues for Consonants.” *Journal of the Acoustical Society of America* 27, no. 4: 769-773.

Dixit, R. P. 1987. In defense of the phonetic adequacy of the traditional term ‘voiced aspirated’. In *Proceedings of the Eleventh International Congress of Phonetic Sciences* 2: 145-148.

Duběda, Tomáš. 2005. *Jazyky a jejich zvuky*. Praha: Nakladatelství Karolinum.

Elston-Güttler, K.E., S. Paulmann, and S.A. Kotz. 2005. “Who's in control? Proficiency and L1 influence on L2 processing.” *Journal of Cognitive Neuroscience* 17, no. 10: 1593-1610.

Evans, B.G., P. Iverson. 2007. “Plasticity in vowel perception and production: A study of accent chase in young adults. ” *Journal of the Acoustical Society of America*, 121, 3814–3826.

Fillmore, Lily Wong. 1991. “When learning a second language means losing the first.” *Early childhood research quarterly* 6, no. 3: 323-346.

Francis, A.L., N. Kaganovich, and C. Driscoll-Huber. 2008. “Cue-specific effects of categorization training on the relative weighting of acoustic cues to consonant voicing in English.” *Journal of the Acoustical Society of America* 124, no. 2: 1234-1251.

Garrels, Scott R. 2005. “Imitation, mirror neurons, and mimetic desire: Convergence between the mimetic theory of René Girard and empirical research on imitation.” *Contagion: Journal of Violence, Mimesis, and Culture* 12, no. 1: 47-86.

Gibson, J. 1966. *The senses considered as perceptual systems*. Boston: Houghton-Mifflin.

Gimson, Alfred Charles. 2001. *Gimson’s Pronunciation of English*. 6th ed. London: Edward Arnold.

Goldinger, Stephen D. 1998. “Echoes of echoes? An episodic theory of lexical access.” *Psychological Review* 105, no. 2: 251-279.

Goldinger, Stephen D., and T. Azuma. 2004. “Episodic memory reflected in printed word naming.” *Psychonomic Bulletin and Review* 11, no. 4: 716-722.

Green, K. P., Kuhl, P. K., Meltzoff, A. N., and Stevens, E. B. 1991. “Integrating speech information across talkers, gender, and sensory modality: Female faces and male voices in the McGurk effect.” *Perception and Psychophysics*, 50, no. 6, 524-536.

Halle, M. 1985. “Speculation about the representation of words in memory.” *Phonetic linguistics*, edited by V. Fromkin. New York: Academic Press: 101-114.

Hayward, Catherine. 2000. *Experimental Phonetics*. London: Longman.

Hintzman, D. L. 1986. “‘Schema abstraction’ in a multiple-trace memory model.” *Psychological Review*, 93, 411-428.

Jacoby, Larry L. 1983. “Remembering the data: Analyzing interactive processes in reading.” *Journal of Verbal Learning and Verbal Behavior* 22, no. 5: 485-508.

Jui-Feng, Peng, Chen Li-mei, and Lee Chia-Cheng. 2014. “VOICE ONSET TIME OF INITIAL STOPS IN MANDARIN AND HAKKA: EFFECT OF GENDER.” *Taiwan Journal of Linguistics* 12, no. 1: 63-79.

Koenig, Laura L. 2000. “Laryngeal Factors in Voiceless Consonant Production in Men, Women, and 5-Year-Olds.” *Journal of Speech, Language & Hearing Research* 43, no. 5: 1211-1228.

Kohler, Klaus J., and H. J. Kunzel. 1978. “The Temporal Organisation of Closing-Opening Movements for Sequences of Vowels and Plosives in German.” *Arbeitsberichte Kiel* 10: 118-166.

Krauss, Robert M., and Jennifer S. Pardo. 2004. “Is alignment always the result of automatic priming?.” *Behavioral and Brain Sciences* 27, no. 2: 203-204.

Kučera, H., and Francis, W.N. 1967. *Computational analysis of present day American English*. Providence, RI: Brown University Press.

Ladefoged, Peter, and Keith Johnson. 2011. *A Course in Phonetics*. 6th ed. Boston: Wadsworth, Cengage Learning.

Leather, Nathan, and Allan James. 1996. “Second language speech.” *Handbook of Second Language Acquisition*, edited by W. Richie and T. Bhatia. New York: Academic Press: 269-316.

Liberman, A. M., and Mattingly, I. G. 1985. “The motor theory of speech perception revised.” *Cognition*, 21, 1-36.

Lisker, Leigh. 1957. “Closure Duration and the Intervocalic Voiced-Voiceless Distinction in English.” *Language*, no. 1: 42-49.

Lisker, Leigh. 1975. “Is it VOT or a first-formant transition detector?.” *Journal of the Acoustical Society of America* 57, no. 6: 1547-1551.

Lisker, Leigh, and Arthur S. Abramson. 1967. “SOME EFFECTS OF CONTEXT ON VOICE ONSET TIME IN ENGLISH STOPS.” *Language and Speech* 10, no. 1: 1-28.

Major, Roy C. 1992. “Losing English as a first language.” *The Modern Language Journal* 76, no. 2: 190-208.

Machač, Pavel. 2006. “Temporální a spektrální struktura českých explozív” Praha: Univerzita Karlova, dissertation.

Morris, R.J., K.D. Herring, and C.R. McCrea. 2008. “Voice onset time differences between adult males and females: Isolated syllables.” *Journal of Phonetics* 36, no. 2: 308-317.

Munro, Murray J.; Derwing, Tracey M.; & Flege, James E. 1999. “Canadians in Alabama: A perceptual study of dialect acquisition in adults.” *Journal of Phonetics* 27: 385–403.

Nearey, Terrance M. 1989. “Static, dynamic, and relational properties in vowel perception.” *Journal of the Acoustical Society of America* 85, no. 5: 2088-2113.

Nielsen, Kuniko. 2011. “Specificity and abstractness of VOT imitation.” *Journal Of Phonetics* 39, no. 2: 132-142.

Palková, Zdena. 1994. *Fonetika a fonologie češtiny*. Praha: Karolinum.

Pardo, Jennifer S. 2006. “On phonetic convergence during conversational interaction.” *Journal of the Acoustical Society Of America* 119, no. 4: 2382-2393.

Pardo, Jennifer S. 2012. “Reflections on phonetic convergence: speech perception does not mirror speech production.” *Language and Linguistics Compass* 6.12: 753-767. *Proceedings of Meetings on Acoustics*. Vol. 19, no. 1. Acoustical Society of America.

Pardo, Jennifer S. 2013. “Reconciling diverse findings in studies of phonetic convergence.” *Proceedings of Meetings on Acoustics* 19, no. Proceedings of Meetings on Acoustics - ICA 2013.

Pardo, Jennifer S., Isabel Cajori Jay, and Robert M. Krauss. 2010. “Conversational role influences speech imitation.” *Attention, Perception, & Psychophysics* 72, no. 8: 2254-2264.

Pardo, Jennifer S., Rachel Gibbons, Alexandra Suppes, and Robert M. Krauss. 2012. “Phonetic convergence in college roommates.” *Journal of Phonetics* 40, 190-197.

Pavlenko, Aneta. 2000. “L2 influence on L1 in late bilingualism.” *Issues in Applied Linguistics* 11, no. 2: 175-205.

Raphael, Lawrence J. 2005. “Acoustic Cues to the Perception of Segmental Phonemes.” *The Handbook of Speech Perception*, edited by D.B. Pisoni, and R. E. Remez. Oxford: Blackwell Publishing: 182-206.

Repp, Bruno H. 1979. “RELATIVE AMPLITUDE OF ASPIRATION NOISE AS A VOICING CUE FOR SYLLABLE-INITIAL STOP CONSONANTS.” *Language and Speech* 22, no. 2: 173-189.

Robb, Michael, Harvey Gilbert, and Jay Lerman. 2005. “Influence of Gender and Environmental Setting on Voice Onset Time.” *Folia Phoniatrica Et Logopaedica* 57, no. 3: 125-133.

Ryalls, John, and Allison Zipprer. 1997. “A preliminary investigation of the effects of gender and race on voice onset time.” *Journal of Speech, Language & Hearing Research* 40, no. 3: 642-645.

Sancier, M.L., and C.A. Fowler. 1997. “Gestural drift in a bilingual speaker of Brazilian Portuguese and English.” Journal of Phonetics 25, no. 4: 421-436.

Simek, Jaroslav. 2010. “Explozívy v češtině: temporální vlastnosti a variabilita při realizaci.” Praha: Univerzita Karlova, diploma thesis.

Sheffert, S. M., and Fowler, C. A. 1995. “The effects of voice and visible speaker change on memory for spoken words.” *Journal of Memory and Language*, 34, 665-685.

Shockley, Kevin, Laura Sabadini, and Carol A. Fowler. 2004. “Imitation in shadowing words.” *Perception & Psychophysics* 66, no. 3: 422-429.

Skarnitzl, Radek. 2011. *Znělostní kontrast nejen v češtině*. Praha: Nakladatelství Epocha.

Smith, B. (1978). “Effects of place of articulation and vowel environment on voiced stop consonant production.” *Glossa*, 12, 163–175.

Stevens, Kenneth N., and Dennis H. Klatt. 1974. “Role of formant transitions in the voiced-voiceless distinction for stops.” *Journal of the Acoustical Society of America* 55, no. 3: 653-659.

van Alphen, Petra M., and Roel Smits. 2004. “Acoustical and perceptual analysis of the voicing distinction in Dutch initial plosives: the role of prevoicing.” *Journal Of Phonetics* 32, no. 4: 455-491.

Warren, Paul, and William Marslen-Wilson. 1988. “Cues to lexical choice: Discriminating place and voice.” *Perception & Psychophysics* 43, no. 1: 21.

Winn, Matthew B., Monita Chatterjee, and William J. Idsardi. 2012. “The use of acoustic cues for phonetic identification: Effects of spectral degradation and electric hearing.” *Journal of the Acoustical Society of America* 131, no. 2: 1465-1479.

Whalen, D. H., Arthur S. Abramson, Leigh Lisker, and Maria Mody. 1993. “F0 gives voicing information even with unambiguous voice onset times.” *Journal of the Acoustical Society of America* 93, no. 4: 2152-2159.

Yavas, Mehmet. 2007. “Factors Influencing the VOT of English Long Lag Stops and Interlanguage Phonology.” *New Sounds 2007: Proceedings of the Fifth International Symposium on the Acquisition of Second Language Speech*, 492-498.

Yavas, Mehmet, and Renée Wildermuth. 2006. “The effects of place of articulation and vowel height in the acquisition of English aspirated stops by Spanish speakers.” *IRAL– International Review of Applied Linguistics in Language Teaching* 44, no. 3: 251-263.

10. ANNOTATION

Subconscious Imitation of Perceived Phonetic Properties and its Relationship to Phonology

(Bachelor Thesis)

Author: Tomáš Sedláček

Field of Study: English Philology

Faculty and Department: Philosophical Faculty, Department of English and American Studies

Supervisor: Mgr. Václav Jonáš Podlipský, Ph.D.

Number of characters: 107 263

Keywords: speech imitation, speech convergence, voicing contrast, word-initial plosive, VOT, prevoicing

Description: The current thesis deals with the comparison of Czech and English VOT. More specifically, it is a replication of Nielsen (2011), who concerned with imitation of extended and reduced VOT in English. The first goal was to offer phonetic background. The second objective was to present the factors causing VOT to vary. Subsequently, the views concerning speech imitation together with Nielsen's research were presented. To test the hypothesis, research was conducted in Czech VOT imitation. Unfortunately, the research did not find significant data for imitation and the hypothesis was thus not proven.

10.1 Anotace v češtině

Podvědomá imitace vnímaných fonetických vlastností a její vztah k fonologii
(Bakalářská práce)

Autor: Tomáš Sedláček
Studijní obor: Anglická Filologie
Fakulta a katedra: Filosofická fakulta, Katedra anglistiky a amerikanistiky
Vedoucí práce: Mgr. Václav Jonáš Podlipský, Ph.D.
Počet znaků: 107 263
Klíčová slova: imitace řeči, řečová konvergence, znělostní kontrast, explozíva v iniciální poloze, VOT, prevoicing
Charakteristika: Bakalářská práce se zabývá porovnáním českého a anglického VOT, přesněji řečeno hlavní část je zaměřena na replikaci výzkumu Nielsen (2011), která se zabývala imitací prodlouženého a zkráceného VOT v angličtině. První část práce poskytuje základní fonetické znalosti. Následně jsou uvedeny faktory, které ovlivňují VOT. Další sekce se věnuje imitaci řeči, a také je zde detailněji představen výzkum Nielsen. Hypotéza byla otestována provedením výzkumu v oblasti imitace českého. Výsledky výzkumu však bohužel danou hypotézu nepotvrdily.