Title: Percepcja angielskich i polski spółgłosek właściwych

Author: Arkadiusz Rojczyk

Citation style: Rojczyk Arkadiusz. (2008). Percepcja angielskich i polski spółgłosek właściwych. Praca doktorska. Katowice : Uniwersytet Śląski

Uniwersytet Śląski

Arkadiusz Rojczyk

PERCEPCJA ANGIELSKICH I POLSKICH SPÓŁGŁOSEK WŁAŚCIWYCH

Rozprawa doktorska napisana pod kierunkiem prof. zw. dr. hab. Janusza Arabskiego

Sosnowiec
2008

University of Silesia

Arkadiusz Rojczyk

PERCEPTION OF ENGLISH AND POLISH OBSTRUENTS

A dissertation presented to the Faculty of Philology, University of Silesia, in partial fulfilment of the requirements for the degree of Doctor of Humanities in the subject of linguistics

Adviser: prof. zw. dr hab. Janusz Arabski

Sosnowiec
2008

Contents

Part Three. Perception of the voicing contrast in English and Polish obstruents: An experimental study

Introduction: the aim of the book

Speech can be investigated from three perspectives. From the point of view of articulation – by observing the manoeuvres of articulators in producing speech sounds. From the point of view of acoustics – by analysing visual renderings of acoustic properties in articulated sounds. Finally, from the point of view of perception, by testing listeners' reactions to presented stimuli in a discrimination or identification paradigm. This is the last perspective, the perception, that we assume in this book.

In our contrastive analysis of the perception of English and Polish obstruents, we concentrate on temporal and spectral parameters defining the voicing contrast in the two languages. The voicing contrast has been found to be one of the most intricate contrasting devices used in languages. Early beliefs that it is implemented by the presence or absence of the vocal cord vibration appear to be essentially oversimplified. What is more, cross-linguistic comparisons have demonstrated that a general phonological division into voiced and voiceless categories is realised by diverse fine-grained phonetic parameters across the world's languages. English and Polish are a good source of such differences. They have been documented to differ in the implementation of the voicing contrast both in terms of temporal organisation and spectral features.

In the experimental part, we use temporal and acoustic manipulation techniques in order to isolate a tested parameter and present it to the listeners. We assume a developmental perspective in that we compare the performance of Polish Beginner Learners, Advanced Learners, and Native Speakers of English. These three groups sketch a cross-sectional path from early stages in learning English, through high L2 proficiency, to the target native performance.

We hope that this book will contribute to a better understanding of English-Polish phonetic differences. Unlike numerous comparisons which concentrate on articulatory or acoustic descriptions, this study looks into the problem of the voicing contrast implementation in the two languages from the point of view of actual perceptual performance by Polish and English listeners.

The book is divided into three parts. In Part 1, we propose a general discussion on how phonetic categories are extracted from the acoustic signal. An amazing variety of speech sounds catalogued in the worlds' languages suggests that human perceptual abilities are remarkable in that an infant must be well-equipped to acquire any speech sounds, depending on the ambient language. We subsequently review the research on invariance – the lack of one-to-one mapping between acoustic information and phonetic categories – and we briefly discuss the speech perception models and the phenomenon of categorical perception. In Chapter 2 of this part, we look at perception form the perspective of Second Language Acquisition. In an L1-L2 contact, the native and new perceptual systems of a listener must inevitably interfere. Different models and learning scenarios predict different potential problems. Finally, we ask whether learners can attain native-like perception in L2 and whether the perception performance in L2 can influence the perception of L1.

In Part 2, we concentrate on different strategies of implementing the voicing contrast in English and Polish. We discuss both temporal and spectral parameters. For example, the Voice Onset Time has been found to be a strong and reliable temporal parameter of voicing in initial stops both in production and perception. The implementation of the voicing contrast can be expansive – it may influence the length of a preceding vowel. It can also affect the timing of a segment which realises the contrast – the closure duration in stops will vary depending on their voicing category. Analogically to stops, the voicing contrast in fricatives is realised by durational parameters but, additionally, fricatives show variation in spectral features of frication noise. Finally, affricates appear to combine elements of the voicing contrast implementation found for both stops and fricatives.

In Part3, we describe the study design, applied manipulation techniques, and group characteristics. For each tested parameter, we specify the used stimuli and provide the obtained results. Finally, we test the hypotheses and present a general discussion.

Part One

Speech perception in Second Language Acquisition

Introduction

"We speak in order to be heard and need to be heard in order to be understood" noted Roman Jakobson and Linda Waugh in *The Sound Shape of Language* (1979: 96). This quote dexterously sketches objectives of the research dealing with human speech communication. Articulatory phonetics investigates how we speak by examining the *modus operandi* of articulators in their struggle for producing vowels and consonants. Acoustic phonetics bridges how we speak and how we hear by looking into inherent spectral parameters of sounds transmitted between the speaker and the hearer. Auditory phonetics explains how we hear by providing an impressive number of experiments on how speech sounds are transformed and decoded from acoustic waves into discrete phonetic categories. Finally, how we understand is undertaken by higher order semantic perception studies which endeavour to shed light on how humans process phonetic input and obtain meaningful units.

In this part, we are interested in how phonetic categories are extracted from the acoustic signal. In Chapter 1, we discuss a general nature of speech recognition and briefly look back at the history of speech perception studies. We try to demonstrate that human abilities to perceive sounds are remarkable in the light of the number of sounds in world's languages that an infant must be equipped to acquire. Next, we touch upon the problem that has always bothered speech scientists, namely the fact that speech signal is invariant, i.e. there is no one-to-one mapping between acoustic information and phonetic categories. It is not surprising then that different theories of speech perception came into being in an attempt to find invariance, be it in articulatory gestures or the speech signal itself. We review three of them: the Motor Theory, the Direct Realist Theory, and the Auditory Enhancement Model. Finally, we turn to the phenomenon of categorical perception whereby listeners divide the acoustic continuum; the phenomenon that has been a core concept of methodological approaches to speech perception.

In Chapter 2, we ask whether the critical period for attaining optimal L2 perception exists and whether the capacities for learning L2 categories decrease with age. Next, we briefly discuss the concept of interlanguage and the L1-L2 transfer that might occur in L2 perception, i.e. how a native language influences the perception of L2 speech sounds. Different L2 sound perception models provide different scenarios for the process of learning L2 sound categories and so we discuss their core proposals and predictions. Finally, we contrast two diverging standpoints on to what extent L2 perception can influence L1 perception.

Chapter 1

Auditory perception

Speech, as noted by Alvin Liberman and colleagues (Liberman et al. 1967), is a code. By performing articulatory manoeuvres, a speaker encodes a message and conveys it via acoustic signals to the listener. The listener has the key and can “unravel the code to reveal the message it contains” (Culter and Clifton 1999: 125). The very process of decoding is, however, of severe complexity.

Speech is presented as sound waves to the ear of the listener but it does not make use of an exclusive channel. In fact, sound waves reaching the ear carry all other noise present in the listener’s environment. It is therefore the listener’s first assignment to separate linguistic input from non-linguistic noise. This process exploits the fact that speech signals differ from background noise by generally having periodic nature, while noise is characterised by being aperiodic (Cutler and Clifton 1999). Human auditory system appears to utilise a sort of grouping mechanism that effectively assigns signals to their sources by analysing their frequency characteristics (Bregman 1990 for a review).

When the incoming speech signals have been isolated from surrounding noise, the listener can begin decoding. The task consists in transforming a constant borderless flow of acoustic input into discrete segments. These segments are recognised in linguistics as phonemes and are simply the smallest units in terms of which spoken language can be described. Thus when hearing the word *deep*, the listener processes the acoustic form (Figure 1.1) into discrete elements. The structure of the phonemes can be further described in terms of linguistic units: /d/ is a voiced alveolar plosive, /i/ is a high-front vowel, and /p/ is a voiceless bilabial plosive.

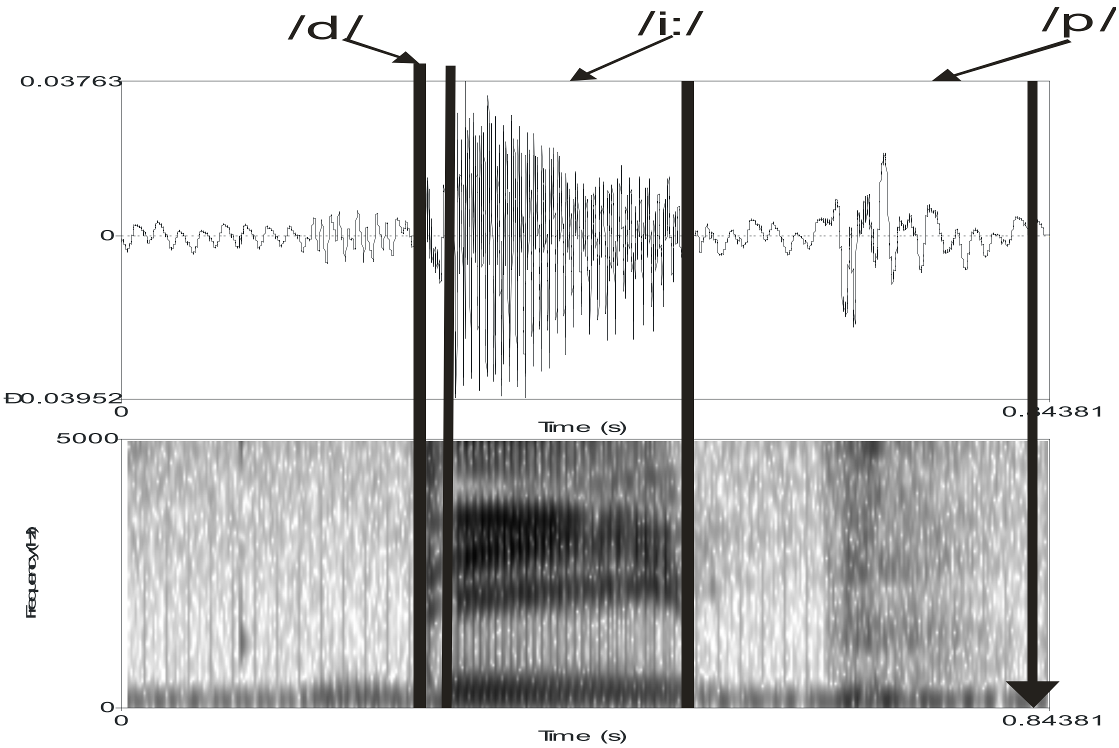


Figure 1.1. Waveform and spectrogram of the word *deep*

The tradition of linguistic studies of speech perception was founded by ancient philosophers and Greek grammarians (Polański 2003). Their descriptions of speech sounds were mainly based on auditory criteria. They were the first to distinguish between voice and noise, as well as between vowels and consonants. The Sanskrit grammarians, on the other hand, focused on vocal anatomy and articulatory processes to the exclusion of acoustic and auditory impressions produced by speech sounds (Allen 1953). The nineteenth century linguists such as Bell (1867) or Sweet (1877) followed suit and focused primarily on speech articulation to describe similarities and differences across languages and in language teaching (discussion in Hume and Johnson 2001). For instance, the Sweet/Bell system of vowel classification, still widely used in teaching a foreign vowel system, is based on speech articulation. The articulatory approach was also a basis for structuralists' description of phonetics and phonology (e.g. Pike 1943).

The rationale for extended emphasis on articulation in linguistic descriptions arguably lies in the fact that articulators are open to inspection. Linguists, equipped with x-ray scanning or electromyographic imaging, can observe the movements of lips, jaw and tongue. The articulatory approach was especially favourable in the classification of consonants, which are produced by the observable contact of two articulators. It was definitely less so with vowels and semi-vowels, however a certain degree of approximation could be noticed.

The invention and development of a sound spectrograph gave speech perception its right place in linguistics (Hume and Johnson 2001, Polański 2003) and gave way to a comprehensive approach to language sound structure in terms of acoustic and auditory properties (Jakobson et al. 1951). Beginning in the early 1950s, researchers at the Haskins Laboratories carried out a series of landmark studies on synthetic speech sounds (Delattre et al. 1952, 1955, 1964, Liberman 1957, Liberman et al. 1952, 1954, 1956).[1] Although probably the most influential publication of the century on the sound structure of language, namely *The sound pattern of English* (Chomsky and Halle 1968), turned its interest into the phonetic study of speech articulation, the last 50 years have witnessed a growing body of perceptual data and their inclusion in linguistic theories, e.g. in Optimality Theory (Prince and Smolensky 1993, McCarthy and Prince 1993)[2], which has allowed for the statement of perceptually grounded constraints that interact with other constraints motivated by other general principles (Hume and Johnson 2001).

1.1 Uniqueness of human speech perception

What must amaze every researcher entering the domain of speech perception is its complex nature. For almost 60 years now there has been a sustained effort to develop machine speech recognition devices and so far no engineering approach to speech perception has achieved a complete success (Benzeghiba et al. 2007, Gerosa et al. 2007, Moore 2007, Scharenborg et al. 2007). What machines, or even other primates lack, is, according to Chomsky (1980), an independent module for language acquisition. This module follows an independent course of development in the first years of life and allows a child to achieve a language competence in their native language that cannot be explained in traditional learning terms. The module matures and develops with experience but the mature system does not simply mirror the experience (Massaro 1994: 220). The language user inherits rule systems of highly specific structure. This innate knowledge allows humans to acquire the rules of the language, which cannot be induced

[1] Ingenious retrospection into early attempts with speech synthesis and perception experiments in Liberman (1996).
[2] A comprehensive application of auditory perception in OT to explain phonological processes in Flemming (2002).

from normal language experience because of the paucity of language input. The advocates argue that the data of language experience are so limited that no process of induction, abstraction, generalisation, or analogy can account for perfect language competence. The universal grammar given by biological endowment allows a child to learn to use language appropriately without conscious learning of its formal rules.[3] Similarly, early speech perception abilities of infants are claimed to be indicative of "finely tuned linguistically relevant perceptual abilities" (Miller and Eimas 1983: 135) or even an "innately given, universal set of phonetic categories" (Eimas 1991: 111), however general perceptual learning is also of great importance (Jusczyk 1993, 1997).

Impressive potential of human auditory system is best manifested by the vastness of speech sounds that an infant must be ready to acquire. The UCLA Phonological Segment Inventory Database, the result of over 100 years of meticulous study by phoneticians, comprises a representative sample of the phonological categories of the world's 451 distinct languages (Kluender 1994, De Boer 1999). What becomes apparent upon careful inspection of the UPSID data is sheer diversity of speech sound categories in the world's languages. Maddieson (1984) classified 869 sound categories occurring in languages: 558 consonants, 260 vowels and 51 diphthongs. An ample testament to the rich variety of sounds used in the world's languages is the fact that he needed as many as 58 phonetic attributes for the classification. Not only are the sound categories diverse themselves but also there is significant heterogeneity in sound inventories across languages. The Rotoka and Mura languages need only 11 phonemes to organise their phonological systems, whereas language !Xuʗ exploits no fewer than 141 phonemes (Maddieson 1984, Kluender 1994, Epstein 2000).

Although a handful of sounds are extremely common, the majority of 869 phonemes are relatively rare. One hundred and sixteen of the languages in UPSID have at least one sound that no other language in the database has, which gives 47% of speech sounds that are unique, i.e. occur only in one language (Epstein 2000). All catalogued languages have stop consonants with three prevailing places of articulation – bilabial, alveolar, and velar. Moreover, over 80% of languages utilise the voiced-voiceless distinction at these three place of articulation (Maddieson 1984). A human

[3] Recent criticism levelled at Chomsky's Language Acquisition Device by Lieberman (2006) consists in questioning speech as a uniquely human phenomenon. Lieberman claims that the body of data on animals' communication (e.g. Gardner and Gardner 1969, 1971, 1994) has not been given sufficient heed and that chimpanzees, despite their articulatory limitations, show elements of auditory speech processing. Also, Sampson (1989, reported in Massaro 1994) documented that language input is not so sparse and chaotic as Chomsky's followers want to believe.

must also be equipped to distinguish between 30 different fricatives, most of which are rare and very few extremely common. In fact, over 90% of all languages use fricative consonants. Alveolar /s/ is found in 80% of the classified languages with palato-alveolar /ʃ/ and labiodental /f/ being fairly frequent as well. In terms of the voiced-voiceless opposition, only roughly 30% of fricatives used in the world's languages are accompanied by the vocal cord vibration, the remaining majority is realised as voiceless (Maddieson 1984). Vowel systems found across languages also abound in diverse categories. Phoneticians have found at least 44 different vowels in the world's languages (De Boer 1999). Some languages make do with only 3 vowels, whereas others use as many as 24 (Maddieson 1984), for example Norwegian utilizes 15 different vowel qualities (De Boer 1999). 21.5% of the UPSID languages have five vowels and most of the five-vowel systems tend to use the same vowels; almost all languages contain /i/, /a/, /u/, /e/, and /o/ (De Boer 1999).

Another traditional argument for uniqueness of human speech perception is that the transmission rate of the speech signal appears to exceed human perceptual capacity. In natural tempo of articulation, a listener processes a rate of between 10 to 20 phonemes per second (Massaro 1994) or even between 20 to 30 phonemes per second (Lieberman and Blumstein 2002). Speed with which listeners can identify phonetic distinctions and put them together to form meaning surpasses their ability to identify non-speech signals. The fastest rate at which non-speech sounds can be identified is about 7 to 9 segments per second (Miller 1956, reported in Lieberman and Blumstein 2002). Sounds transmitted at a rate of 20 segments per second merge into an undifferentiated tone.[4]

An explanation to this seemingly puzzling fact lies in a widely recognised observation that sounds are not articulated as separate entities but rather are strongly coarticulated. Although coarticulation is identified, for the most part, as "destructive, of some essential properties of phonological segments, in particular their discreteness, their static nature and their context-invariance" (Fowler and Galantucci 2005: 635), it has a blessing effect on the efficacy of speech recognition. According to Hockett's (1955) famous metaphor; in the flow of speech vowels and consonants are like Easter

---

[4] Moore (1997) reports that listeners can identify brief sequences of sounds at rates of up to 100 per second (one every 10 ms) but, as noted by Hawkins (1999: 201), "at these short durations it seems that the listeners learn the overall sound pattern rather than perceiving each item separately." On higher than phoneme processing level the research into speech rate indicates that natural speech rates range from 125-255 words per minute (Nelson 1948, Harwood 1955). Above 225 words per minute, there is an accelerating decline in comprehension by native listeners (Foulke and Sticht 1969, discussion in Jones et al. 2007).

Eggs and articulation is a clothes wringer that breaks the eggs and mixes them with each other. The perceiver is conceived of as an observer of a conveyor belt on which the broken and wrung eggs are carried and whose job is to “identify the original phonetic eggs from an acoustic mess of brightly colored shell, yolk, and albumin (Remez 1994: 165). Although coarticulation is viewed both as distortion of phonetic segments (Ohala 1981) and elimination of the possibility of articulatory and acoustic invariants corresponding to consonants and vowels (Liberman and Mattingly 1985), it has an important role in speech perception. It is thought to increase the efficiency of perception by speeding up the rate at which phonemes can be transmitted. Since information about more than one phoneme is normally transmitted simultaneously, each articulation effectively lasts longer than the acoustic segment most closely associated with the ‘pure’ phoneme (Hawkins 1999). In consequence, listeners have more time to decode each separate gesture.

Although coarticulation has an undisputed global beneficiary impact on the efficacy of speech perception, it has nevertheless long posed one of the most difficult problems for speech researchers – namely the uncertain relationship between properties of the speech signal and a given perceptual (phonemic) category. No simple mapping between units of phonetic structure and units of acoustic structure is commonly termed in the literature as *the lack of invariance*.

1.2. Lack of invariance

As acoustic information specifying a particular phonetic segment varies dramatically with a change in the identity of surrounding segments, it is very often the case that a single phonetic segment is realised by different acoustic signals. Classic examples with speech synthesis (Cooper et al. 1952, Liberman et al. 1954, Delattre et al. 1955, Liberman et al. 1967) demonstrated that acoustic correlates of the place of articulation of a stop consonant depend on the following vowel. The primary acoustic cue for /d/ in the syllable /di/ is a rising second formant transition, while /d/ in the syllable /du/ is signalled by a falling transition. A single burst of noise at a frequency of 1440 Hz will be heard as /p/ in one phonetic context – when followed by /i/ – but as /k/ in another context – when followed by /a/. The question that the study of speech perception must answer is where in the signal listeners find cues for phonetic categories and how they cope with no one-to-one mapping between acoustic information and sound categories. The quest for the solution gave way to the theories of speech perception that endeavour to find invariance either in articulatory gestures or the signal itself. They are briefly discussed in the following subsections (detailed discussion in Kluender 1994, Remez 1994, Appelbaum 1996, Lotto et al. 1997a, Hawkins 1999, Lieberman and Blumstein 2002, Fowler 2003, Diehl et al. 2004, Fowler and Galantucci 2005).

1.2.1. Motor Theory of Speech Perception

The essence of Motor Theory (Liberman et al. 1967, Liberman and Mattingly 1985) is that listeners interpret the acoustic signal in terms of the articulatory patterns that would produce auditory patterns heard in the signal. Due to the lack of invariance in the acoustic signal (see Section 1.2.) that would allow listeners to identify sound categories, the authors propose that invariance can be found in neuromotor commands to the articulators (e. g. tongue, lips, and vocal folds) which are recovered by human listeners from the acoustic signal. In the earliest version of the Motor Theory (Liberman et al. 1967), these were the vocal tract movements themselves that were thought to be reconstructed from the acoustic patterns. In the most recent version, called the Revised Motor Theory (Liberman and Mattingly 1985), listeners are conjectured to reconstruct the speaker's intended gestures and not realised gestures. These intended gestures are thought to be abstract control units that can give rise to linguistically relevant vocal tract movements. Accordingly, the listener perceives the articulatory plans that control the vocal tract movements that would produce a perfect rendition of the speaker's intended utterance. As "[g]estures are the objects of perception" (Liberman and Mattingly 1985: 10) and "[t]he invariants of speech perception are the phonetic gestures" (Liberman and Mattingly 1985: 29), there is an indissoluble link between production and perception. Speech is special according to the Motor Theory in that, of all the phenomena human beings perceive, speech sounds are the only ones that they also produce. Since humans are not only perceivers but also producers, they are said to have tacit knowledge of how speech sounds are produced.

Because "[t]he objects of speech perception are the […] phonetic gestures of the speaker, represented in the brain as invariant motor commands that call for movements (Liberman and Mattingly 1985: 2), for motor theorists coarticulation between adjacent phonemes occurs during the execution of movements. Consequently, coarticulation is not essential to the linguistic structure of the signal and is not represented at the level of abstract gestures. It is largely seen as a smoothing process between successive gestures that occurs inevitably during the execution of movements, because the nature of the vocal tract means that there must be movement between successive targets, and those movements must be relatively smooth. In this view, coarticulation destroys the purity of the underlying phoneme string.

The Motor Theory has been criticised on different grounds, best summarised by Ohala (1986). First, a ventriloquist produces requisite properties of speech without active movements of jaw and lips. The intelligibility of speech produced in this manner suggests that the perceiver is indifferent to the peculiarities of articulation, though not to the acoustic effects. In Kijak and Rojczyk (forthcoming), we demonstrated that the bilabial approximant /w/ can be changed into velarised /l/ by manipulating formant transitions of the following vowel. We replicated auditory experiments demonstrating that the difference between /b/ and /w/ in sequences /ba/ and /wa/ is signalled by the following vowel, and more precisely, by the length of its formant transitions (Shinn and Blumstein 1984, Walsh and Diehl 1991). By exchanging the vowel in both syllables, we obtained the stimuli: /b/ + the vowel with longer transitions from /wa/, and /w/ + the vowel with short transitions from /ba/. Indeed, the first percept gave very strong auditory impression of /gwa/ due to the longer formant transitions cuing the preceding /w/, even if it was acoustically absent. The second percept, on the other hand, was heard as /la/ with strongly velarised /l/. It is interesting to note, in the light of proposals by motor theorists, who claim that humans perceive gestures encoded in acoustic signal, that motor configuration for /w/ gave the percept of velarised /l/ only by modifying the following vowel. It seems that the listener is not able to read intended gestures for /w/ and thus hears velarised /l/. The second argument raised by Ohala (1986) refers to the fact that birds such as the mynah replicate speech sounds by means of syrinx and beak, rather than larynx, tongue and lips, but still they are understood even though people are oblivious of their anatomical characteristics. In fact, evidence from experiments with animals has always been somewhat problematic for the Motor Theory. While it can be accepted that human listeners read from the acoustic signal human speakers' vocal gestures, the problem arises how birds could possibly do it. A number of perception experiments with birds (e.g. Kluender et al. 1987, Lotto et al. 1997b) showed that birds can respond to changing formant transitions between /ba/ and /ga/ in a similar fashion like humans. Ohala's (1986) next point emphasises the fact that several key diachronic phenomena in phonetic inventories are well explained by appealing to the acoustic similarity of historically related phonetic manifestations, the transformations being only weakly constrained by articulation (Flemming 2002 for a discussion in the Optimality Theory model). Last, the clinical literature shows that talkers compensate for disorders of articulation by approximating the acoustic properties of sounds of speech and not their articulation (Lieberman 2006 for details).

1.2.2. Direct Realist Theory of Speech Perception

The Direct Realist Theory of Speech Perception (Fowler 1981, 1984, 1994, 1996, 2003) claims that, similar to the Motor Theory, the objects of speech perception are articulatory rather than acoustic events – in Fowler's (2003: 256) own words, "listeners […] perceive gestures, because gestures cause the structure in stimulation to the ear." However, unlike the Motor Theory, the Direct Realist Theory asserts that the articulatory objects of perception are actual, phonetically structured, vocal tract movements, or gestures, and not plans for these movements, such as neuromotor commands or intended gestures.

The Direct Realist Theory puts speech perception in a universal context of biological function of perception. Perceptual systems constitute the only means that animals and humans have to know their world (Gibson 1966, 1979, reported in Fowler 2003). This view is succinctly summarised by Fowler (1996: 1732) in the following passage:

> Perceptual systems have a universal function. They constitute the sole means by which animals can know their niches. Moreover, they appear to serve this function in one way: they use structure in the media that has been lawfully caused by events in the environment as information for the events. Even though it is the structure in media (light for vision, skin for touch, air for hearing) that sense organs transduce, it is not the structure in those media that animals perceive. Rather, essentially for their survival, they perceive the components of their niche that caused the structure.

Thus, according to the Direct Realist Theory, a talker's gestures (e.g. the closing and opening of the lips during the production of /pa/ or compression of the tongue against the alveolar ridge in the production of /ta/) structure the acoustic signal, which then serves as the informational medium for the listener to recover the gestures. The gestures of the Direct Realist Theory are very similar to those of articulatory phonology (Browman and Goldstein 1986, 1992, 1995). They refer to the place and degree of constriction of active articulators in the vocal tract.

The way in which the Direct Realist Theory attempts to account for coarticulation appears to be relatively simple. Talkers produce gestures, one for each phonetic segment. Adjacent gestures are co-produced, i.e. they overlap one another in time so that at any given point, the acoustic signal is likely to show influences of two or more phonetic gestures. Accordingly, each gesture lasts for longer than the acoustic segment with which it is mainly associated. At first, it is only weak, co-occurring with another stronger gesture, then it is the main information of the acoustic segment, and finally it lessens and wanes away. Overlapping gestures are not mixed together thus losing their original character. Quite the opposite, each gesture remains separate and coherent, which is both reflected in the acoustic signal and is perceived by the listener. As Hawkins (1999: 236) put it, "[i]t is as if you had three pieces of clay of different colors arranged in a line. When you smear the three together, you have a single larger piece of clay, but you can still see the component clays at the boundaries, and how they fit together, because each maintains its own color."

Unfortunately, the Direct Realist Theory, as a model set in a gesture approach to speech perception, suffers the same criticism as the Motor Theory (see Ohala's (1986) arguments in Section 1.2.1.). Moreover, prosody has been barely addressed and there has been little discussion on how units higher than the gesture can be organised into linguistic units (Hawkins 1999, Diehl et al. 2004).

1.2.3. Auditory Enhancement Theory

According to the Auditory Enhancement Theory (Diehl and Kluender 1989, Diehl et al. 1990), "the sound systems of language communities have adapted to be fairly robust signalling devices by exploiting general characteristics of auditory systems whenever possible" (Kluender 1994: 180). It can be achieved by developing an inventory of phonemes so as to optimise phonetic distinctiveness acoustically and auditorily. For auditory enhancement theorists, units of speech perception include distinctive features and not gestures like in the Motor Theory or Direct Realist Theory (see Sections 1.2.1. and 1.2.2.).

The proponents argue that, although a single acoustic property may correspond to a single auditory property, it is more typical that a feature contrast is conveyed by a number of different acoustic distinctions. For example, one dominant tendency among languages is that back, but not front, vowels are produced with the lips rounded. More precisely, there are 254 languages that have the high back rounded vowel /u/ but only 20 languages that have the high back unrounded vowel /✶/. Similarly, 271 languages have the high front unrounded vowel /i/ and only 21 have the high front rounded vowel /y/ (Maddieson 1984). No apparent articulatory constraint would explain why front vowel /i/ is unrounded and its back counterpart /u/ is rounded. The acoustic experiments have demonstrated, however, that the effect of lip rounding lowers Formant 2 and makes the vowels acoustically more backlike, thereby enhancing the contrast between /i/ and /u/ (Stevens et al. 1986, Diehl and Kluender 1989, Diehl et al. 1990, Rojczyk 2006 for a spectrographic analysis of Polish). Another argument supporting the theory comes from the process of nasalization. Across languages for which vowel nasalization is not phonemic, low vowels in words such as *cot* and *cat* (also *caught* in American English) tend to be nasalized much more often than high vowels (Ohala 1974). Again, any purely mechanical explanation for this tendency based on some articulatory coupling between the tongue height and velum appears to be ruled out by electromyographic evidence (Lubker 1968 reported in Kluender 1994). However, one of the acoustic consequences of nasalization is effective raising of Formant 1 of low vowels and thereby making them even lower (House and Stevens 1956, Stevens et al. 1987, Rojczyk 2006 for a spectrographic analysis of Polish). The explanation to this observation is that nasalization serves to enhance the high-low distinction by effectively lowering low vowels and providing listeners with sufficient auditory contrast between high and low vowels. This acoustic assumption was confirmed in perceptual studies where listeners identified nasalized and nonnasalised vowels of varying vowel heights, which showed that nasalization served to make vowels sound lower (Wright 1986, Krakow et al. 1988).

The Auditory Enhancement Theory sets itself against the Motor Theory and the Direct Realist Theory in that it does not seek invariance in articulatory gestures but rather in trading relations of different acoustic features in speech signal. Nevertheless, voices of criticism can be heard questioning strong emphasis on combining acoustic properties. Nearey (1995: 36) notes that "[s]peakers must be both acrobats and magicians. They must learn to do articulatory cartwheels to produce perceptual illusions”, and reasons that, although experimental data are probably reliable, it is not necessary to assume that speakers intentionally use trading relation of two or more features to enhance the contrast. The theory is, nevertheless, unique in providing at least partial explanation of the robustness of natural speech that makes listening and understanding possible even in very difficult listening conditions.

1.3. Categorical perception

It is impossible to discuss the acoustic manipulation techniques employed in this study without reference to the phenomenon of categorical perception. Early studies at the Haskins Laboratories (Liberman 1957, Liberman et al. 1961a,b) reported that changes along some dimension of the speech signal are not perceived continuously but in a discrete manner, i.e. categorically. When listening to series of steps in the acoustic continuum (e.g, /bV/ - /dV/ - /gV/), a change from one stimulus to the next sometimes caused no change in the consonant heard, while at other points in the continuum the same change was heard as an abrupt change in the place of articulation (e.g. /bV/ to /dV/). This led to the conclusion that listeners are limited in their ability to discriminate differences between different sounds belonging to the same phoneme category.[5] Two patterns are evident in the results (discussion in Hawkins 1999, Fowler 2003, Diehl et al. 2004). First, labelling functions exhibited abrupt boundaries between phoneme categories. Second, discrimination accuracy was close to chance for stimulus pairs within a phoneme category but nearly perfect for stimulus pairs separated by an identification boundary. Consequently, speech perception is closely related to the presence or absence of functional phonemic differences between sounds.

A significant portion of research on categorical perception concentrates on the Voice Onset Time, which shows highly categorical discrimination (e.g. Lisker and Abramson 1964, 1970, Abramson and Lisker 1970, Eimas et al. 1971, Lasky et al. 1975,) both in natural speech and by means of nonspeech analogs (Miller et al. 1976,

[5] Massaro (1994: 225) is, however, critical on methodological grounds. He notes that “the categorical model usually provides an inadequate description of the relation between identification and discrimination, and has not been shown to provide a better description than continuous models” (p. 225). Although he is critical about textbooks that describe speech perception as categorical (Miller 1981, Eimas 1985, Flavell 1985, Anderson 1990), his campaign appears to have been futile and categorical speech perception is an accepted fact (e.g. recently Diehl et al. 2004). Additionally, categorical perception of manual contrast has also been documented for sign languages (Emmorey et al. 2003, Baker et al. 2005, Emmorey 2007).

Pisoni 1977). We postpone a detailed discussion on the Voice Onset Time perception to later parts of this study (see Part 2, Section 3). It is also due to note that not all continua evince strong categorical perception. Vowel continua, for example, are less categorical (Fry et al. 1962, Pisoni 1973), showing good discrimination both across and within categories. It is speculated that the difference between categorical and less categorical perception between consonants and vowels might be related to a difference in the motor conditions of articulation. The consonants are produced by discrete motions that must attain certain targets, for example closures at certain places for the stop consonants. For the vowels, on the other hand, the tongue position can assume a large number of different positions within the front-back and high-low continuum (Repp 1984 for a comprehensive discussion).

1.3.1. Infants categorise the acoustic continuum

The observation that speech perception is categorical is hugely buttressed by the data obtained from experiments with infants' speech perception. One of the most important issues in speech perception is how listeners come to perceive sounds in a manner that is particular to their native language. Undoubtedly, in order to communicate proficiently, a listener must discriminate acoustic continuum of the speech signal in a fashion that is linguistically relevant to their ambient language (Eimas et al. 1971, Jusczyk 1982, Eimas et al. 1987, Kuhl 1987, Goodman and Nusbaum 1994, Damper 2000, Serniclaes 2005, Werker and Yeung 2005).

"Young infants can discriminate nearly every phonetic contrast on which they have been tested – including those that do not occur in their language-learning environment" (Werker and Pegg 1992: 285), and it "appears that the ability to discriminate the sounds of language is grounded on raw perceptual abilities of the mammalian auditory system" (MacWhinney 1998: 202). Extensive research has demonstrated that infants can categorise changes in formant transitions (Moffitt 1971, Morse 1972, Eimas 1974), frequency of release bursts (Miller et al. 1977, Jusczyk et al. 1990), Voice Onset Time (Eimas et al. 1971, Jusczyk et al. 1989, Lasky et al. 1975, details on VOT categorisation in Part 2, Section 3.1.), place of articulation for fricatives (Eilers et al. 1977, Holmberg et al. 1977 reported in Kluender 1994), nasals (Eimas and Miller 1977), glides (Jusczyk et al. 1978), and liquids (Eimas 1975).

The fact that infants can distinguish categorically nearly every phonetic contrast they have been tested on, whether or not that contrast is phonemic in their ambient language, materialised in the claim that infants have a specialised biological predisposition to discriminate a universal set of phonetic contrasts (Eimas 1991). By this view, the process of learning a language involves either a decline or reorganisation of this universal sensitivity. This point of view is not, however, immune to criticism. Firstly, Kluender (1994: 200), adducing Jacob (1977), writes;

> Cross-linguistic phonetic data call into question whether there could ever have been adequate selective pressure for a universal set of phonetic segments to become supported by specialized biological predispositions. The reason for such doubt is that there is simply too much diversity in the phonetic inventories used in languages. Innately specified processes, as products of selective pressure, should instead give rise to much greater conformity in phonetic inventories, and languages would generally share a collection of speech sounds that have been primed by the biological substrate. Furthermore, one would suspect that this collection should be relatively modest in size, for after accommodating an inventory that is adequate for successful communication […] there would be little pressure to increment the size of the universal set of phonetic segments.

Secondly, Pisoni et al. (1994) review results from experiments which show that there is no significant loss in auditory discriminative ability and suggest that both loss and subsequent regaining of discriminatory powers is due to changes in selective attention based on experience of what is important. Category-defining attributes of the signal become perceptually more distinctive and attributes that do not define the category become less distinctive.

For every teacher of foreign language pronunciation, it is an unquestionable fact that properly designed and conducted pronunciation training improves both pronunciation and perception. Whether the perception in L2 can match native speakers' performance is a matter of debate and one of the objectives of this study. Fortunately for all foreign language learners, categories which are absent in one's L1, and hence suppressed, can be regained by adequate stimulation and learning. In Chapter 2, we look at how new L2 perceptual categories come into being and try to answer the question of how successful L2 learners can be.

Chapter 2

Auditory perception in Second Language Acquisition

> It must of course be recognised that the mechanism of comprehension are no less part of linguistic performance than are mechanism of production, and it is likely that the perceptual capacities of the learner must be sufficiently developed in order for phonological development to proceed (Brown and Matthews 1997: 82).

Every adult learner of a foreign language must face the difficulties of learning sound categories absent in their native language. Sadly, despite their efforts, they are outperformed by infants and young children when the task is to learn the sound system of language. Every healthy child is able to learn native categories for the sounds in their ambient language, while adult learners struggle to attain native-like performance and commonly are not successful even after long exposure to L2.

L2 perception has always been taken seriously in speech perception research (discussion in Strange 1995, Iverson et al. 2001, Flege 2003, Escudero 2005, Cho and McQueen 2006). Polivanov (1931) provided several anecdotal examples of how the phonemes of L2 are perceived through the L1 system. They describe the difficulties which arise from L1 inluences on L2 perception. Trubetzkoy (1929, 1969) also suggested that the inadequate production of L2 results from L1 phonology working as a sieve through which L2 vowels and consonants must pass. Cross-linguistic speech perception research conducted in the 1960s showed that L2 learners have "perceptual foreign accents" (Strange 1995: 22), which were believed to be a result of their perceptual system being shaped by their first language.

2.1. Critical period in acquiring L2 perception?

In L2 production, it is common to observe divergences from target norms in terms of vowels, consonants, consonant clusters, words, or whole sentences (Leather and James 1996 for a review). While the controversy exists whether speakers who began learning their L2 in childhood – the so-called early learners – will differ from native speakers, there is agreement that the differences between native versus nonnative are greater for late learners, i.e. individuals who began learning in adolescence or adulthood (see Long

1990). This observation is commonly explained by claiming that humans possess an innate biological clock for language learning that allows direct learning from the input until approximately the age of nine whereupon acquisition begins to result in poorer attainment levels (Penfield and Roberts 1959). Similarly, Lenneberg (1967) formulated his well-known Critical Period Hypothesis, which states that this loss of predisposition for language learning has a biological basis since it is due to the completion of hemispheric lateralisation around puberty. Accordingly, only before puberty can learners acquire L2 from mere exposure to the input without conscious and laboured effort.[1] Adherents of the Critical Period Hypothesis suggest that the capacity for successful speech and language learning declines beyond the critical period. For example, De Keyser (2000: 518-519) suggested that:

> Somewhere between the ages of 6-7 and 16-17, everybody loses the mental equipment required for the abstract patterns underlying a human language, and the critical period really deserves its name […] It may be that the severe decline of the ability to induce abstract patterns implicitly is an inevitable consequence of fairly general aspects of neurological maturation and that it simply shows up most clearly in language acquisition.

However, adult learners can find some hope for future success in L2 perception in the fact that the primacy of age-related constraints has been recently debated (hence a question mark in the title of this subchapter). Scovel (1988: 62) suggests that if L2 acquisition is constrained by a critical period, it may affect production and perception differently:

> Pronunciation is the only part of language which is directly "physical" and which demands neuromuscular programming. Only pronunciation requires an incredible talent for sensory feedback of where the articulators are and what they are doing. And only pronunciation forces us to time and sequence motor movements. All other aspects of language are entirely "cognitive" or "perceptual" in that they have no physical reality.

Similar objections are articulated by Kuhl (1998, 2000a), who observes that the critical period for language acquisition results more from the interference of previous experience than from age. The changes in perceptual processing due to language experience may be self-reinforcing because initial exposure to language will alter how

[1] Other less radical notions such as 'sensitive period' have also been proposed (see Long 1990).

all subsequent sounds are perceived. Even though an adult learning a second language could be exposed to the same acoustic distribution of speech sounds as an infant acquiring the same language, the auditory distribution of those sounds would be different for adults due to prior perceptual changes caused by their L1 sound pattern. The decline in L2 acquisition abilities from childhood through puberty reflects a stronger and stronger neural commitment to one's L1 which is enhanced by continuous and incremental exposure to L1 sound pattern. Although this loss of perceptual sensitivity may be difficult to reverse, it is not precluded by age limitations. Likewise, Hyltenstam and Abrahamsson (2003), in a recent review and proposal on maturational constraints in L2 acquisition, suggest that there is a continuous maturational period that predicts that acquisition will be increasingly difficult with age but they remain neutral with respect to the exact extent it may hinder the attainment of native L2 perception because other non-maturational constraints can influence the end result. Cutler and Boersma (2005) also seem to reject the critical period factor in learning L2 perception. They claim that phonological reorganisation necessary for effective L2 perception is blocked by various language-specific constraints that are present in the phonological system of the listener's native language. Although improving performance of adult listeners has been found very difficult, appropriate training can improve performance to a certain extent regardless of the fact that it starts after puberty (e.g. Bradlow et al. 1997). Finally, Werker and Tees (1984) conclude that when given enough practice and adapted testing procedures adult listeners can regain their ability to distinguish nonnative contrasts.

2.2 L1-L2 transfer in speech perception

Early research on L2 acquisition acknowledged that the L1 system plays a substantial role in the process of learning a second language and led to the formation of *interlanguage* understood as a separate linguistic system based on the observable output which results from a learner's attempted production of a target language norm (Selinker 1972). Performance-oriented approaches such as the Contrastive Analysis Hypothesis (e.g. Lado 1957, reported in Escudero 2005) suggest that L1 habits are used in the process of learning a second language and that they have a negative or positive results depending on whether they are similar to or different from the habits used in the target

L2. The concept of transfer understood as copying of L1 features into target L2 system has been thoroughly studied both for syntax and semantics (e.g., Arabski 1968, 1979a,b, 1997, 2006, 2007). In a second language acquisition, a transfer can be full or partial depending on L1-L2 structural differences and the proficiency of learners (Archibald and Young-Scholten 2003). It can also have different meanings so that it could refer to a learner's conscious or unconscious strategy, to the process of transferring L1 knowledge onto L2 learning, or to the result of such a process (Hammanberg 1997).

Every L2 learner when confronted with new L2 perceptual categories will be forced to use their native sound system, which is the only one available, to sieve new L2 sounds. Learners will thus use L1 sound mappings to and from the signal (Schwartz and Sprouse 1996, Escudero and Boersma 2004, Escudero 2005). For example, English /d/ has positional variants of retroflex /d↩◉/ and dental /d🗀/. The former is realised in the environment of the retroflex continuant /r/ as in *adroit*, while the latter can be found in the environment of the dental fricative /❄/ in words like *width* (Polka 1991). In English, however, they do not have a phonemic value like they do in Hindi, which uses the two variants to change the meaning. Werker and Lalonde (1988 reported in Polka 1991 and Kluender 1994) showed that the stimuli which are identified by native Hindi listeners as dental or retroflex are all assimilated into the set of stimuli identified as alveolar by native English listeners. Therefore, it seems that the contrast which is not phonemic in a language loses its perceptual salience – a process referred to as single category assimilation (Best et al. 1988).

Category goodness assimilation (Best et al. 1988) occurs when attributes of one category of a two-category nonnative contrast can be well correlated with attributes of a single native category, while attributes of the other category of the nonnative contrast are less well correlated with attributes of the native category. One example of this is the Farsi distinction between velar and uvular stops. Native English listeners do not lose the ability to discriminate Farsi velars from uvulars. Instead, they perceive the Farsi voiced velar and uvular stops as being good or poor instances of the same category /g/ (Polka 1992). In this case, Farsi velar stops are perceived as relatively good English velar stops because they share most of the acoustic and auditory attributes of the English /g/

category. Farsi uvular stops, on the other hand, share fewer attributes with those of English /g/ or have attributes that are loosely similar but not identical with those for /g/.[2]

Another way native and nonnative contrasts can interact, in the literature termed as two-category assimilation (Best et al. 1988), can be found in cases where the native language does not exactly share a contrast with a nonnative language but the native language does have an analogous contrast that facilitates perception of the nonnative contrast. For example, French does not include a voicing distinction for dental fricatives such as /👎/ - /❄/. Nevertheless, native French listeners can discriminate voiced from voiceless English fricatives, perceiving them as versions of French dental stops /d🗁/ and /t🗁/ respectively (Jamieson and Morosan 1986). The results show that French listeners perceive the English fricatives as versions of French stops because the acoustic and auditory attributes of the dental fricatives are well correlated with attributes of the French dental stops – a similar scenario will certainly also hold for Polish, which has dental stops with an implemented voicing contrast. Similarly, Michaels (1974, reported in Flege 2003) noted that Russians tend to substitute /t/ for English /❄/ whereas Japanese learners substitute /s/ even though both Russian and Japanese have /t/ and /s/. He hypothesised that Russians' perception of "non-stridency" in English /❄/ leads them to substitute the closest non-strident Russian sound /t/, whereas Japanese speakers' perception of "continuancy" in English /❄/ leads them to substitute the closest continuant sound in Japanese, which is /s/.[3]

2.3. L2 sound perception models

The enormous body of data collected from cross-linguistic perception research called for systematisation and typology. This led to the formation of several models attempting to explain the processes present in L2 perception, as well as predict the course of L2 perception learning. Considering their approach to L2 sound perception and the claims they make, they can be divided into phonological models represented by Phonological Interference Model (Brown 1998, 2000) and Ontogeny Phylogeny Model (Major 2001, 2002a,b), and phonetic models. In general, phonological approaches try to account for L2 acquisition by assuming that learners have a formal knowledge that underlies their

---

[2] Cognitive linguists adduce this example to speculate on prototypes and category formation proposed by Rosch (1978, Rosch et al. 1976) (see Taylor 1989).

[3] Polish learners seem to be less consistent, substituting both /t/ and /s/ for English /❄/.

observable linguistic behaviour and performance. They base this assumption on the general proposal of generative linguistics (Chomsky 1957) that performance is not always equal to competence because it can be constrained by non-linguistic factors that may be sociological or psychological in nature. The knowledge that underlies performance in the area of segmental phonology can be viewed as a system of structures that is represented in learners' minds. Also, phonological proposals consider distinctive features to be units of analysis for describing phonological systems. Since within the interest of the present study are the fine-grained phonetic features, we shall leave these models aside and concentrate on the phonetic approach (see Escudero 2005 for a review of phonological models).

Phonetic approaches do not rely on abstract systems that shape the learner's performance. Instead, they consider the actual phonetic components of the acoustic signal that form sound categories.

2.3.1. Speech Learning Model

Although the Speech Learning Model (SLM) has been primarily concerned with the ultimate attainment of L2 production (Flege 1988, 1992, 1995, 1999, 2002) it has recently begun to show interest in the ultimate attainment of L2 perception (Flege 2003). It focuses explicitly on L2 acquisition. SLM starts with two broad assumptions. First, "bilinguals cannot fully separate their L1 and L2 phonetic subsystems (Flege 2003: 326). Second, the capacities underlying successful L2 speech acquisition remain intact across the life span (Flege 2003: 327). The second assumption stands in contrast to the view that speech learning is constrained by the critical period (Lenneberg 1967).

SLM "does not discount the proposals […] regarding the filtering or warping of L2 input" (Flege 2003: 327). Therefore, in early stages of L2 speech learning, learners filter out phonetic features that are used to distinguish L2 but are absent in their L1. Munro (1993) found that native Arabic speakers who had lived in the United States learnt to produce a native-like spectral difference between English /i:/ and /ɪ/ which differ acoustically from the closest vowels in Arabic. Nevertheless, they exaggerated the temporal difference between English /i:/ and /ɪ/ trying to produce phonologically long and short vowels rather than "tense" and "lax" English vowels. In perception, Flege and Hillenbrand (1986) showed that identifying fricatives as /s/ and /z/ in American English, native English listeners used two well-known phonetic cues to the syllable-final voicing contrast (discussed in detail in Chapter 2, Sections 4 and 6), i.e. fricative duration (longer for /s/ than for /z/) and preceding vowel duration (shorter before /s/ than before /z/). Nonnative, Swedish and Finnish listeners, who have no phonemically contrastive /s/ - /z/ pairs in their native language but do have contrastive long and short vowels, used only vowel length differences to differentiate /s/ from /z/. Accordingly, Swedish and Finish listeners might have reinterpreted the role of phonologically contrastive vowel duration in their L1 as a cue to the voicing contrast in nonnative listening. More recently, Broersma (2005) suggested that Dutch listeners may have used other than vowel duration cues present in their L1 in perceiving the fricative voicing distinction.

SLM proposes that native versus nonnative differences are more likely to arise as the result of interference from prior phonetic learning than from a loss of neural plasticity. Therefore, even adults retain capacities used by infants in acquiring L1 to establish new phonetic categories for vowels and consonants in L2. However, according to SLM, formation of native-like L2 categories decreases with age. Phonetic categories develop in L1 through childhood and adolescence.[4] When they fully mature, the L2 sound categories are blocked and suppressed.

In SLM, phonetic categories interact through mechanisms called "phonetic category assimilation" and "phonetic category dissimilation" (Flege 2002). When a new category is established for an L2 speech sound in a phonetic space occupied by and L1 sound, the new L2 category and the old L1 category will dissimilate. As a result, neither the L1 category nor the new L2 category will be identical to the categories possessed by monolinguals. Category assimilation, on the other hand, is predicted to occur when a new L2 sound is significantly different from the closest L1 sound and a new category has not been established. In such cases, an L2 learner will "develop a 'composite' category that merges the properties of the L1 and L2 categories that have been perceptually equated, in proportion to the input received" (Flege 2003: 330). Consequently, the L2 sound will remain L1-like and the L1 sound will eventually become L2-like (Flege 1987).

SLM argues that the state of development of L1 categories at the time of L2 acquisition will affect the native-like attainment of L2 perception. It results from the

[4] Children are believed to fully acquire the phonemes of their L1 by the age of 8, however, the motor control development continues well into adolescence (Hazan and Barrett 1999, Johnson 2000, Walley and Flege 2000).

fact that the more L1 categories are developed, the more likely they are to block the formation of new categories for L2 sounds. Consequently, native-like L2 perception will be more likely to be found in learners that have an early age of arrival in the L2 community than in learners with a late age of arrival (Flege and MacKay 2004 reported in Escudero 2005, Waniek-Klimczak 2005 for production). Moreover, learners who frequently use their L1 will be less likely to attain native-like L2 perception than those who experience full submersion in L2 society (Flege et al. 1999, Piske et al. 2001, Flege and MacKay 2004, Waniek-Klimczak 2005 for production).

2.3.2. Perceptual Assimilation Model

The Perceptual Assimilation Model (PAM) (Best 1995, Best et al. 2001) proposes that adult listeners have no mental representations or mental perceptual mappings for sound perception and that they directly seek and extract the invariants of articulatory gestures. This proposal is based on Articulatory Phonology (Browman and Goldstein 1986, 1989) and the Direct Realist Approach (see Section 1.2.2.). In the beginning, infants hear and detect every ariculatory gesture and later on they learn to detect only high-level features, i.e. those that signal sound contrasts in their native language. Once a child is able to process the high-level features that form the phonological system of their language, the task of perceiving L1 sounds becomes easier and more adult-like.

In the L1-L2 scenario, PAM proposes that the accuracy with which L2 speech sounds are discriminated depends on how they are perceptually assimilated by L1 speech sounds. Distinct L2 categories that are not perceptually assimilated or suppressed by any L1 category will be discriminated well, even in the absence of prior experience. However, when L2 speech sounds are perceptually assimilated by an L1 category, as is often the case at the beginning of a learning process, their discrimination will significantly decrease. PAM hypothesises that L2 speech sounds will be discriminated more accurately if they are assimilated by two distinct L1 speech sounds than if they are assimilated by a single L1 speech sound category. In other words, accuracy in the discrimination of nonnative sounds depends on the way they are assimilated to the L1 sounds. L2 speakers have already tuned their linguistic perceptual device to particular features in their native system and will have difficulty detecting other features in the new language.

In the developmental process, according to PAM, learners will be able to perceive a nonnative contrast by splitting their L1 categories. For example, Best and Strange (1992) suggest that exposure to L2 input may lead to the reorganisation of assimilation patterns in cross-language perception. PAM does not, however, address the problem of how successful L2 learners ultimately can be.

2.3.3. Second Language Linguistic Perception Model

The most recent proposal, the Second Language Linguistic Perception Model (L2LP) (Escudero 2005, 2006), is based on the Linguistic Perception Model (Boersma 1998, Boersma et al. 2003, Escudero and Boersma 2003) which is a phonological proposal for explaining speech sound perception. The L2LP model provides a rigorous phonetic and phonological description of L1 and target L2 perception. The emphasis is put on the optimal perception hypothesis which states that an optimal listener matches perception with production, which means that their use of auditory dimensions matches the use of the same dimensions in production.

The L2LP model proposes that the description of optimal L1 perception leads to predicting the initial state for L2 acquisition, i.e. the perceptual system that learners initially use in their L2. The learner automatically uses their entire L1 perception categories when starting to learn their L2.[5] When the learning process begins, the learner can encounter tasks that differ in both number and type, depending on how the initial L2 perception compares to the target L2 perception. The model provides an explicit and comprehensive account of how L2 learners develop the linguistic knowledge that will turn them into optimal L2 listeners. L2 learning is predicted to be governed by the same mechanisms that are present in the acquisition of L1 sound categories. That is, L1 learning device, which is responsible for the perception and recognition learning in L1, also applies to L2 acquisition.

The model assumes that L2 learners will be confronted with learning tasks that depend on the cross-language differences between their L1 and L2 optimal perceptions. The number and nature of the tasks will determine the learner's L2 sound perception scenario and the level of difficulty in the pursuit of optimal L2 perception. The **new**

[5] The L2LP model puts great stress on testing L2 perception with beginners. Only then can the complete L2 perception learning scenario be obtained.

scenario, in which a representational learning task is either to create new categories or to split already existing ones, is the most difficult. Learners who face the new scenario do not reach optimal performance because they show clear signs of creating new categories without having the necessary cue integration to optimally perceive the target L2. According to the authors, this does not mean that these learners will not attain optimal perception but only that they will do so with great difficulty. For the **subset** scenario, the L2LP proposes that lexicon-driven perceptual learning will be initiated by recognition learning. That is, the learning task is to reduce the number of perceptual categories. Learning starts when recognition has to change due to a semantic-driven error, i.e. too many sound categories impede correct lexical recognition. This recognition-perception mismatch results in the gradual reduction of pre-existing sound categories. This process is found to be medium difficult. The **similar** scenario occurs when the learner perceives the same number of sounds as those produced in the target language because their L1 has the same number of sound categories. The perception of an L2 contrast that has a corresponding contrast in the L1 and phonetic differences in the L1 and L2 sound categories result in slight differences in sound categorisation. Because the similar scenario only presupposes a perceptual task, i.e. adjustment of perceptual boundaries and not creation or suppression of other categories, it is considered to be the least difficult. It is contrary to the SLM (Section 2.3.1.), which claims that the perception of similar sounds in L2 poses the greatest challenge and the acquisition of similar sounds will hardly ever result in native-like L2 perception.

2.3.4. Native Language Magnet Model

The Native Language Magnet Model (NLM) (Kuhl 1991, 1993, 2000a,b) is discussed as the last proposal because it is mainly interested in L1 perceptual acquisition by infants, however recently it has turned its interest into cross-linguistic processes. It accounts for the transition from auditory to language-specific perceptual processing. The NLM proposes that perception of the acoustic properties of speech sounds is defined by early experience. Infants perceptually sort segment-sized units into categories based on the recurrence of features they have detected in speech input. Kuhl (2000a) puts forward a body of evidence (Kuhl et al. 1992, Goodsitt et al. 1993, Saffran et al. 1996) showing that infants acquire sophisticated information from the signal through the detection of the distributional and probabilistic properties of the ambient language.[6] It is argued that infants' perception becomes language specific through the categorisation, statistical processing, and perceptual warping of acoustic dimensions, all of which take place within their first year.

The NLM proposes that infants' perceptual mapping of ambient language speech sounds crates a "complex network, or filter, through which language is perceived" (Kuhl 2000a: 11854). Therefore, L1 language-specific filter will make the acquisition of L2 much more difficult because future learning is constrained by the initial mental mappings that have engaged neural structure. In other words, learning to perceive L2 sounds is constrained by the initial mapping, i.e. the native language sound mapping that has taken place. Moreover, this constraint operates independently of any critical period. However, Kuhl (2000b) also suggests that early in life, interference effects are minimal so that two different mappings can be acquired, whereas when a second language is learnt after puberty another form of separation between the two perceptual systems may be required to avoid interference. This difference has been shown in brain imaging studies which have found that adult bilinguals who acquire both languages early in life (or dialects, Abutalebi et al 2006) activate overlapping regions of the brain when processing the two languages, whereas late learners activate two distinct regions of the brain (Kim et al. 1997, Abutalebi et al. 2001, for most recent reviews of the research on bilingual brain using the EEG, PET, and FMRI techniques see also Dijkstra 2007, Indefrey 2007).

2.4. Can L2 perception influence L1 perception?

Whether the experience with L2 sound categories can influence L1 perception is a debatable matter and depends on the assumed model. The two models that give it the most attention and represent totally contrasting standpoints are the Speech Learning Model (SLM) (see Section 2.3.1.) and the Second Language Linguistic Model (L2LP) (see Section 2.3.3.). The SLM proposes that L1 and L2 phonetic categories are represented in a common phonological space so that both systems mutually influence one another. As a consequence, it is predicted that when a new phonetic category is

[6] There is some counter evidence though (Sussman and Lauckner-Morano 1995, Lively and Pisoni 1997, Lotto et al. 1998 Frieda et al. 1999). The problem with finding information about typical speech input distributions for infants is the lack of control over the quality and quantity of language experience.

established for an L2 sound that is close to an L1 sound, it will dissimilate (Flege 2002) or it will cause a change in a feature weighting (Gottfried and Beddor 1988, also Francis and Nusbaum 2002). As a result, the L1 and L2 categories of bilinguals will be different from those of native speakers of the two languages thus leading to different L1 and L2 perception (Flege et al. 2003).[7] In a situation when a new category is not established for an L2 sound that differs audibly from the closest L1 sound, experienced L2 learner will be expected to develop a composite or merged category that contains both the L1 and L2 categories, the situation which results from assimilation (Flege 1987, MacKay et al. 2001 reported in Escudero 2005). Additionally, Flege (2002) argues that the principles of assimilation and dissimilation as well as the existence of a common system may underlie Grosjean's (1989, 2000) claims that the bilingual's two systems are always engaged at the same time so that the mixing of L1 and L2 is inevitable. For example, in Rojczyk (forthcoming), we primed the production vowel duration differences in Polish among Polish-English late bilinguals. The subjects showed increased durations in vowels preceding phonologically devoiced consonants in Polish when presented with English stimuli of a similar phonological form in a priming experiment (300 ms intervals between the presentation of English and Polish stimuli). We concluded that the interference of L2 in a code-switching paradigm leads to a temporal reorganisation of L1 under the influence of L2.

A different prediction is proposed by the L2LP model (see Section 2.3.3), which claims that for both L1 and L2 categorisation to be optimal and for the two languages not to influence each other in their representations, they must be separate systems (Boersma et al. 2003, Escudero and Boersma 2003, Escudero 2005). If two separate systems underlie the perception of two languages, it is proposed that "L2 development need not affect the already optimal L1 perception provided that sufficient input for both languages is received" (Escudero: 2005: 313). Contrary to the SLM, the L2LP model proposes that learners who use their two languages to similar extents will exhibit L2 development as well as L1 stability. It is argued that an L2 learner can attain optimal L2 perception and maintain their optimal L1 perception because the two languages have separate perception grammars.

[7] The effect is weaker in phonological processing when code switching (e.g., Meuter and Allport 1999) – "[t]he asymmetry in the effects of language mixture suggests that normally when processing L2, L1 is active and influences performance. When processing in L1, L2 may or may not be active, but the time course of processing the more dominant language may allow selection to occur at an earlier point in time" (Sebastián-Gallés and Kroll 2003: 305).

2.5. Summary

This part provides general characteristics of human auditory perception and its role in second language acquisition. The fact that humans are able to perceive, recognise, and process acoustic signal underlies the whole verbal communication and makes it an essential property of human language. Every infant must be equipped to acquire some of hundreds of different sounds, depending on the language it is born into. It is true though that there is no invariance in the speech signal and hence no one-to-one mapping between the acoustic form and a segment. Whether the defining properties of speech sounds are in neuromotor commands to the articulators, actual phonetically structured vocal tract movements, or trading relations between different properties of the acoustic signal has been a matter of heated debates for many years in the psycholinguistic literature. A commonly accepted fact is, however, that humans possess capacity to divide acoustic continuum and perceive speech sounds discretely, which gives rise to sound categories that can perform phonological functions.

Problems that every second language learner faces when learning L2 speech sounds have been of great interest in speech perception research since the very beginning. The learners, circumscribed by their L1 perceptual map, filter and warp L2 categories, which will inevitably result in impaired recognition. Common observations indicate that learners who begin to learn L2 will have problems attaining native-like perception, however recent results demonstrate that it is not impossible, which means that learners can readjust their perceptual device and are not necessarily limited by the critical period. The very process of learning L2 speech sounds is seen differently depending on the assumed model. The Speech Learning Model claims that learners filter out phonetic features that are used to distinguish L2, but are absent in their L1, and even when proficient learners create L2 categories, they will hardly ever be native-like due to constant interference from L1. The Perceptual Assimilation Model proposes that, at early stages of learning process, L2 categories may be assimilated by an L1 category and only when L1 categories are split will learners be able to discriminate L2 contrasts. The Second Language Linguistic Perception Model emphasises that learners of L2 are equipped with the same learning mechanisms they used in acquiring L1 and, depending on the cross-linguistic scenario, they will have to create new L2 categories, or reduce or readjust pre-existing L1 categories. The Native Language Magnet Model suggests that L2 learners, like infants acquiring native language, extract statistical recurrence of features in the speech signal to create new sound categories, the only difference being that L2 learners are constrained by their L1 mapping, which will necessarily impede L2 perception by category interference. Finally, we have addressed the problem of whether L2 perception can influence L1 perception and concluded that proposals diverge depending on the assumed model. The SLM proposes that L1 and L2 phonetic categories are represented in a common phonological space so that both systems mutually influence one another. To the contrary, L2LP argues that an L2 learner can attain optimal L2 perception and maintain their optimal L1 perception because the two languages have separate perception grammars.

Part Two

Temporal and spectral properties of the voicing contrast in obstruents

Introduction

The voicing contrast has long enjoyed widespread attention in the phonetic and phonological literature. As one of the most powerful contrastive tools, it has found a prominent place in all phonological models. However, early proposals classifying the voiced-voiceless distinction as presence or absence of vocal fold vibration turn out to be far from exhaustive. The voicing contrast is not only expansive in that it affects neighbouring sounds but also its implementation differs across manners of articulation. Nor are all languages unified in the same realisation of the voiced-voiceless opposition. This is what acoustic phonetics has clearly demonstrated and phonological models seem to have long overlooked.

In this part, we look at different voicing implementation strategies operative in English and Polish. We begin with a review of a phonological approach to voicing and discuss the concept of a fortis-lenis opposition which, albeit very useful in phonological descriptions, finds little support in phonetic and acoustic experiments. Next, we proceed to the Voice Onset Time, which has been found to be a strong and reliable temporal parameter of voicing in initial stops both in production and perception. That the voicing contrast realised in one segment can influence the production of neighbouring segments will be demonstrated by the discussion of preceding vowel duration where the voicing status of an ensuing obstruent can change the temporal duration of a preceding vowel. It is compensated, however, by the duration of a consonant itself, as shall be demonstrated in subsequent sections. Like in stops, the voicing contrast in fricatives is realised by durational parameters but, additionally, fricatives show variation in acoustic features of frication noise, which will be discussed in a separate section. Finally, affricates, the most underresearched group of obstruents, combine elements of the voicing contrast implementation found both in stops and fricatives.

1. Voicing and voicelessness

The voicing contrast in obstruents is a ubiquitous phenomenon in the world's languages (Maddieson 1984, Lotto and Kluender 2002). For example, English is heavily dependent on this feature (Denes 1963 reported in Pickett 1999) and it is the only laryngeal feature that it employs distinctively.[1] Generally, however, not very precisely as shall be seen in subsequent sections, the term 'voicing' refers to the vibratory action of the vocal folds - caused by an adequate pressure drop (van den Berg 1958) - which produces voicing periodicity in the speech wave. Voicing is usually present during the constriction of voiced consonants and absent during the constriction of voiceless consonants. This difference is controlled by muscles in the larynx that hold the vocal folds in either a closed position for voicing or in an open position for voicelessness (Baer 1975, Jansen 2004, Heffner 1964 for physiological details).

Although the word 'voicing' is meant to be a technical term in linguistics, "yet it is beset with a certain amount of confusion" (Abramson 2000: 25). Phonologists working within a particular theoretical framework use the term as a label for an abstract phonological feature that is said to play a distinctive role in grammar. For instance, traditional phonological descriptions differ in their treatment of voicing in English initial stops (discussion in Healy and Levitt 1980, Keating 1980, 1984, Lisker 1984, 1986, Westbury and Keating 1986). Heffner (1964) considers any contrast of the type /b, d, g/ - /p, t, k/ to be one of voicing, even though English stops are not voiced throughout their total length. The aspiration of initial voiceless stops does not find its way into Heffner's description of the basic contrast.

Trubetzkoy (1969) used three phonetic features; [+/-voice], [+/-tense], and [+/-aspirated] to describe various voicing categories. It was believed that while the same phonetic features could be used for various languages, the distinctive feature for each language must be determined independently by phonological evidence. Slavic languages were described as having distinctive [+/-voice], redundant [+/-tense], and no aspiration. French, English, and German had co-varying [+/-voice] and [+/-tense] but it was not specified which was distinctive and which was redundant. In English, [+/-aspiration] was an allophonic feature associated with voiceless tense stops.

[1] Other types of laryngeal and glottal distinctions for consonants in the world's languages are breathy voice, slack voice, creaky voice, and stiff voice (Stevens 1977, Henton et al. 1992).

Jakobson, Fant, and Halle (1951) decided that tenseness should be distinctive, at least for English and French, so that those two languages had redundant voicing, while Slavic languages had distinctive voicing. Tenseness and aspiration were said to be related and tenseness was used in describing three-way contrasts together with voicing.

Abercrombie (1967) proposed to describe voicing contrasts as a choice of two or three of the following categories: 'voiced', 'voiceless unaspirated', and 'voiceless aspirated'. These categories were given an articulatory and acoustic basis by Lisker and Abramson's (1964) Voice Onset Time (see Section 3) with five phonetic categories along a single continuum. These five phonetic categories were combined into the three contrastive categories. In this view, aspiration was not an independent feature but a natural concomitant of one of the voicing categories. Similarly, Ladefoged (1971) suggested five phonetic categories: 'fully voiced', 'partly voiced', 'voiceless unaspirated', 'voiceless slightly aspirated', 'voiceless aspirated'. Accordingly, French voiced stops were 'fully voiced', while English ones were 'partly voiced'.

Although Chomsky and Halle (1968) were aware of the five-category division of the Voice Onset Time (Keating 1980), they wished to maintain binary features rather than multi-valued scales at the phonological level. They chose to describe four of the five Voice Onset Time categories with four binary features based on articulatory configurations: [+/-stiff vocal cords], [+/-slack vocal cords], [+/-spread vocal cords], [+/-constricted vocal cords]. Although the obtained output was fairly precise, it definitely lacked simplicity (Sommerstein 1977). Later work in generative phonology used some version of [+/-voice] plus [+/-aspiration] to represent voicing contrasts and phonetic forms (Jassem 1983).

In a fine-grained phonetic approach towards the description of voicing contrast in languages, it has long been observed (e.g. House and Fairbanks 1953) that various acoustic properties are often found in conjunction with voicing distinctions (Abramson 2000). They are the Voice Onset Time, preceding vowel duration, closure duration, and frication noise for fricatives.[2] They constitute an interesting domain of comparison

[2] We are acquainted with the phenomenon of fundamental frequency variation (Lehiste and Peterson 1951, Haggard et al. 1970, Umeda 1981, Kohler 1982, 1984, Ohde 1984, Kingston 1986, Diehl 1991, Kingston and Diehl 1994, Diehl and Mollis 1995, Holt et al. 2001) and first formant cutback (Cooper et al. 1952, Liberman et al. 1958, Haggard et al. 1970, Fujimura 1971, Stevens and Klatt 1974) as a concomitant of the voiced-voiceless distinction. We did not include them in the present study for two reasons. First, it is impossible to manipulate those parameters using natural speech stimuli, which are a methodological foundation of this study. Second, the data suggest that both parameters are only subsidiary to the Voice Onset Time. In comparison to fundamental frequency, "voice onset time is clearly the dominant cue (Abramson and Lisker 1985: 32) and the results "suggest f0 does not exert an obligatory

between English and Polish and it is for this reason that we discuss them in great detail in the following sections and indicate their usefulness in cross-linguistic perception research. First, however, we look at the concepts of 'fortis' and 'lenis', which have long accompanied the voiced-voiceless distinction.

2. Elusiveness of the fortis-lenis distinction

Because the concept of voicing is phonologically problematic in allophonic realisation, i.e. for segments undergoing positional partial devoicing, some linguists - as early as Rousselot (1924, reported in Bell-Berti 1975) - rejected voicing as a phonologically relevant feature and concluded that some other feature must distinguish members of the phonemic category from its allegedly voiceless counterparts. Therefore, the opposition fortis-lenis or tense-lax came into being for English or German (Jakobson and Halle 1962, Chomsky and Halle 1968). It is claimed that the compressions of voiceless consonants are articulated with more force or tension than for the voiced consonants. In other words, when voicing occurs in the lax category, it is but a secondary or concomitant effect of articulatory effort. The fortis-lenis or tense-lax features are believed to be sufficient in distinguishing voiced and voiceless obstruents.

However, physiological experiments looking for the difference in tension have found it elusive (discussion in Malécot 1970, Bell-Berti 1975, Slis 1975, Pickett 1999, Abramson 2000). Measures of the mechanical pressure of the lip in /p/ and /b/ did not show significant differences in tenseness (Malécot 1966, Lubker and Parris 1970). On the other hand, the area of tongue contact on the hard palate for alveolar stops is reported to be greater for /t/ than for /d/, however the study was conducted with Japanese as the language base (Fujimura et al. 1973). Although measures of intraoral air pressure have generally shown greater values for /p/ than for /b/ (Tatham and Morton 1969, Warren and Hall 1973), Lisker (1970) concludes that intraoral air pressure is not sufficient for categorising /p, t, k/ from /b, d, g/. Moreover, Netsell (1969) demonstrated

influence on categorisation of consonants as [+/-VOICE]" (Holt et al. 2001: 764). Likewise, the first formant cutback is "neither necessary nor sufficient to elicit voiced stop judgements" (Lisker 1978a: 375), and the Voice Onset Time is both "the more effective cue" (Lisker 1978a: 375) and "does emerge as most potent perceptual cue" (Summerfield and Haggard 1977: 436), even though it can push VOT boundaries in perception (Cooper et al. 1952, Liberman et al. 1958, Sawusch and Pisoni 1974, Lisker et al. 1977, Parker 1988, Kluender 1991).

negligible differences in subglottal pressure between the English voiced /d/ and aspirated voiceless /t/.

Electromyographic studies undertaken to determine the differences in the strength between voiced and voiceless obstruents have not found consistent variation in the EMG strength. Harris and colleagues (1965), and Fromkin (1966) reported finding no consistent difference between /p/ and /b/. Although Tatham and Morton (1973) found small but significant differences in EMG signal strength at the point of release of /p/ and /b/, Bell-Berti's data (1975) reveal different patterns of muscular activity for individual subjects – they used a different arrangement of muscle activities to achieve a pharyngeal cavity expansion necessary for the continuation of glottal pulsing during voiced stop consonant occlusion.

The fact that articulatory force "has no agreed-upon physical meaning" (Lisker and Abramson 1967: 3) led to questioning *raison d'etre* of the fortis-lenis distinction. Lisker and Abramson (1964: 385) write:

> No one of the physical measures, whether physiological or acoustic, that have been proposed as correlates of the fortis/lenis dimension, has been shown *not* to be significantly connected with voicing or aspiration. And in fact an examination of the phonetic literature generally fails to turn up any language which is said to possess stop categories that differ only in force of articulation. For languages in which the fortis/lenis difference is invoked, it is too often the case to be accidental that voiceless and aspirated stops are discovered to be fortis, while voiced and unaspirated ones are at the same time lenis […] The ambiguous status of the terms "fortis" and "lenis" (or "tense" and "lax") is also reflected in statements by several writers to the effect that a number of phonetic features, *among them voicing and aspiration*, may be taken as manifestations of an underlying division of stops on the basis of a fortis/lenis opposition.

Likewise, Bell-Berti (1975: 460) observes:

> It is clear, then, that the feature [+/-tense] is inadequate for describing the pharyngeal volume changes concomitant with voicing distinctions, as that feature at best explains the larger portion of some speakers' pharyngeal adjustments and never explains the full measure of enlargement.

However, recent opinions on the fortis-lenis opposition are more conciliatory and do not reject the articulatory force altogether from the phonological level. Abramson (2000: 27) notes:

> I hasten to add here that being skeptical of the foregoing argument [that obstruents are differentiated by force and not voicing] does not require the dismissal of the physiological possibility of using level of effort for phonological distinctions. For example, a language can use extra contraction of the thyroarytenoid muscle for systematic shifting of voice quality in the vowel following the release of members of a particular consonant class[3]: the phonologist might then reasonably invoke a feature of tensity. In the case of the absence of voicing in certain "voiced" segments, however, those phonologists leapt to a conclusion without good phonetic evidence.

Similarly, Pickett (1999: 125) adds in a footnote:

> Some linguists prefer the terms *lax* or *lenis* for voiced and *tense* or *fortis* for unvoiced, referring to evidence that the constrictions of unvoiced consonants are articulated with greater force or tension than for the voiced consonants. This is true, but the present author believes this may be only a secondary, synergistic effect, necessary to contain the higher air pressure in the mouth that occurs on unvoiced consonants because of the wide-open posture of the vocal folds. The primary factor is believed to be the open or closed posture of the vocal folds.

Recent publications dealing with obstruent voicing are wary enough to signal the distinction between phonological and phonetic voicing. While the feature [+/-tense] might be useful in phonology, it finds little support in phonetics and speech perception research. Jansen (2004: 4) uses the criterion of force to highlight the distinction between the phonological and phonetic analysis: "[n]ote that this is intended solely to keep the distinction between phonological and phonetic categories maximally clear; it is certainly not meant that tense and lax are useful concepts in dealing with phonetic substance."

[3] [AR] Slis (1975) proposes a neuromuscular theory of voicing in stops which states that the acoustic features of unvoiced voiceless stops and tense (long) vowels are due to stronger neural commands to the articulators than for the voiced stops and lax (short) vowels.

3. Voice Onset Time

The Voice Onset Time (VOT) introduced by Lisker and Abramson (1964)[4,5] is defined as the single production dimension, the time interval between the release of a stop occlusion and the onset of vocal cord vibration, or in the authors' own words "the time interval between the burst that marks release and the onset of periodicity that reflects laryngeal vibration"(Lisker and Abramson 1964: 422). In a survey of 23 languages, they found that word-initial stops fall into three broad categories that show little cross-linguistic variation:

1. voicing lead or negative VOT – voicing starts well before the release of the plosive (approximately -30 ms or more). It is present acoustically as "low-frequency harmonics of a buzz source" (Keating et al. 1981: 1264) or simply "laryngeal buzz" (Lisker 1986: 8) (see Figure 3.1.).[6]

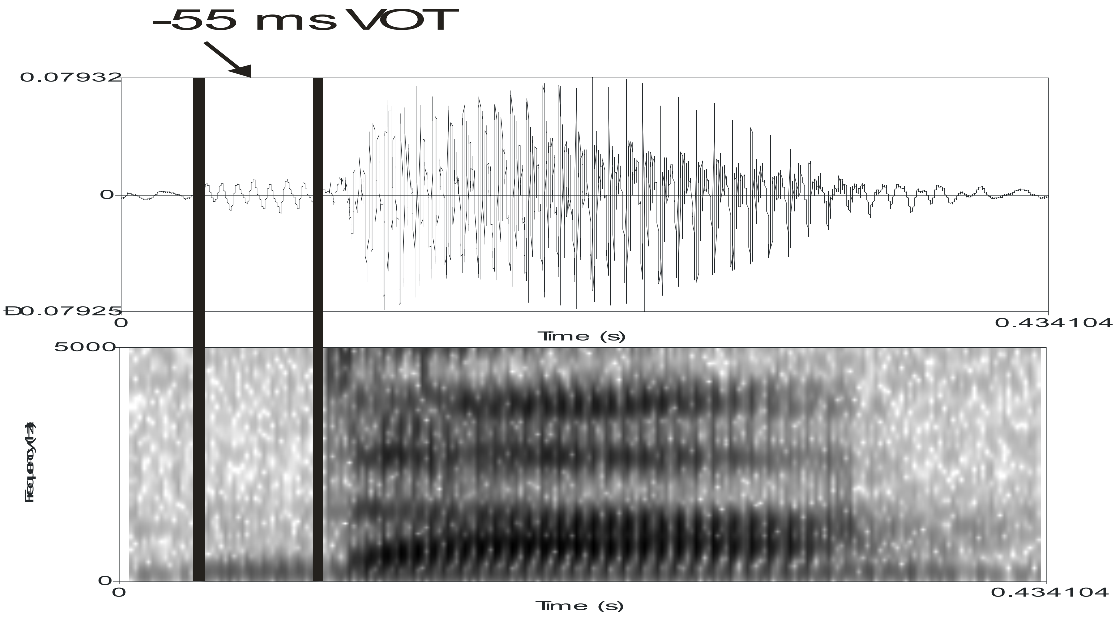


Figure 3.1. Waveform and spectrogram of the syllable *da*. Voicing lead -55 ms VOT

[4] With later refinements (Lisker and Abramson 1965, 1971, Abramson and Lisker 1965, 1985, Abramson 1977)

[5] Research in this direction started as early as in 1958 (Liberman et al. 1958) with the proposal of F1 cutback as a cue to the voicing contrast (Liberman 1996). However, Damper and Harnad (2000: 844) report a personal communication with Michael Studdert-Kennedy, who pointed out that F1 cutback was viewed as a purely acoustic variable, while VOT was originally an articulatory and temporal variable.

[6] Measurement criteria originally defined by Lisker and Abramson (1964: 389); "the point of voicing onset was determined by locating the first of the regularly spaced vertical striations which indicate glottal pulsing, while the instant of release was found by fixing the point where the pattern shows an abrupt change in overall spectrum. Oral closure is marked spectrographically by the total or almost total absence of acoustic energy in the formant frequency range; oral release is marked by the abrupt onset of energy in the formant frequency range".

2. a short lag or zero onset – voicing starts at or shortly after the stop release (approximately 0 to +30 ms, maximum + 35 ms (Keating 1984)) (see Figure 3.2.).

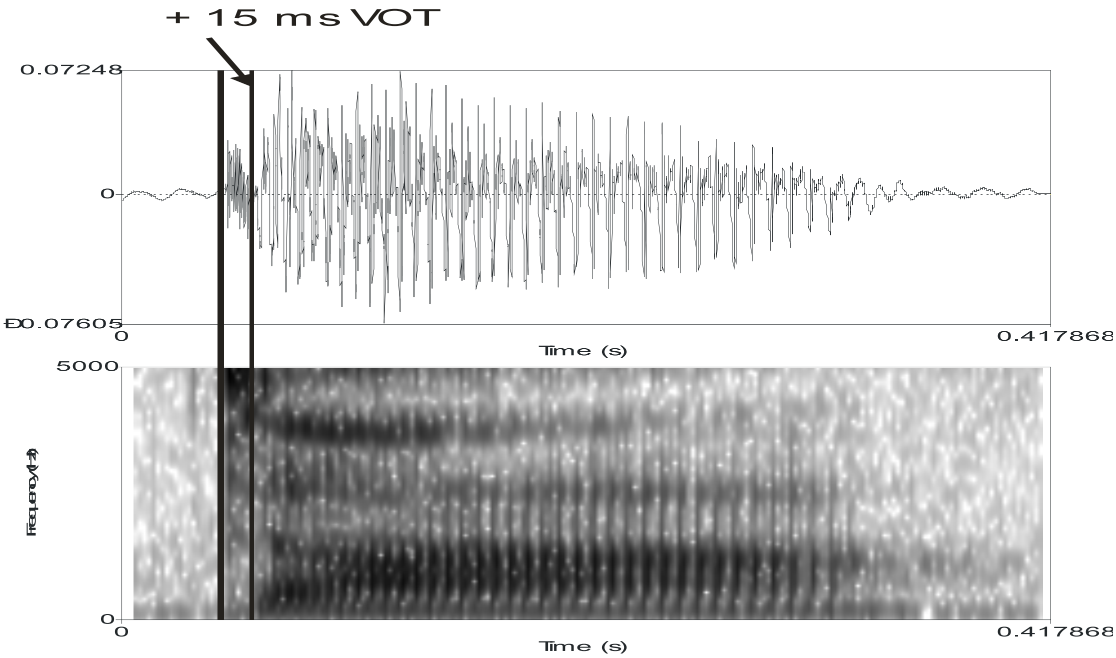


Figure 3.2. Waveform and spectrogram of the syllable *da*.[7] Short lag +15 ms VOT

3. a long lag – voicing starts well after the release of the plosive (approximately +50 ms or more). It is accompanied by silence (Klatt 1975, Lisker 1986) or aspiration if "the vocal tract resonates to turbulent air passing through the open glottis (Lisker and Abramson 1964: 416). Aspiration is acoustically registered as "noise (i.e. random stippling), mostly at frequencies of the second and third formant" (Lisker and Abramson 1964: 386), "a large glottal abduction that peaks around the release of a stop" (Jansen 2004: 41), "turbulent excitation of the upper vocal tract" (Abramson 1977: 296), "turbulence formed aerodynamically […] at the somewhat open glottis" (Abramson 2000: 8), or "friction noise generated at the still-open glottis by the flow of air through the vocal tract after stop release" (Keating 1984: 295) (see Figure 3.3.).

[7] We use English phonemic labelling of VOT categories. Short lag category will have different labels in English and Polish (see Section 3.3.).

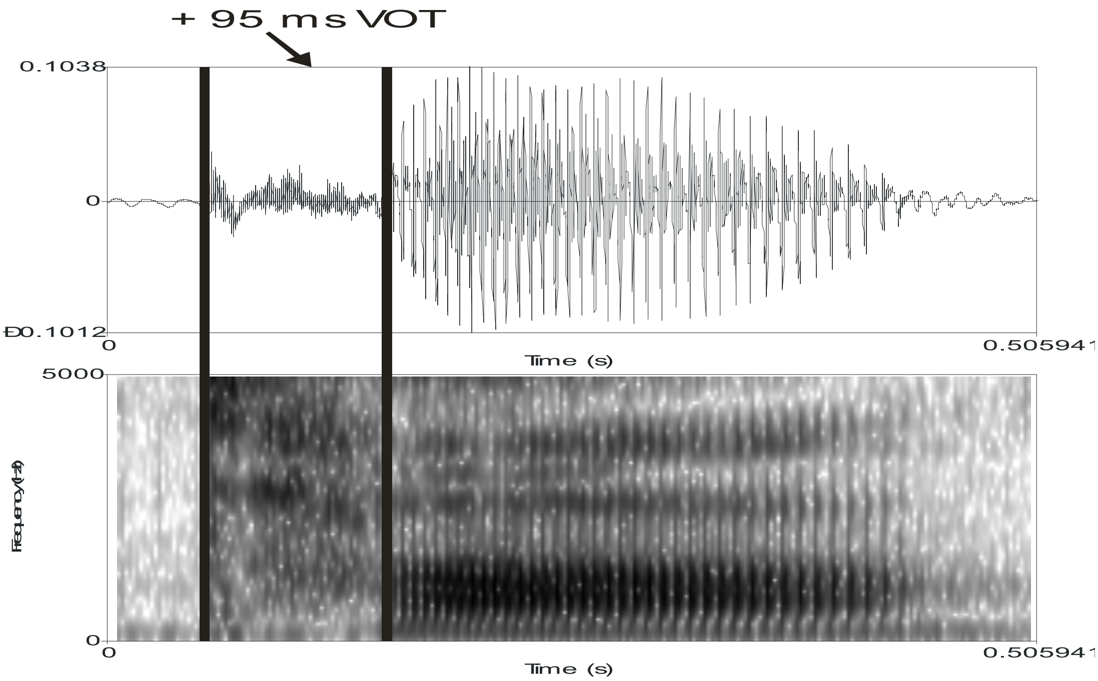


Figure 3.3. Waveform and spectrogram of the syllable *ta*. Long lag +95 ms VOT

Attempts to provide a unified phonetic conception of the voiced-voiceless distinction based on the VOT continuum (e.g. Keating 1984, Kohler 1984, Kingston and Diehl 1994, 1995) proposed the division into voicing and aspirating languages. Voicing languages contrast prevoiced plosives with short lag plosives. This type of languages dominates in eastern and southern Europe, comprising virtually all varieties of Romance and Slavonic as well as the Baltic languages and Hungarian (Jansen 2004). If a language has a single series of stops, these belong almost always to this category. 49 out of 50 languages with a single series of oral stops sampled by Maddieson (1984) have a short lag VOT.[8] Aspirating languages, on the other hand, contrast short lag with long lag stops. Languages belonging to this category are Danish, Faroese, Icelandic, Norwegian, Swedish (Jessen 1998, Jansen 2004) and standard varieties of English and German (for dialects see Wells 1982, Docherty 1992, Gimson 1994, Hughes et al. 2005).

In general, the three above-mentioned VOT categories are sufficient for the description of contrasts used in languages and even for their allophonic variation. The 51 languages surveyed by Keating and colleagues (1983) all use at least some kind of short lag stops in virtually every position. Voicing lead – short lag contrasts and short lag – long lag contrasts are equally common across the surveyed languages. There are,

[8] Language *Aleut* seems to be an exception. It has a single series of long lag stops (Cho and Ladefoged 1999).

however, languages that cannot be satisfied with only two contrasts, i.e. voicing lead-short lag or short lag-long lag. They include Thai and Eastern Armenian (Lisker and Abramson 1964), which have three contrasts. Most recently Riney and colleagues (2007) signalled that Japanese might need another category between short lag and long lag (see also Shimizu 1996). The same has been reported for Hebrew (Raphael et al. 1995). There are even languages which are claimed to exploit all three categories plus voiced aspirates such as Hindi or other languages of India (Keating 1984). In the light of all this evidence, Raphael and colleagues (1995) propose that Lisker and Abramson's (1964) original short lag class should be divided into two categories, one for the voiced stops of aspirating languages and one for the voiceless stops of voicing languages. Cho and Ladefoged (1999) (see also Cho and Ladefoged 1997, Ladefoged and Cho 2000) even identify four degrees of positive VOT.

VOT values differ across the place of articulation. Labial stops are consistently shorter than alveolar and velar stops. There appears, however, some speaker variation as to whether alveolar stops have shorter VOT than velars (Lisker and Abramson 1967, Zue 1976, Weismer 1979, Nearey and Rochet 1994) or whether they are the same (Crystal and House 1988a, Cooper 1991 reported in Whalen et al. 2007). Docherty (1992), in his study of VOT in British English, finds a distinction between labials and non-labials but does not find a robust difference between alveolars and velars. A similar pattern is found in Cho and Ladefoged (1999), who report significant differences in mean VOT values between velars and coronals (dental and alveolars) but labials and coronals are not significantly different. Most recently Whalen and colleagues (2007) found the labial<alveolar<velar pattern in babbling infants learning English and French, and Cole and colleagues (2007) found the same pattern in American radio announcers. The explanation of VOT differences depending on the place of articulation lies in articulatory principles (discussion in Cho and Ladefoged 1999) which say that the further back the closure, the longer the VOT values (Fischer-J↓rgensen 1954, Peterson and Lehiste 1960), the more extended the contact area, the longer the VOT values (Stevens et al. 1986), and the faster the movement of the articulator, the shorter the VOT (Hardcastle 1973).

Vowel quality has been found to influence VOT values (Klatt 1975, Summerfield 1975a, Waniek-Klimczak 2005). High tense vowels increase the VOT duration – Klatt (1975: 691) reports that the average VOT of long lag stops before /i, u/ is about 15% greater than before /↔, —/. At the same time, VOT categories can influence the voice quality of the vowel. A laryngographic study by Abberton (1972) showed that the onset of vowels after long lag stops has some characteristics of creaky voice with a long closed phase and a slow opening phase. On the other hand, Han (1998 reported in Cho et al. 2002) showed that vowels after prevoiced stops have a breathy voice. Moreover, changes in speech rate affect the range of VOT values, particularly for voiceless stops (Miller et al. 1986, Miller and Volaitis, 1989, Volaitis and Miller 1992, Kessinger and Blumstein 1997, 1998, Lim et al. 2001, Waniek-Klimczak 2005). VOT values decrease as the speech tempo increases and the perceptual boundary between voiced and voiceless stop consonants shifts accordingly towards shorter VOT values (Summerfield 1981, Miller et al. 1986, Miller and Volaitis 1989, Volaitis and Miller 1992).

3.1. Categorisation of VOT in perception

The continuum of VOT with the three-category distinction identified by Lisker and Abramson (1964) shows a strong categorisation pattern in perceptual labelling of voiced and voiceless stops (Abramson and Lisker, 1965, 1967, 1970, 1973, Lisker and Abramson 1967, 1970). They used a parallel-resonance synthesiser to obtain synthetic stimuli varying in small steps of VOT from -150 ms through 0 to +150 ms for labial, alveolar, and velar stops. Identification function was flat within categories appropriate for a given language but became very steep at the boundary between voiced and voiceless categories. This led to the conclusion that VOT is perceived categorically, i.e. the discrimination performance is discontinuous. Categorical boundaries depended on the place of articulation – they moved back in the vocal tract from labial through alveolar to velar with the boundaries for English from about 25 ms VOT through about 35 ms VOT to approximately 42 ms respectively. What is more, VOT categorisation depends on the rate of production and syllable structure. Miller and Volaitis (1989) generated VOT continua ranging from /bi/ to /pi/ and asked subjects to rate the goodness of the consonants as /p/. The results showed that the VOT boundaries had a dynamical character regarding the tempo of presentation. The same effect was obtained by Summerfield (1981), who demonstrated shifts in perception of VOT cues for stop voicing with changes in the speaking rate of a preceding carrier sentence. When the preceding carrier sentence rate was slow, the phonetic boundary was located at longer VOT values than when the sentence was fast.

Interesting and surprising results have been obtained with VOT perception tests administered to animals, which demonstrated the same boundaries as adult English speakers on the same stimuli. For example, Kuhl and Miller (1975) tested chinchillas' discrimination between /da/ - /ta/ syllables that varied along the VOT continuum (0 ms - +80 ms). They found the boundary values of 33 ms for the chinchillas and 35 ms for the humans. Later, Kuhl and Miller (1978) showed that chinchillas' VOT categories change with place of articulation just as they do for adult humans. The boundary for /ba/ - /pa/ was at about 25 ms for animals and humans and for /ga/ - /ka/ it was located at 42 ms for both groups. Similar experiments replicated with budgerigars showed parallel results (Dooling et al. 1989 reported in Hawkins 1999) – the birds not only showed evidence for categorical sensitivity to VOT but also their boundaries were shortest for labials and longest for velars. A less detailed study with rhesus monkeys (Waters and Wilson 1976 reported in Keating 1980) used large VOT steps (70 ms) from -140 to +140 ms. In the forced-choice format, the subjects showed the best discrimination in the 0 – 70 ms region. More recently, Holt et al. (2001) evidenced that Japanese quail can be taught to categorise VOT with different fundamental frequency in a similar fashion obtained for human listeners (Diehl 1991, Diehl and Molis 1995). Finally, both humans and birds experience the effect of varying F1 frequency on labelling VOT continuum (Kluender 1991, Kluender and Lotto 1994).

Investigations of infant perception show VOT categorisation abilities which are independent from linguistic experience. Eimas and colleagues (1971) tested whether four-week-old infants could discriminate a difference in voicing between stop consonants. They synthesised a set of syllables with VOT ranging from -20 ms to +80 ms in 10 ms steps. They obtained a category boundary at about +25 ms VOT, which was very similar to the boundary obtained for adult English speakers (see also Jusczyk et al. 1989). Lasky et al. (1975) found that infants raised in a Spanish-speaking environment can discriminate differences of VOT of the English voice contrast without experience with this language, the same was reported for infants whose environment was a language such as Kikuyu (Streeter 1976). Generally, up to about six months of age, infants discriminate three voicing categories, separated by two VOT boundaries (Lasky et al. 1975, Aslin et al. 1981, discussion in Serniclaes 2005). After six months of age, only the positive VOT boundary remains active in languages with a single distinction between short and long positive VOT categories (Eilers et al. 1979).

The afore-mentioned experiments with animals and infants point to a natural psycho-acoustic boundary located at around +35 ms (Keating 1980). The short lag - long lag voicing distinction seems stronger and more salient than prevoiced - short lag boundaries. Serniclaes (2005) argues that infants raised in an ambient language such as Spanish or French, where the perceptual boundary is located at around 0 ms (Serniclaes 1987), must learn this boundary in the course of development, while the natural psycho-acoustic positive boundary must be deactivated. Werker and Tees (1984) argue that language experience tends to maintain or even enhance natural boundaries that coicinde with phonemic boundaries (e.g. in English) and to downgrade natural boundaries that are linguistically not functional (e.g. in Spanish or French). To date there has been no convincing argument why so many languages do not take advantage of the "English" boundary if it is so natural and why they divide the VOT continuum at different places (Lisker and Abramson 1964, Ladefoged and Maddison 1996, Cho and Ladefoged 1999).[9]

3.2. VOT in a cross-linguistic scenario

Generally, learners of a target language which uses different VOT categories than their L1 produce values intermediate between L1 and L2. Flege (1987) observed that native French adults who had learned English and native English adults who had learned French produced L2 stop consonants with VOT values differing from the VOT values produced by native English and French speakers respectively. The native French learners managed to increase VOT in English long lag stops but not sufficiently to match English monolinguals. To the contrary, the native English learners decreased VOT in French stops but not sufficiently to match French monolinguals.

The produced values also depend on the obtained input. Spanish learners who had learned English primarily from native speakers in the United States produced voiceless stops with the long lag VOT typical for English (Flege 1991). However, participants who learned English primarily from native speakers of Spanish in Puerto Rico (Flege and Eefting 1987) produced English stops with VOT values that were intermediate to the VOT values typical for voiceless stops in Spanish and English. MacKay et al. (2001) found that native Italian speakers who had lived in Canada for many years

[9] An interesting and critical, albeit a bit outdated, discussion in Ehret (1987).

produced English /b/ with prevoicing less often than is typical for Italian /b/ but more often than was observed for English monolinguals. Waniek-Klimczak (1993, 1996 reported in Waniek-Klimczak 2005) demonstrated that Polish speakers of English in the UK, whose L1 contrasts prevoiced and short lag stops (see Section 3.3.), used intermediate values of English long lag stops depending on the level of L1 everyday use. The age of learning has also been shown to be critical among speakers of Polish living in the USA (Waniek-Klimczak 2005).

The VOT distinction identified in cross-linguistic production by Lisker and Abramson (1964) has been demonstrated to be operative in perception studies across languages. Lisker and Abramson (1970), and Abramson and Lisker (1973) report on identification experiments with synthesised stops, which show that native speakers of Spanish and American English place the category boundaries between voiced and voiceless stops at different places along the VOT continuum. The Spanish subjects put the category boundary between /d/ and /t/ at +22 ms, whereas English speakers placed this boundary at 35 ms. Caramazza and colleagues (1973) observed that the voiced-voiceless VOT boundaries of French learners of English were intermediate to French and English monolinguals' boundaries. They concluded that the early bilinguals would probably never match English monolinguals due to the continued influence of Fench stops. Similar findings were obtained for Spanish-English bilinguals by Williams (1980), who found that bilinguals develop compromise VOT categories reflecting the properties of phonetically different realisations of the voiced-voiceless contrast in L1 and L2. Most recent experiments demonstrated that even a brief exposure to VOT categories atypical for one's L1 may result in adjustment of VOT categorisation to handle new stimuli (Clarke and Garrett 2004, Clarke and Luce 2005).

3.3. VOT in English and Polish

English is known to partition the VOT continuum into two categories: short lag for voiced and long lag for voiceless, however prevoiced values may also occur for a voiced category[10] and short lag values for a voiceless category depending on positions

[10] Prevoiced values in English have been found to be conditioned by the place of articulation, vowel context, and speaker's sex. Other studies have shown that prevoicing is realised mainly in hyperspeech and that with increased speaking tempo voiced categories attain short lag values (Miller et al. 1986, Kessinger and Blumstein 1997, Magloire and Green 1999).

and speakers (Keating 1984). Original measurements obtained by Lisker and Abramson (1964) for initial stops show a definite boundary for short lag and long lag stops.

	VOICED	VOICELESS
labial /b, p/	+1 msec	+58 msec
alveolar /d, t/	+5 msec	+70 msec
velar /g, k/	+21 msec	+80 msec

Table 3.1. Mean VOT values for English initial stops (after Lisker and Abramson 1994: 394)

Kopczyński (1977) noted higher values for American English stops but still they show a clear division into short lag and long lag categories.

	VOICED	VOICELESS
labial /b, p/	+18 msec	+82.5 msec
alveolar /d, t/	+14 msec	+84 msec
velar /g, k/	+31 msec	+71 msec

Table 3.2. Mean VOT values for English initial stops (after Kopczyński 1977: 72)

Polish, on the other hand, contrasts voicing lead and short lag categories for voiced and voiceless stops respectively (Keating 1980, 1984, Keating et al. 1981). Voiced stops are located in negative VOT values while voiceless stops are produced with moderate positive VOT values.

	VOICED	VOICELESS
labial /b, p/	-88.2 msec	+21.5 msec
alveolar /d, t/	-89.9 msec	+27.9 msec
velar /g, k/	-66.1 msec	+52.7 msec

Table 3.3. Mean VOT values for Polish initial stops (after Keating et al. 1981: 1262)

	VOICED	VOICELESS
labial /b, p/	-78 msec	+37.5 msec
alveolar /d, t/	-72 msec	+33 msec

velar /g, k/	-61 msec	+49 msec

Table 3.4. Mean VOT values for Polish initial stops (after Kopczyński 1977: 72)

While in English a fair amount of overlap is found for voiced and voiceless stops in running speech, especially in casual conversation (e.g. Lisker and Abramson 1967, Moslin 1978 reported in Keating 1980), Polish shows remarkably little overlap in VOT values even in running speech (Keating 1980).

Polish learners of English face the task of learning to produce long lag values for English stops. It is therefore fully justified that English pronunciation coursebooks tailored for Polish learners (e.g. Jassem 1973, 1974, Bałutowa 1974, Reszkiewicz 1981, Arabski 1987, Sobkowiak 2001) encourage the learners to produce English voiceless stops with a puff of air ensuing plosion. It is aimed to move VOT to higher values by imposing intervening aspiration noise. That Polish learners have problems with mastering English long lag values has been demonstrated by Waniek-Klimczak (1993, 1996, 2005). Polish speakers of English were reported to produce intermediate values higher than Polish short lag but not high enough to match native speakers.[11]

Perception experiments in English showed that perceptual categories match the production categories. They accurately divide the English VOT continuum into short lag and long lag regions.

labial	+22 msec
alveolar	+37 msec
velar	+40 msec

Table 3.5. Labelling VOT boundaries for English initial stops (after Lisker and Abramson 1970: 565)

labial	+27 msec
alveolar	+35 msec
velar	+42 msec

Table 3.6. Labelling VOT boundaries for English initial stops (after Kuhl and Miller 1978: 910)

[11] It is interesting to note that short lag voiceless stops are considered to be articulatorily simpler than long lag stops (Westbury and Keating 1980 reported in Keating 1984) and easier to acquire (Kewley-Port and Preston 1974, Scobbie et al. 1996). However, the prevoiced category is acquired relatively late (Macken and Barton 1980, Eilers et al. 1984, Allen 1985, Gandour et al. 1986).

labial	+32 msec
alveolar	+27 msec
velar	+66 msec

Table 3.7. Labelling VOT boundaries for English initial stops (after Zlatin 1974 : 989)

Experiments with Polish point to the fact that VOT is not such a stable perceptual dimension as it is for English, being subject to strong range effects (Keating 1980). Mikoś and colleagues (1978) used VOT stimuli with ranges covering Polish VOT in production and found boundaries near the production categories typical for Polish. However, as the authors admit, they used Polish speakers who either spoke English or were exposed to it constantly. As a result, the effect of bilingualism or language contact could not be precluded. In another perception experiment with Polish monolinguals, Keating et al. (1981) observed that English boundaries are uniformly higher than any of the Polish boundaries, reflecting the fact that Polish and English use different VOT contrasts. Because in Polish the voicing contrast is between voicing lead and voicing lag, the only information that Polish subjects needed to identify a stop as voiced or voiceless was the negative or positive VOT and not its numerical values. The fact that Poles are sensitive to differences in VOT around 0 msec means that they will have difficulties with categorising English short lag versus long lag stop categories. Indeed, Kopczyński (1977) reports a strong confusion rate between American English /b, d, g/ and Polish /p, t, k/. Keating et al. (1981) conclude that the Polish type of VOT boundary around 0 msec is not due to predispositions of the auditory system – which is claimed to be set default to contrast short lag with long lag values (e.g. Eimas et al. 1971, Eimas 1975, Lasky 1975) - but rather Polish listeners must acquire a set of discrimination functions typical for Polish. The problem for Polish learners of English lies in the fact that they “may never need to establish a precise VOT category boundary” (Keating et al. 1981: 1268), which causes a cross-linguistic perceptual obstacle.

4. Vowel duration

A great many languages - but not all (Keating 1985)[12] and not in all contexts (Davis and Summers 1989) - are documented to exhibit differences in preceding vowel duration as

[12] According to Keating (1985), this regularity is completely absent in Polish, Czech, and Saudi Arabic (more discussion on Polish in Section 4.1.)

a cue to consonant voicing. Vowels preceding voiceless consonants are shorter than vowels preceding voiced consonants (Chen 1970, Klatt 1973, Luce and Charles-Luce 1985, Summers 1987, de Jong 1991, 1995, 2004, Solé 2007), which has a perceptual significance for the consonant voicing distinction (Denes 1955, Raphael 1972, Summerfield 1975b, Kluender et al. 1988). Kluender and colleagues (1988) speculate that, in the context of a preceding long vowel, a consonant closure sounds shorter than in the context of a preceding short vowel. Therefore, languages may exploit the tendency of the auditory system to exhibit durational contrast in order to enhance the subtle difference between voiced-voiceless final obstruents. (Kluender et al. 1988: 161) write:

> We suggest that the principle of durational contrast provides a natural explanation both of the speech and non-speech results. Specifically, a long initial segment makes a given medial gap seem shorter by contrast and hence, in the case of speech, more like a voiced segment. A short initial segment, on the other hand, makes a medial gap seem longer (i.e. in speech, more voiceless). Thus vowel-length differences are a means of enhancing the perceptual distinctiveness of the closure-duration cue for consonant voicing contrast.

The perceptual dependence between vowel duration and closure duration of the following consonant led to the proposal of incorporating both durations into a single measure such as the vowel-consonant duration ratio (Kohler 1979, Port and Dalby 1982). Similarly, Maddieson (1997) suggests that the difference in vowel duration compensates for the difference in duration between voiced and voiceless stops, resulting in a more constant total duration for vowel plus a following stop. However, Fowler (1992) presented counter-evidence by demonstrating that contrast does not affect judgements either of closure duration or of vowel duration. To the contrary, Fowler's results show that long vowels are associated with increased judgments that a closure is long and, compatibly, long closures are associated with increased judgments that a vowel is long. In Fowler's own words (1992: 143); "[w]hatever the reason may be for vowel durations and closure durations to vary inversely in voiced and voiceless obstruents, it is not because language communities are exploiting durational contrast in the auditory system".[13]

[13] An interesting debate on the durational contrast and the stimulus length effect in Fowler (1990, 1991), Diehl et al. (1991), and, more recently, Kluender et al. (2003).

A mechanistic approach to vowel duration differences proposes that vowels are lengthened before voiced obstruents (Chomsky and Halle 1968) or shortened before voiceless obstruents (Belasco 1953) due to the muscular activity governing the articulatory gestures. It is claimed that muscular activity for a vowel preceding a voiced consonant is greater than for a vowel preceding a voiceless consonant. For example, in electromyographic experiments, Raphael (1975: 32) found that "the acoustically measured durational differences long observed between vowels preceding voiced and voiceless consonants are primarily controlled physiologically by motor commands to the muscles governing the articulators which are active in the formation of vowels". Other studies have demonstrated differences in the timing of the onset of muscular activity of the following consonants in relation to the offset of preceding vowel activity (Ohala et al. 1968, Leanderson and Lindblom 1972, MacNeilage 1972).

4.1. Vowel duration in English and Polish

That vowels are shorter before voiceless consonants than before voiced consonants in English was noted as early as in 1950 by Daniel Jones (1950: 121), who noted that "words like heed and heat […] are distinguished solely by the length of the vowel".[14] More precisely, there is evidence that the process involves shortening rather than lengthening. Although Chen (1970), and Kluender and colleagues (1988) suggest that the vowel preceding a voiced stop is lengthened, House and Fairbanks (1953), and Lehiste (1970) report that in English the duration of vowels before voiced stops and sonorants is roughly equal. Roach (2003) claims that we should rather speak of shortening before voiceless consonants than lengthening before voiced consonants. In support of this, Raphael (1971 reported in Raphael et al. 1975) showed that a vowel preceding a voiceless consonant will be from two-thirds to one-half of the duration of that same vowel in an open syllable or preceding a voiced consonant.[15]

Perception experiments with English subjects indicate that the pattern found in production corresponds to listeners' expectations. In one of the first experiments, Denes

[14] In fact, there are even earlier studies on the subject, e.g. Rositzke (1939), Heffner (1941).

[15] English children, however, are reported to exaggerate the increase in vowel duration before voiced obstruents (Krause 1982). Only when they are roughly by the age of three, do "children produce differences in vowel duration before voiced and voiceless final consonants which are of a magnitude similar to that found in adults' production" (Raphael et al. 1980: 340, see also Smith 1978, Weismer et al. 1981, Smit and Bernthal 1983).

(1955) manipulated durations of a preceding synthetic vowel and found that the perception of voicing of the final consonant increases with lengthening of the preceding vowel. Similarly, Malécot (1970) pointed out that the duration of vowels before final consonants is both a powerful and sufficient cue to voiced and voiceless pairs. Raphael (1972) manipulated the duration of synthetic vowels as a cue to the perception of word-final stops and fricatives. The results indicate that the preceding vowel duration is a sufficient and necessary cue to the perception of voicing; however it is less so before fricatives than before stops. Later studies (Parker 1974, O'Kane 1978, Hogan and Rozsypal 1980, Port 1980, Port and Dalby 1982, Luce and Charles-Luce 1985) confirmed the significance of vowel duration as a cue to the voicing of following obstruents. For example, Hillenbrand et al. (1984) used a computer editing technique and observed an increase in voiceless stop responses after removing a portion of the preceding vowel-to-consonant transition.[16]

Unlike English, Polish belongs to a group of languages which are recognised for devoicing word-final obstruents and thus neutralising phonologically the voiced-voiceless distinction (Wierzchowska 1980, Sawicka 1995, Ostaszewska and Tambor 2000). In other words, an underlying voice contrast at the end of words is realised as voiceless during production. However, phonetic studies applying temporal measurements have indicated that the word-final underlying voice distinction is maintained by one or a combination of temporal parameters. It has been attested for Catalan (Dinnsen and Charles-Luce 1984, Charles-Luce and Dinnsen 1987, Charles-Luce 1993), German (Dinnsen and Garcia-Zamor 1971, Port et al. 1981, Fourakis and Iverson 1984, Port and Crawford 1989), Dutch (Warner et al. 2004), or French (Snoeren et al. 2006).[17] Unlike in English, however, incomplete neutralisation in languages that devoice word-finally seems to be strongly influenced by orthographic distinctions (Jassem and Richter 1989, Charles Luce 1993, Kopkalli 1993, Warner et al. 2006). Perception results point to a limited role of temporal cues for distinguishing between underlying voiced and voiceless final obstruents. Port and colleagues (1981) found that German listeners could correctly identify 72% of all the presented items. Port and O'Dell (1985), and Port and Crawford (1989) yielded a similar identification

---

[16] However, in CVNC (where N stands for a nasal consonant) utterances, Raphael and colleagues (1975) observe that nasal duration appears to be a stronger cue than vowel duration itself for the word-final voiced-voiceless consonant distinction.

[17] In contrast, Turkish shows no significant vowel duration differences before word-final stops (Kopkalli 1993 reported in Warner et al. 2004, Wilson 2003).

rate; 60% and 69% respectively. Janker and Piroth (1999) demonstrated that vowel duration is not a significant cue for the obstruents-final voiced-voiceless distinction in German. Although, as mentioned by Slowiaczek and Szymanska (1989), Catalan listeners could also, to a certain extent, differentiate between voiced and voiceless final obstruents, Dinnsen (1985) and Charles-Luce (1985) remark that possible acoustic residuals need not necessarily contribute to the perceptual decision.

The studies of Polish voicing contrast implementation showed some variation in vowel duration despite word-final neutralization. Although in her study of voicing contrast in Polish, Keating (1980: 179) concluded that "[t]he data for Polish clearly show that vowel duration does not vary systematically according to the voicing of the following consonant", Slowiaczek and Dinnsen (1985: 334) found that "vowels are approximately 10% longer before final obstruents that are underlyingly voiced compared with those that are underlyingly voiceless". Tieszen (1997) reported that the variation of vowel duration between underlyingly voiced and voiceless obstruents depends on which of the two major Polish dialects is spoken. For speakers from Warsaw (Northeastern Dialect), in all environments and all places of articulation, the duration of a vowel was significantly longer for underlyingly voiced than voiceless stops. The results obtained from Kraków (Southwestern Dialect) show small, statistically non-significant, durational differences in voicing cues, suggesting that the voicing distinction word-finally is fully neutralized. Jassem and Richter (1989) found no significant differences in vowel duration between underlyingly voiced and voiceless obstruents in Polish when minimizing the role of orthography by prompting subjects with questions to which the target words formed obvious one-word answers. They also concluded that any differences in vowel duration in Polish result from hyperarticulation and that in more naturalistic speech, the durational measures of the vowel portion reveal no contrast. Waniek-Klimczak (2005) showed that Polish speakers of English are not able to control vowel duration as a cue to the voicing contrast of a following stop in English. Before voiced stops, Polish speakers did not produce sufficient durations, whereas, before voiceless stops, they did not sufficiently reduce the vowel length as compared to native speakers.

The perception study by Slowiaczek and Szymanska (1989), which used pairs of words obtained by Slowiaczek and Dinnsen (1985) as stimuli, revealed that Polish "appears to be 'perceptually-neutralizing' the word-final segments […] [Polish subjects] are at the very least not attending consistently to the 55% difference in vowel durations for underlyingly voiced and voiceless obstruents […] and presumably would not consistently attend to the 10% difference in vowel durations obtained in the Slowiaczek and Dinnsen (1985) production study" (p. 211). The fact that Poles are not able to perceive vowel duration differences as a cue to the voicing contrast means that it is one of the perceptual learning tasks that they must accomplish in learning English.

5. Closure duration

Closure duration in stop consonants has been found to strictly correspond to preceding vowel duration as a cue the voicing contrast (see Section 4 for references). Voiced stops are characterised by shorter closure, while voiceless stops by longer closure. It is speculated that the rationale for this regularity lies in the fact that maintaining voicing during stops is difficult (Westbury and Keating 1986), so voiced stop closures are shorter to facilitate maintenance of voicing throughout the closure (Flemming 2002). The amount of the variation of closure duration appears to be dependent on the place of articulation with a velar<alveolar<labial pattern (Umeda 1977, Luce and Charles-Luce 1985), however Crystal and House (1988a) reported an alveolar<labial and velar hierarchy. Although closure duration is strongly correlated with the duration of a preceding vowel (Kohler 1979, Port and Dalby 1982), closure duration has been found to change less with speaking rate than the surrounding vocalic portions (Gay 1978). However, other studies point to the VC ratio which remains fairly constant across changes in a global speaking rate (Barry 1979, Port 1980).

Perception experiments have demonstrated that longer closure durations will yield perceptual judgments of voicelessness and shorter ones of voicing (Lisker 1957, Port 1979, Port and Dalby 1982, Repp and Williams 1985), provided that this durational difference is not overridden by closure voicing (Lisker 1981) or release burst (Raphael 1981).

5.1. Closure duration in English and Polish

As in the case of vowel duration (see Section 4.1.), the production of shorter closure duration in voiced and longer closure duration in voiceless stops in English (House and Fairbanks 1953, Peterson and Lehiste 1960) reveals the same pattern in perception.

Perceptual experiments with native English speakers have shown that closure duration differences are a fairly consistent cue to the voiced-voiceless distinction. Lisker (1957) used synthesised stimuli with bilabial /p, b/ in an intervocalic position and concluded that "[t]he experimental results in sum support the view that closure-durational differences play a major role in the voiced-voiceless stop distinction" (p. 48). His subjects needed about 75 msec of closure to hear a stop as voiced and 130 msec to hear it as voiceless. A later study by Lisker (1978b) demonstrated that reducing the closure of word-medial /p/ in *rapid* led to consistent *rabid* responses. The same pattern was found by Raphael (1981), who noted, however, that "extending the [g] closures to durations appropriate to [k] closures produced little or no perceptual change so long as the quantity of voicing during the closure was proportionately appropriate to that found in productions of [g]" (p. 134-135). Repp and Williams (1985) used natural-speech samples and studied perceptual efficacy of closure duration in stop-cluster sequences. They noted that the shorter the closure, the more likely subjects were to report a voiced stop consonant, which means that "listeners have incorporated tacit knowledge about these temporal regularities into their perceptual criteria for the voiced-voiceless distinction" (Repp and Williams 1985: 455). Their study also showed, however, that the closure duration cue contributes only to the perception of stops as voiced, not as voiceless.

Unlike English, Polish is phonologically a word-final devoicing language which, nevertheless, shows traces of incomplete neutralisation (see Section 4.1.). Slowiaczek and Dinnsen (1985: 332) found that the mean closure durations were statistically longer for underlyingly voiceless obstruents than for underlyingly voiced obstruents. Moreover, a greater difference between contexts (test words followed by a consonant or vowel) was revealed for underlyingly voiced than for underlyingly voiceless obstruents – a difference between consonant contexts and vowel contexts was 3 msec for underlying voiceless obstruents and 14 msec for underlying voiced obstruents. Labial stops were additionally differentiated by voicing into closure of the final consonant. Similar results were obtained by Tieszen (1997) for speakers of Northeastern Dialect of Polish, who revealed longer closure durations for voiceless than for voiced stops and detectable glottal pulsing for voiced compression. No statistically significant effects were found for speakers of Southwestern Dialect. Keating (1980: 177) studied Polish dental stops in the intervocalic position and noted that the mean duration for /t/ was 130.1 ms and for /d/ 91.5 ms. The difference between the /t/ and /d/ closure was statistically significant. Durations less than 90 msec were almost uniformly /d/ and greater than 140 ms were /t/, however there was an overlap between /t/ and /d/ in the ranges between 90 ms to 140 ms.

As noted by Keating (1984: 303), English and Polish differ to the point that the voiced closure durations are shorter in English than in Polish. This difference has been proved to pose a problem for Polish speakers of English (Waniek-Klimczak 1993, 2005). Polish speakers have difficulties with reducing the closure duration of voiced stops in English. While native speakers and early Polish-English bilinguals share a similar durational pattern in the case of voiceless plosives, late bilinguals tend to use longer closure durations in voiced stops.

We have no knowledge of any perception studies that would attempt to ascertain the perceptual efficacy of closure duration in Polish. Slowiaczek and Szymanska (1989) did not isolate the closure duration effect on the perception of underlyingly voiced and voiceless stops. It is not surprising though, considering the fact that even the preceding vowel duration did not appear to be a perceptual cue. Therefore, we might assume with almost certainty that Poles will not read closure durations as a cue to the voicing contrast, which is quite unlike English speakers, who have been demonstrated to effectively detect this temporal parameter.

6. Frication noise in fricatives

One of the main phonetic correlates of the voiced-voiceless distinction in fricatives is frication duration (Cole and Cooper 1975, Stevens et al. 1992, Pirello et al. 1997, Jessen 1998, Kuzla et al. 2007). In English, in word-final position, this parameter appears to be correlated with the preceding vowel duration. For the *base – bays* opposition, the lengthening of the vowel by the voicing of the following consonant is 120 ms in utterance-final position and 30 ms in the non-final but pre-boundary position. The shortening of the fricative constriction by voicing is 80 ms in utterance-final position and 35 msec in non-final position (Pickett 1999). Steven et al. (1992) found a 30 ms frication duration difference between voiced and voiceless fricatives in English, with preceding vowels being longer before shorter voiced than longer voiceless fricatives. In utterance-final position, vowel duration differences increase to 41 ms. Crystal and House (1988b) report a mean difference in frication duration of 39 ms. Stevens et al.

(1992) have suggested that the frication duration differences between English voiceless and voiced fricatives are mechanical by-products of the voicing distinctions between them. Voiced fricatives have shorter frication intervals because they are produced with a smaller glottal abduction gesture, which fulfils the aerodynamic requirements for turbulence noise generation for a relatively short interval in comparison to the large abduction gesture that accompanies voiceless fricatives.

Perceptual studies have confirmed frication duration as a cue to the voiced-voiceless opposition in fricatives. Cole and Cooper (1975) showed that shortening the duration of frication of fricatives word-initially produced a change in the percept from voiceless to voiced. Steven et al. (1992) reported that English listeners base their voicing judgments of intervocalic fricatives on an assessment of time interval in the fricative during which there is no glottal vibration. This time interval must exceed about 60 msec if the fricative is to be judged as voiceless. Similarly, Flege and Hillenbrand (1986) found that, in identifying fricatives as /s/ or /z/ in American English, native English listeners used fricative duration as a cue to the syllable-final voicing contrast.

Another established cue to the voiced-voiceless distinction in fricatives is vocal cord vibration. Voiced fricatives are signalled by a lower relative frication intensity (Strevens 1960, Balise and Diehl 1994) due to "some of the energy […] being used up by the larynx tone generator" (Fry 2001: 122). English fricatives tend to be partially devoiced both word-initially and word-finally (Ladefoged 1971, 2000, Stevens et al. 1992, Gimson 1994). This tendency towards partial devoicing can be explained on articulatory and aerodynamic grounds (discussion in Pickett 1999, Kuzla et al. 2007). For vocal fold vibration to be initiated word-initially, a critical pressure difference must be created between subglottal and supraglottal air pressure (Westbury and Keating 1986, Baer 1975). Due to the oral impedance in obstruents, oral pressure increases over time and vocal fold vibration ceases unless compensatory articulatory strategies are used to maintain the transglottal pressure. Therefore, in voiced fricatives, transglottal airflow and airflow through the oral constriction must be balanced to produce vocal fold vibration and frication noise at the same time (Stevens et al. 1992, Stevens 1998). Failure to do so leads to partial devoicing. This difficulty is confirmed by evidence that the vocal folds are held more open during voiced fricatives than during voiced stops (Sawashima 1968, Sawashima and Miyazaki 1973). Similarly, Hirose and Gay (1972) found that there is less activity in the abductor muscles in the fricatives than in the stops, indicating that the folds are positioned farther apart for fricatives than for stops to allow more airflow and thus produce frication sound.

In Polish, word-initially, fricatives are reported to be fully voiced or voiceless (Wierzchowska 1967, 1980). On the other hand, word-final fricatives are only voiceless due to the word-final devoicing rule operative in Polish (see Section 4.1.). Although Slowiaczek and Dinnsen (1985) found traces of non-complete neutralisation in that there were some remnants of voicing into friction, this pattern was not demonstrated for all subjects. Nowocień (2000 reported in Gonet 2001) found 20% prevoicing preceding the formation of the constriction for voiced fricatives word-initially. Kopczyński (1977) studied the perception of partially devoiced English obstruents word-initially by speakers of Polish and found 22% misidentification of partially devoiced fricatives as voiceless by Poles. As for the frication duration in Polish fricatives as a perceptual cue to the voicing contrast, we are not acquainted with any studies that would attempt to find in Polish a pattern similar to English, where shorter fricatives produce voiced percepts.

7. Release burst in affricates

In comparison to stop consonants, affricates are severely underesearched in the literature. Phonetic literature suggests that the voicing distinction in affricates is generally cued in the same fashion as for plain stops in a given language. Accordingly, Polish affricates are considered either fully voiced or voiceless word-initially, and actively devoiced word-finally (Wierzchowska 1967, 1980). Indeed, Nowocień (2000) found a strong pre-affricate prevoicing in Polish which constitutes 70% of the duration of the whole voiced segment. Similarly, English affricates are described in roughly the same aspirated vs. unaspirated terms as plain stops (Jones 1956, Gimson 1994). The duration of the stop segment in affricates tends to be about the same as for a simple stop in similar positions and the voicing contrast is cued by longer closure for voiceless and shorter for voiced affricates with greater frication noise intensity for voiceless than for voiced sounds (Fry 2001). The comparison with a plain stop voicing contrast, however, does not appear to be well grounded. Jessen (1998) finds that the release stage of the English voiced affricate /dʒ/ is markedly longer than that of the corresponding plain stops and the aspiration of /tʃ/ can be partially or fully overlapped by its release stage.

It is concluded therefore that duration and the quality of the fricative release in signalling voiced-voiceless affricates is a more powerful cue than VOT continuum observed in stops.

8. Summary

This part has provided a discussion of temporal and acoustic parameters of the voicing contrast implementation in English and Polish obstruents. Although the voicing contrast has been believed to be manifested phonetically as the presence or absence of vocal fold vibration, this view turns out to be essentially oversimplified. Similarly, the fortis-lenis opposition, applied by phonological models, is not supported by acoustic and articulatory observations of speech production. On the contrary, phonetic descriptions point to different temporal and acoustic implementation techniques of the voicing contrast which differ not only across manners of articulation but also across languages.

The Voice Onset Time has been found to be a powerful cue to the voicing or voicelessness of initial stops both in production and perception. Cross-linguistic differences are demonstrated by the comparison of English and Polish. English contrasts short lag vs. long lag VOT values, whereas Polish, as all Slavic languages, locates its VOT boundary between voicing lead and short lag values. Production measurements have been fairly well correlated with perception results in discrimination and identification experiments, with the voicing contrast boundaries oscillating around +35 positive VOT and 0ms VOT for English and Polish respectively.

Preceding vowel duration is an example of the expansive nature of the voicing contrast where the implementation of voicing or voicelessness in obstruents word-finally influences temporal organisation of a preceding vowel. English is reported to be heavily dependent on this phenomenon – it significantly reduces the length of the vowel followed by a voiceless stop. This articulatory tendency finds support in perception experiments where a single parameter of vowel duration can cue the voicing contrast of the following obstruents. Polish, on the other hand, makes a different case. It is one of the word-final neutralising languages and, although it has been found not to completely neutralise the voicing contrast word-finally and to vary, to a certain extent, preceding vowel duration, perceptual studies have demonstrated that any vowel duration differences are perceptually suppressed and do not serve as a cue to the voiced-voiceless distinction.

Closure duration has been found to be well correlated with the preceding vowel duration in that it is shorter in voiced than in voiceless stops. This compensatory shortening and lengthening has been well documented in English and was found to be a sufficient cue to the perception of the voicing contrast. In Polish, the closure duration has been observed to vary to a limited extent and thus indicate the non-complete status of a word-final neutralising process. However, its potential to cue the voicing contrast has not been confirmed in perception studies.

The implementation of voicing distinction in English fricatives is realised both temporally and acoustically. Shorter frication noise in fricatives obtains voiced percepts for speakers of English. Also, English fricatives are not stable in spectral implementation of voicing word-initially and word-finally, where they are partially devoiced. Polish voiced fricatives are considered to be fully voiced or voiceless word-initially, and voiceless word-finally, even though word-final neutralising rule may not be complete for all speakers and leave some voicing into friction traces.

Finally, affricates seem to rely on the release burst as a cue to the voicing distinction. Although they are claimed to follow the implementation pattern of plain stops, this does not appear to be a fully precise description. VOT boundaries in affricates do not match those obtained for plain stops, probably due to a superimposed frication noise. Voicing implementation in English and Polish affricates seems to be in agreement with a pattern obtained for fricatives in those languages in that English voiced affricates are partially devoiced word-initially and word-finally, and Polish voiced affricates are either fully voiced or voiceless word-initially and neutralised word-finally.

Part Three

Perception of the voicing contrast in English and Polish obstruents: An experimental study

1. Study design

1.1. Objectives and hypotheses

In the first part of an experimental study, we attempt to answer the question of how Polish learners of English perceive English temporal and spectral parameters of the voicing contrast and how this perception develops with language experience. In the second part, we attempt to see whether experience with L2 English can influence the perception of the voicing contrast parameters typical for L1 Polish. We apply digital manipulation techniques to obtain precisely-controlled parameters on natural speech samples. They enable a researcher to extract a single cue and play down the others in order to single out a desired feature and test it in a cross-linguistic paradigm.

We compare three groups of subjects:

1. Native Speakers of English who serve as a control group for setting optimal L1 English perception
2. Polish Early Beginner Learners of English who provide information about an initial stage in L2 perception development
3. Polish Advanced Learners of English who are expected to allow us to locate the characteristics of a highly developed stage in L2 perception

The parameters controlled for the L1 Polish – L2 English influence are as follows:

1. VOT continuum in initial stops
2. Frication noise in initial fricatives
3. Frication duration in initial fricatives
4. Release burst in initial affricates
5. Vowel duration as a cue to the voicing contrast in final stops
6. Closure duration in final stops
7. Frication noise in final fricatives
8. Closure duration in final affricates

For the L2 English – L1 Polish influence, the following parameters were chosen:

1. VOT continuum in initial stops
2. Frication duration in initial fricatives
3. Release burst in initial affricates

The discussion of temporal and spectral parameters which contribute differently to the implementation of the voicing contrast in English and Polish (see Part 2) allows us to put forward the following hypotheses for testing:

1. Polish learners will not match native speakers in categorising positive VOT values between short lag for voiced and long lag for voiceless stops
2. Polish learners will not match native speakers in recognising partially devoiced initial fricatives as voiced

3. Polish learners will not match native speakers in reading reduced frication noise in initial fricatives as a cue to the voiced category
4. Polish learners will not match native speakers in categorising the reduced release burst in initial affricates as a cue to the voiced category
5. Polish learners of English will not match native speakers in reading increased vowel duration as a cue to the voiced category of a following stop
6. Polish learners will not match native speakers in recognising decreasing closure duration as a cue to the voiced category of a final stop
7. Polish learners will not match native speakers in recognising partially devoiced final fricatives as voiced
8. Polish learners of English will not match native speakers in reading reduced closure duration in final affricates as a cue to the voiced category

Testing the performance of Beginner Learners and Advanced Learners on manipulated Polish stimuli is expected to provide an answer on whether fluency in L2 influences perception mechanisms in L1. If this is the case, the following hypotheses should be validated:

9. 0 ms VOT value will be perceived as voiced by Advanced Learners and as voiceless by Beginner Learners
10. When confronted with a hybrid initial stop comprising both voicing lead and voicing lag, Beginner Learners will attend more readily to the voicing lead, whereas Advanced Learners will attend to the voicing lag
11. Advanced Learners will be more sensitive to the shortening of frication noise in Polish fricatives as a cue to the voiced category
12. Advanced Learners will be more sensitive to the shortening of the release burst as a cue the voiced category in Polish initial affricates

In the following sections, we discuss in detail specification of the stimuli, recording and manipulation techniques, and group characteristics.

1.2. Stimuli

1.2.1. Natural speech samples

In the experiments, we chose to use natural speech samples due to the most recent trends in speech perception studies which are beginning to favour real speech to synthesised speech. Although a large body of perception experiments have used speech synthesisers to "isolate a given cue and see whether it influences listeners' decisions" (Fry 2001: 131), most recently one can observe a recurring interest in natural speech stimuli. Abramson (2000: 9-10) writes:

> Testing is then done by making incremental changes […] and playing the resulting stimuli to native speakers of the language for labelling or discrimination. In the early decades of such experimentation, it was difficult or impossible to create such stimuli with natural speech while not allowing anything to change but the parameter of interest, so terminal analog synthesisers were used under the control of carefully drawn schematic spectrograms or, later, programme instructions to the parameters of the synthesisers […] Nowadays, instead of using pure synthesis, it is possible to make carefully controlled spectral and temporal changes in natural speech to make stimuli for experiments in speech perception.

What is more, Hawkins (2003) discusses experiments using synthetic speech and concludes that they could not get the best performance out of their participants due to an artificial nature of presented stimuli. Similarly, Duffy and Pisoni (1992) observe that real speech is more memorable than laboratory-standard synthetic speech because it is more perceptually coherent.

1.2.2. Nonsense syllables

The rationale for our choice of nonsense syllables lies in a well-documented fact that any semantic influence will distort speech perception results on a level of segments, let alone single extracted parameters. A precise control for lexical influences in the study like ours would be extremely difficult, if possible at all. For example, common words can be both produced (Dell 1990, Jescheniak and Levelt 1994, Caramazza et al. 2001, Dell and Gordon 2003) and recognised (Soloman and Postman 1952, Oldfield and Wingfield 1965, Luce and Pisoni 1998) with greater facility than rare words. Other lexical variables also affect recognition similarly. Concrete words are favoured over abstract words in both production (Martin et al. 1996) and recognition (Strain et al. 1995). Also, predictable words can be produced (Griffin and Bock 1998) and recognised (Morton and Long 1976) more quickly than those that are less congruent with their context.

As early as in 1963, Denes and Pinson (1963: 146) noted that "[a]s a supplement to ambiguous acoustic cues, linguistic information serves as a powerful aid in speech recognition". In fact, both word knowledge and the knowledge of the frequencies with which phones follow one another can affect how phones are identified. One of the most recognised influences is the Ganong effect. Ganong (1980) showed that lexical knowledge can affect how a phoneme is identified by creating pairs of continua in which the phoneme sequence at one end was a word but the sequence at the other end was a nonword. In one pair of continua, VOT was varied to produce a *gift*-*kift* continuum and *giss*-*kiss* continuum. He found that listeners provided more /g/ responses in the *gift*-*kift* continuum than in the *giss*-*kiss* continuum. They tended to give responses suggesting that they identified the words they knew preferentially. This leads to the conclusion that lexical information feeds down and affects perceptual processing of phonemes and that when the processor yields an ambiguous output, lexical knowledge is brought to resolve the ambiguity (Fowler 2003).

A second finding of lexical effects is phonemic restoration (Warren 1970, Samuel 1981, 1996). When a phoneme is excised from a word, e.g. /s/ from *legislature*, and is replaced with noise, listeners report hearing the missing phoneme. Moreover, listeners asked to make a judgement whether the phoneme is present or absent in the noise show lower perceptual sensitivity to the phonemes in words than in nonwords.

A final lexical effect occurs in experiments on compensation for coarticulation. Elman and McClelland (1988), using the compensation for coarticulation paradigm, demonstrated lexical influence on perceptual processing of consonants. They generated continua ranging from /t/ to /k/ (*tapes* to *capes*). These sounds were placed after the words *Christmas* and *Spanish*. The results showed that there were more judgments of *capes* following *Christmas* than following *Spanish*. Accordingly, the only thing that made the final fricative of *Christmas* an /s/ was the listeners' knowledge that *Christmas* is a word and *Christmash* is not. Also, lexical knowledge was responsible for making the final fricative of Spanish an /♠/.

More recent studies (Samuel 2000, Dahan and Gaskell 2007, Ventura, Kolinski et al. 2007, Ventura, Morais et al. 2007) have corroborated lexical effects on phoneme perception and the influence of lexical knowledge on phoneme identification. Sebastián-Gallés and Kroll (2003: 292) provide methodological guidelines by saying that "[m]ost of the [perception] studies […] have examined performance for simple syllables or nonsense strings of phonemes. This seems to be a reasonable approach, since the interest is centred on how phonemes are processed and lexical or other higher-level influences are in principle undesired".

The choice for nonsense syllables is even more justified in the case of the present study. The comparison of Early Beginners with Advanced Learners cannot eschew a strong lexical bias on the part of a former group. Limited lexical knowledge of Early Beginners would definitely strengthen the semantic influence of a handful of words they might know. The continuum between syllables such as, e.g. *den* - *ten* would definitely yield significantly more *ten* judgments irrespective of actual phonetic processing since a numeral *ten* is learnt much earlier than the lexeme *den*. Even for Advanced Learners and Native Speakers the *den* – *ten* continuum would be distorted due to decidedly higher frequency of *ten* over *den*.

Finally, neurophysiological evidence shows that the production and perception of nonsense syllables activates secondary auditory cortex and other brain regions responsible for language (Paus et al. 1996).

All the stimuli used in this study are composed of a CVC sequence due to the fact that sounds uttered in a dynamic consonant-vowel-consonant context are perceived more accurately than sounds produced in isolation (Strange et al. 1976, Shankweiler et al. 1977, Bailey and Summerfield 1980).

1.2.3. Recording procedures

All recordings of the samples for manipulation were conducted up to the highest standards set by Lieberman and Blumstein's (2002) tape recording techniques. Speech signals were recorded without distortions with the signal-to-noise ratio over 20dB. English samples were read and recorded by an educated male speaker of American

English. Polish samples were read and recorded by an educated female speaker of Polish. Neither of the speakers had any reported history of a speech disorder or any detectable articulation impediments. Both speakers were instructed to read each syllable with a flat intonation.

A Media Tech MT385 USB microphone with a flat response between 100 and 16000 Hz was positioned 20 centimetres from a speaker's mouth. The speech input was processed and recorded by an external Sound Blaster X-Fi X-MOD sound card with a 24 bit sampling rate, frequency range 140 – 20000 Hz and sensitivity 112 dB +- 3 dB. The recording was sampled at 22.05 kHz (16 bit resolution) (e.g. Wright 2001, Cho and McQueen 2006, but see also Kuzla et al. 2007 for 16 kHz, and Clarke and Luce 2005, Mani and Plunkett 2007 for 44.1 kHz). All samples were subsequently stored in a notebook hard drive memory as WAV files ready for manipulation.

1.2.4. Measurement criteria

Prior to the manipulation, all individual parameters were measured using a Praat 4.6.18 speech-analysis software package (Boersma 2001, Boersma and Weenink 2007) by means of a spectrographic display and waveforms. VOT was measured as a temporal span between the release burst and the beginning of regular vertical striations corresponding to the quasi-periodic voice pulses, i.e. from the first peak of the stop release burst up to the zero crossing nearest to the onset of the second formant of the following vowel (e.g. Abramson 1977, Lisker 1978, Keating 1980, Keating et al. 1981, Cho et al. 2002, Cole et al. 2007). Vowel duration was measured from the onset of periodicity showing clear formant structure to the end of periodicity signalled by a drop in amplitude (e.g. Peterson and Lehiste 1960, Raphael et al. 1980, Slowiaczek and Dinnsen 1985, Fowler 1992, Waniek-Klimczak 2005)[1]. Closure duration was delimited as the interval from the offset of a vowel to the release burst, typified by a sudden increase in amplitude in the waveform (e.g. Lisker 1957, Keating 1980, Slowiaczek and Dinnsen 1985, Cole et al. 2007). The duration of frication noise was sought in the interval from the beginning to the cessation of aperiodic noise (Cole and Cooper 1975).

[1] Our durational measurements of vowel duration included formant transitions. As demonstrated by Raphael and colleagues (1980), the effective duration of a vowel extends over all parts of the acoustic signal, including especially the transitions that reflect the consequences of the coarticulation of vowel and consonant. For earlier alternative proposals to measure only steady-state formant duration, see Sholes (1959), Denes (1955), Raphael (1972), Raphael et al. (1975).

The voiced portion of fricatives was identified as the interval representing periodic information in the waveform (voice bar) concurrent with the noise in the waveform (Slowiaczek and Dinnsen 1985). The duration of release burst of affricates was defined as the time-span between the rise in amplitude at the release point to the onset of the second formant of the following vowel (Jessen 1998).

1.2.5. Manipulation technique: PSOLA

For temporal manipulation of stimuli, we applied a PSOLA technique, available in Praat 4.6.18 speech-analysis toolkit. PSOLA (the time-domain pitch-synchronous overlap and add) works pitch-synchronously in that each frame is centred around a pitch-mark in the speech, rather than at regular intervals as in normal speech signal processing (Jurafsky and Martin 2000). The concatenated waveform is split into a number of frames, each centred around a pitchmark and extending a time period on either side. Speech is made longer by duplicating frames and shorter by leaving frames out. PSOLA makes possible to compress or expand the time base with very few changes in pitch and spectral information (Moulines and Charpentier 1990, Moulines and Verhelst 1995, Quené 2007). The technique is so effective because it separates each frame first and then decreases the distance between the frames. The internals of each frame are not changed, therefore the frequency of the components is hardly altered and the resultant speech sounds the same as the original, except for different durational values.

Figures 1.1. and 1.2. demonstrate the lengthening manipulation of vowel duration in the syllable *theep* /θi:p/. Figure 2.1. shows an originally obtained syllable with the vowel length of 152 ms.

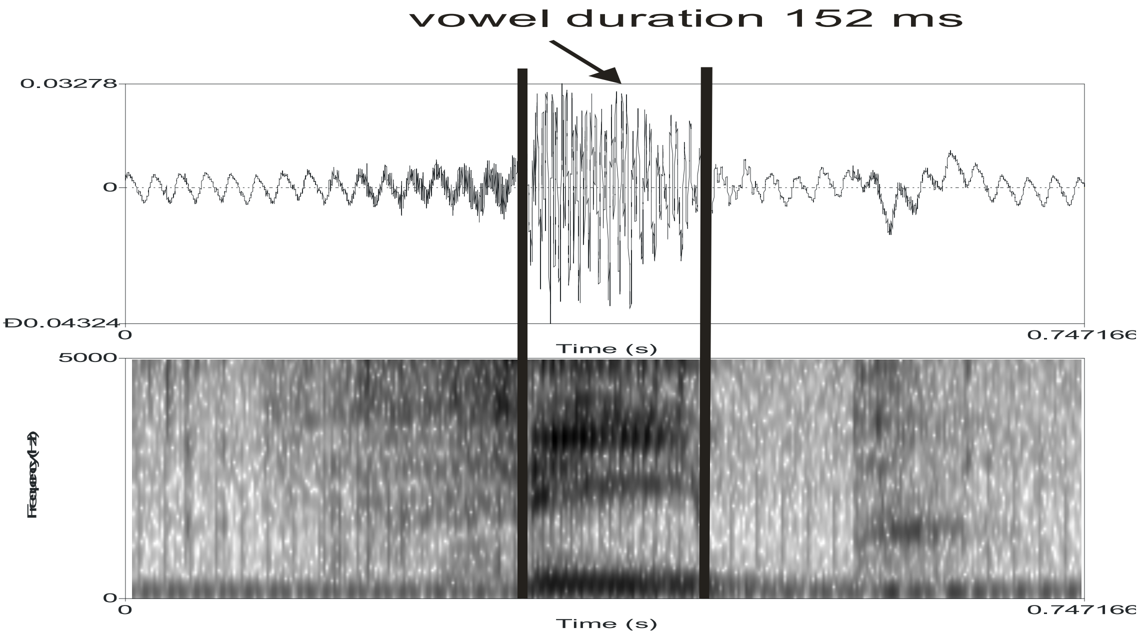


Figure 1.1. Original syllable *theep* /θi:p/ with 152 ms vowel duration

Next, we script the PSOLA device to manipulate the durational tier and to increase the vowel duration to 182 ms, without modifying the durational parameters of remaining segments.

```
Create DurationTier... lengthen 0 3.674
Add point... 0.0 2188/2188
Add point... 2.188 2188/2188
Add point... 2.189 182/152
Add point... 2.341 182/152
Add point... 2.342 1332/1332
Add point... 3.674 1332/1332
```

Line 2 and 3 set the cursors at the beginning of a signal and the ending of the fricative /θ/ and specify that this time span should not be modified. Line 4 and 5 delimit the vowel and change proportion of its duration from 152 ms to 182 ms. Line 6 and 7 input the information that the closure duration and release burst of /p/ must not be manipulated. The obtained stimulus, with 182 ms vowel duration, is displayed in Figure 1.2. No other temporal parameters have been altered.

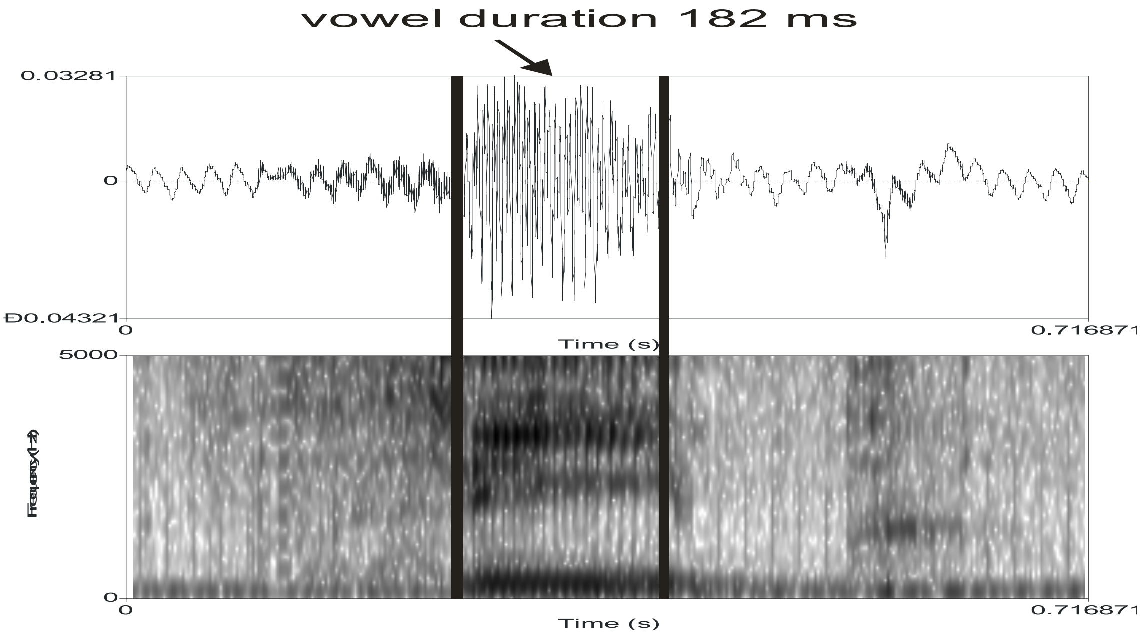


Figure 1.2. Modified syllable *theep* /θi:p/ with 182 ms vowel duration

For the manipulation of voicing in fricatives, we used standard editing technique for cutting and pasting portions of speech signal available in Praat 4.6.18.

1.3. Subjects

1.3.1. Native Speakers

We invited 11 Native Speakers (referred to hereafter as NS) to participate in the study, 6 speakers of American English and 5 speakers of British English. They ranged in age from 23 to 56 years (Mean: 32, Std. Dev.: 9.05). All subjects volunteered and were not paid for their participation. They were all naïve to the object of the study. A preliminary interview revealed that they were all monolinguals and did not speak fluently any second language. None of the subjects had any reported history of a speech disorder or hearing loss. Neither did they report any current hearing disorders.

1.3.2. Advanced Learners of English

A total of 24 Polish Advanced Learners of English (referred to hereafter as AL) participated in the study. They were all 3rd-year students of English Philology, University of Silesia. Their skills had been repeatedly confirmed by annual practical examinations. Additionally, we had had a brief interview in English with each subject to confirm their proficiency. They ranged in age from 21 to 25 years (Mean: 22, Std. Dev.: 1.2). They had all had long experience with learning English (Mean: 12.4 years, Min: 8 years, Max: 16 years, Std. Dev: 2.48 years). All subjects volunteered and were not paid for their participation. They were all naïve to the object of the study. None of the subjects had any reported history of a speech disorder or a hearing loss. Neither did they report any current hearing disorders.

1.3.3. Beginner Learners of English

The group comprised 26 Polish Beginner Learners of English (referred to hereafter as BL). They were all students in beginner groups in a language school and had six-month experience with learning English. A preliminary interview revealed that they did not speak any other foreign languages fluently. All subjects were participants of an English course sponsored by the European Social Fund. This programme provided a new opportunity for people who had never learnt English. They ranged in age from 26 to 47 years (Mean: 39, Std. Dev.: 6.51). All subjects volunteered and were not paid for their participation. They were all naïve to the object of the study. None of the subjects had any reported history of speech disorder or a hearing loss. Neither did they report any current hearing disorders.

1.4. Experimental procedures

The experiments took place in a quiet room. The stimuli were presented via high-quality powered loudspeakers at a comfortable level. Special care was taken to provide the same acoustics for all subjects. Each stimulus was presented twice and each presentation was followed by a two-second pause.

Prior to the presentation, the subjects were instructed in a target language about the methodology of the study. Even BLs were instructed in English, in simplified language with a following brief summary in Polish. In order to activate a desired language mode, the presentation of English stimuli was preceded by a short conversation in English. Analogically, prior to the presentation of Polish stimuli, the conversation was in Polish.

In order to provide the highest presentation standards, the AL and BL groups were further divided into subgroups of 6-8 people. The were presented with randomised 38 English stimuli and 4 Polish stimuli. The NSs did not participate in the Polish part of the study. Each session took approximately 20 minutes.

The subjects were asked to circle the sound they heard in each syllable in a forced-choice identification format (e.g., Liberman et al. 1980, Keating et al. 1981, Fowler 1992), i.e. the subjects were given two alternatives, voiced and voiceless, in each syllable. The NSs were given English orthographic approximation of elicited sounds (Appendix 1) and the BLs were given Polish representations thereof (Appendix 2). The ALs were provided with transcribed options because they could read and write transcription proficiently (Appendix 3).

Prior to the experiment, the subjects were encouraged to ask questions to elucidate any uncertainties. Before the experiment proper, they were presented with 5 trials in a training session. They were strongly urged to make an identification in each syllable heard, even though in some cases the judgment might represent no more than a guess (Liberman et al. 1952).

1.5. Statistical devices

For testing the significance of the in-group effect, we used a Cochran Q test, which is an extension of McNemar's *Chi*-square test for changes in frequencies or proportion of more than two dependent samples. It is a non-parametric test which measures nominal variables. Specifically, it tests whether several matched frequencies or proportions differ significantly among themselves. When Cochran Q test was not applicable because there were only two variables, we used McNemar's *Chi*-square for comparing two nominal proportions. For the between-group effect, we used a *Chi*-square test, which evaluates the relationship between two dichotomous nominal variables (details in Scholfield 1991, Oakes 1998, Howell 1999).

2. Results

2.1. English stimuli

2.1.1. VOT

2.1.1.1. Stimuli

From a recorded syllable *keef* /ki:f/ (+70 ms VOT in initial /k/), we created 8 stimuli with partitioned VOT continuum. The syllable in which a velar stop is followed by a high vowel was motivated by the fact that a velar followed by a high vowel obtains the longest VOT continuum (Cho and Ladefoged 1999, Chang et al. 2001). We modified the syllable to obtain 10ms-step stimuli across the VOT continuum (for 10 ms steps see Abramson and Lisker 1967, Lisker 1978, Keating et al. 1981, Clarke and Luce 2005). /k/ with 0 ms VOT was obtained by removing an /s/ segment from syllable *skeef* /ski:f/ (Lotz et al. 1960, Reeds and Wang 1961, Davidsen-Nielsen 1969, Imsri 2002, Lisker 2002). As a result, we obtained the following stimuli:

1. *keef* /ki:f/, /k/ +70 ms VOT
2. *keef* /ki:f/, /k/ +60 ms VOT
3. *keef* /ki:f/, /k/ +50 ms VOT
4. *keef* /ki:f/, /k/ +40 ms VOT
5. *keef* /ki:f/, /k/ +30 ms VOT
6. *keef* /ki:f/, /k/ +20 ms VOT
7. *keef* /ki:f/, /k/ +10 ms VOT
8. *keef* /ki:f/, /k/ 0 ms VOT

Figures 2.1. to 2.7. show waveforms and spectrograms of all the stimuli.

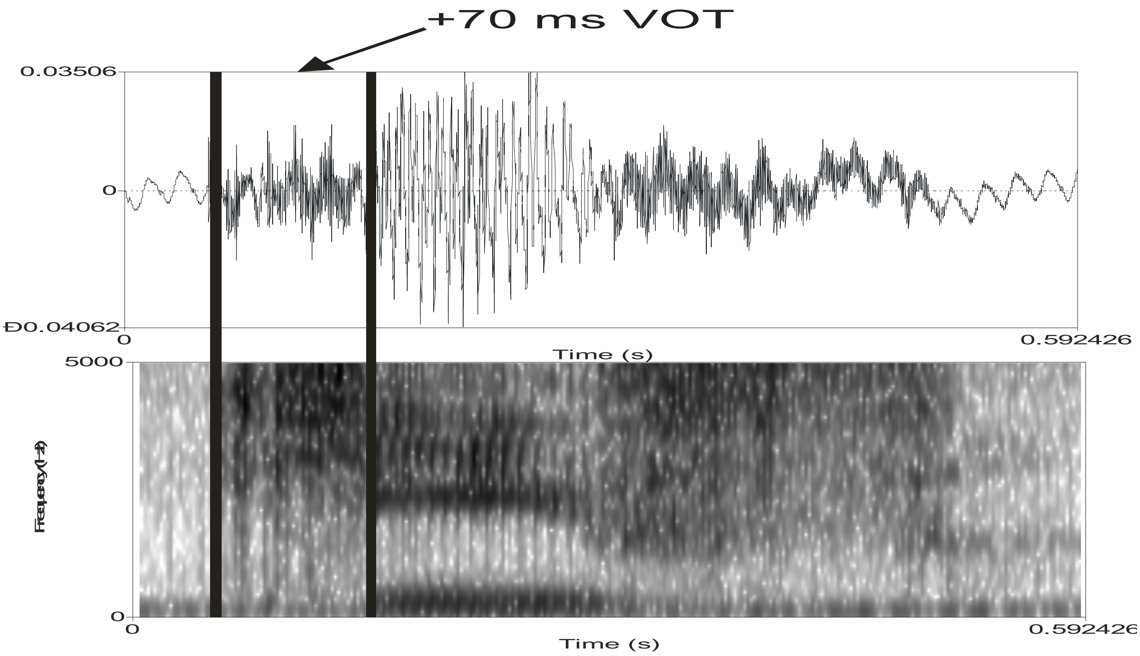


Figure 2.1. Waveform and spectrogram of syllable *keef*, /k/ +70 ms VOT

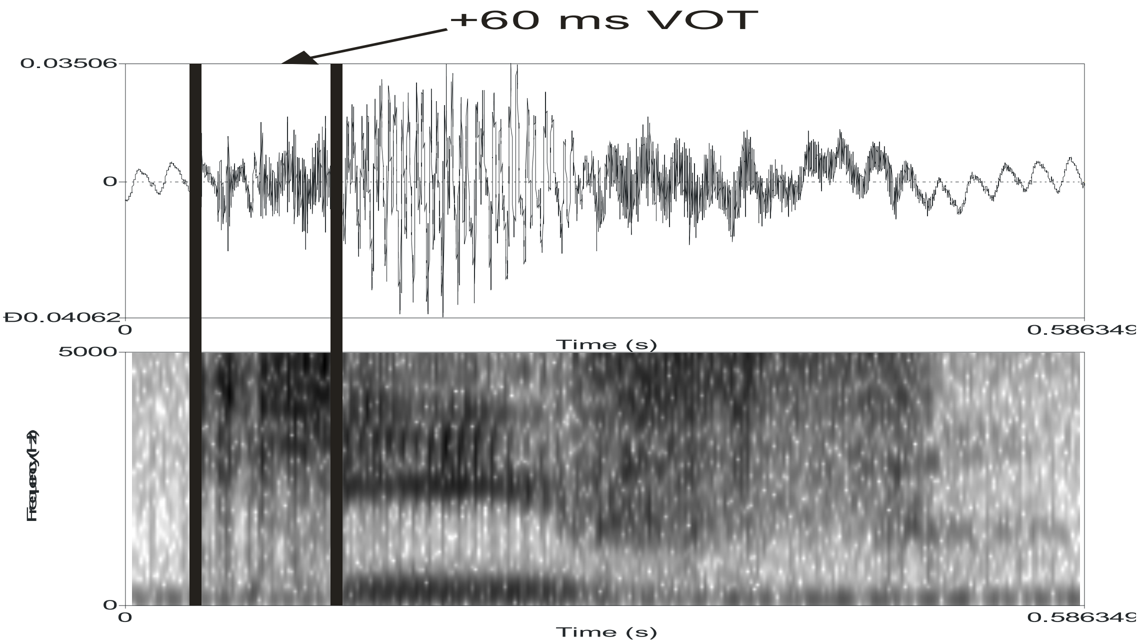


Figure 2.2. Waveform and spectrogram of syllable *keef*, /k/ +60 ms VOT

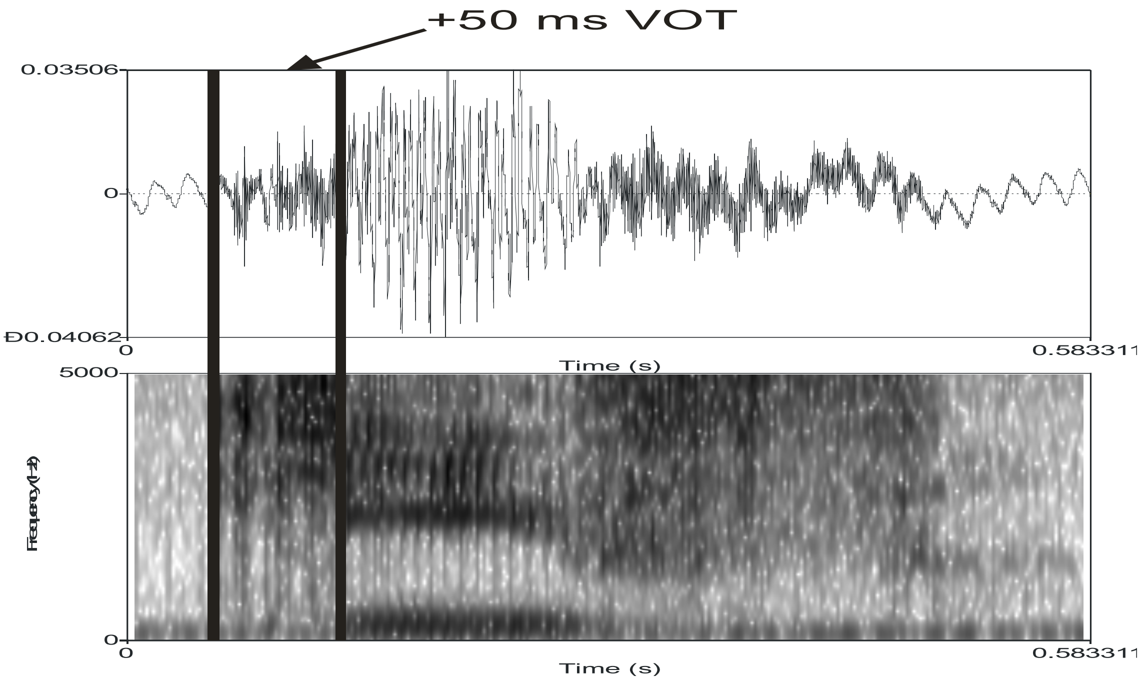


Figure 2.3. Waveform and spectrogram of syllable *keef*, /k/ +50 ms VOT

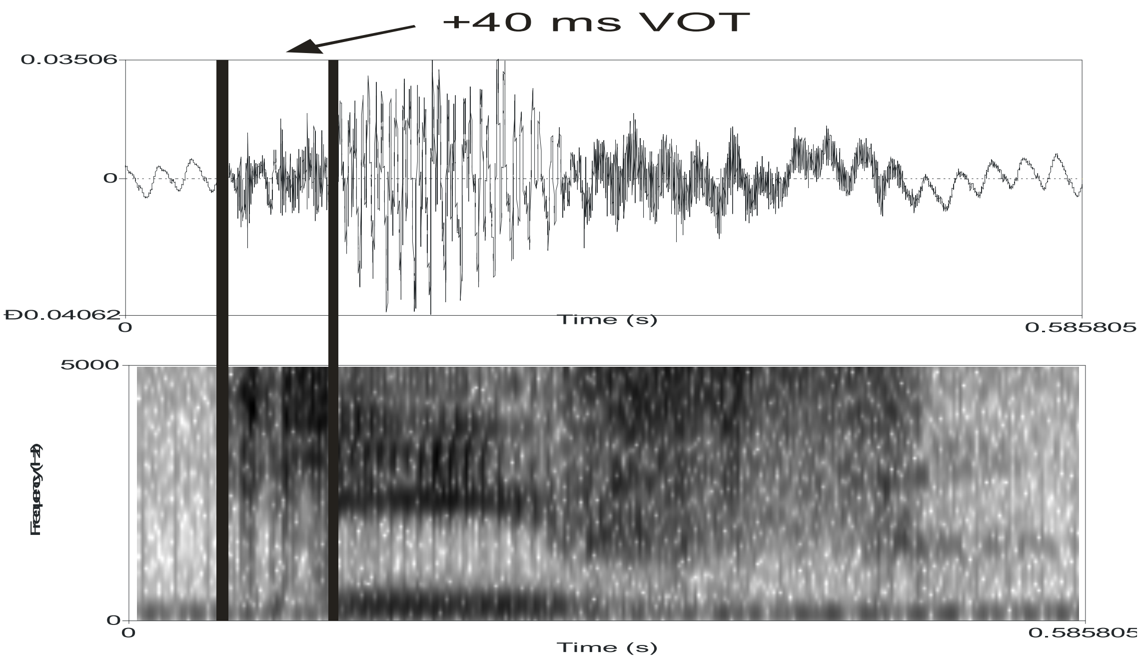


Figure 2.4. Waveform and spectrogram of syllable *keef*, /k/ +40 ms VOT

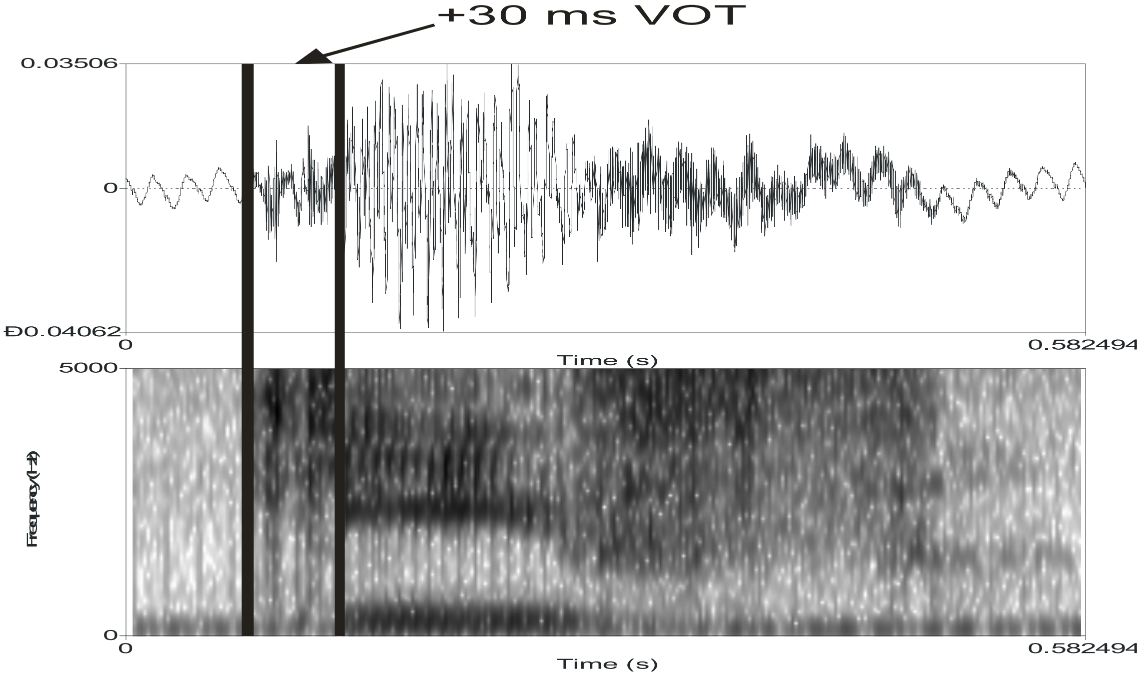


Figure 2.5. Waveform and spectrogram of syllable *keef*, /k/ +30 ms VOT

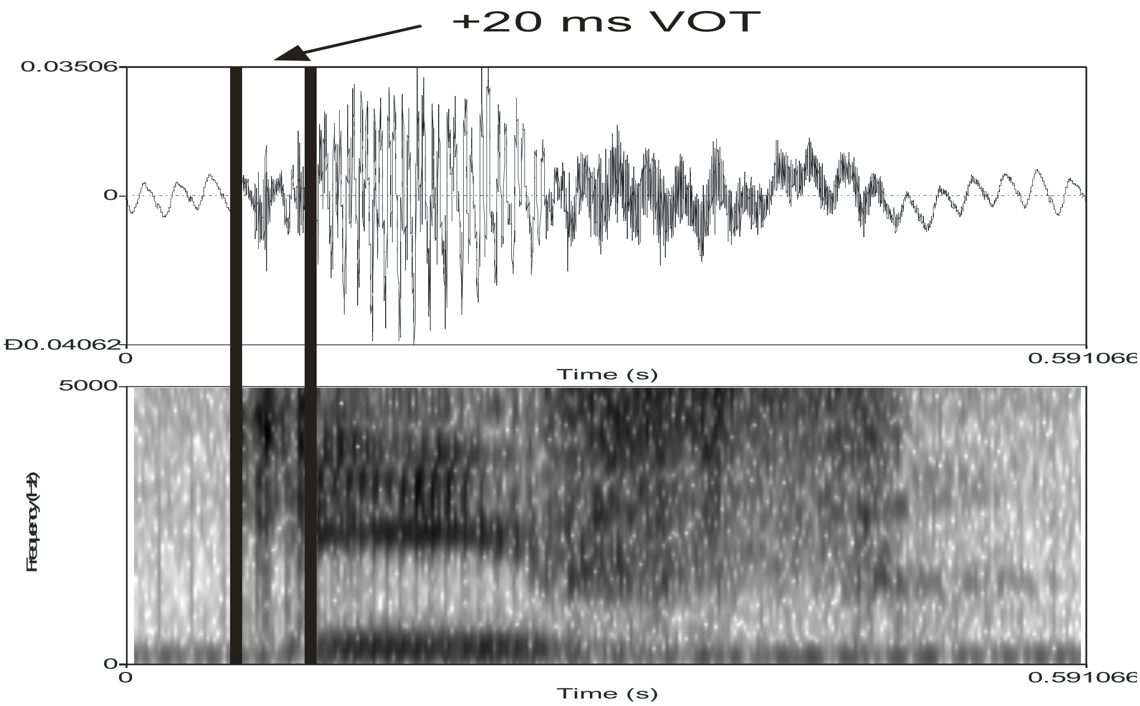


Figure 2.6. Waveform and spectrogram of syllable *keef*, /k/ +20 ms VOT

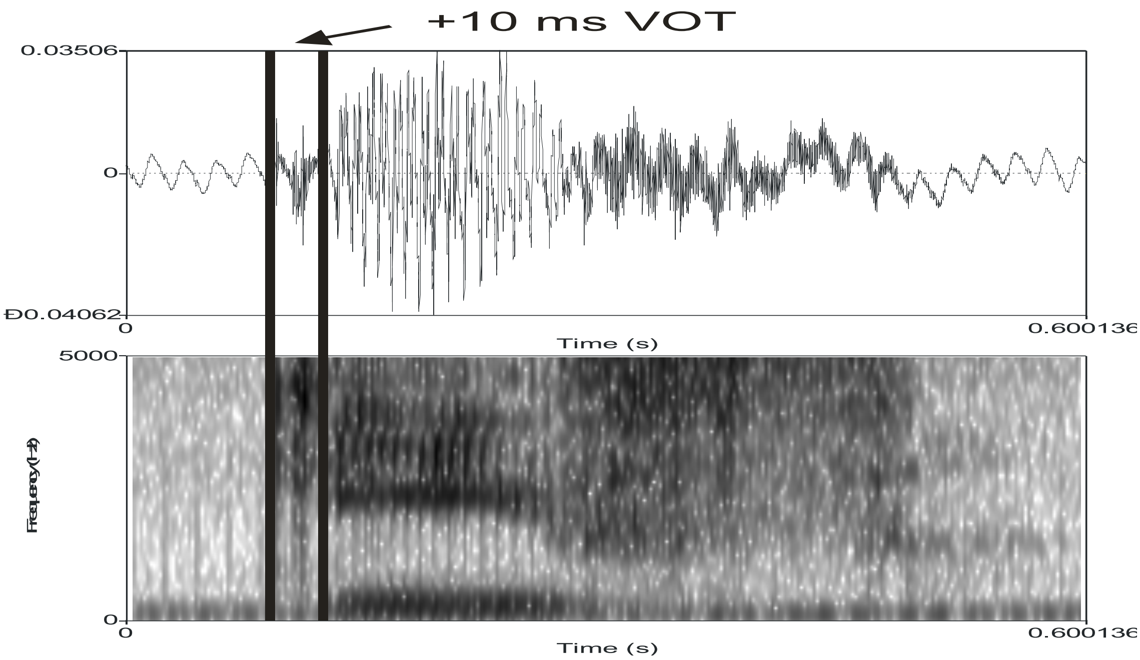


Figure 2.7. Waveform and spectrogram of syllable *keef*, /k/ +10 ms VOT

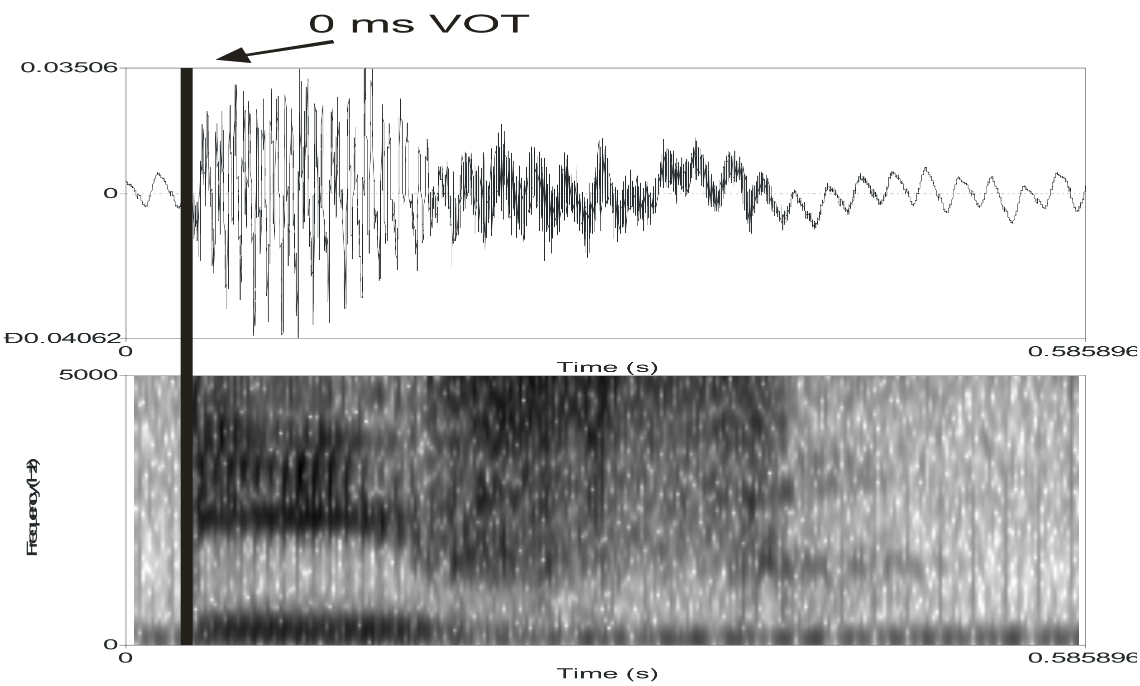


Figure 2.8. Waveform and spectrogram of syllable *keef*, /k/ 0 ms VOT

2.1.1.2. Beginner Learners – results

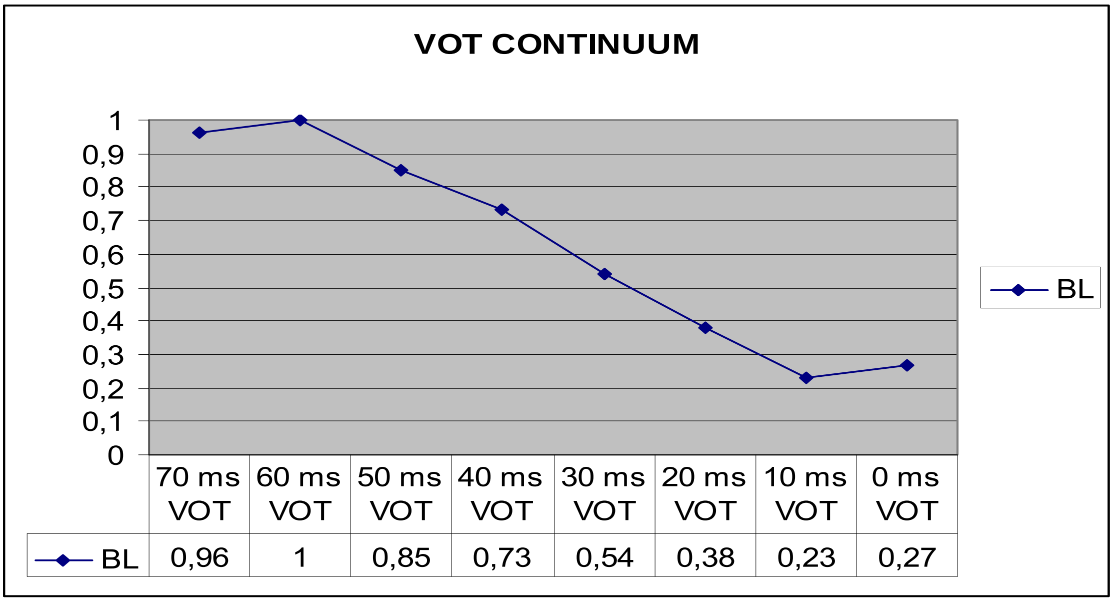


Figure 2.9. Recognition of an initial sound in *keef* as /k/ across the VOT continuum by Beginner Learners

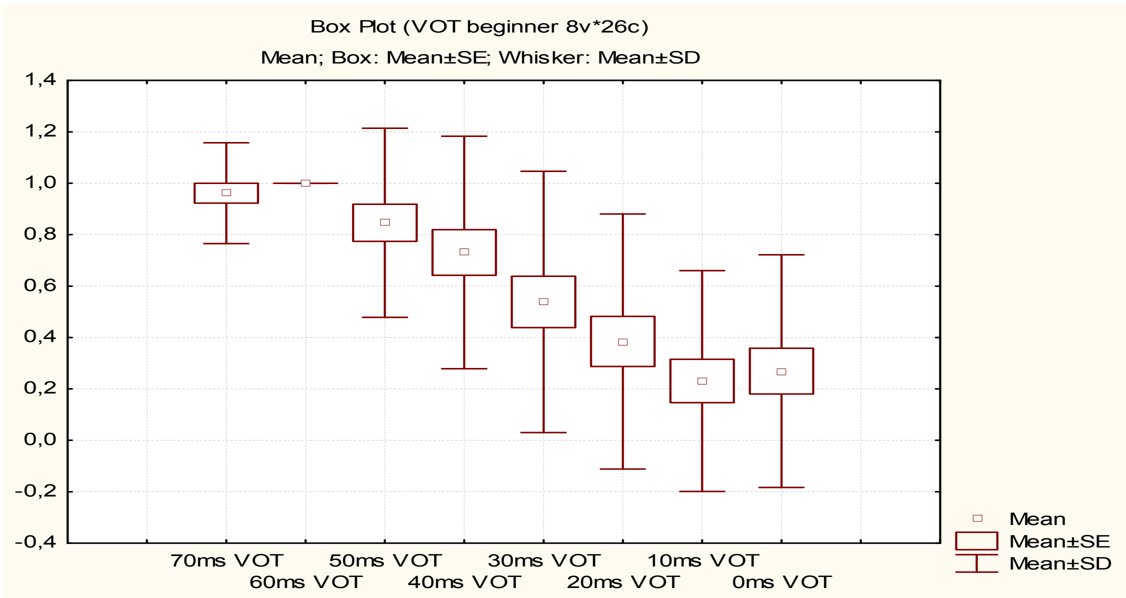


Figure 2.10. Box Plot: Recognition of an initial sound in *keef* as /k/ across the VOT continuum by Beginner Learners

The results show that Beginner Learners reported a gradual change from /k/ to /g/ along the decreasing VOT values with a highly statistically significant effect (Q=85.997, p=0.000**) However, there is no sudden categorisation peak typical for Native

Speakers. Moreover, the BLs were not consistent in judging VOT values – not all subjects reported extreme +70 ms VOT as voiceless and there is a slight and unexpected rise in voiceless judgements for the 0 ms VOT stimulus. A steady decrease in voiceless responses begins at +50 ms VOT and stops at +10 ms VOT.

2.1.1.3. Advanced Learners – results

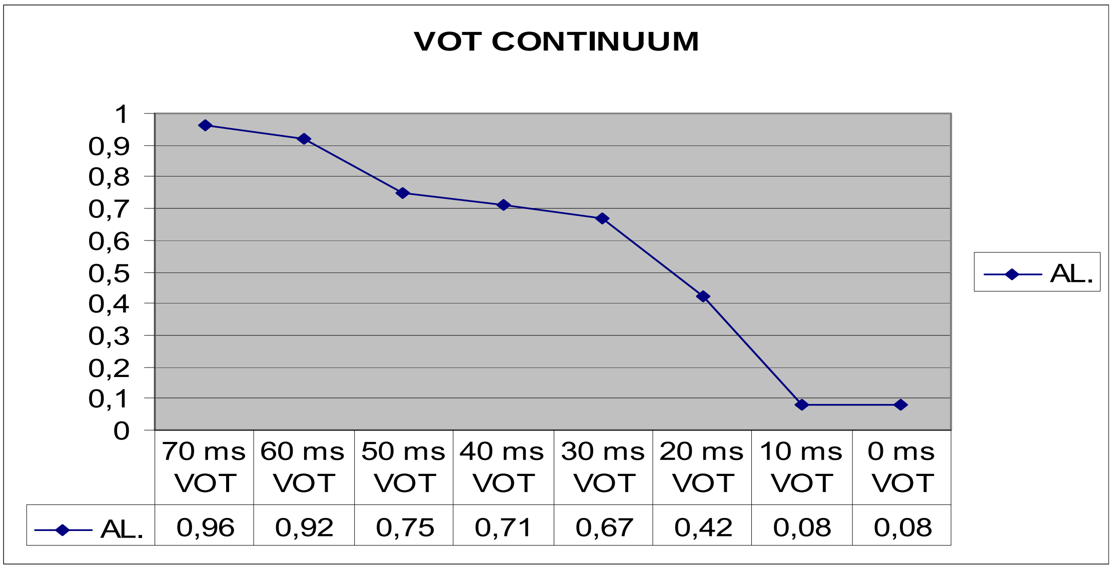


Figure 2.11. Recognition of an initial sound in *keef* as /k/ across the VOT continuum by Advanced Learners

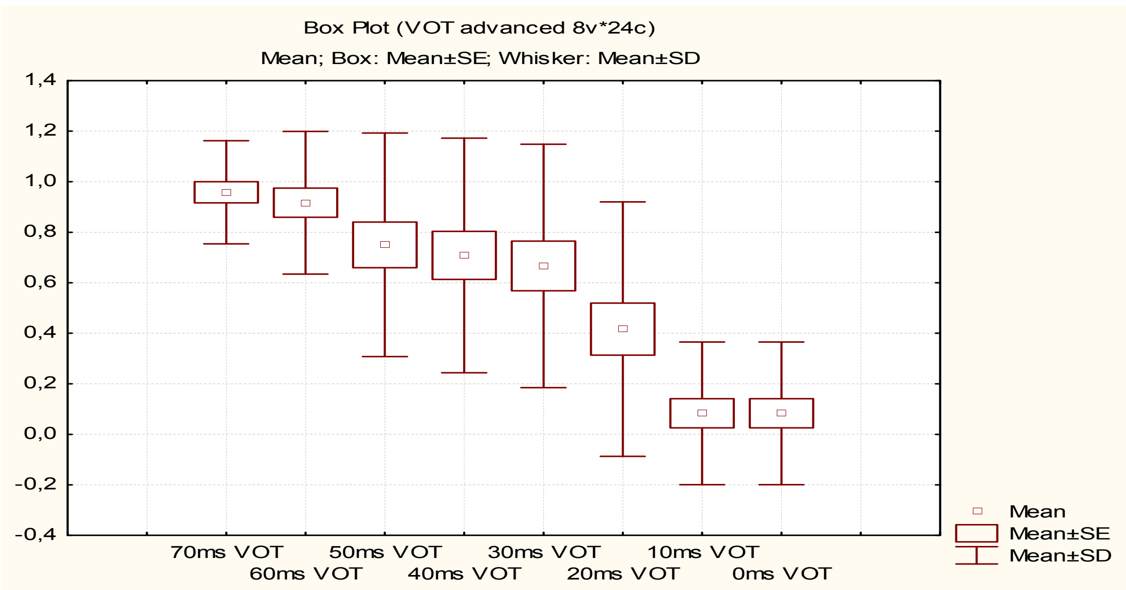


Figure 2.12. Box Plot: Recognition of an initial sound in *keef* as /k/ across the VOT continuum by Advanced Learners

An effect of the VOT continuum was highly significant (Q=84.620, p=0.000**) in the AL group. Advanced Learners demonstrated a categorisation peak at around +20 VOT. It is interesting to note, however, that values at +10 ms VOT and 0 ms VOT were not categorised as voiced by all the subjects. A second slight peak can be observed at +50 ms VOT with subsequent levelling at +40 ms VOT and +30 ms VOT.

2.1.1.4. Native Speakers – results

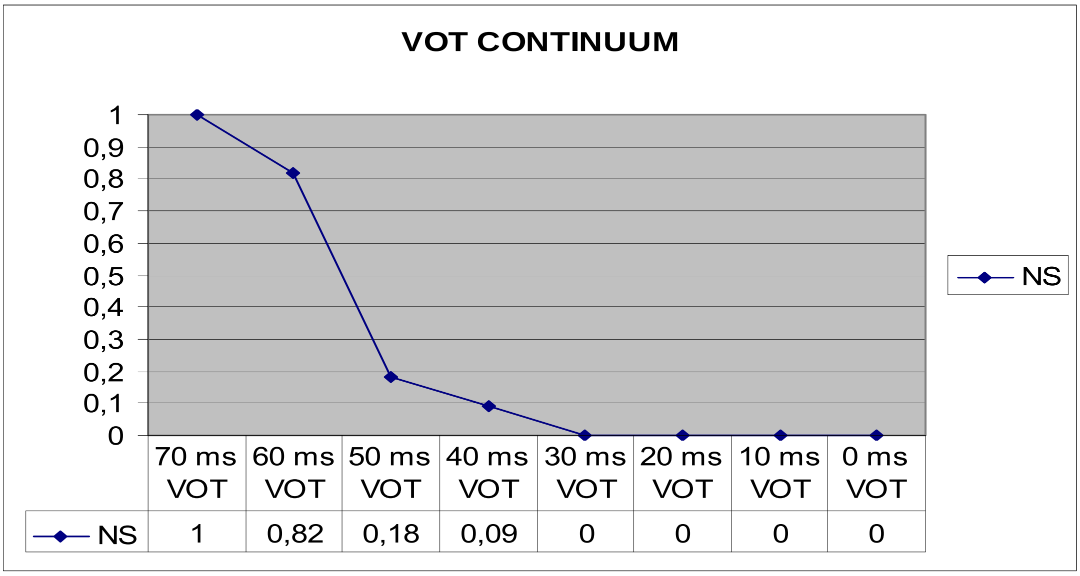


Figure 2.13. Recognition of an initial sound in *keef* as /k/ across the VOT continuum by Native Speakers

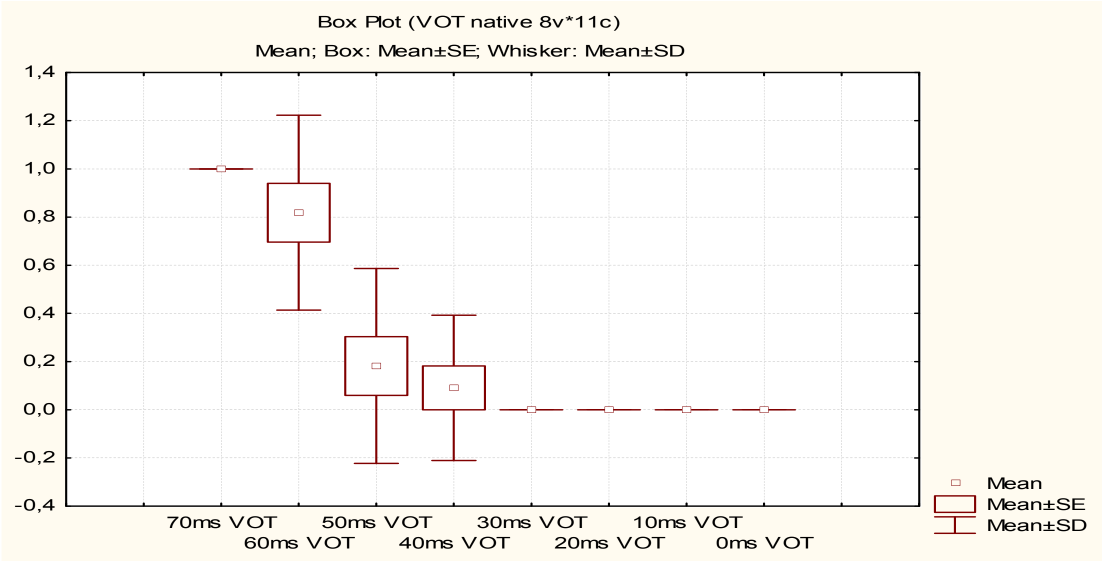


Figure 2.14. Box Plot: Recognition of an initial sound in *keef* as /k/ across the VOT continuum by Native Speakers

As in the case of the BL and AL groups, the stimulus effect was highly significant (Q=60.221, p=0.000**) for the NSs. However, unlike the Polish groups, Native Speakers of English showed a strong partitioning peak of the VOT continuum. Strong categorisation from /k/ to /g/ judgments commences at high VOT values, i.e. at +50 ms VOT. It is completed by a gradual decrease down to +30 ms VOT. All stimuli ranging from +20 ms VOT to 0 ms VOT were consistently reported as voiced.

2.1.2. Partially devoiced initial fricatives

2.1.2.1. Stimuli

From naturally obtained syllables *zeef* /zi:f/ with fully voiced initial /z/ and *seef* /si:f/ with fully voiceless initial /s/, we generated 5 stimuli varying in devoicing degree:

1. *zeef* /zi:f/, /z/ 160 ms voiced
2. *zeef* /zi:f/, /z/ 80 ms voiceless + 80 ms voiced
3. *zeef* /zi:f/, /z/ 120 ms voiceless + 40 ms voiced
4. *zeef* /zi:f/, /z/ 140 ms voiceless + 20 ms voiced
5. *zeef* /zi:f/, /z/ 150 ms voiceless + 10 ms voiced

Figures 2.15. to 2.19. show waveforms and spectrograms of all the stimuli.

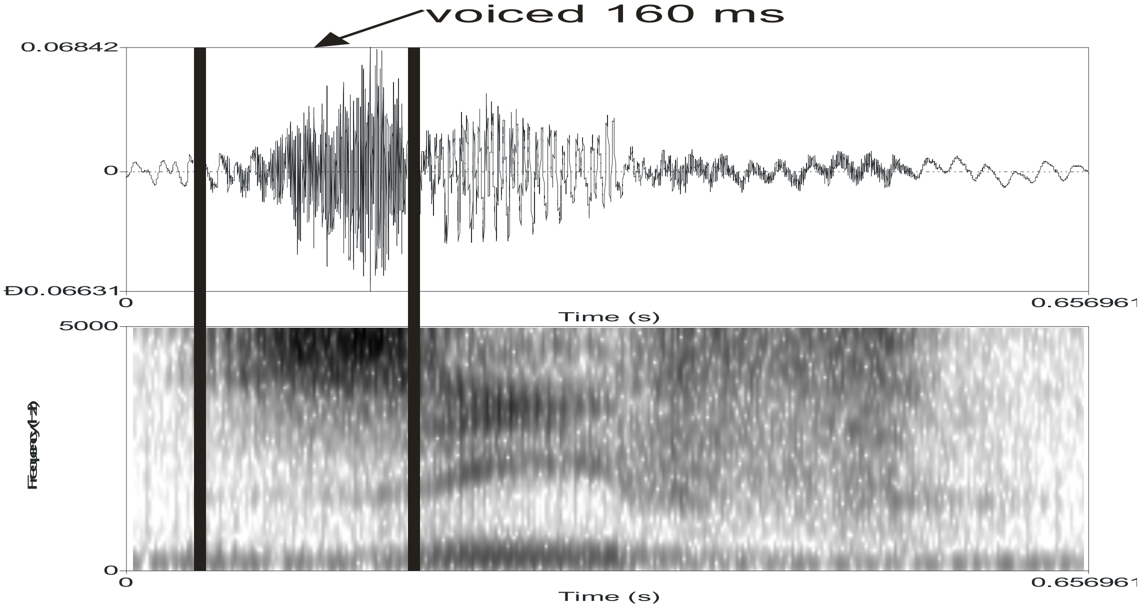


Figure 2.15. Waveform and spectrogram of syllable *zeef*, /z/ 160 ms voiced

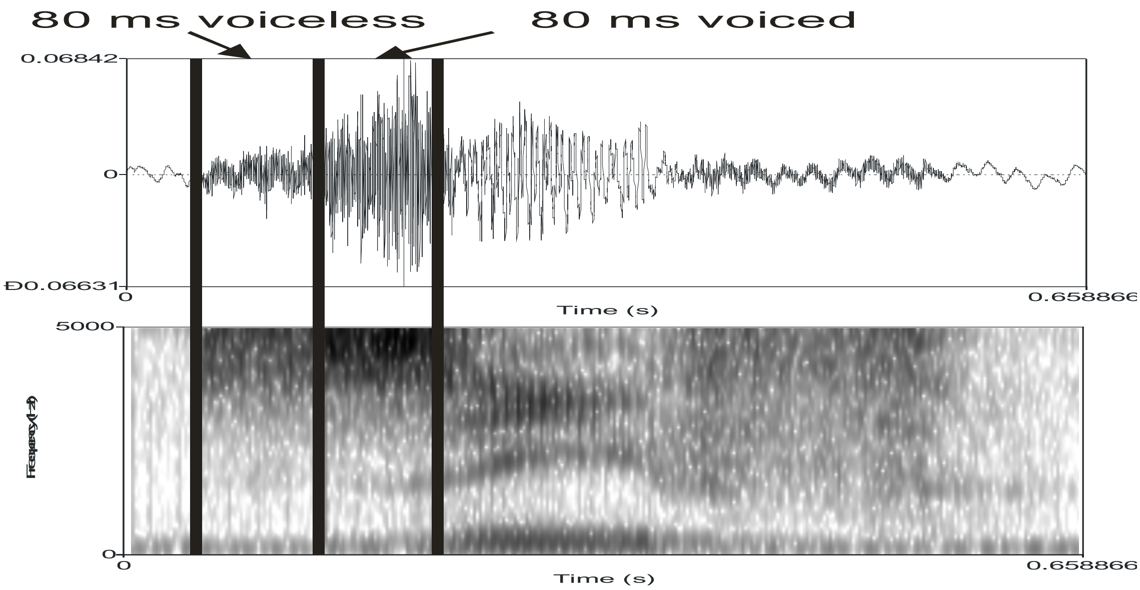


Figure 2.16. Waveform and spectrogram of syllable *zeef*, /z/ 80 ms voiceless + 80 ms voiced

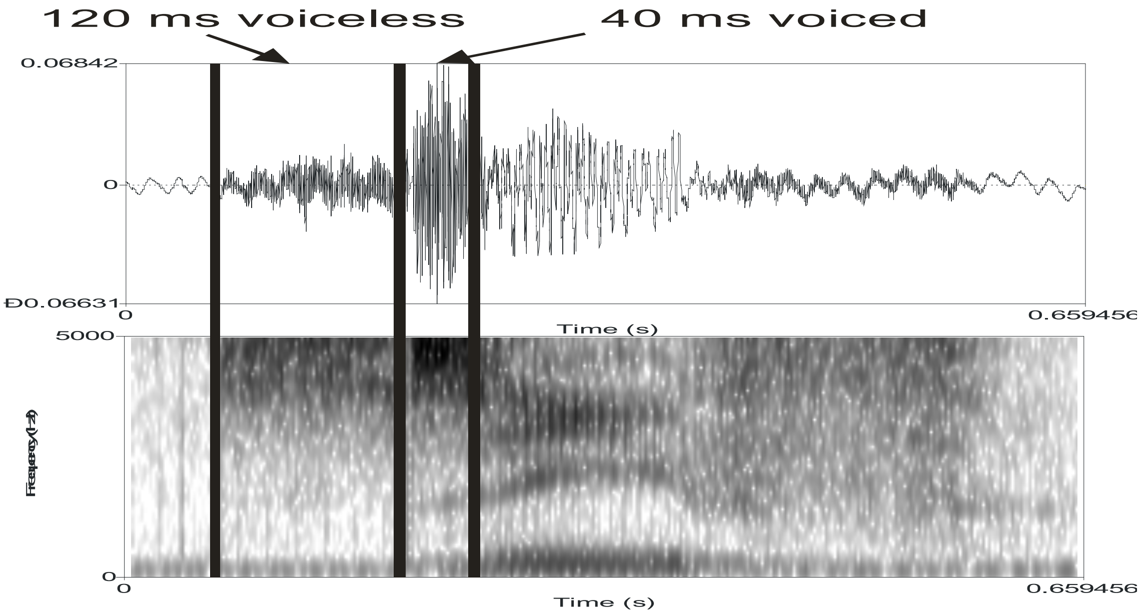


Figure 2.17. Waveform and spectrogram of syllable *zeef*, /z/ 120 ms voiceless + 40 ms voiced

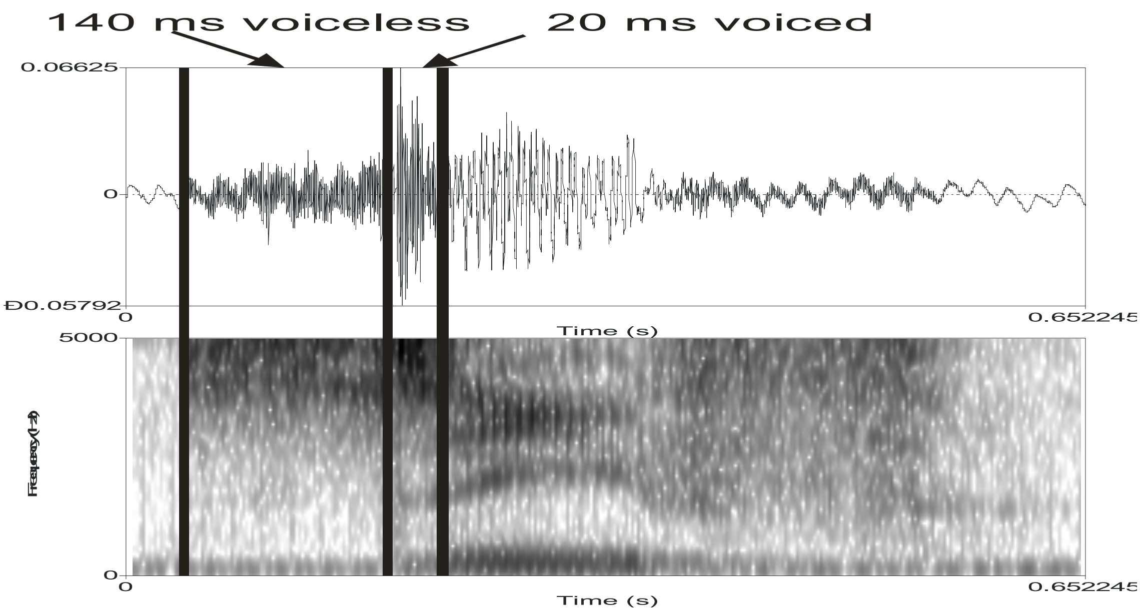


Figure 2.18. Waveform and spectrogram of syllable *zeef*, /z/ 140 ms voiceless + 20 ms voiced

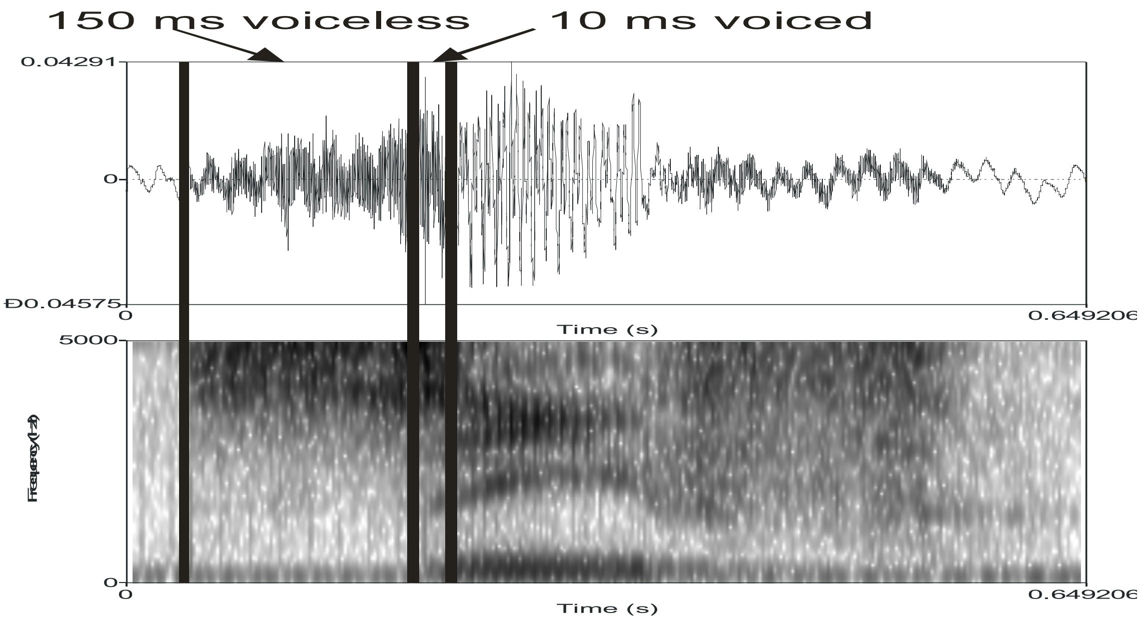


Figure 2.19. Waveform and spectrogram of syllable *zeef*, /z/ 150 ms voiceless + 10 ms voiced

2.1.2.2. Beginner Learners – results

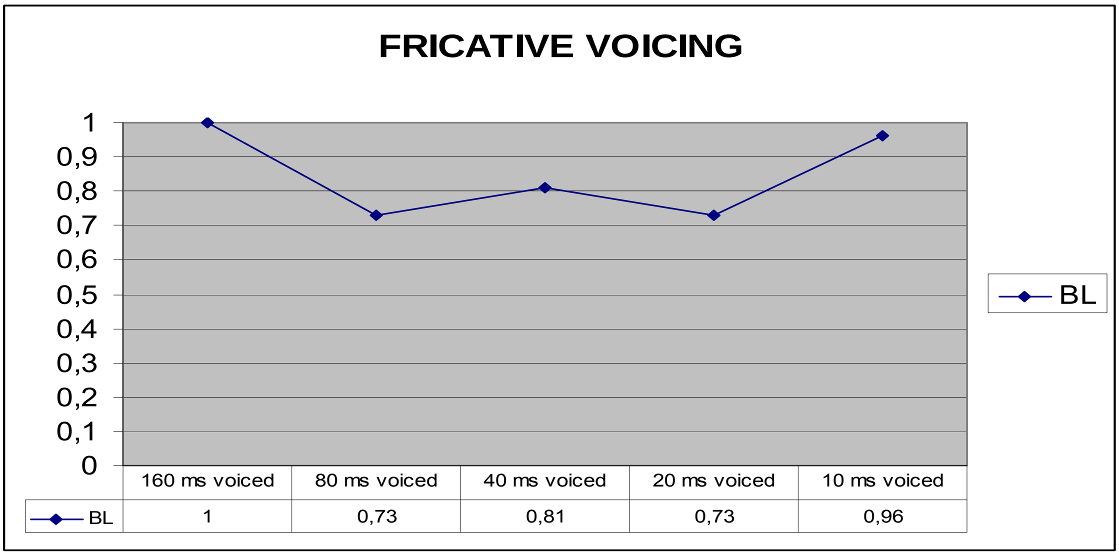


Figure 2.20. Recognition of an initial sound in *zeef* as /z/ across varying degrees of devoicing by Beginner Learners

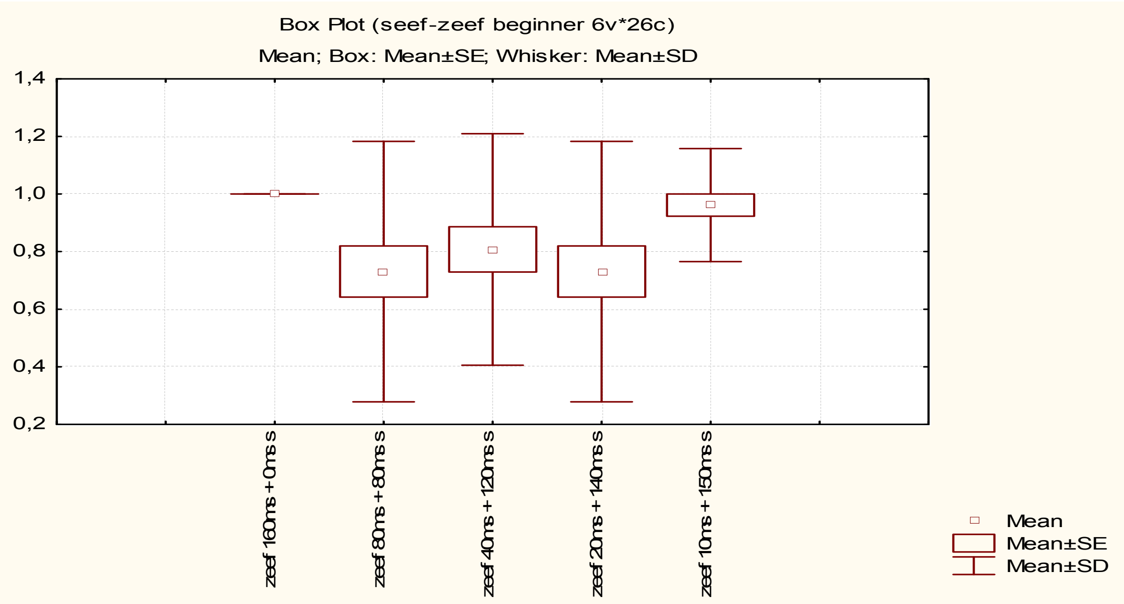


Figure 2.21. Box Plot: Recognition of an initial sound in *zeef* as /z/ across varying degrees of devoicing by Beginner Learners

Although the stimulus effect was statistically significant (Q=14.667, p=0.005**), the BL subjects demonstrated a puzzling recognition tendency. As expected, the inclusion of an 80 ms voiceless element brought about increased voiceless judgments. However, a further extension of a voiceless period caused, contrary to the expectations, an increase in voiced responses. Most surprisingly, almost all Beginner Learners reported the most devoiced stimulus (150 ms voiceless) as belonging to the voiced category.

2.1.2.3. Advanced Learners – results

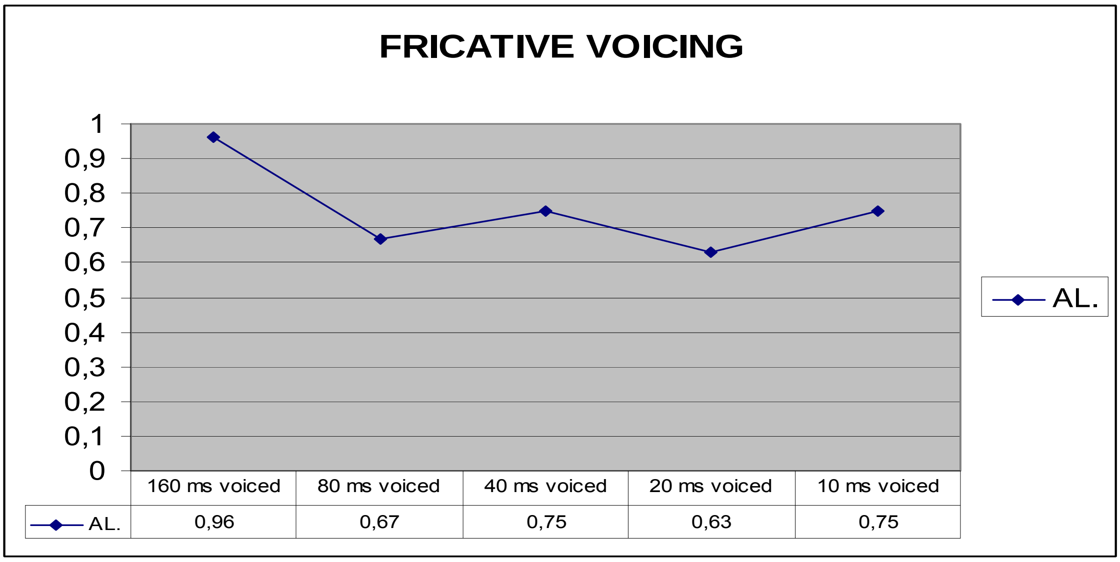


Figure 2.22. Recognition of an initial sound in *zeef* as /z/ across varying degrees of devoicing by Advanced Learners

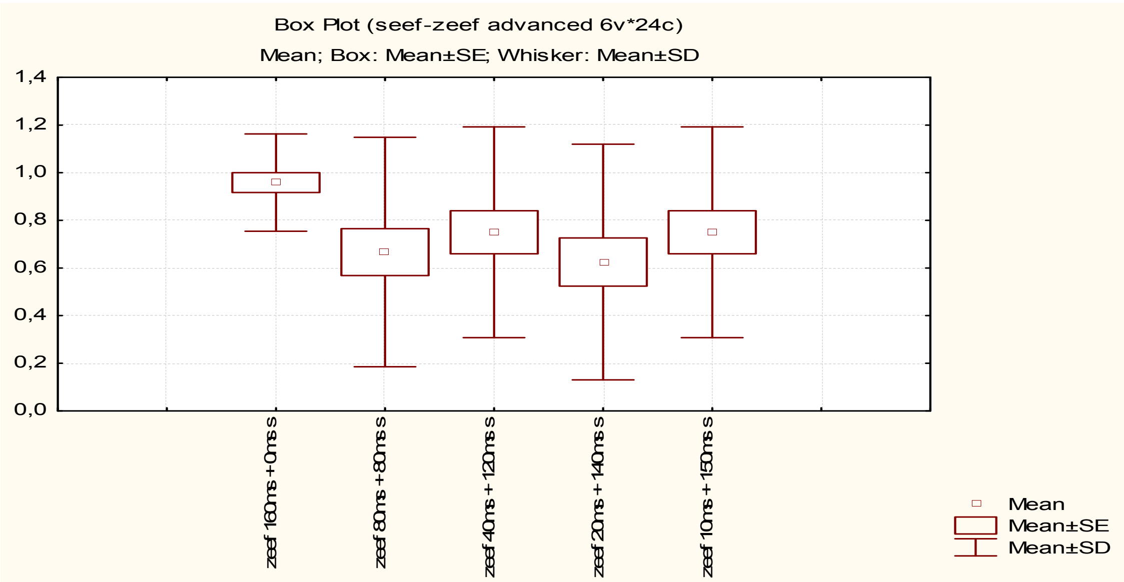


Figure 2.23. Box Plot: Recognition of an initial sound in *zeef* as /z/ across varying degrees of devoicing by Advanced Learners

The results obtained in the AL group show a similar pattern to those obtained for the BLs, even though, for this group, the distribution of responses across the stimuli did not meet the criteria of statistical significance (Q=8.444, p=0.077), which means that the null hypothesis could not be rejected and the effect might have been obtained by chance. It is interesting to note, however, that the ALs did not respond consistently to increasing devoicing, as demonstrated by an increase in voiced judgments for the stimuli with only 40 ms and 10 ms of voiced portion. Most surprisingly, the most devoiced stimulus was recognised as voiced 75% of the time.

2.1.2.4. Native Speakers – results

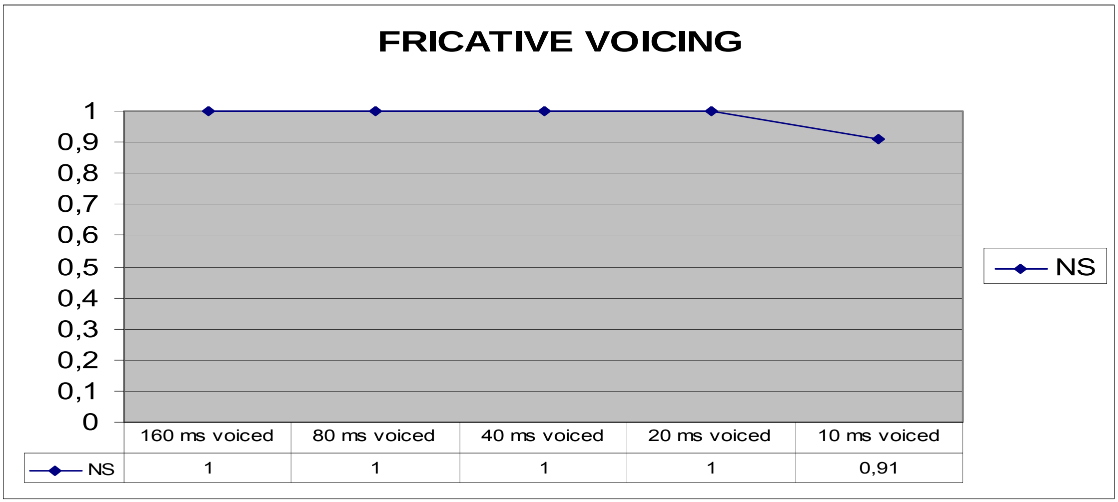


Figure 2.24. Recognition of an initial sound in *zeef* as /z/ across varying degrees of devoicing by Native Speakers

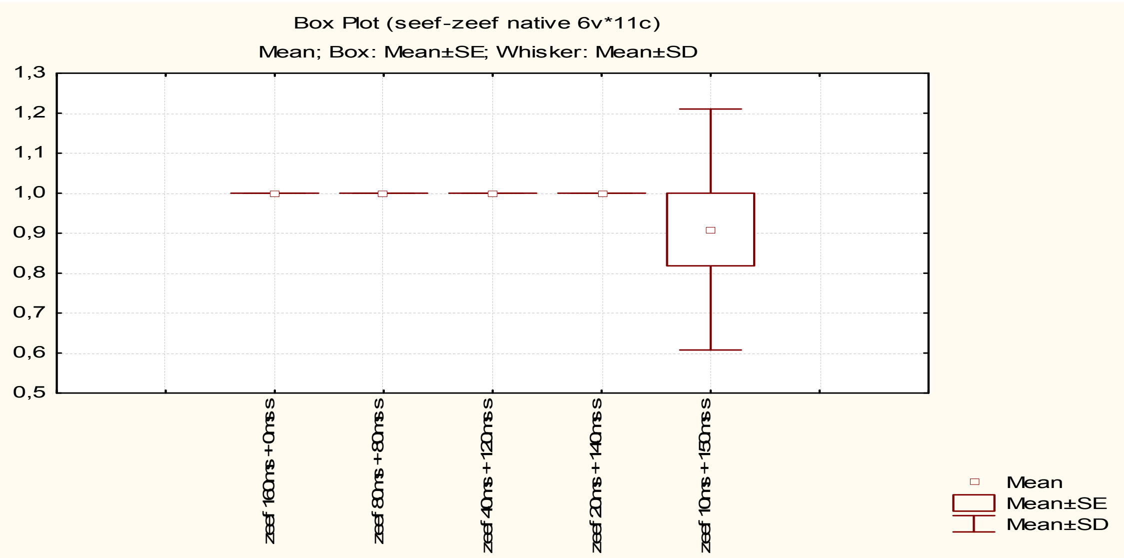


Figure 2.25. Box Plot: Recognition of an initial sound in *zeef* as /z/ across varying degrees of devoicing by Advanced Learners

Native Speakers did not show any effect of the influence of partial devoicing on voiced judgements, except for the most devoiced stimulus with only 10 ms of voicing period. Not surprisingly, the stimulus effect was far from significant (Q=4.000, p=0.4).

2.1.3. Frication duration in initial fricatives

2.1.3.1. Stimuli

In a recording session, we obtained a syllable *foss* /f①s/ with 137 ms duration of initial /f/. Next, we created another stimulus by reducing initial frication duration by half. The reduction was expected to bring about a change from a voiceless to voiced percept (see Part 2, Section 6) Consequently, the subjects were presented with two stimuli:

1. *foss*, /f①s/, /f/ 137 ms duration
2. *foss*, /f①s/, /f/ 68 ms duration

Figures 2.26. and 2.27. show waveforms and spectrograms of the two stimuli.

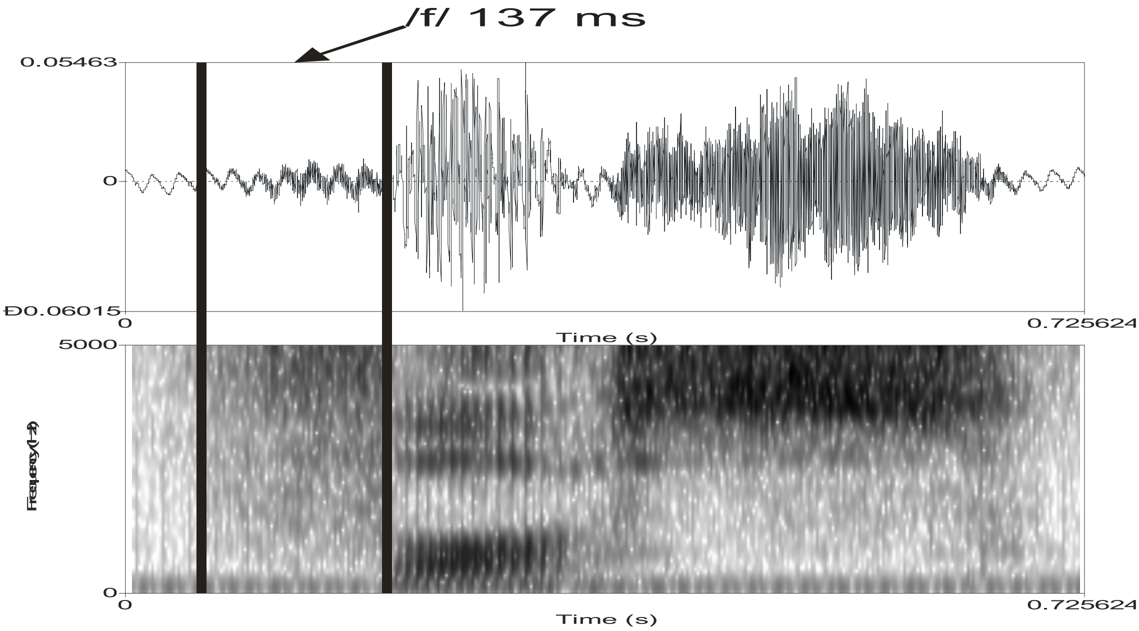


Figure 2.26. Waveform and spectrogram of syllable *foss*, /f/ 137 ms duration

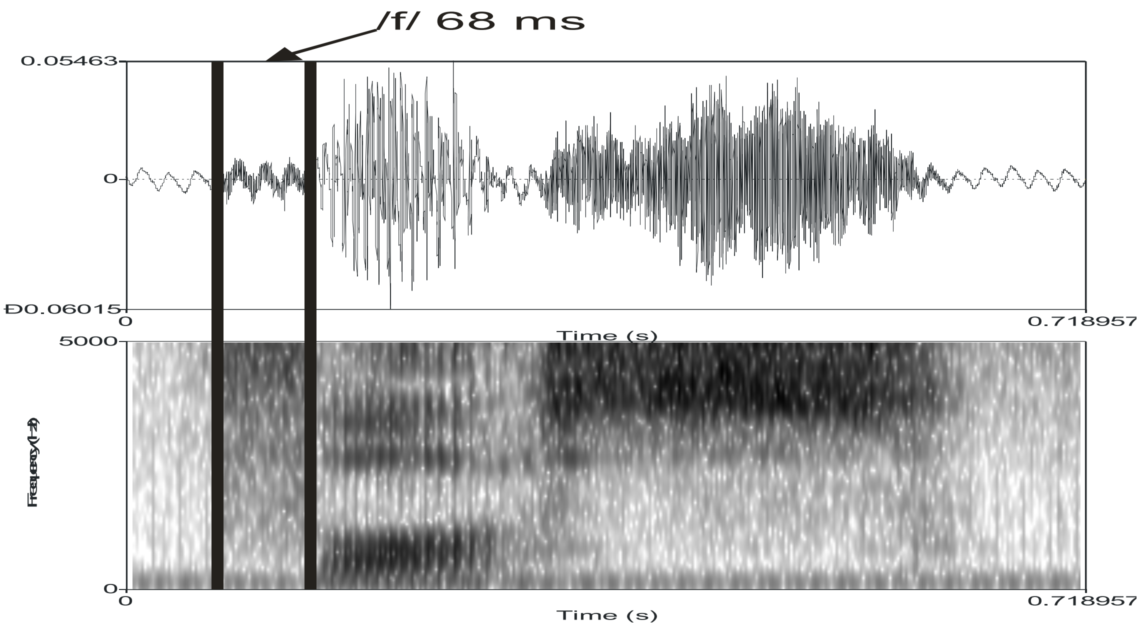


Figure 2.27. Waveform and spectrogram of syllable *foss*, /f/ 68 ms duration

2.1.3.2. Beginner Learners – results

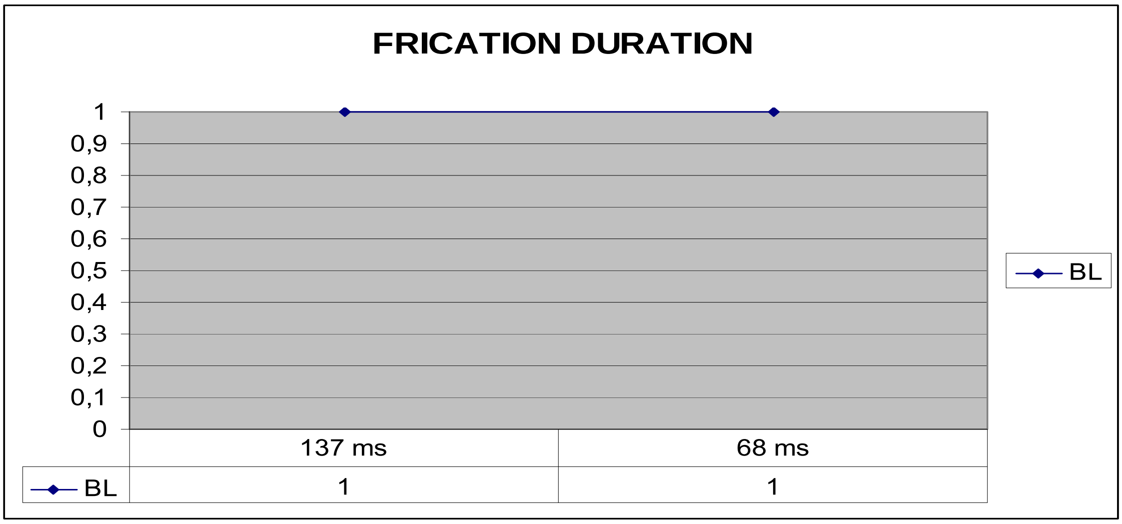


Figure 2.28. Recognition of an initial sound in *foss* as /f/ across varying frication duration by Beginner Learners

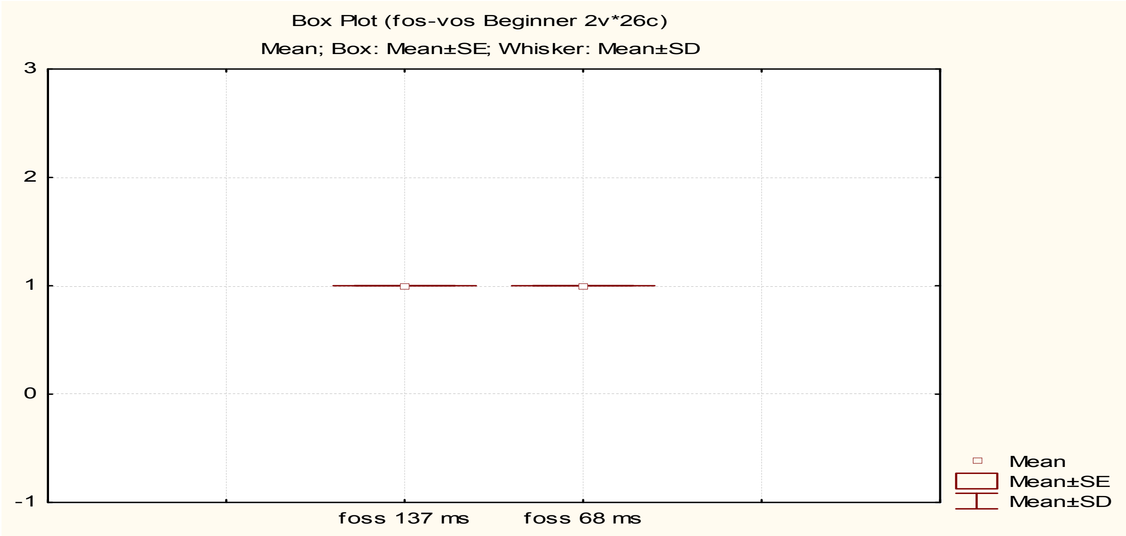


Figure 2.29. Box Plot: Recognition of an initial sound in *foss* as /f/ across varying frication duration by Beginner Learners

Beginner Learners did not react to the reduced duration of an initial fricative. The reduction from 137 ms to 68 ms did not result in a change to a voiced percept. All the subjects reported the reduced segment as voiceless.

2.1.3.3. Advanced Learners – results

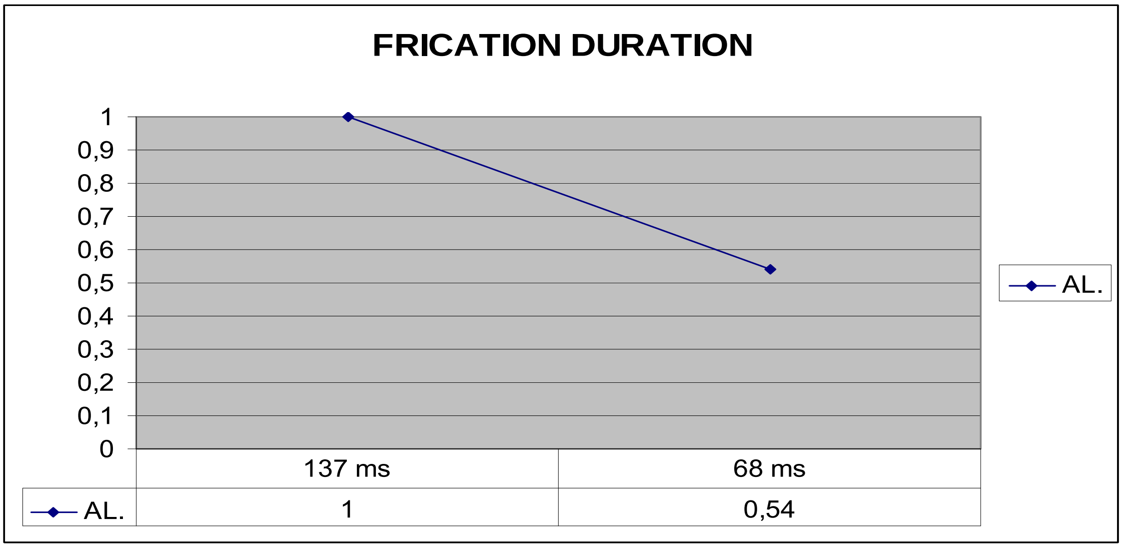


Figure 2.30. Recognition of an initial sound in *foss* as /f/ across varying frication duration by Advanced Learners

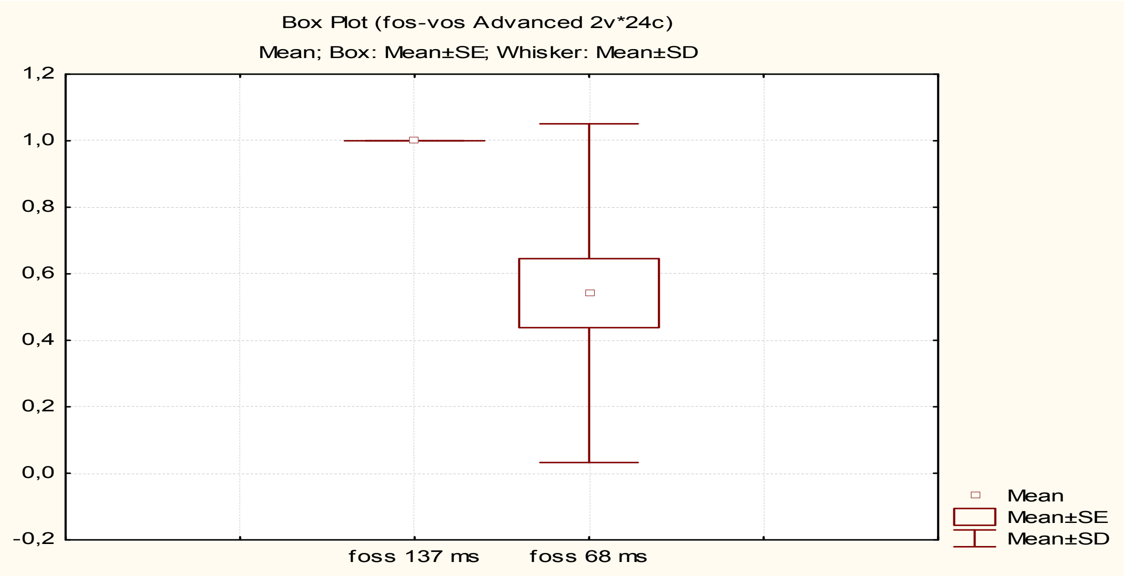


Figure 2.31. Box Plot: Recognition of an initial sound in *foss* as /f/ across varying frication duration by Advanced Learners

The results show that Advance Learners were sensitive to the reduction of frication duration in an initial fricative, however the obtained results did not meet the significance criteria (McNemar *Chi*=2.70, p>0.05). Nevertheless, it is interesting to note that almost 50% of the subjects in this group reported hearing a voiced percept when the length of an initial segment was 68 ms.

2.1.3.4. Native Speakers – results

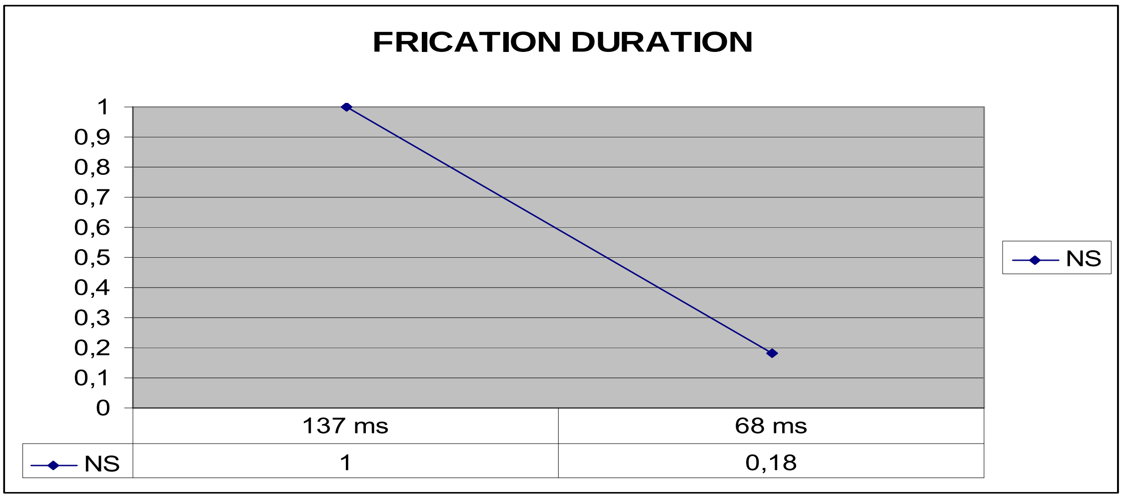


Figure 2.32. Recognition of an initial sound in *foss* as /f/ across varying frication duration by Native Speakers

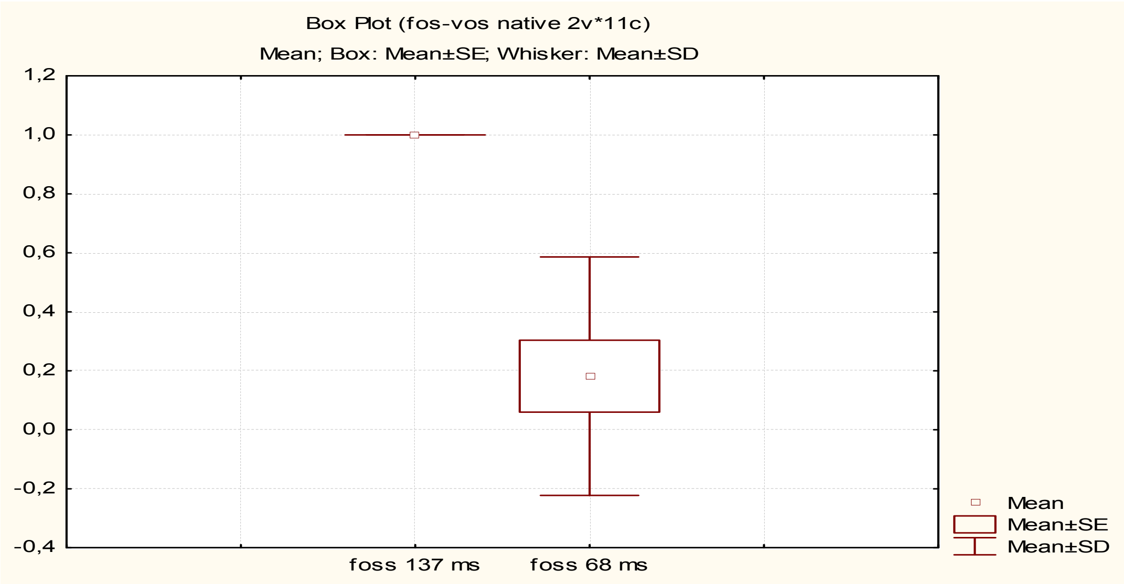


Figure 2.33. Box Plot: Recognition of an initial sound in *foss* as /f/ across varying frication duration by Native Speakers

Native speakers showed a significant (McNemar *Chi*=4.92, p=0.0265*) shift in voicing judgments across the decreasing frication duration. When presented with the shortened frication noise, the subjects reported hearing a voiced percept more than 80% of the time.

2.1.4. Duration of the release burst in initial affricates

2.1.4.1. Stimuli

From a naturally obtained syllable *cheeth* /t♠i:θ/ with 120 ms release duration of the initial /t♠/, we obtained another 2 stimuli by reducing the release duration by 40 ms steps. The subjects were presented with the following stimuli:

1. *cheeth* /t♠i:θ/, /t♠/ 120 ms release burst
2. *cheeth* /t♠i:θ/, /t♠/ 80 ms release burst
3. *cheeth* /t♠i:θ/, /t♠/ 40 ms release burst

Figures 2.34 to 2.36. demonstrate waveforms and spectrograms of the three stimuli.

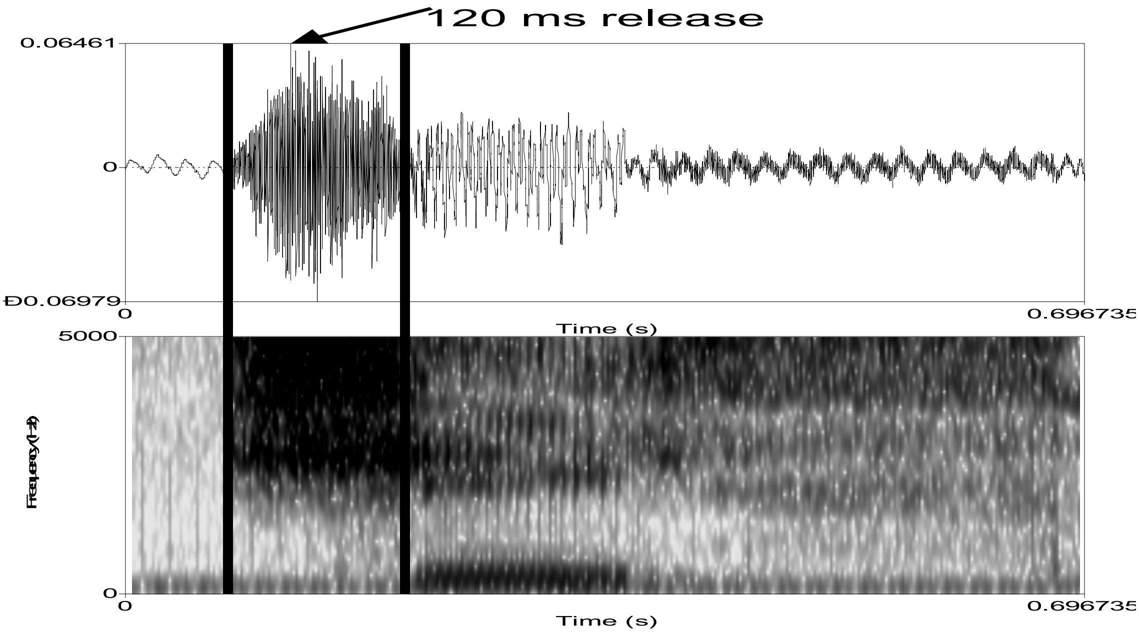


Figure 2.34. Waveform and spectrogram of syllable *cheeth*, /t♦/ 120 ms release burst

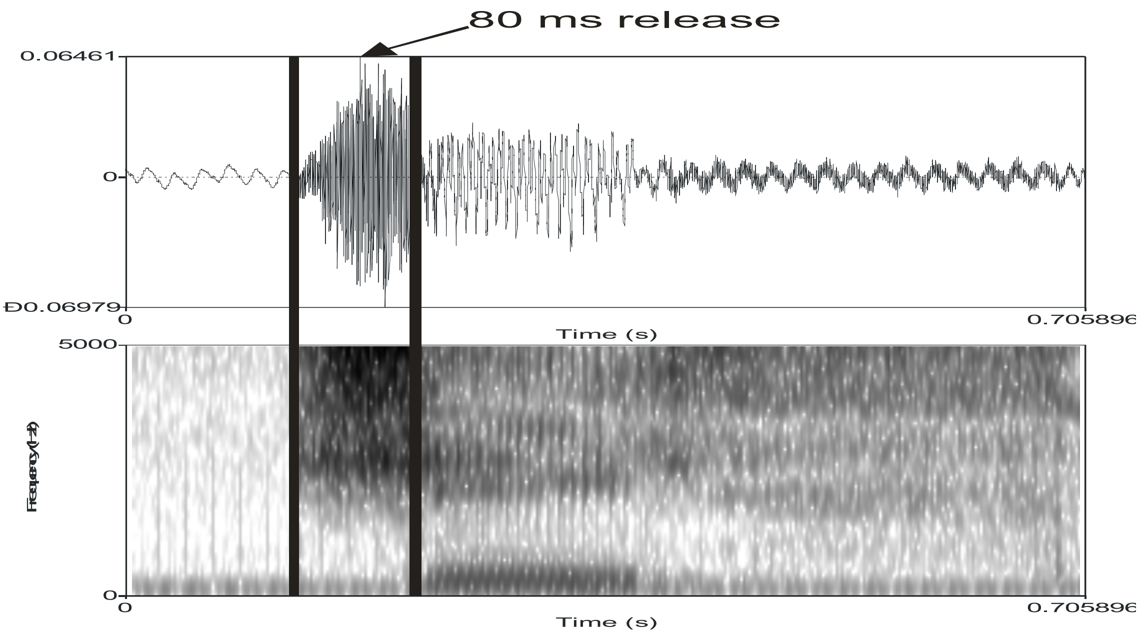


Figure 2.35. Waveform and spectrogram of syllable *cheeth*, /t♦/ 80 ms release burst

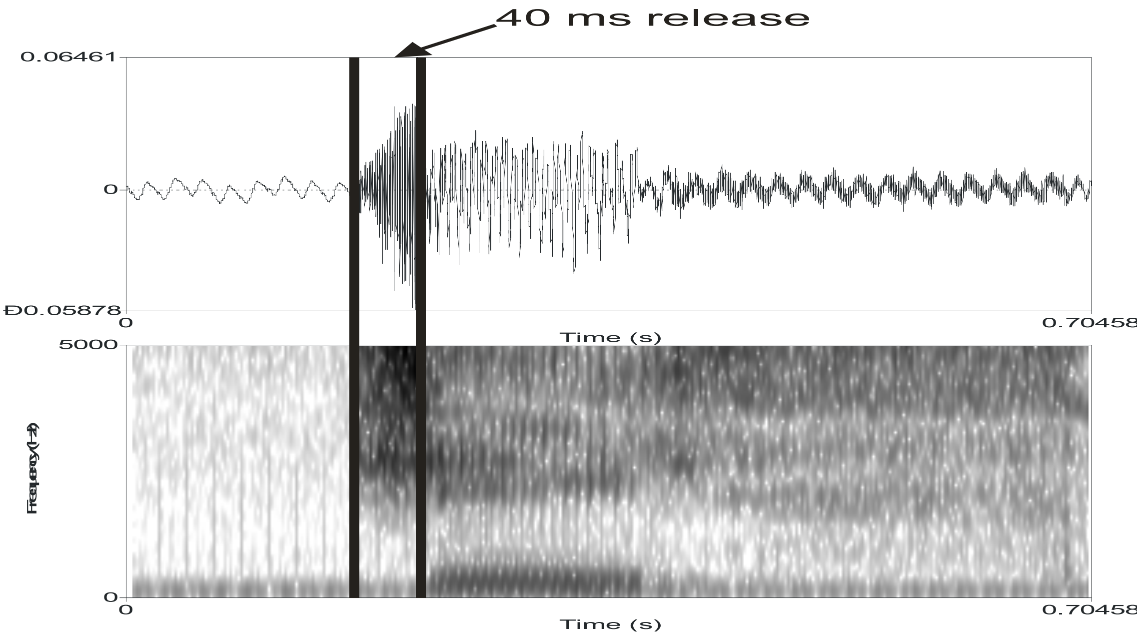


Figure 2.36. Waveform and spectrogram of syllable *cheeth*, /t♠/ 40 ms release burst

2.1.4.2. Beginner Learners – results

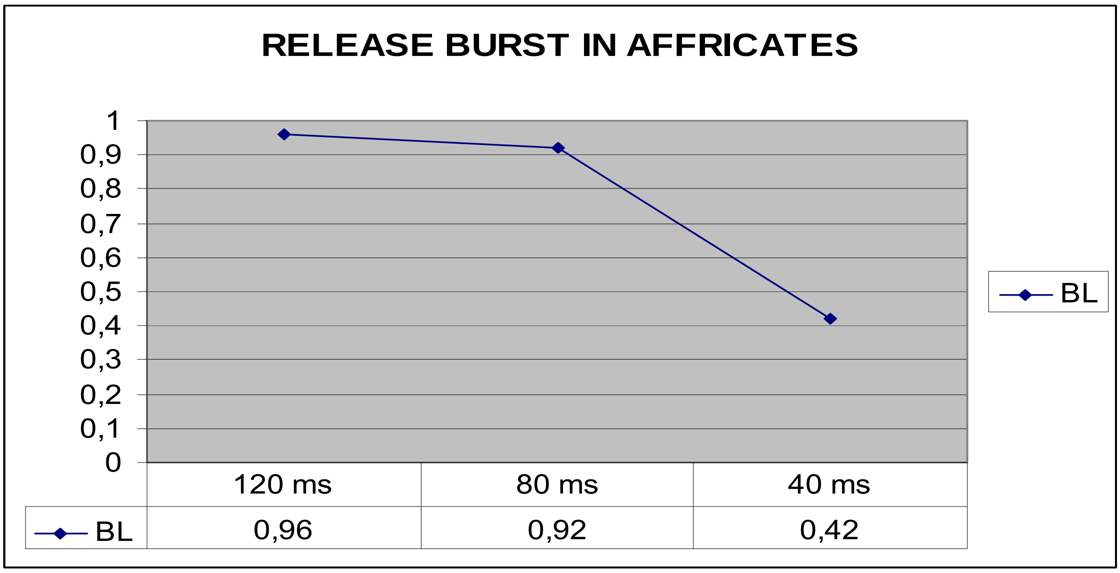


Figure 2.37. Recognition of an initial sound in *cheeth* as voiceless /t♠/ across varying duration of the release burst by Beginner Learners

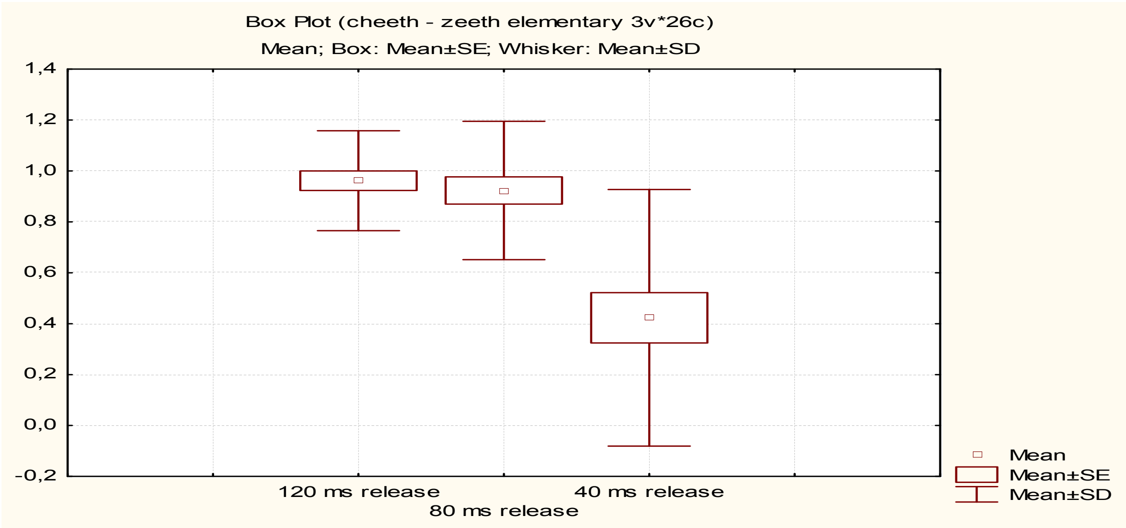


Figure 2.38. Box Plot: Recognition of an initial sound in *cheeth* as voiceless /t♦/ across varying duration of the release burst by Beginner Learners

The results show that the BLs were sensitive to the duration of the release burst as a cue to the voicing contrast in affricates. The change in a voicing status was statistically highly significant (Q=24.400, p=0.000**). Although the first shortening by 40 ms resulted only in a minor shift in voicing judgments, the stimulus with the shortest release burst, only 40 ms, was reported as voiced by almost 60% of the subjects.

2.1.4.3. Advanced Learners – results

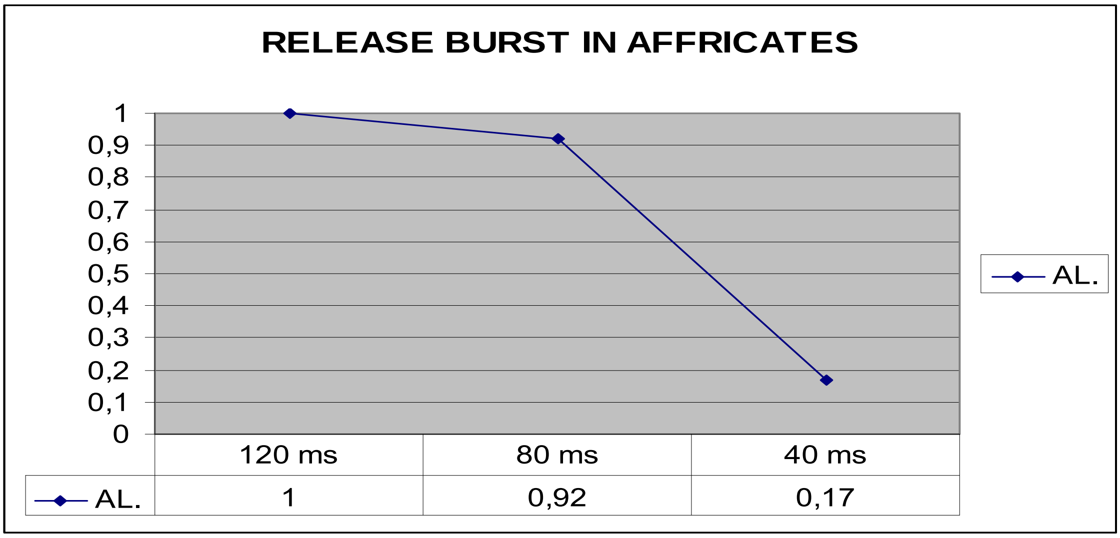


Figure 2.39. Recognition of an initial sound in *cheeth* as voiceless /t♦/ across varying duration of the release burst by Advanced Learners

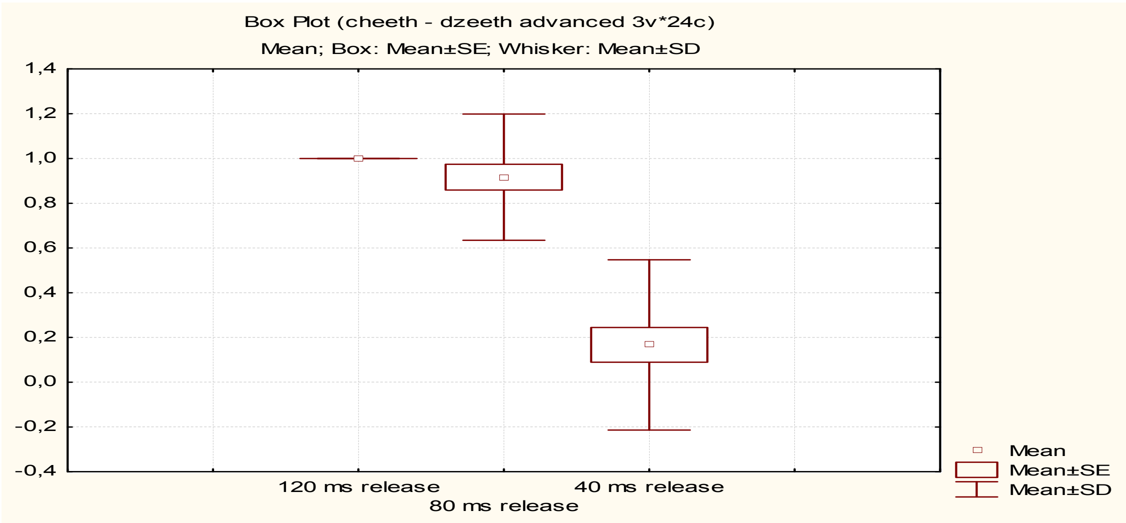


Figure 2.40. Box Plot: Recognition of an initial sound in *cheeth* as voiceless /t♦/ across varying duration of the release burst by Advanced Learners

Similar to the pattern observed for the BL group, the ALs responded to the shortening of duration of the release burst. The difference was found to be highly significant (Q=36.400, p=0.000**). Again, the first shortening by 40 ms did not result in a significant shift, however, when the release was reduced by 80 ms, the subjects reported hearing a voiced segment more than 80% of the time.

2.1.4.4. Native Speakers – results

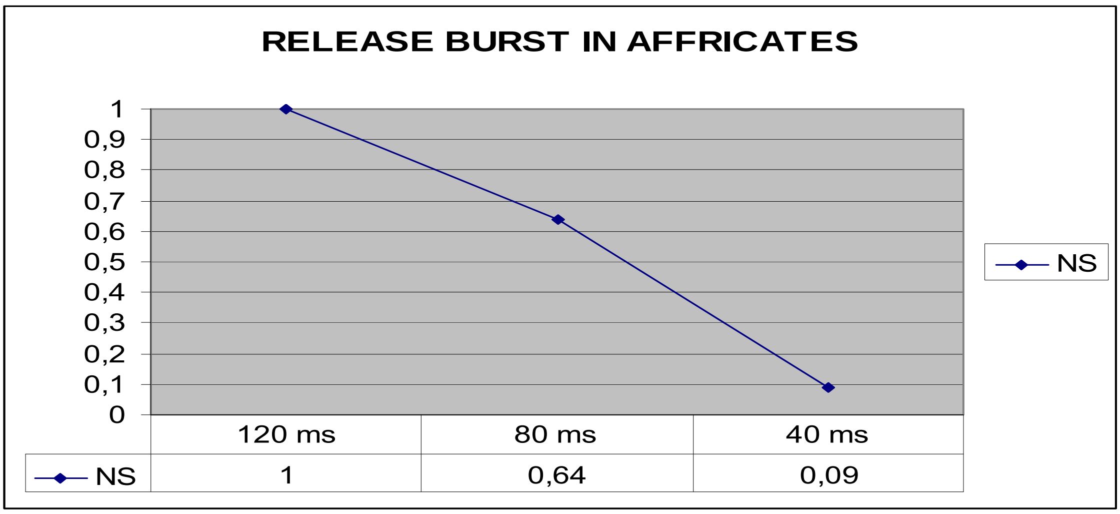


Figure 2.41. Recognition of an initial sound in *cheeth* as voiceless /t♦/ across varying duration of the release burst by Native Speakers

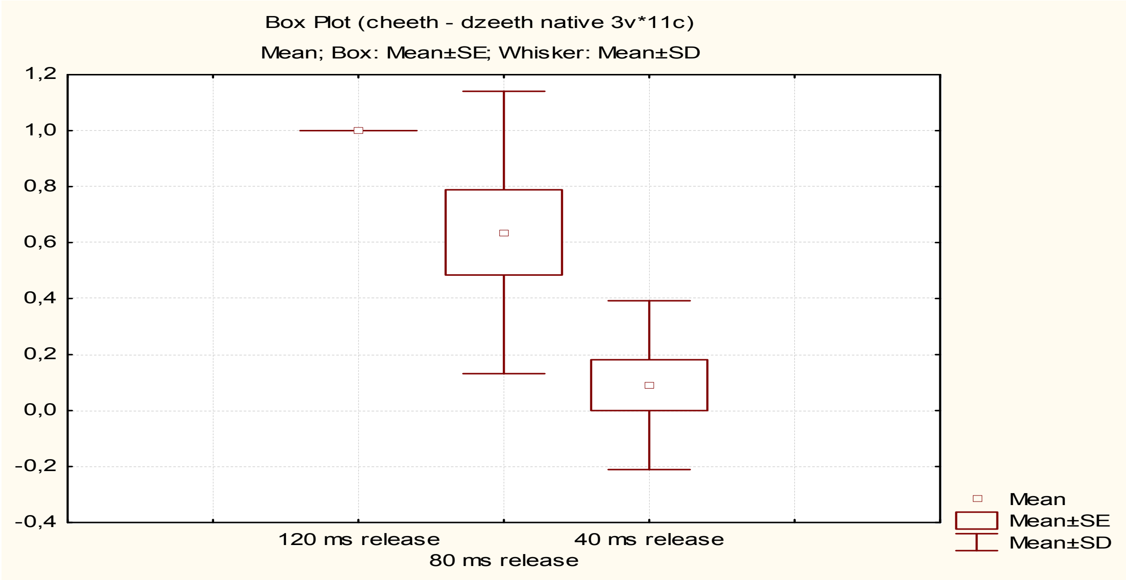


Figure 2.42. Box Plot: Recognition of an initial sound in *cheeth* as voiceless /tʃ/ across varying duration of the release burst by Native Speakers

The group of Native Speakers was more sensitive to the reduction of the release burst than the Polish groups. The stimulus effect was found to be highly significant (Q=15.200, p=0.000**). The first reduction by only 40 ms brought about an almost 40% increase in voiced judgments. Shortening of the release burst by 80 ms resulted in subjects reporting hearing a voiced affricate nearly most of the time (91%).

2.1.5. Vowel duration

2.1.5.1. Stimuli

By manipulating a naturally obtained syllable *theep* /θi:p/ with 142 ms of the vowel duration, we lengthened the vowel duration as a cue to the voicing contrast of the following stop. We used 30 ms steps and generated 6 stimuli:

1. *theep* /θi:p/, vowel duration 142 ms
2. *theep* /θi:p/, vowel duration 172 ms
3. *theep* /θi:p/, vowel duration 202 ms
4. *theep* /θi:p/, vowel duration 232 ms
5. *theep* /θi:p/, vowel duration 262 ms

6. *theep* /θi:p/, vowel duration 292 ms

Since we were interested in vowel duration as the only factor conditioning the perception of the voicing contrast, we decided to modify the release burst of the final plosive. Because the stimulus template was the syllable closed by a voiceless plosive /p/, we weakened its release burst by removing the highest energy excitation point, so that it would not serve as an overriding cue to the vowel duration. The research has shown that the final release burst is conditioned by the identity of the preceding vowel (Parker and Walsh 1981), gender of the speaker (Byrd 1992, 1993, 1994), place of articulation (Crystal and House 1988a), speaking style (Picheny et al. 1985, 1986, Bond and Moore 1994) and the position of the stops within an utterance (Halle et al. 1957 reported in Tsukada et al. 2004). More importantly, it has been found to be perceptually informative about the place and voicing both for native (Householder 1956, Malécot 1958, Wang 1959) and nonnative listeners (Bent and Bradlow 2003, Smith et al. 2003).

Figures 2.43. to 2.48. show waveforms and spectrograms of the 6 stimuli.

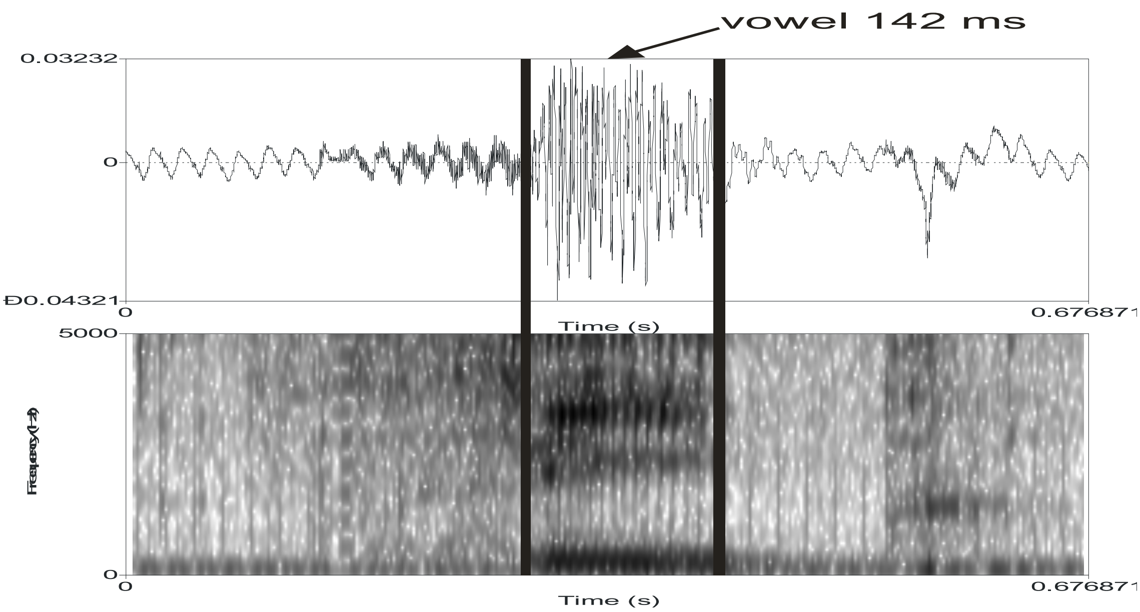


Figure 2.43. Waveform and spectrogram of syllable *theep*, vowel duration 142 ms

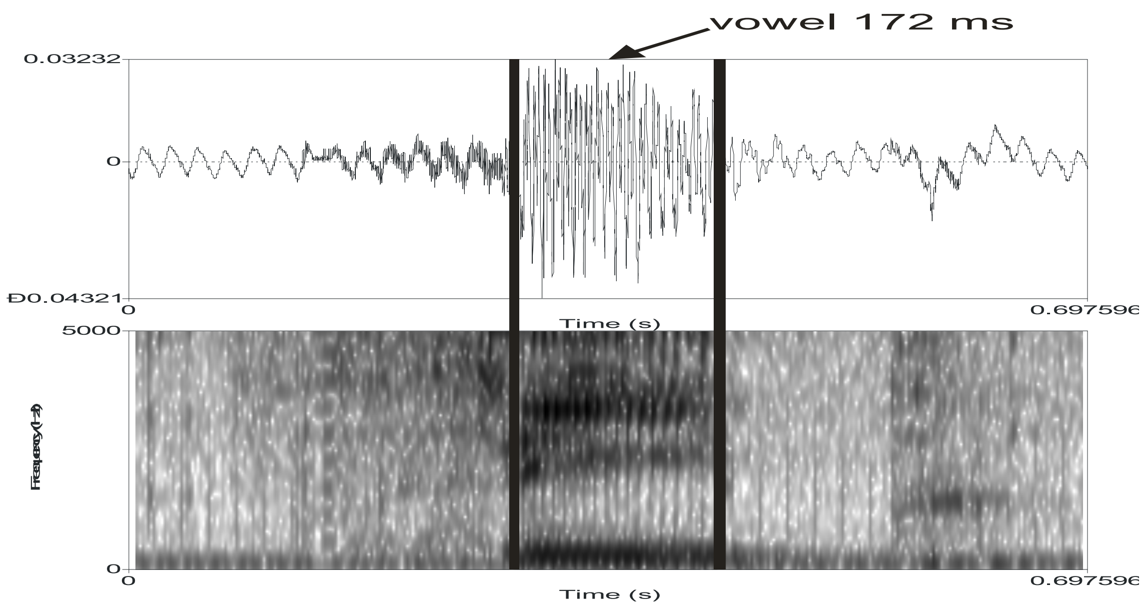


Figure 2.44. Waveform and spectrogram of syllable *theep*, vowel duration 172 ms

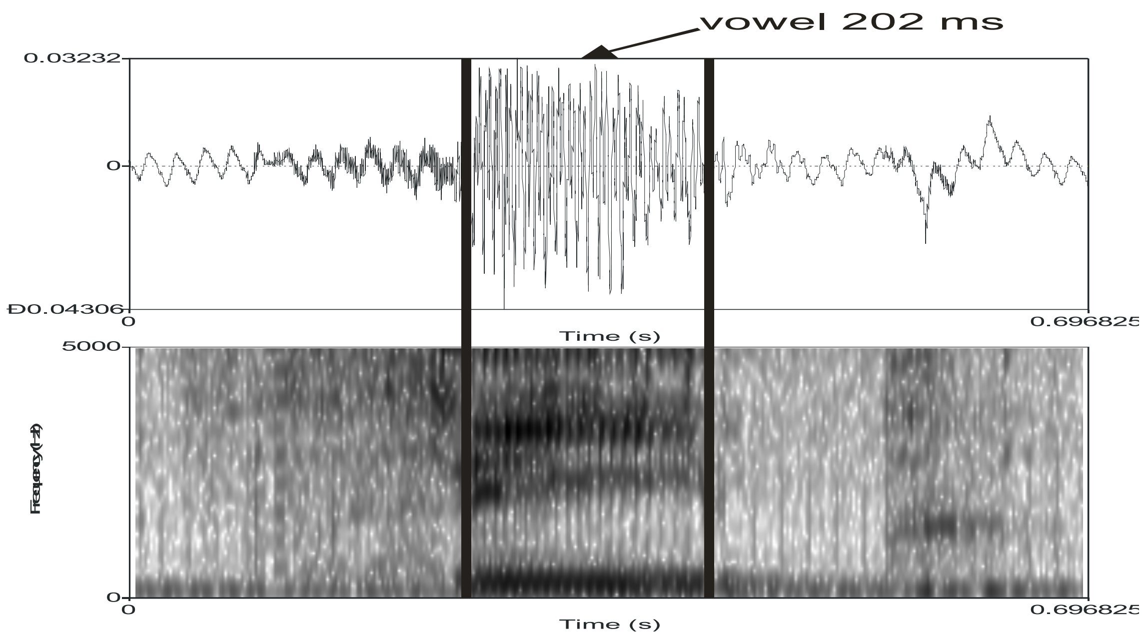


Figure 2.45. Waveform and spectrogram of syllable *theep*, vowel duration 202 ms

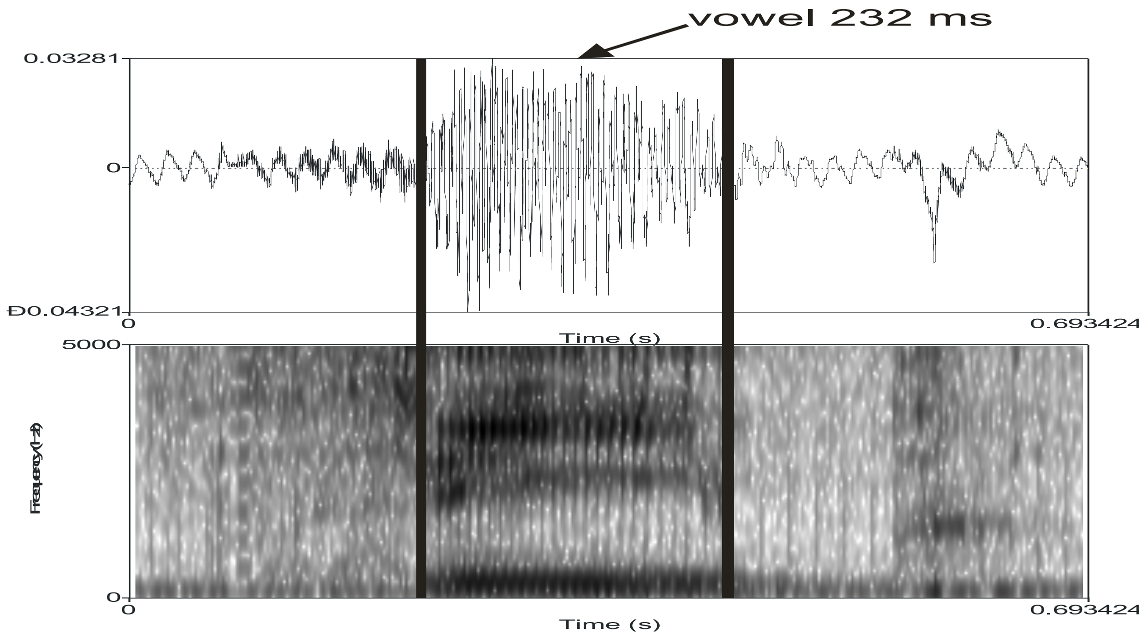


Figure 2.46. Waveform and spectrogram of syllable *theep*, vowel duration 232 ms

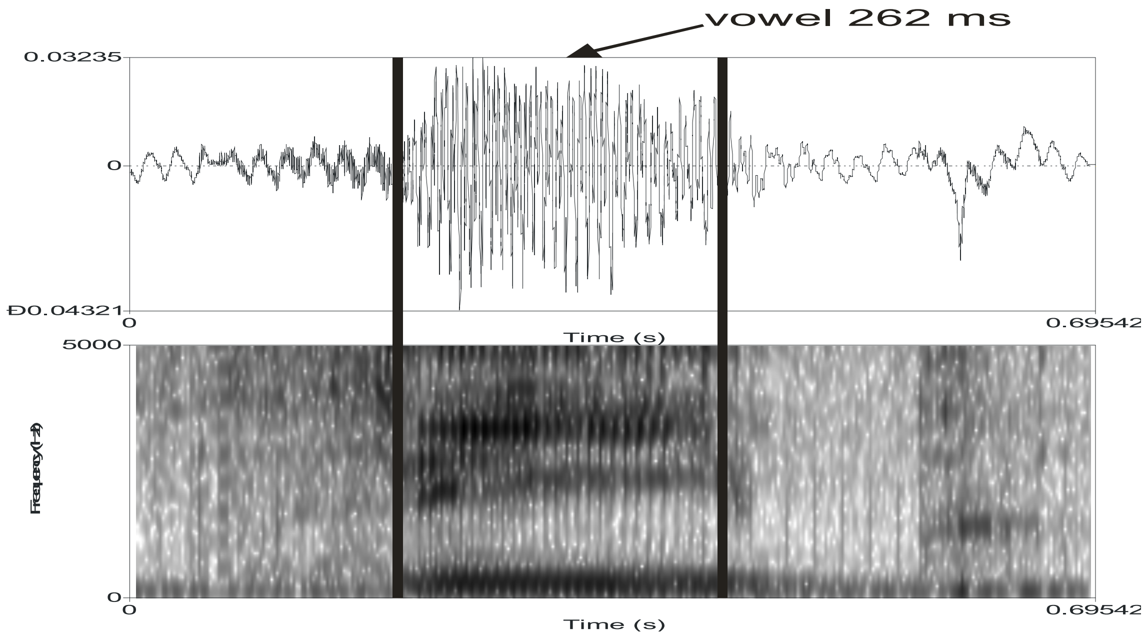


Figure 2.47. Waveform and spectrogram of syllable *theep*, vowel duration 262 ms

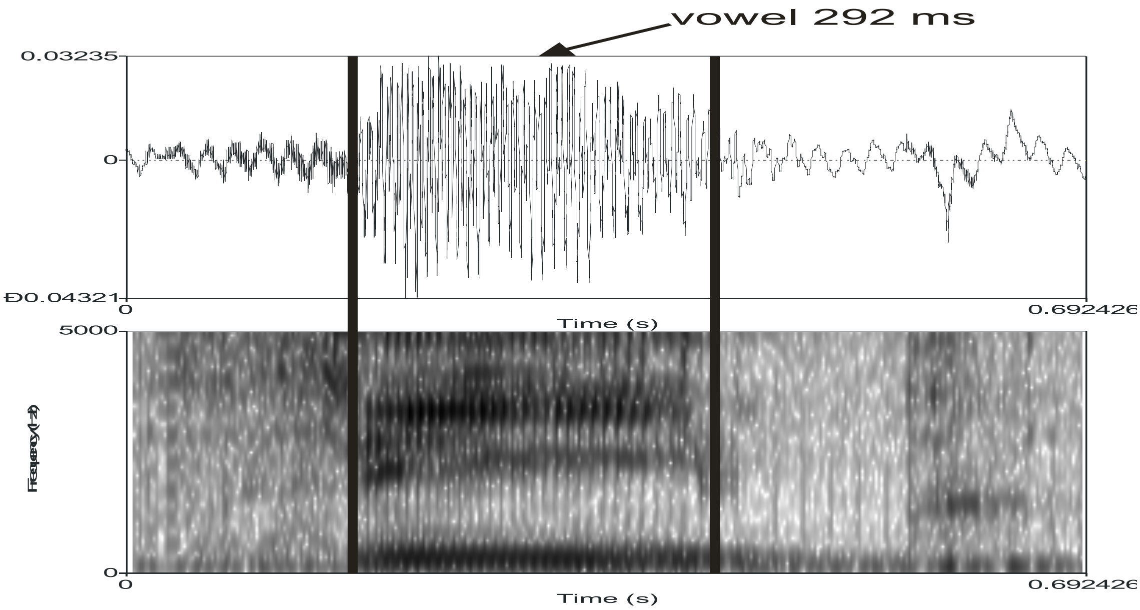


Figure 2.48. Waveform and spectrogram of syllable *theep*, vowel duration 292 ms

2.1.5.2. Beginner Learners – results

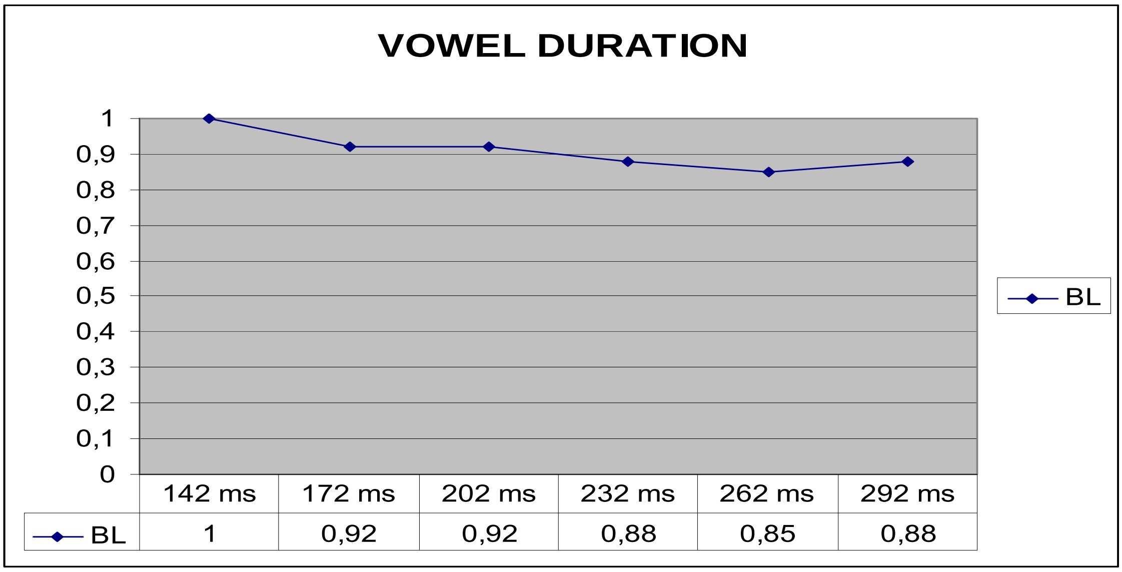


Figure 2.49. Recognition of a final sound in *theep* as voiceless /p/ across varying vowel duration by Beginner Learners

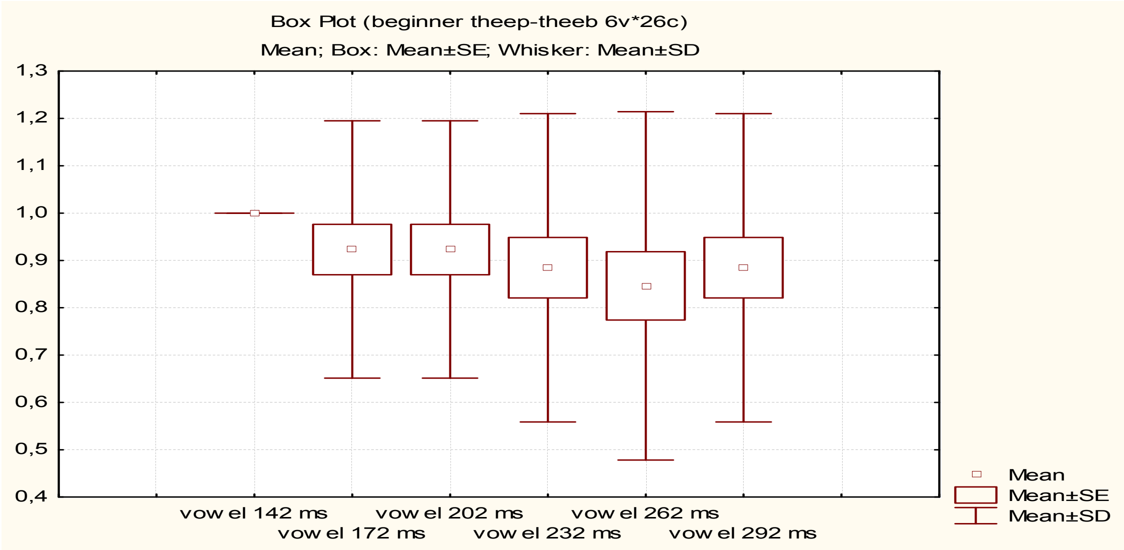


Figure 2.50. Box Plot: Recognition of a final sound in *theep* as voiceless /p/ across varying vowel duration by Beginner Learners

Beginner Learners did not read increasing vowel duration as a cue to the voicing contrast of the final plosive. The stimulus effect was not statistically significant ($Q=5.600$, $p=0.347$). Although the stimuli 172 ms – 262 ms caused a slight decrease in voiceless judgments, the consistency in reactions to the whole vowel duration span was not maintained and the subjects reported increased voiceless judgements for the longest, 292 ms vowel, stimulus.

5.1.5.3. Advanced Learners – results

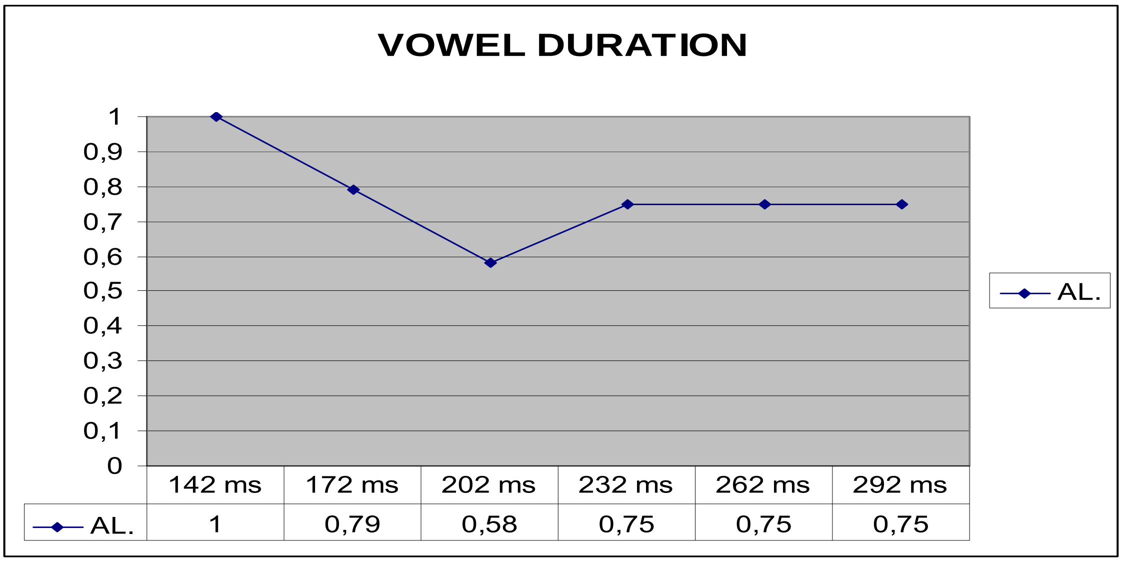


Figure 2.51. Recognition of a final sound in *theep* as voiceless /p/ across varying vowel duration by Advanced Learners

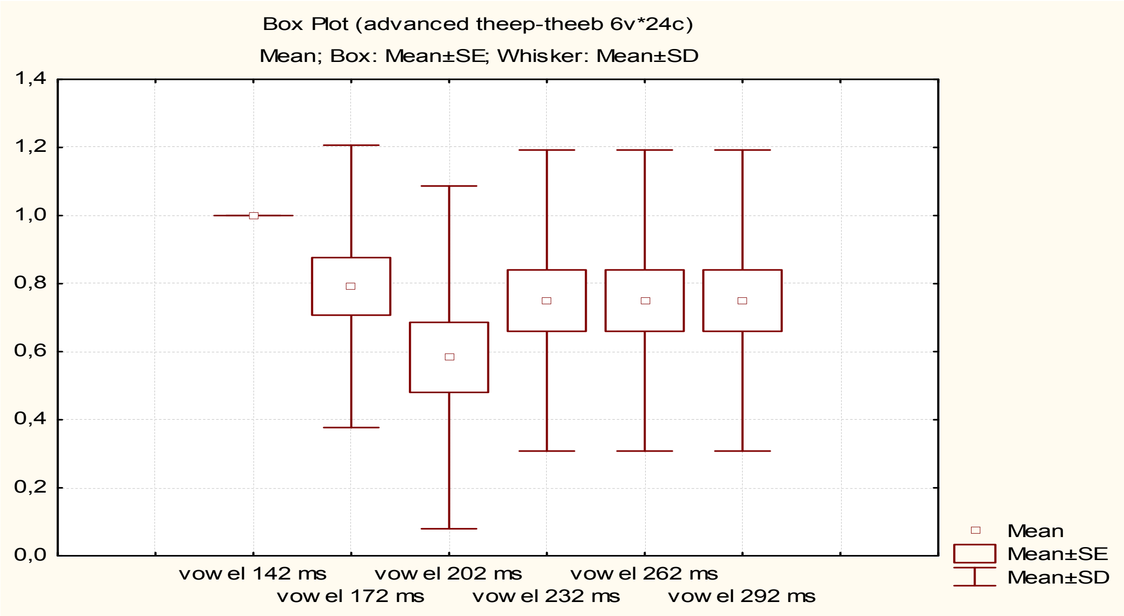


Figure 2.52. Box Plot: Recognition of a final sound in *theep* as voiceless /p/ across varying vowel duration by Advanced Learners

The stimulus effects in the AL group met the requirements of statistical significance (Q=12.560, p=0.027*), however, as in the case of the BLs, the ALs demonstrated a surprising lack of consistency. The stimuli 172 ms and 202 ms brought about a steady decrease in voiceless reports, but for the stimuli ranging from 232 ms to 292 ms, the voiceless reports levelled and the subjects reported hearing a voiced percept only 25% of the time.

2.1.5.4. Native Speakers – results

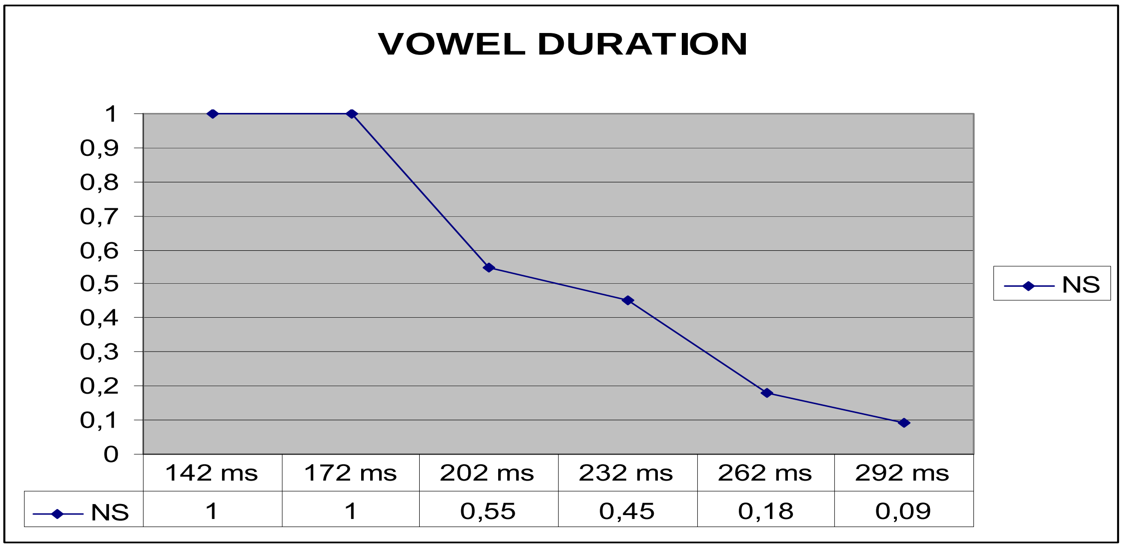


Figure 2.53. Recognition of a final sound in *theep* as voiceless /p/ across varying vowel duration by Native Speakers

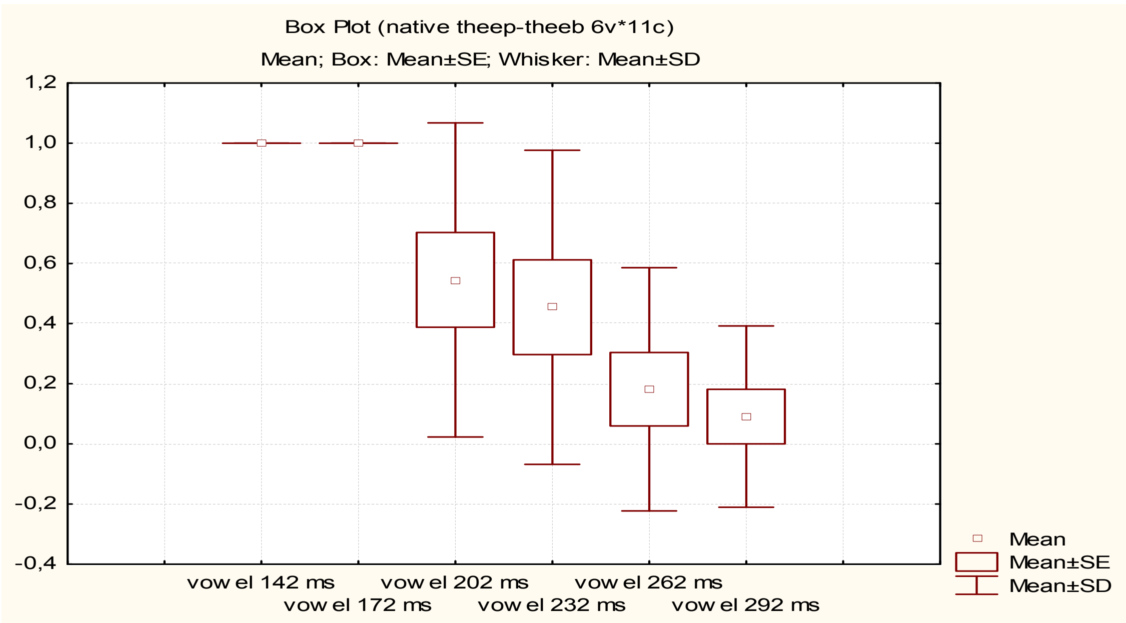


Figure 2.54. Box Plot: Recognition of a final sound in *theep* as voiceless /p/ across varying vowel duration by Native Speakers

The group of Native Speakers showed a highly significant reaction (Q=30.000, p=0.000**) to the stimulus effect by reporting a shift to a voiced stop as an effect of increasing vowel duration. The most drastic, almost categorical, shift into voiced judgments was caused by the 202 ms stimulus. The stimuli 232 ms to 292 ms brought about a subsequent increase in voiced reports, resulting in an almost complete shift for the 292 ms stimulus, which more than 90% of the subjects perceived as ending with a voiced stop.

2.1.6. Closure duration in stops

2.1.6.1. Stimuli

Unlike earlier studies on the perception of closure duration as a cue to the voicing contrast in stops (e.g. Lisker 1957, Port 1979), which concentrated on the intervocalic position, we chose to examine the stop closure duration in a word-final position, i.e. when followed only by the release burst. The choice was motivated by the findings which show that, in a VCV sequence, the lead-in vowel transitions carry considerable perceptual weight for the voiced-voiceless opposition in preceding stops, since their presence is a natural and unavoidable consequence of stop release in the intervocalic context (Raphael 1981, Hillenbrand et al. 1984, Hawkins 1999). As in the case of the vowel duration, we weakened the release burst by removing the highest energy excitation point, so that it would not serve as an overriding cue to the closure duration. Additionally, in order to single out the closure duration as the only cue to be judged by the subjects, we precluded the influence of the preceding vowel duration by averaging its length (Mean: 176 ms) between *thog* /θ①g/ (215 ms) and *thock* /θ①k/ (137 ms), so that it would be ambiguous to the voicing status of the following stop and would not perceptually override the closure duration.

Form a naturally obtained syllable *thock* /θ①k/ (125 ms closure duration of /k/), we generated 5 stimuli by decreasing the closure duration in 25 ms steps. As a result, we obtained the following stimuli:

1. *thock* /θ①k/, closure duration 125 ms
2. *thock* /θ①k/, closure duration 100 ms
3. *thock* /θ①k/, closure duration 75 ms
4. *thock* /θ①k/, closure duration 50 ms
5. *thock* /θ①k/, closure duration 25 ms

Figures 2.55. to 2.59. present waveforms and spectrograms of the 5 stimuli.

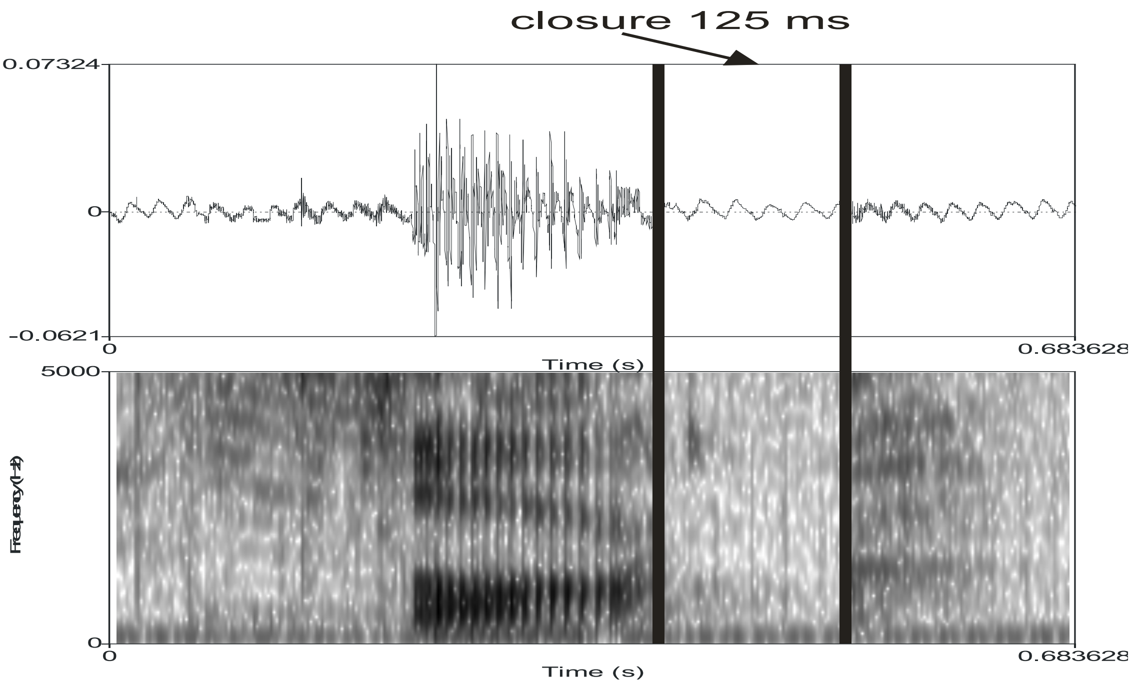


Figure 2.55. Waveform and spectrogram of syllable *thock*, closure duration 125 ms

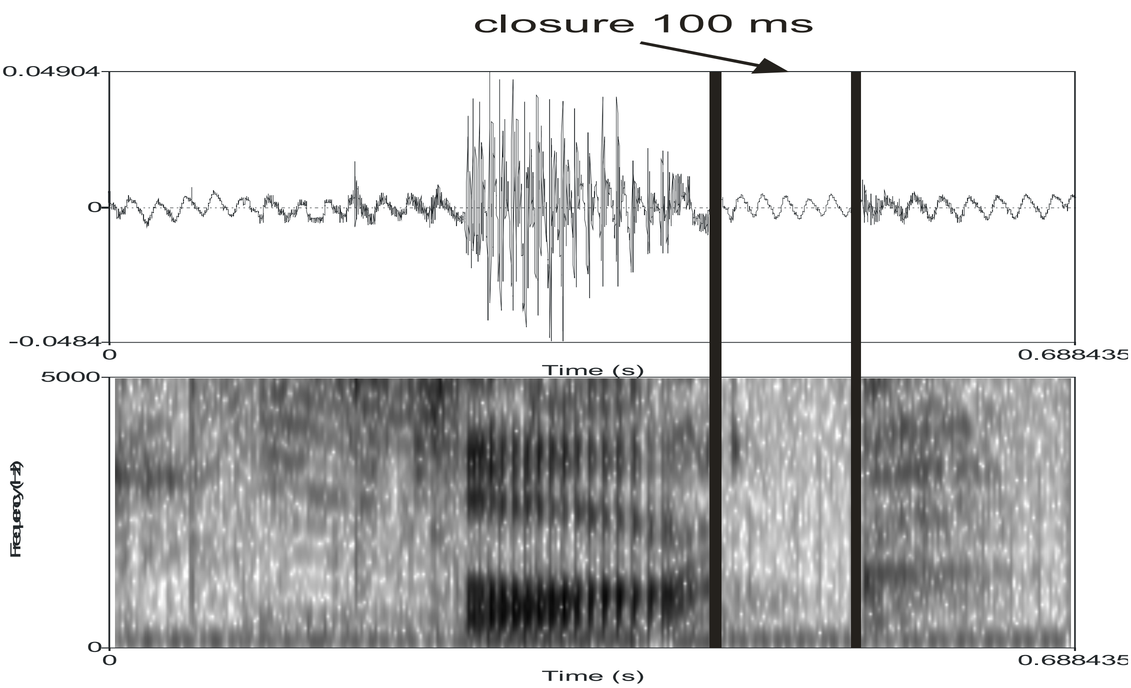


Figure 2.56. Waveform and spectrogram of syllable *thock*, closure duration 100 ms

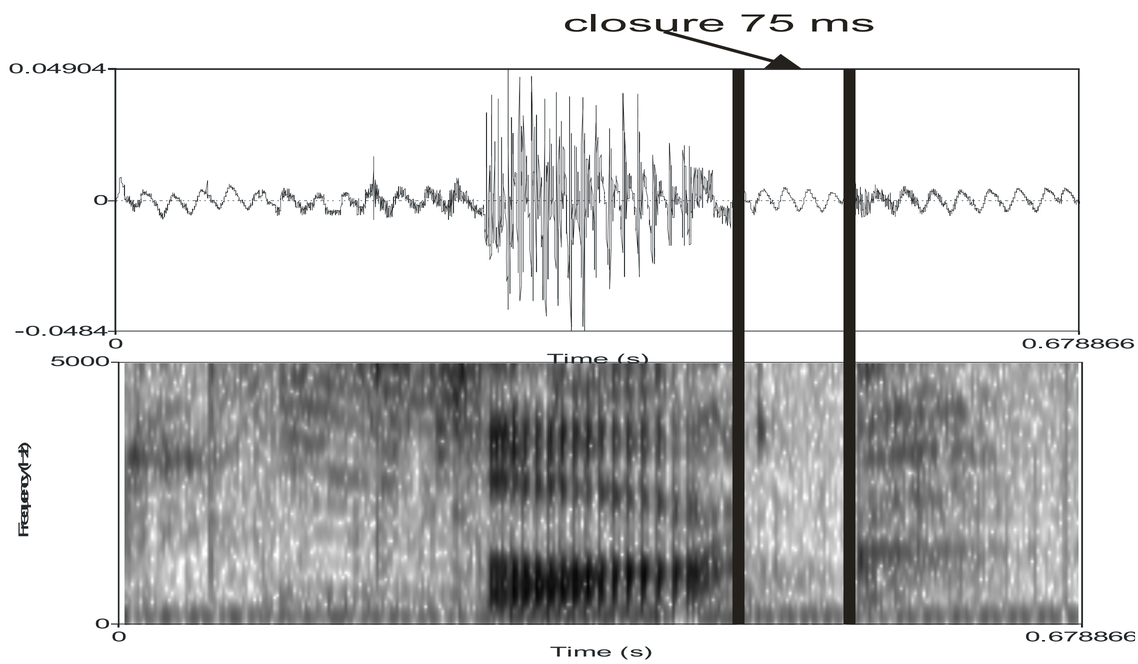


Figure 2.57. Waveform and spectrogram of syllable *thock*, closure duration 75 ms

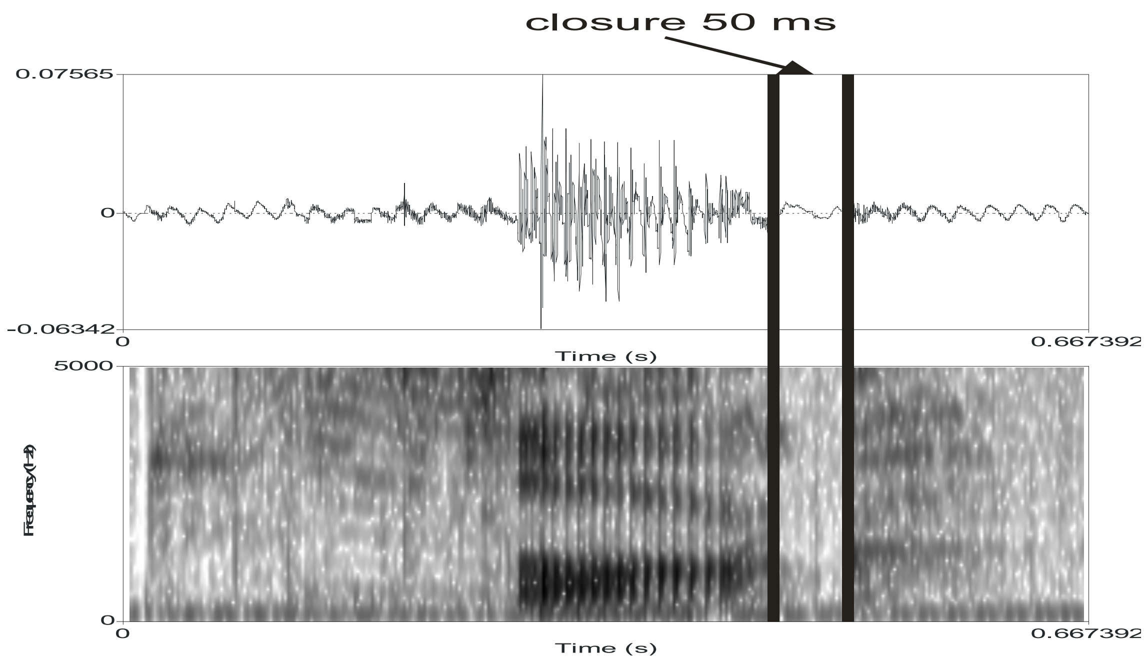


Figure 2.58. Waveform and spectrogram of syllable *thock*, closure duration 50 ms

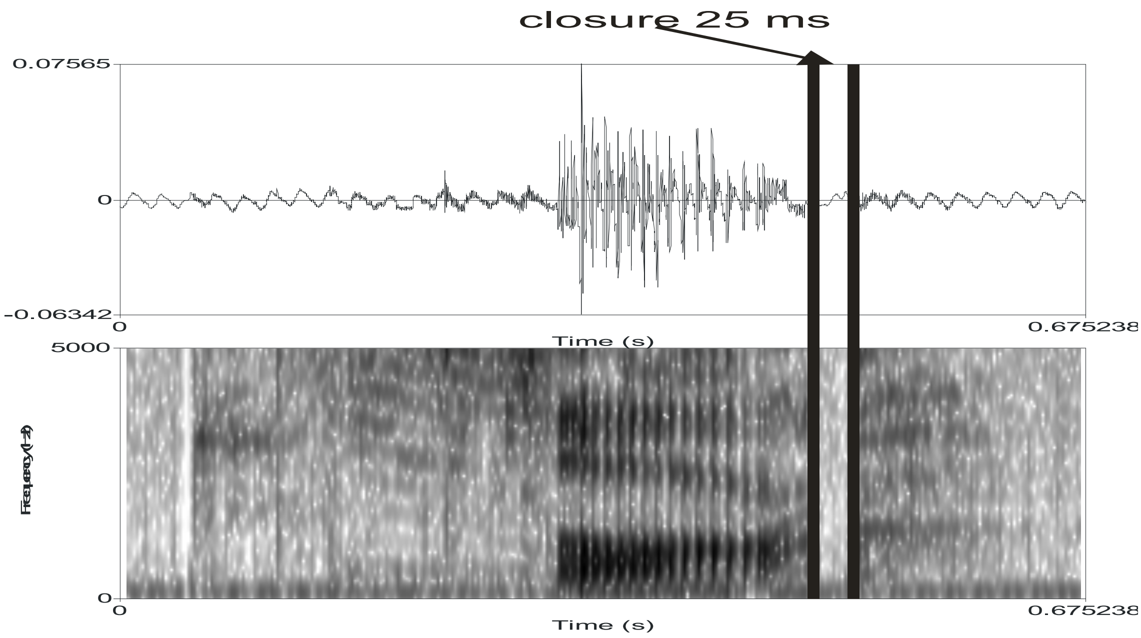


Figure 2.59. Waveform and spectrogram of syllable *thock*, closure duration 25 ms

2.1.6.2. Beginner Learners – results

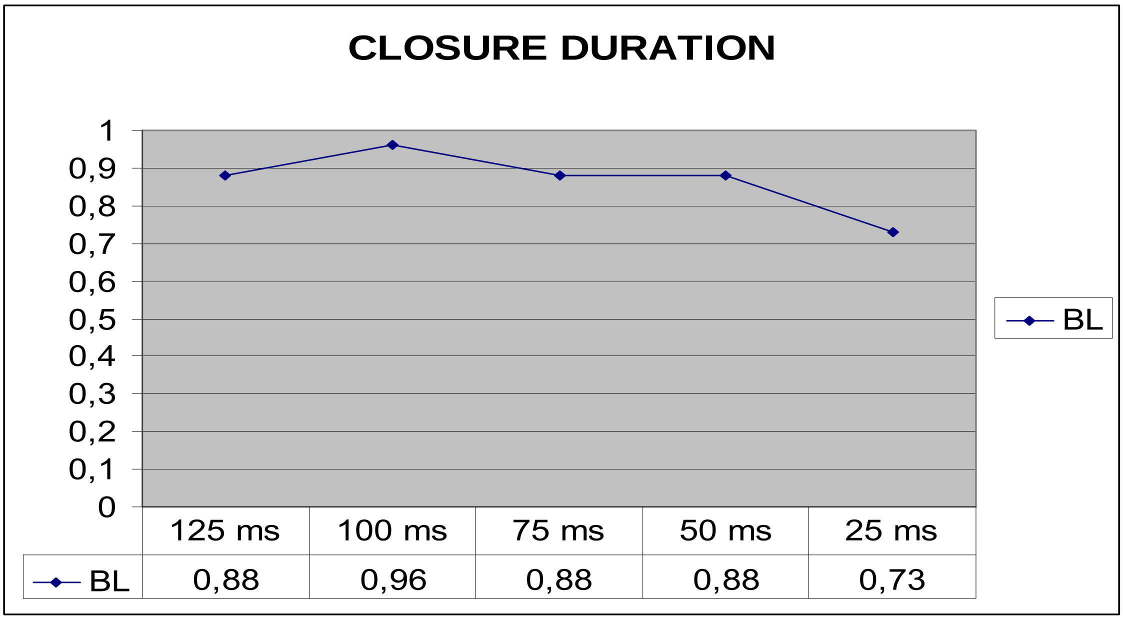


Figure 2.60. Recognition of a final sound in *thock* as voiceless /k/ across varying closure duration by Beginner Learners

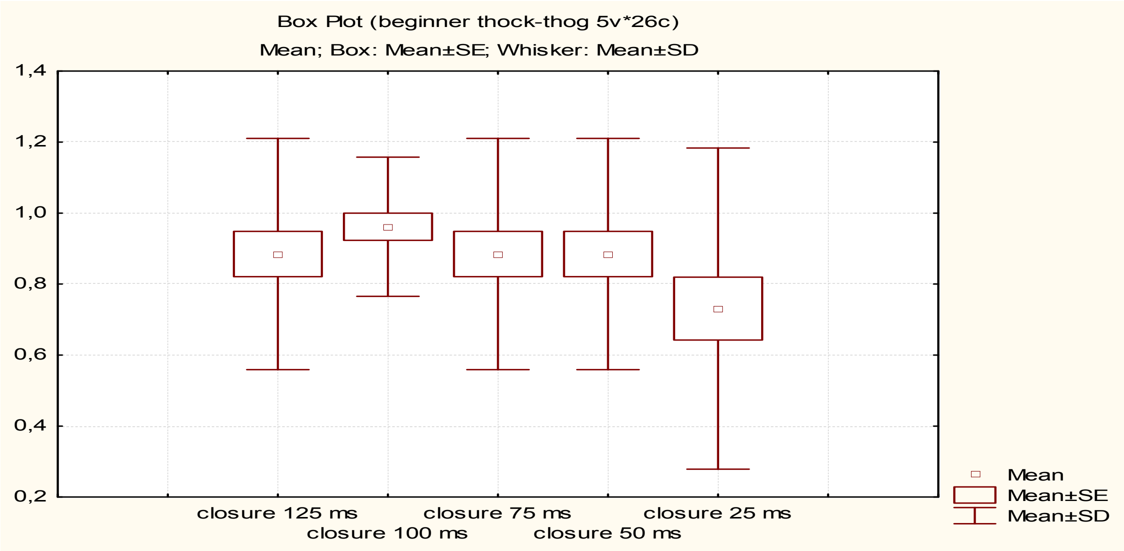


Figure 2.61. Box Plot: Recognition of a final sound in *thock* as voiceless /k/ across varying closure duration by Beginner Learners

The effect of closure duration was statistically significant in the BL group (Q=9.600, p=0.048*). The decreasing closure duration, however, did not result in a consistent shift from voiceless to voiced judgments. Moreover, it is interesting to note that the 100 ms stimulus even caused an unexpected increase in voiceless reports. The stimuli from 75 ms to 25 ms brought about a gradual but mild change towards voiced judgments.

2.1.6.3. Advanced Learners – results

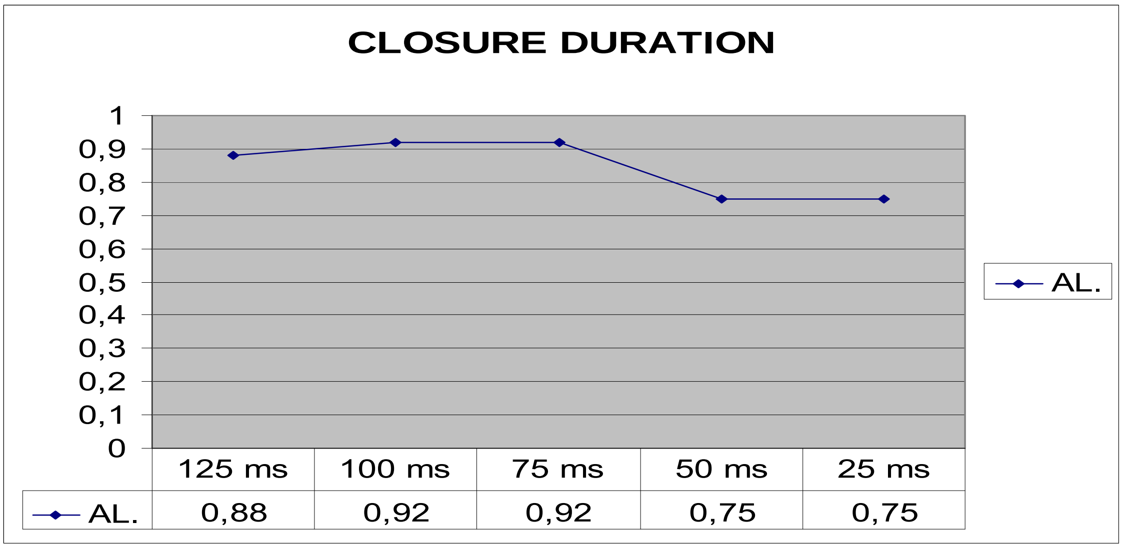


Figure 2.62. Recognition of a final sound in *thock* as voiceless /k/ across varying closure duration by Advanced Learners

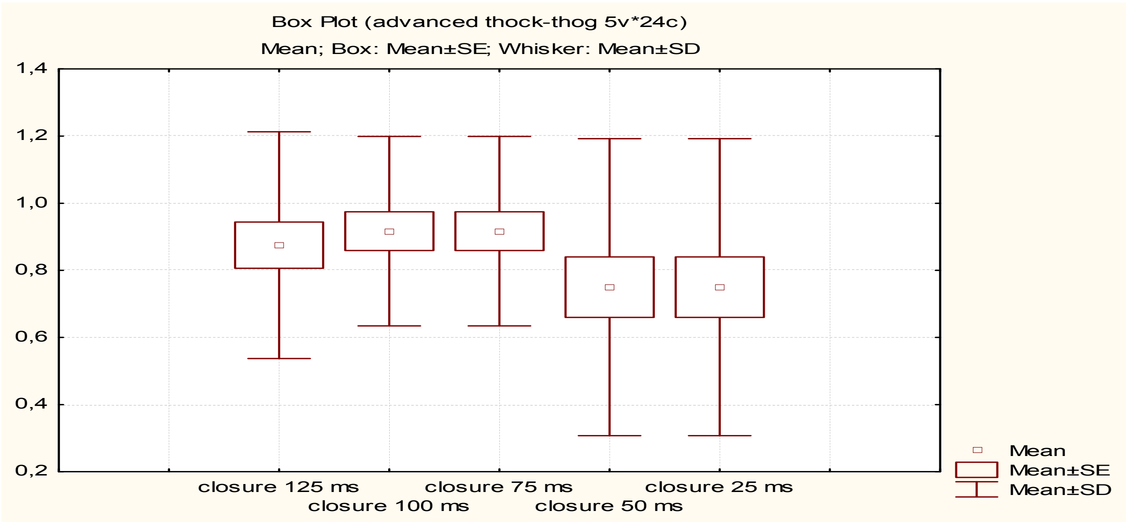


Figure 2.63. Box Plot: Recognition of a final sound in *thock* as voiceless /k/ across varying closure duration by Advanced Learners

For the AL group, the stimulus effects was not found to be statistically significant (Q=5.419, p=0.247). Nevertheless, what comes to the fore is, similar to the tendency found for the BLs, an unexpected increase in voiceless reports for the 100 ms stimulus. The only shift from voiceless to voiced judgments was observed for the 50 ms stimulus. However, since the subjects did not perform below a chance level, a random interference cannot be discounted.

2.1.6.4. Native Speakers – results

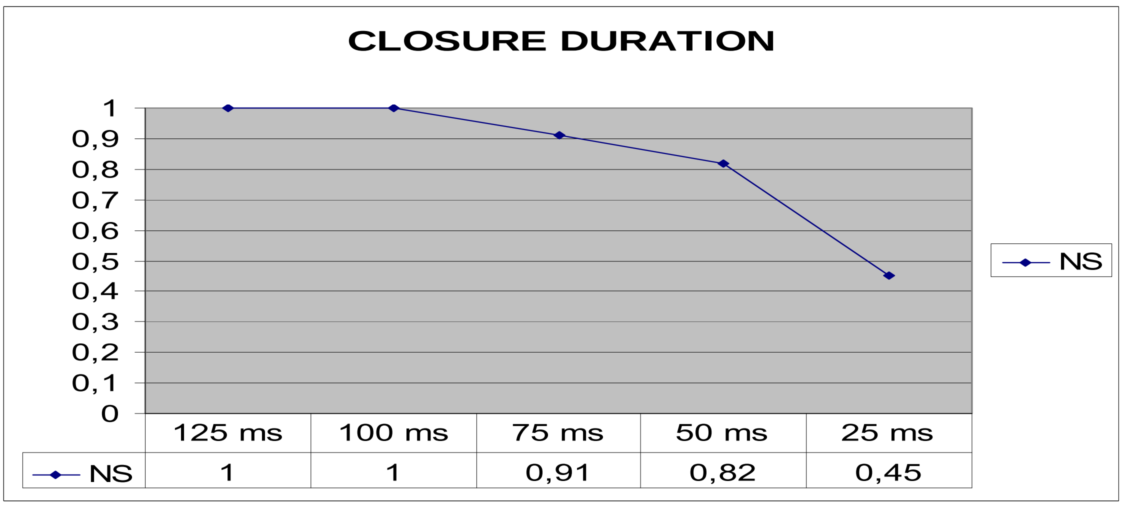


Figure 2.64. Recognition of a final sound in *thock* as voiceless /k/ across varying closure duration by Native Speakers

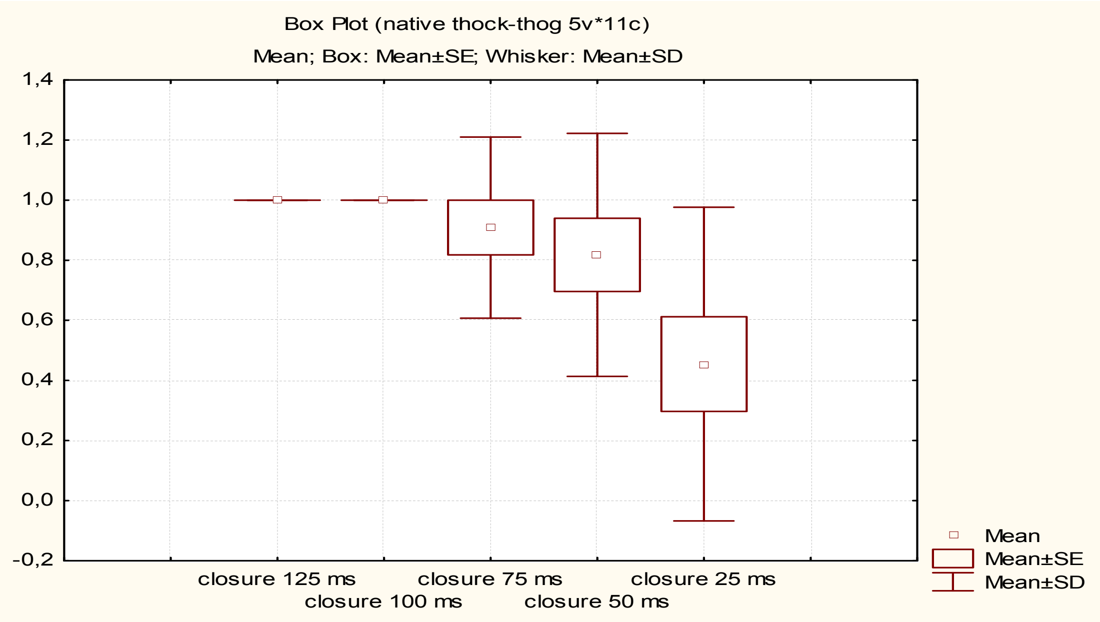


Figure 2.65. Box Plot: Recognition of a final sound in *thock* as voiceless /k/ across varying closure duration by Native Speakers

The group of Native Speakers showed a consistent and highly statistically significant reaction to the decreasing closure duration (Q=17.714, p=0.001**). Starting with the 75 ms stimulus, the subjects increased their voiced judgments. The most conspicuous effect of the decreasing closure duration was observed for the shortest, 25 ms stimulus, for which 55% of the subjects reported hearing a voiced stop.

2.1.7. Partially devoiced final fricatives

2.1.7.1. Stimuli

From naturally obtained syllables *heez* /hi:z/ with fully voiced final /z/ and *hees* /hi:s/ with fully voiceless final /s/, we generated 4 stimuli varying in devoicing degree. In order to exclude the influence of the preceding vowel duration, we averaged its length (Mean: 251 ms) between the values measured for *heez* /hi:z/ (311 ms) and *hees* /hi:s/ (190 ms), so that it would be ambiguous to the voicing status of the following fricative and would not serve as an overriding cue to the frication voicing. Consequently, we obtained the following 4 stimuli:

1. *heez* /hi:z/, /z/ 215 ms voiced
2. *heez* /hi:z/, /z/ 165 ms voiced + 50 ms voiceless
3. *heez* /hi:z/, /z/ 115 ms voiced + 100 ms voiceless
4. *heez* /hi:z/, /z/ 65 ms voiced + 150 ms voiceless

Figures 2.66. to 2.69. show waveforms and spectrograms of all the stimuli.

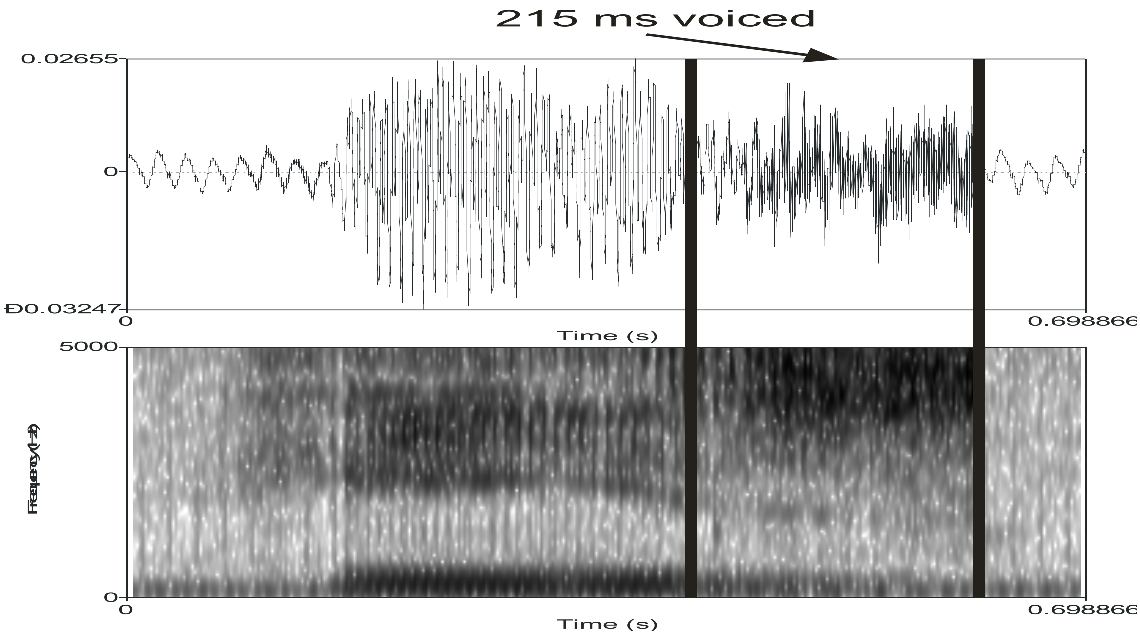


Figure 2.66. Waveform and spectrogram of syllable *heez*, /z/ 215 ms voiced

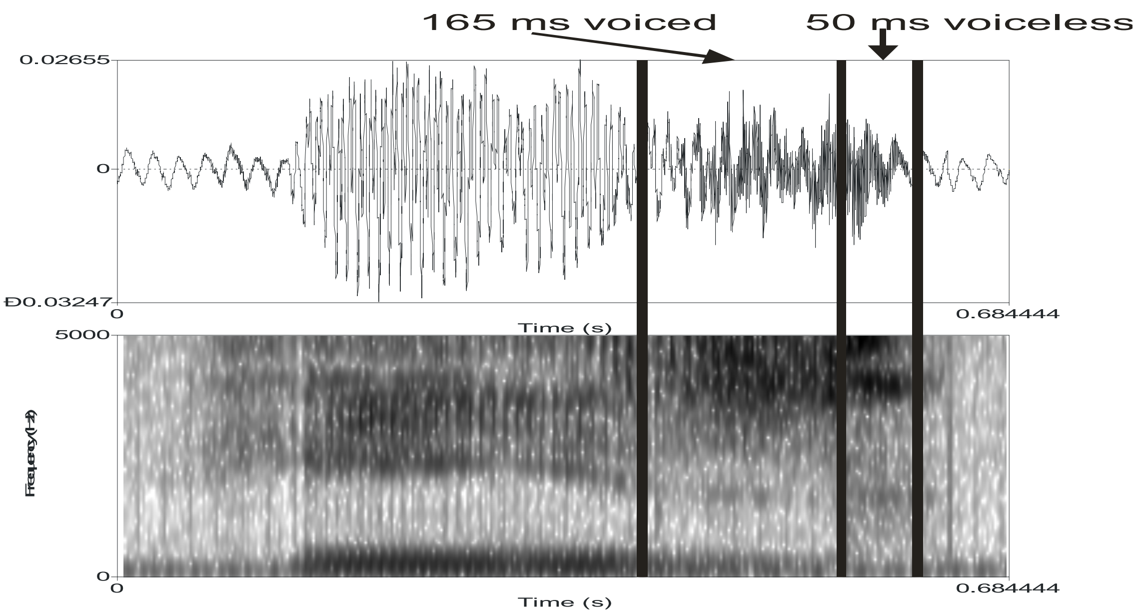


Figure 2.67. Waveform and spectrogram of syllable *heez*, /z/ 165 ms voiced + 50 ms voiceless

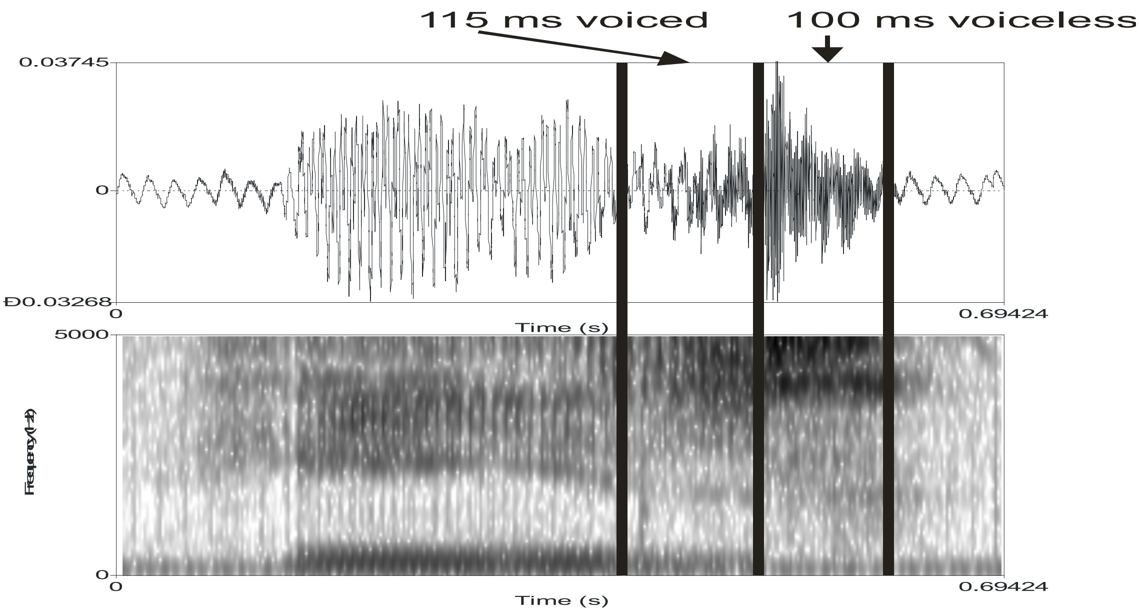


Figure 2.68. Waveform and spectrogram of syllable *heez*, /z/ 115 ms voiced + 100 ms voiceless

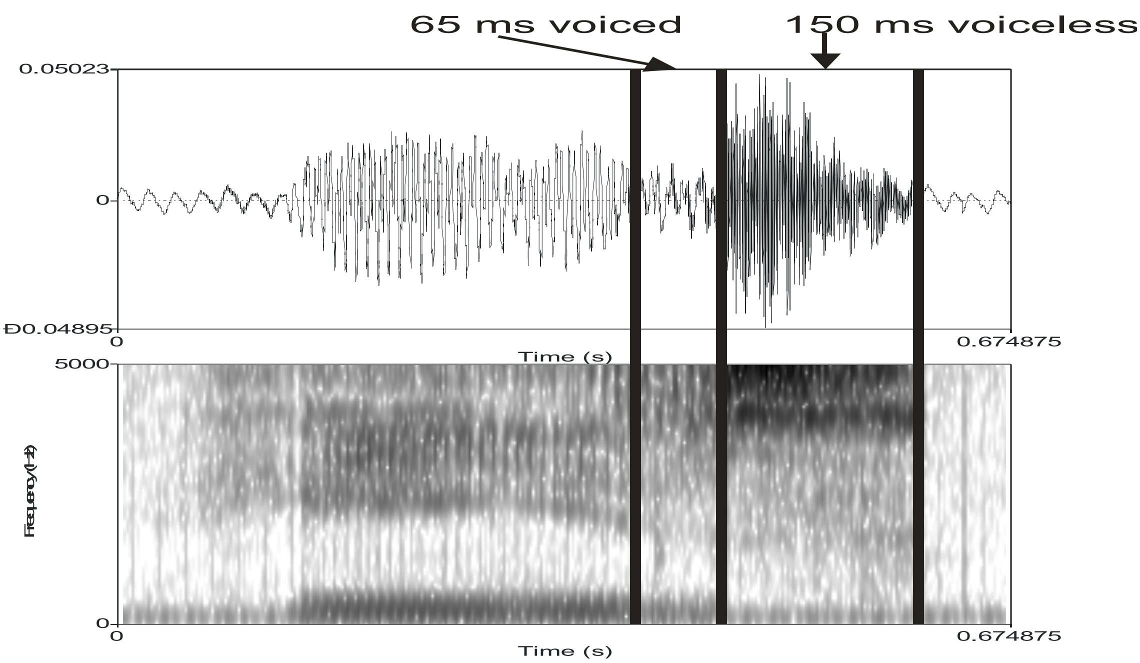


Figure 2.69. Waveform and spectrogram of syllable *heez*, /z/ 65 ms voiced + 150 ms voiceless

2.1.7.2. Beginner Learners – results

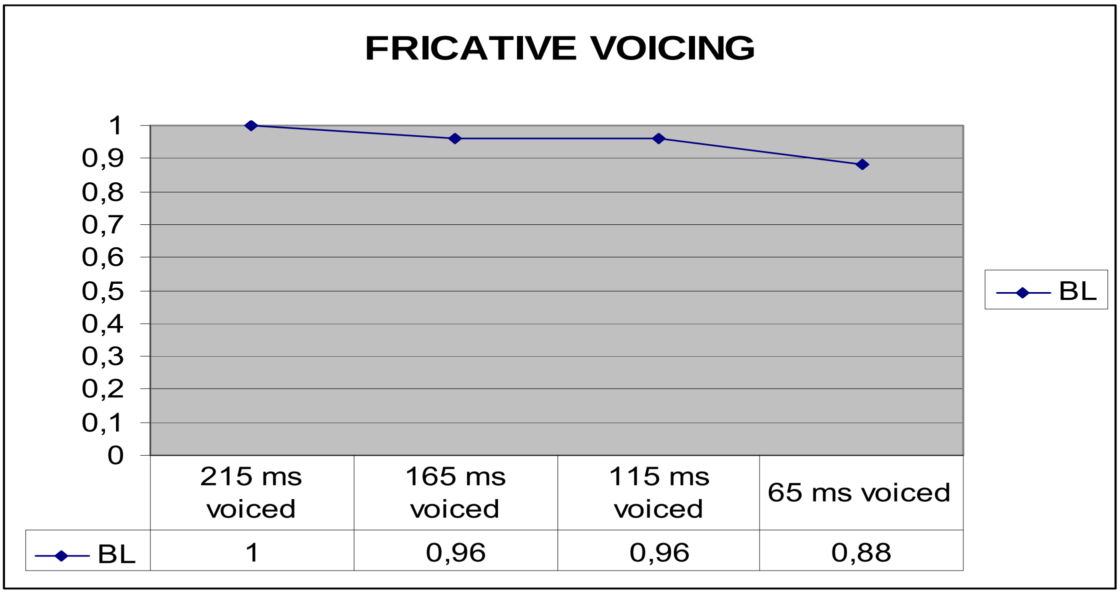


Figure 2.70. Recognition of a final sound in *heez* as voiced /z/ across varying degrees of devoicing by Beginner Learners

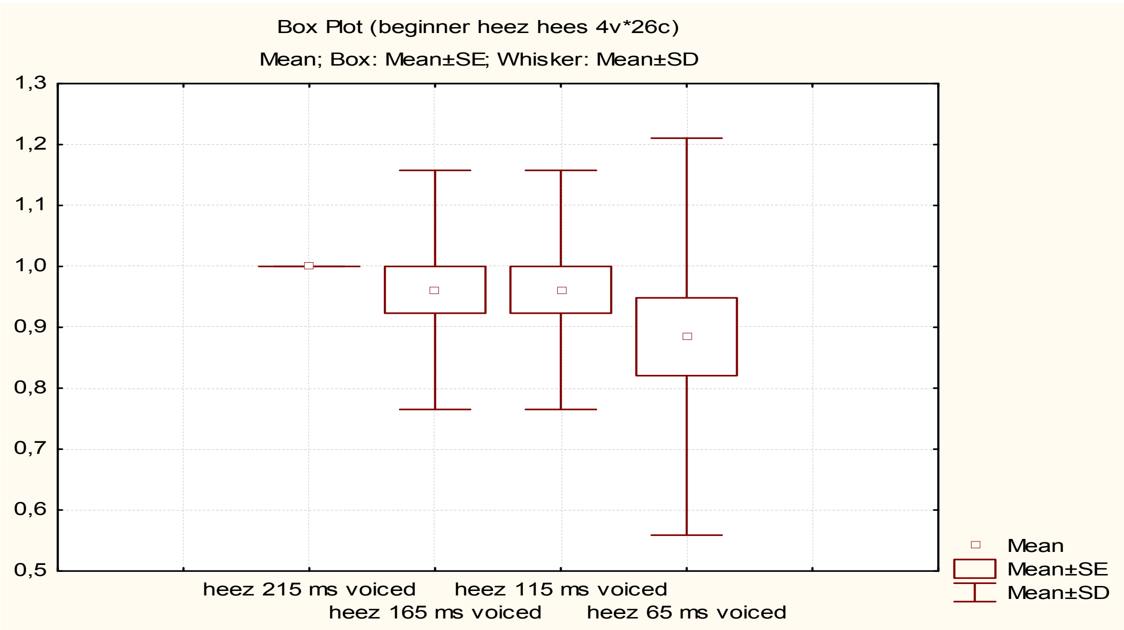


Figure 2.71. Box Plot: Recognition of a final sound in *heez* as voiced /z/ across varying degrees of devoicing by Beginner Learners

The effect of partial devoicing of the final fricative was not statistically significant for the BLs (Q=4.385, p=0.223). Even when the fricative was voiceless in final 150 ms of its portion, the subjects recognised it as voiced 88% of the time.

2.1.7.3. Advanced Learners – results

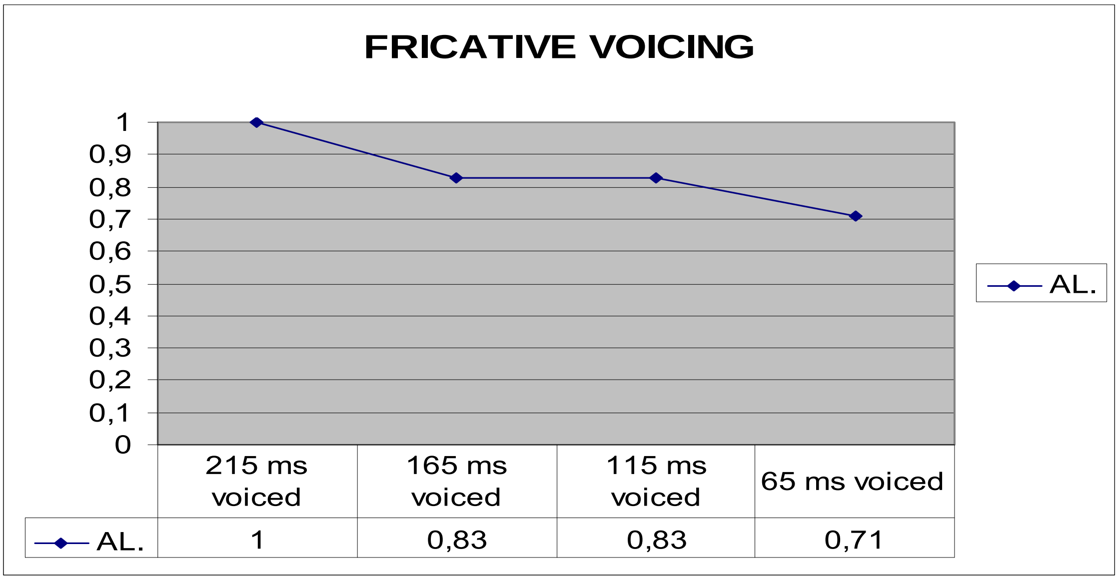


Figure 2.72. Recognition of a final sound in *heez* as voiced /z/ across varying degrees of devoicing by Advanced Learners

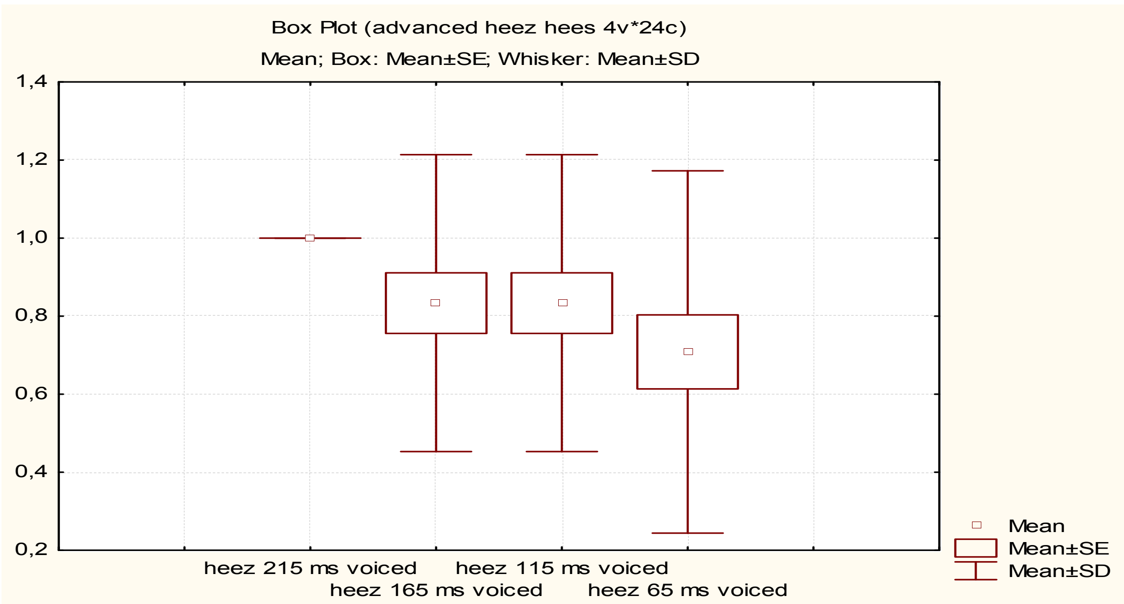
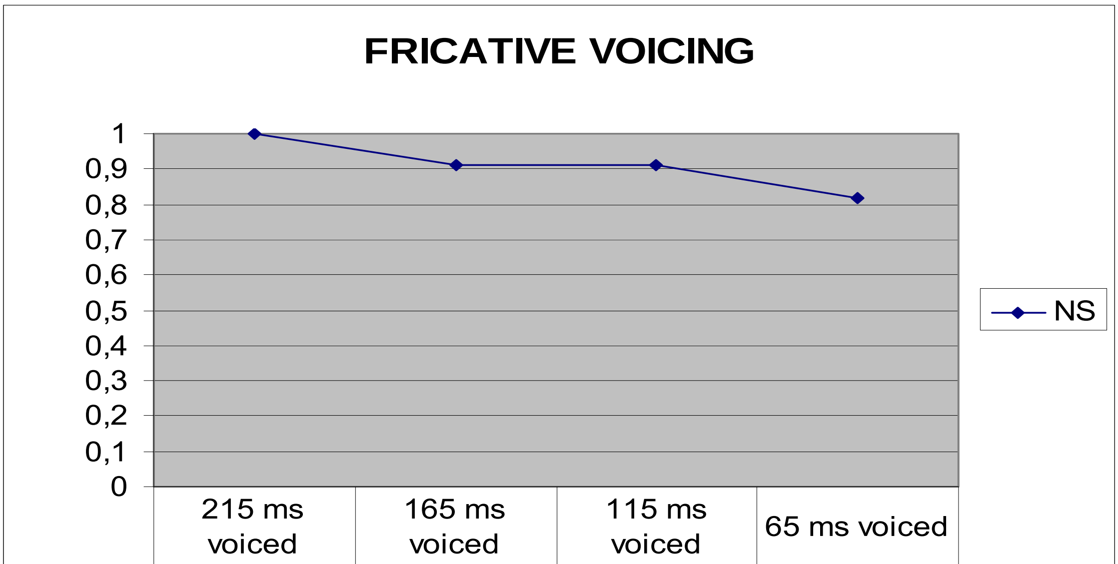


Figure 2.73. Box Plot: Recognition of a final sound in *heez* as voiced /z/ across varying degrees of devoicing by Advanced Learners

The effect of partial devoicing of the final fricative was found to have a statistically significant influence on the ALs' voicing reports. (Q=9.580, p=0.022*). For the 165 ms voiced stimulus, the subjects reported hearing a voiceless segment 17% of the time. Further decrease in voiced reports was observed for the 65 ms voiced stimulus – the ALs perceived it as voiceless 29% of the time.

2.1.7.4. Native Speakers – results



	215 ms voiced	165 ms voiced	115 ms voiced	65 ms voiced
NS	1	0,91	0,91	0,82

Figure 2.74. Recognition of a final sound in *heez* as voiced /z/ across varying degrees of devoicing by Native Speakers

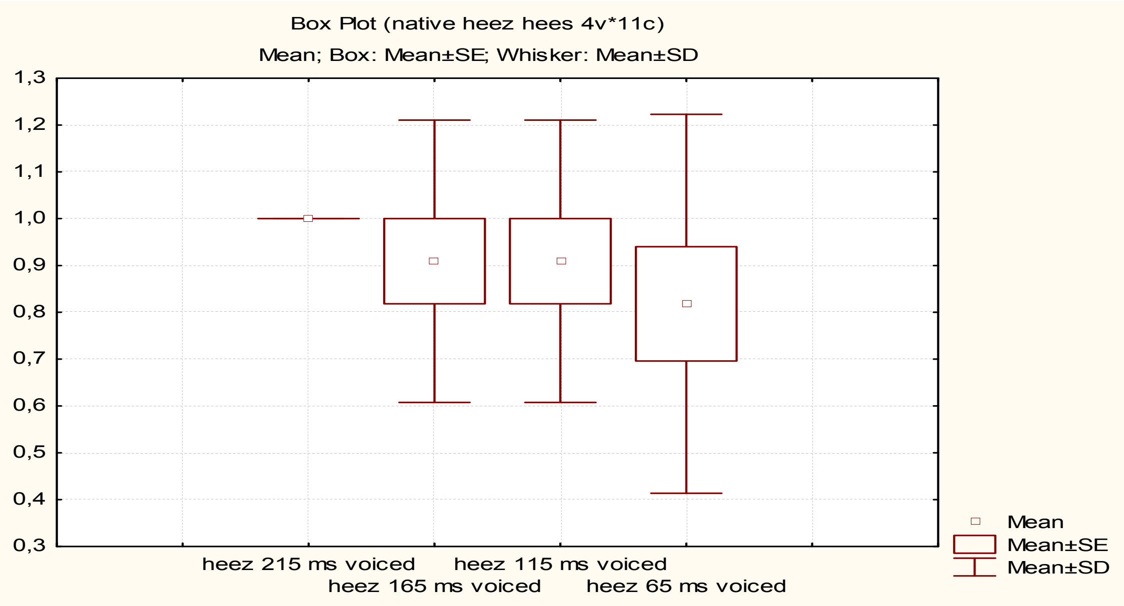


Figure 2.75. Box Plot: Recognition of a final sound in *heez* as voiced /z/ across varying degrees of devoicing by Native Speakers

As in the case of the BL group, and unlike the AL group, the partial devoicing effect was not statistically significant for the NSs (Q=2.400, p=0.494). However, similar to the results obtained for the previous two groups, one can observe a decreasing tendency in voiced percepts along an expanding voiceless period in the tested segment.

2.1.8. Closure duration in affricates

2.1.8.1. Stimuli

When preparing the stimuli for the experiments, we observed an intriguing regularity that the reduction of the closure duration of the final affricate in syllable *heedge* /hi:dʒ/ gave a strong auditory impression of voicelessness in this segment. Not only are we unacquainted with any research that would confirm this regularity, but it is also opposite to the tendency found for stops, which are perceived as voiced, not voiceless, with decreasing closure length (see Part 2, Section 5 and Part 3, Section 2.1.6.). Accordingly, we used a naturally obtained syllable *heedge* /hi:dʒ/ with 147 ms of its closure and reduced it by half, hence generating a stimulus *heedge* /hi:dʒ/ with 73 ms of closure duration. The preceding vowel length was not reduced – if the shortening of closure duration is a sufficiently strong cue to the voicelessness of the affricate, it was expected to override the vowel duration which was typical for following voiced affricates. The subjects listened to the following stimuli:

1. *heedge* /hi:d│/, / d│/ closure duration 147 ms
2. *heedge* /hi:d│/, / d│/ closure duration 73 ms

Figures 2.76. and 2.77. show waveforms and spectrograms of both stimuli.

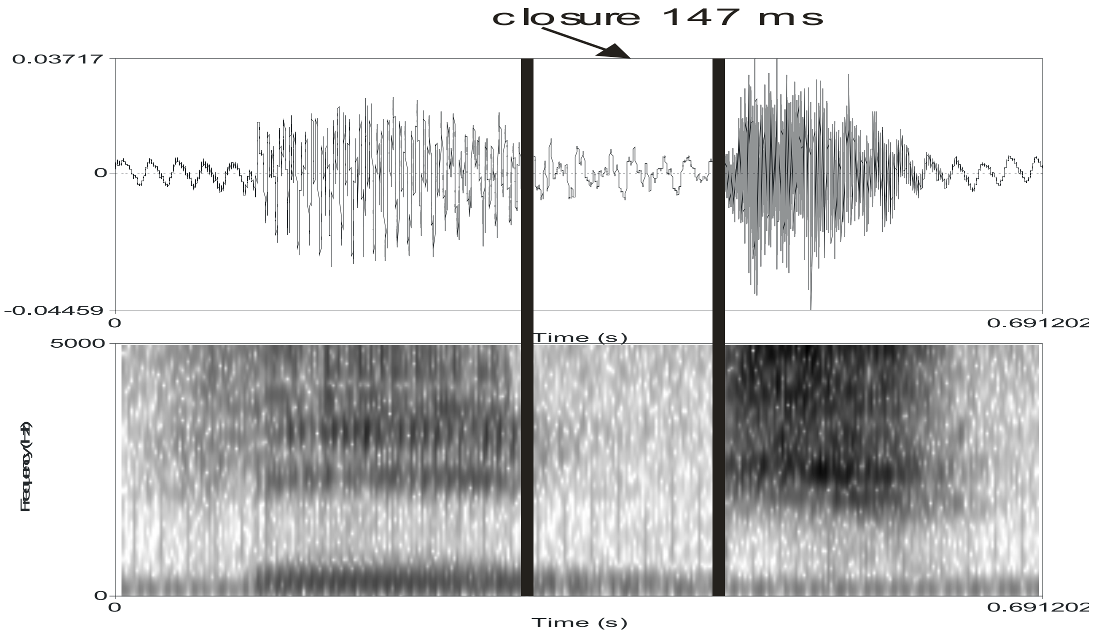


Figure 2.76. Waveform and spectrogram of syllable *heedge*, closure duration 147 ms

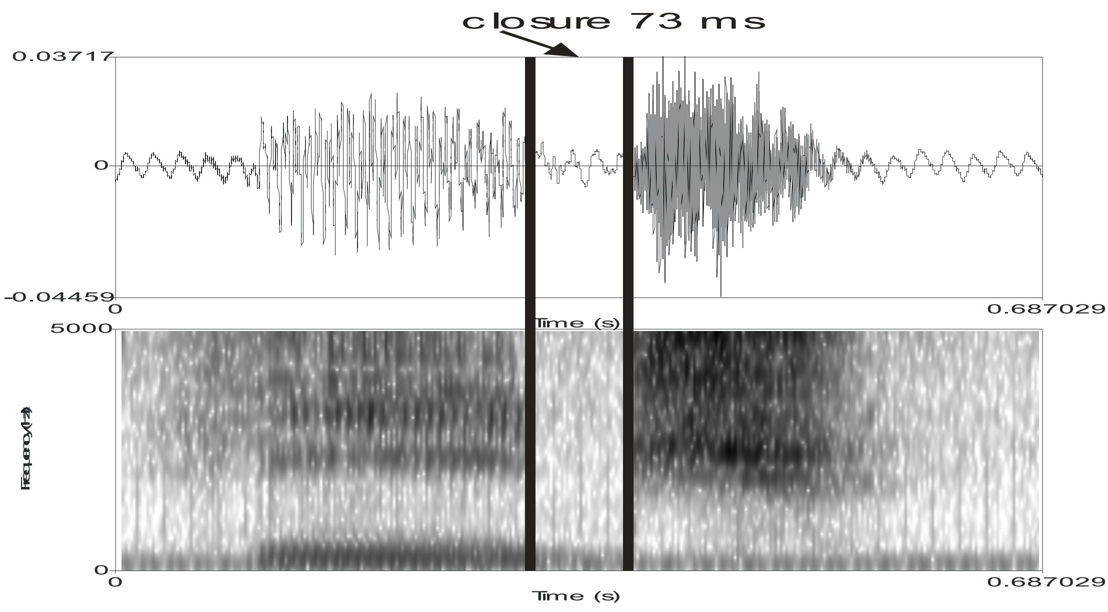


Figure 2.77. Waveform and spectrogram of syllable *heedge*, closure duration 73 ms

2.1.8.2. Beginner Learners – results

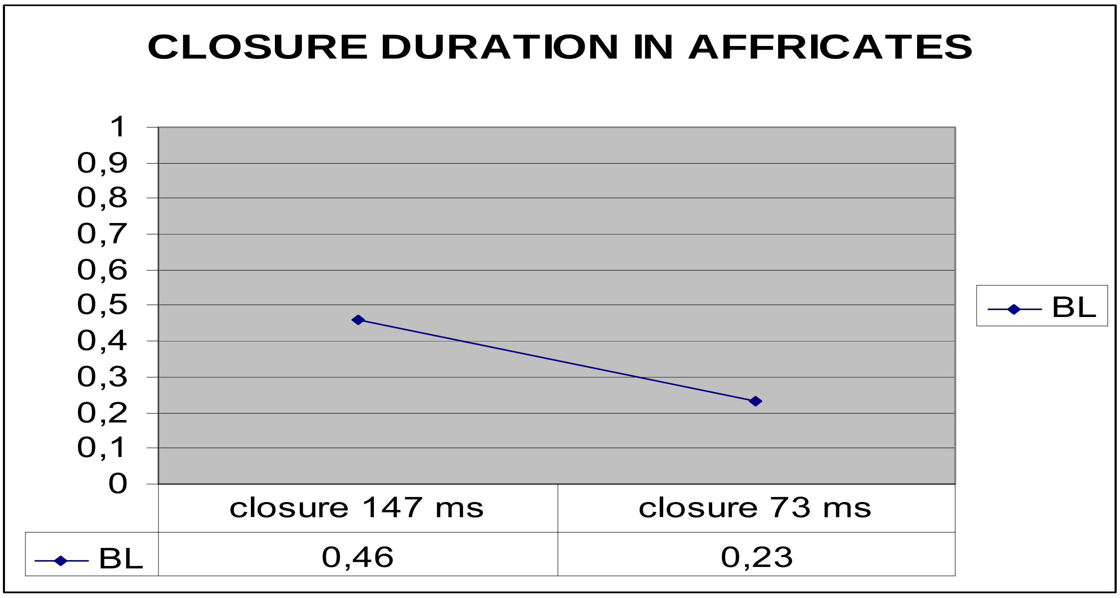


Figure 2.78. Recognition of a final sound in *heedge* as voiced /d|/ across varying closure duration by Beginner Learners

Figure 2.79. Box Plot: Recognition of a final sound in *heedge* as voiced /d|/ across varying closure duration by Beginner Learners

Although the stimulus effect was not significant in the BL group (McNemar *Chi*=1.39, p=0.239), the tendency to judge the affricate with shortened closure as voiceless is shown by the decreasing number of voiced reports. It is interesting to note that only

46% of the BL subjects heard a naturally obtained syllable *heedge* /hi:dʒ/ as ending with a voiced affricate. This fact seems to underlie statistical insignificance of the shortening effect.

2.1.8.3. Advanced Learners – results

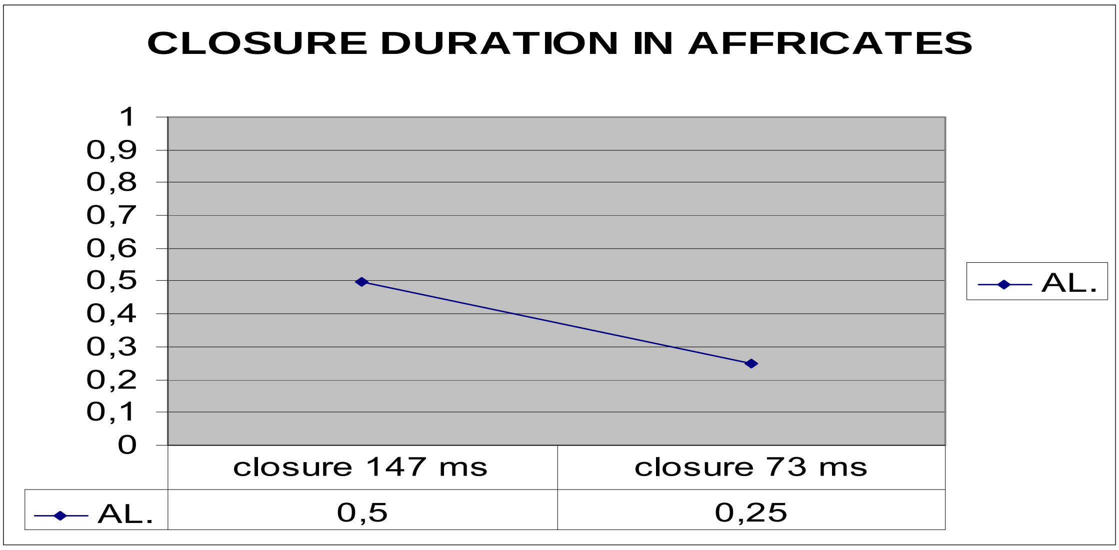


Figure 2.80. Recognition of a final sound in *heedge* as voiced /dʒ/ across varying closure duration by Advanced Learners

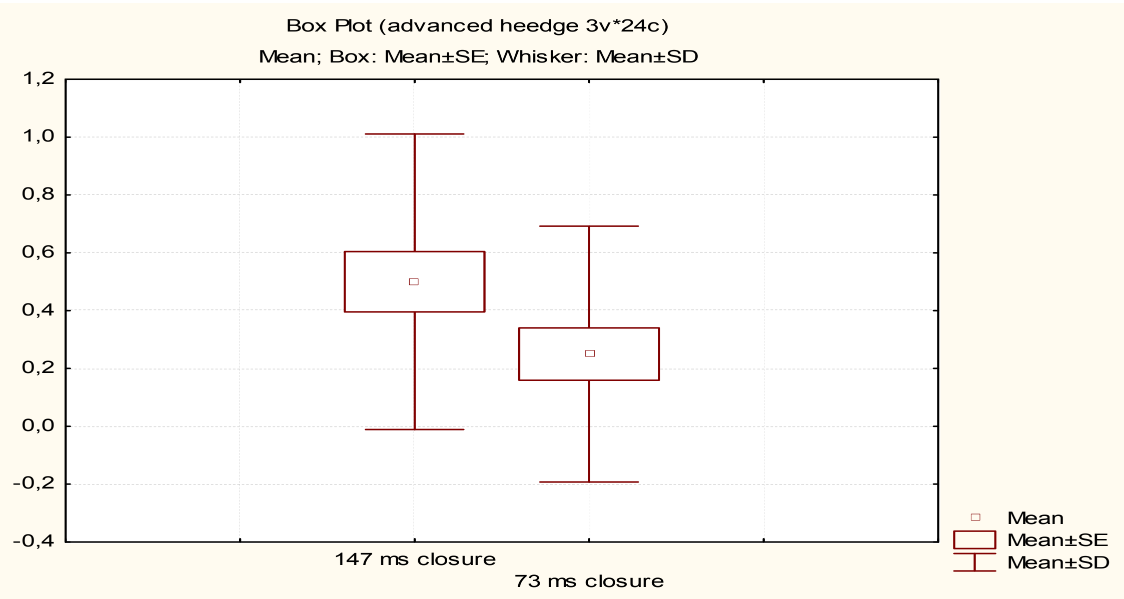


Figure 2.81. Box Plot: Recognition of a final sound in *heedge* as voiced /dʒ/ across varying closure duration by Advanced Learners

Like in the BL group, the ALs reacted to the shortened closure duration by reporting hearing more voiced segments, however the results did not meet the criteria of statistical significance (McNemar *Chi*=2.12, p=0.146). It is again striking that only 50% of the AL subjects judged the natural syllable *heedge* /hi:d|/ as ending with a voiced affricate.

2.1.8.4. Native Speakers – results

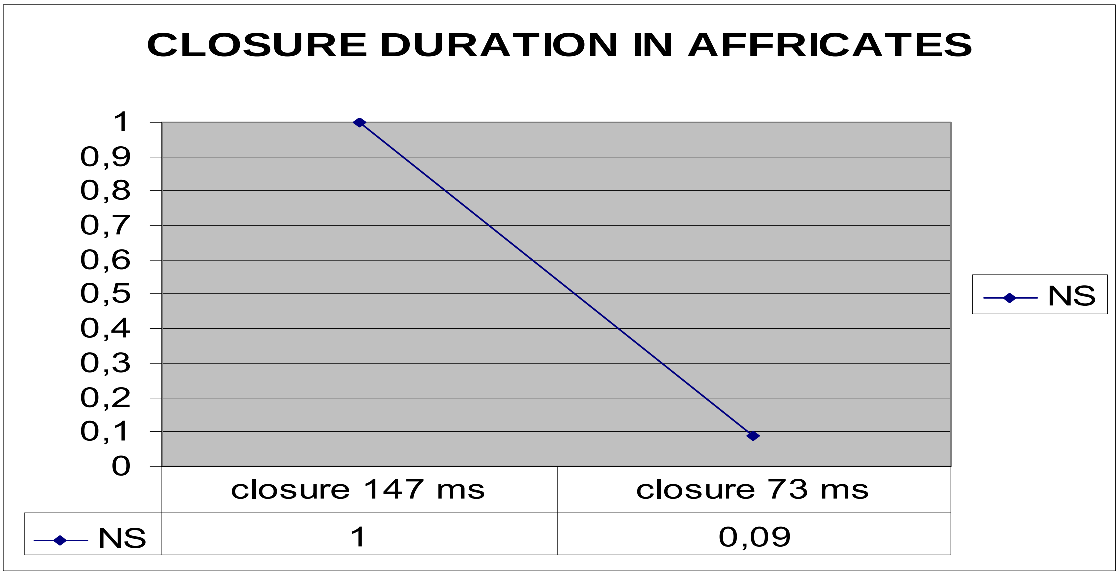


Figure 2.82. Recognition of a final sound in *heedge* as voiced /d|/ across varying closure duration by Native Speakers

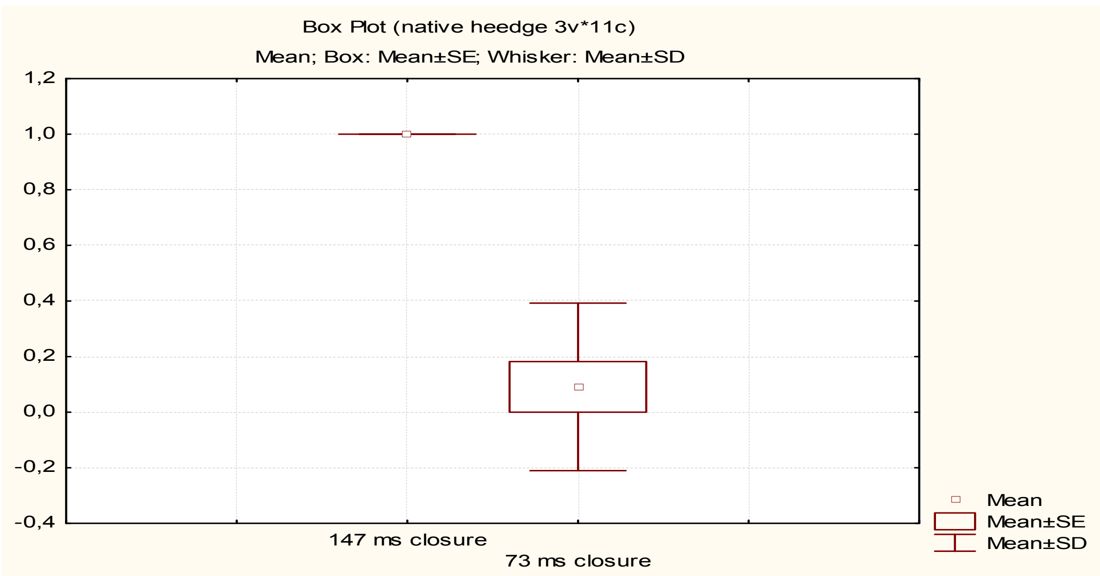


Figure 2.83. Box Plot: Recognition of a final sound in *heedge* as voiced /d|/ across

varying closure duration by Native Speakers

The NSs showed a strong shift in the voicing judgments along the shortened closure duration. The stimulus effect was highly significant (McNemar *Chi*=6.75, p=0.009**). For the reduced closure, more than 90% of the NSs reported hearing a voiceless percept. It should be emphasised that a shift occurred even though the vowel duration was indicative of the voiced status of the following affricate.

2.2. Polish stimuli

2.2.1. 0 ms VOT

2.2.1.1. Stimuli

We manipulated a Polish syllable *pir* /pir/ and obtained a syllable beginning with 0 ms VOT /p/. The stimulus was presented to the BL and AL groups in order to ascertain whether the two groups would categorise the 0 ms VOT point differently. The rationale lies in the fact that if L2 (English) perceptual system interfered with L1 (Polish) system in the ALs, they would report hearing more voiced stops than the BLs. The 0 ms VOT point is covered by a voiced label in English, whereas in Polish it is predominantly referred to as voiceless (see Part 2, Section 3).

Figure 2.84. shows a waveform and spectrogram of the stimulus.

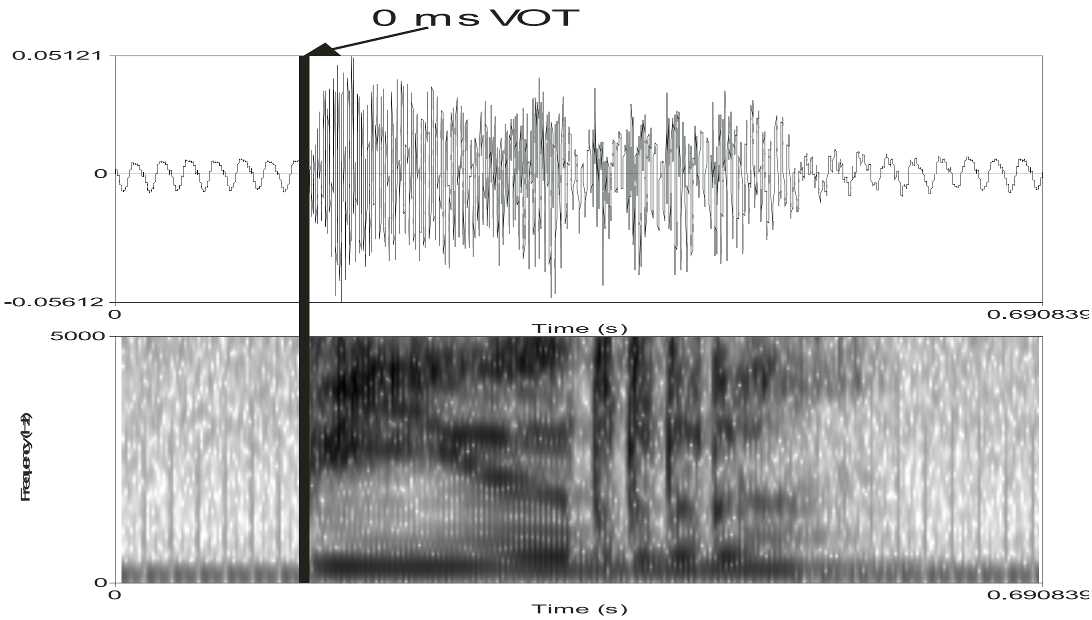


Figure 2.84. Waveform and spectrogram of syllable *pir*, /p/ 0 ms VOT

2.2.1.2. Results

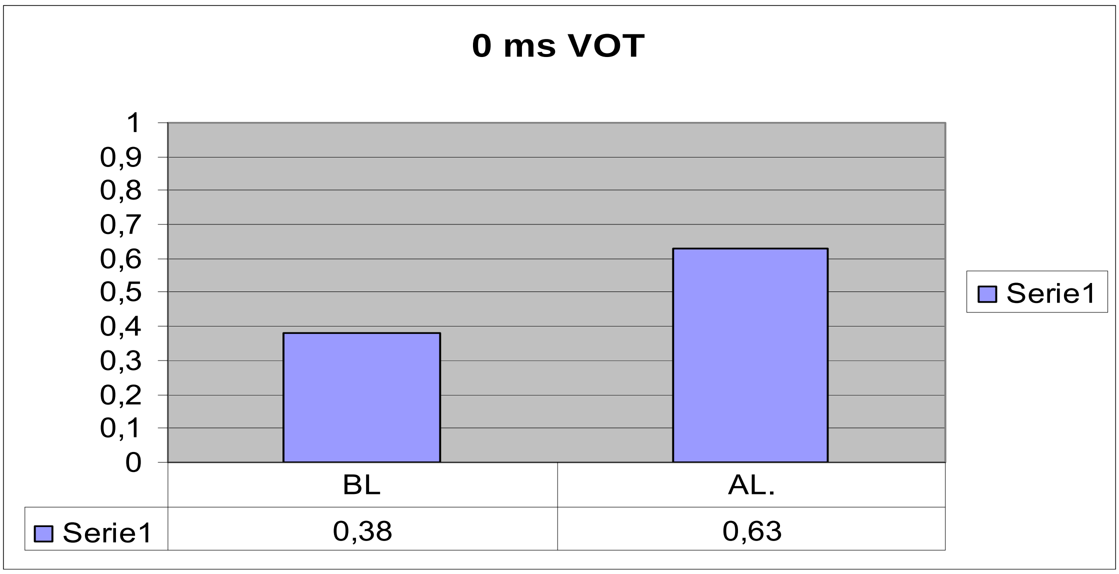


Figure 2.85. Recognition of an initial sound in *pir* with 0 ms VOT as /p/ by Beginner and Advanced Learners

The results show that the BLs were more willing in categorise the 0 ms VOT value as voiced than the AL subjects. It runs counter to the expectations that these would be Advance Learners who would have more 'English-like' VOT boundaries in Polish. The ALs, and not the BLs, were hypothesised to recognise low VOT values as voiced. Since the results were not statistically significant (Yates corrected *Chi*=2.00, p=0.157), they show that Advanced Learners are definitely not more willing to categorise the 0 ms VOT value as voiced. Accordingly, in this case, any L2 interference can be precluded.

2.2.2. Voicing lead versus voicing lag

2.2.2.1. Stimuli

From recorded syllables *kir* /kir/ and *gir* /gir/, we created a hybrid syllable *gir* /gir/ with -64 ms prevoicing and additionally appended +75 ms VOT. There are no stops in the world's languages that are composed of both a voicing lead and voicing lag, hence the term 'hybrid'. The subjects were expected to attend to different types of information in the signal. It was expected that, if the L2-L1 transfer was at play, the ALs would concentrate on the positive values of VOT, ignoring the voicing lead, and thus recognise the segment as voiceless. On the other hand, the BLs were hypothesised to be more sensitive to the voicing lead period.

Figure 2.86. shows a waveform and spectrogram of the stimulus.

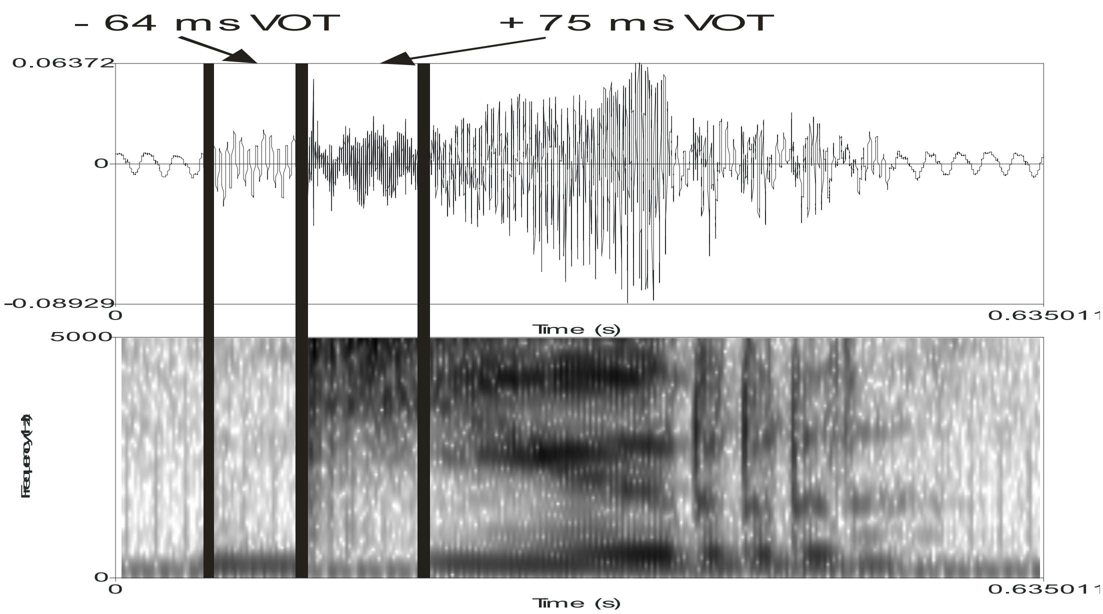


Figure 2.86. Waveform and spectrogram of hybrid syllable *gir*, /g/ -64 ms VOT voicing lead and +75 ms VOT voicing lag

2.2.2.2. Results

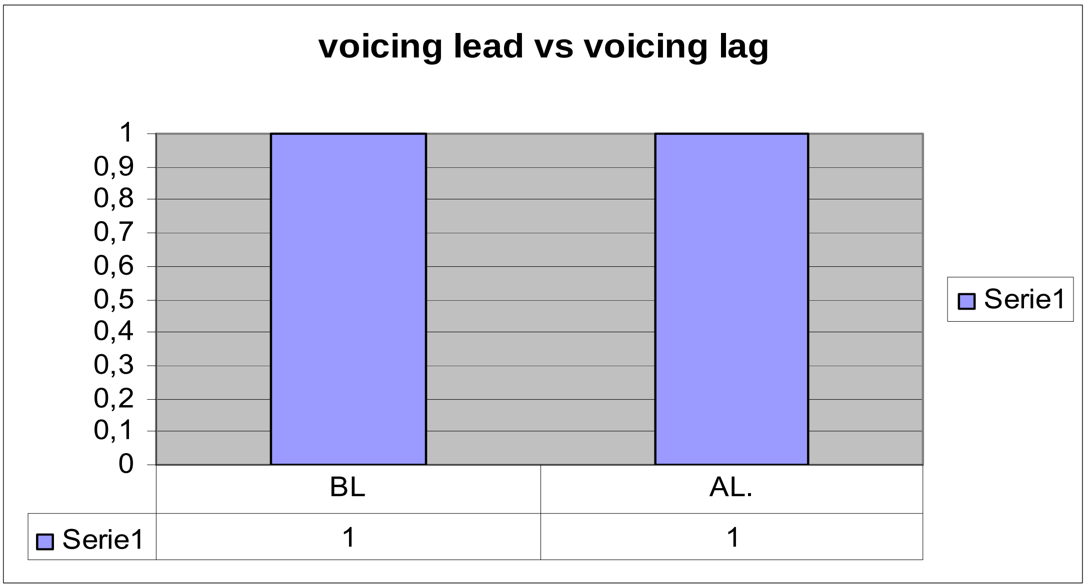


Figure 2.87. Recognition of an initial sound in *gir* with -64 ms voicing lead and +75 ms voicing lag as /g/ by Beginner and Advanced Learners

There were no differences in the perception of the initial sound between the two groups. Both the BL and AL subjects attended to the voicing lead period and reported hearing a voiced segment. The voicing lag appears to have been completely ignored.

2.2.3. Frication duration in initial fricatives

2.2.3.1. Stimuli

By manipulating a recorded Polish syllable *fos* /fos/ with 162 ms of initial fricative duration, we reduced the duration of initial /f/ by half and obtained a stimulus *fos* /fos/ with 81 ms of initial /f/. If L2-L1 interference was at play, the AL subjects were expected to be more sensitive than the BLs to reduced frication duration as a cue to the voicing contrast.

Figure 2.88. shows a waveform and spectrogram of the stimulus.

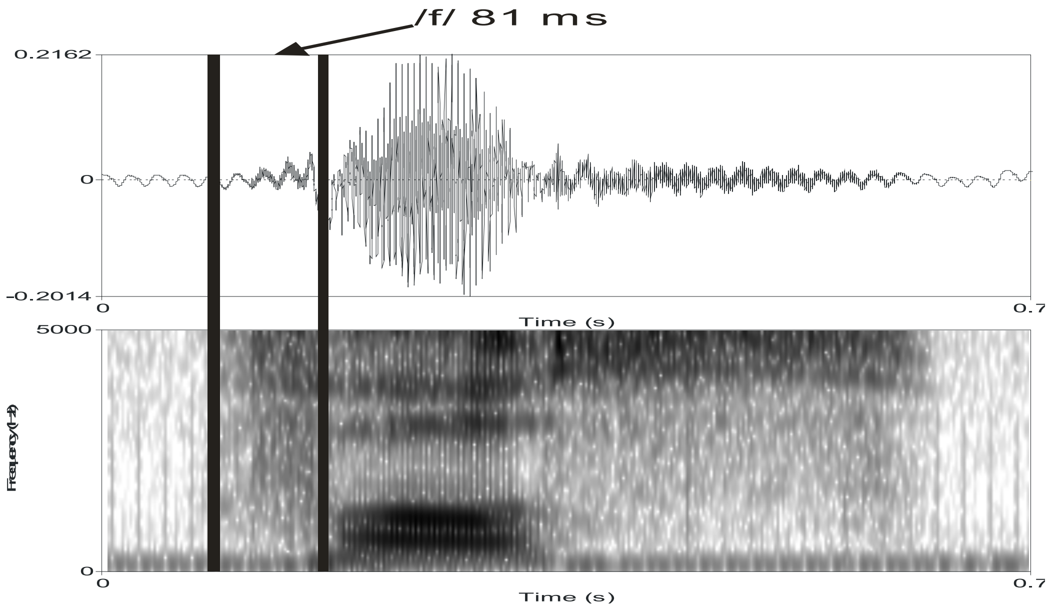


Figure 2.88. Waveform and spectrogram of syllable *fos*, /f/ 81 ms duration

2.2.3.2. Results

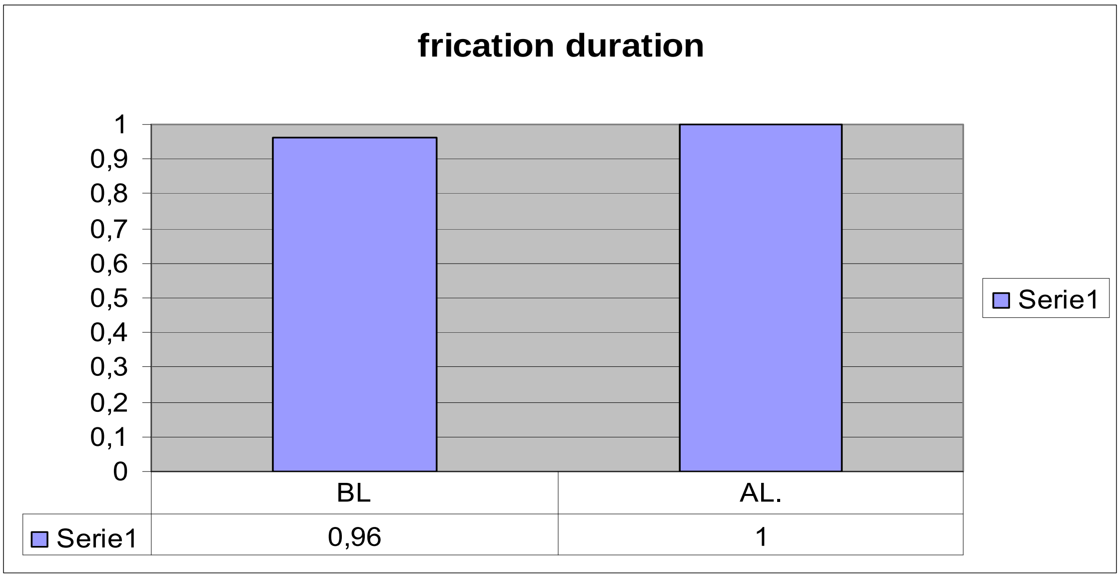


Figure 2.89. Recognition of an initial sound in *fos* as /f/ by Beginner and Advanced Learners

The reduction of the duration of an initial Polish fricative did not result in a shift into a voiced percept. Both groups recognised a shortened fricative as voiceless.

2.2.4. Duration of the release burst in initial affricates

2.2.4.1. Stimuli

From a naturally obtained Polish syllable *czir* /t♦ir/ with 66 ms duration of the release burst of /t♦/, we generated a stimulus by reducing the length of the release burst in /t♦/ by 66%. As a result we obtained a syllable *czir* /t♦ir/ with 22 ms of the release burst in initial /t♦/.

Figure 2.90 presents a waveform and spectrogram of the stimulus.

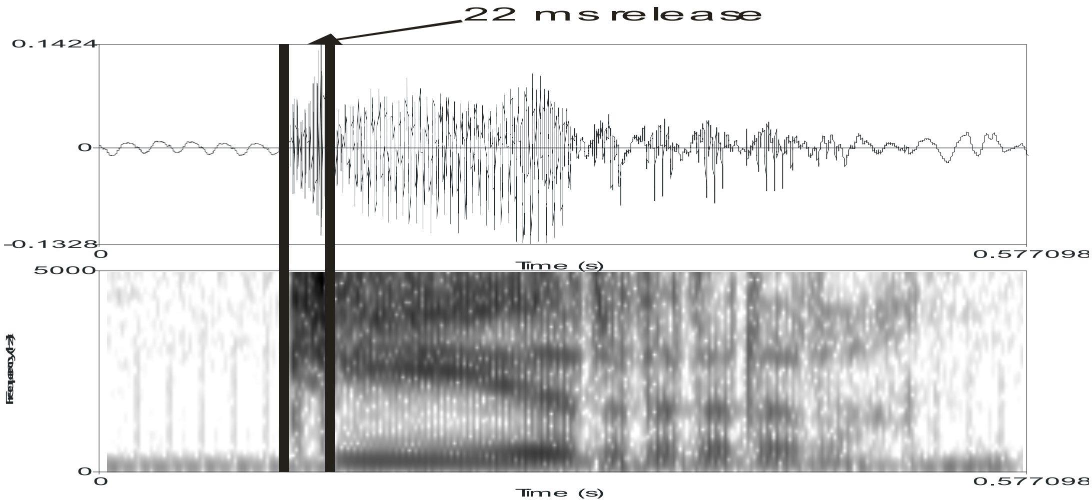


Figure 2.90. Waveform and spectrogram of syllable *czir*, /t♦/ 22 ms release burst

2.2.4.2. Results

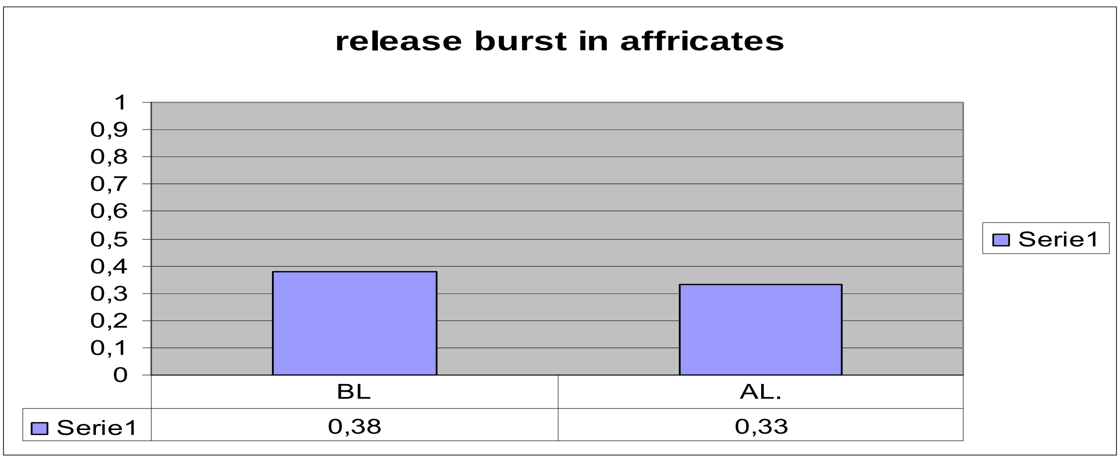


Figure 2.91. Recognition of an initial sound in *czir* as /t♦/ by Beginner and Advanced Learners

Both groups reacted to the reduced duration of the release burst – 62% and 67% of the BLs and ALs respectively reported hearing voiced /d|/ However, there were no statistically significant differences between the two groups (Yates corrected *Chi*=0.01, p=0.934).

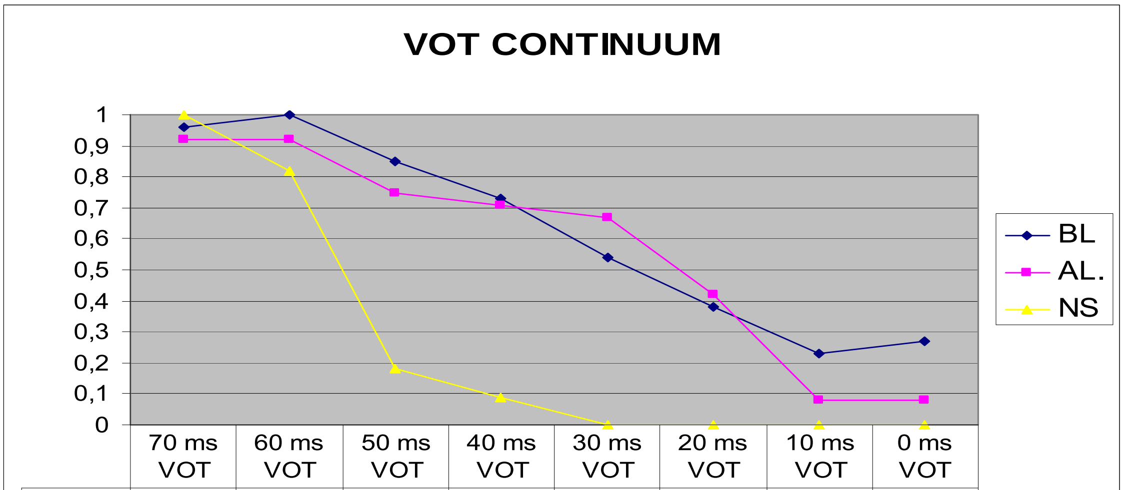
3. Testing hypotheses and discussion

3.1. Transfer from L1 to L2

3.1.1. Voice Onset Time

Hypothesis: Polish learners will not match native speakers in categorising positive VOT values between short lag for voiced and long lag for voiceless stops

The hypothesis has been confirmed. The analysis of perception of the VOT continuum reveals different patterns in the Polish groups and Native Speakers. Figure 3.1. shows the VOT perception patterns for all the three groups.



	70 ms VOT	60 ms VOT	50 ms VOT	40 ms VOT	30 ms VOT	20 ms VOT	10 ms VOT	0 ms VOT
BL	0,96	1	0,85	0,73	0,54	0,38	0,23	0,27
AL.	0,92	0,92	0,75	0,71	0,67	0,42	0,08	0,08
NS	1	0,82	0,18	0,09	0	0	0	0

Figure 3.1. Recognition of an initial sound in *keef* as /k/ across the VOT continuum by Beginner Learners, Advanced Learners, and Native Speakers

70ms	60ms	50ms	40ms	30ms	20ms	10ms	0m
p=0.68	p=0.78	p=0.00**	p=0.00**	p=0.00**	p=0.03*	p=0.84	P=0.84

Table 3.1. Comparison of recognition of an initial sound in *keef* as /k/ across the VOT continuum between Advanced Learners and Native Speakers. Yates corrected *Chi*-square.

70ms	60ms	50ms	40ms	30ms	20ms	10ms	0m
p=0.51	p=0.44	p=0.62	p=0.89	p=0.53	p=0.95	p=0.30	P=0.18

Table 3.2. Comparison of recognition of an initial sound in *keef* as /k/ across the VOT continuum between Beginner Learners and Advanced Learners. Yates corrected *Chi*-square.

As expected, the NSs had a strong categorisation effect along the decreasing VOT values. The 50 ms VOT point brought about the most drastic shift from voiceless to voiced judgments. Values lower than 30 ms VOT were consistently categorised as voiced. The observed pattern reflects the division between short lag and long lag values typical for English.

The group of the BLs reported hearing gradually more voiced percepts along the decreasing VOT. Starting with the 60 ms VOT point, the identification line falls steadily down to the 10 ms VOT point, where only 23% of the BL subjects reported hearing a voiceless segment. There was no categorisation peak – voiceless judgments decreased proportionally along the reduced VOT continuum. Unexpectedly, at the 0 ms VOT point, the BL subjects increased their voiceless judgments and reported hearing a voiceless percept 27% of the time.

The pattern observed for the ALs appears to be an intermediary between the ones found for the NSs and BLs. Around high VOT values (from 70 ms to 50 ms), it levels and falls slowly towards a voiced category. The values from 40 ms to 30 ms show a short levelling tendency. The points at 20 ms and 10 ms seem to be a categorisation point for this group. However, the 0 ms point, as in the case of the BLs, shows another levelling and, unlike the NSs, the ALs do not attain a complete perceptual shift into a voiced category.

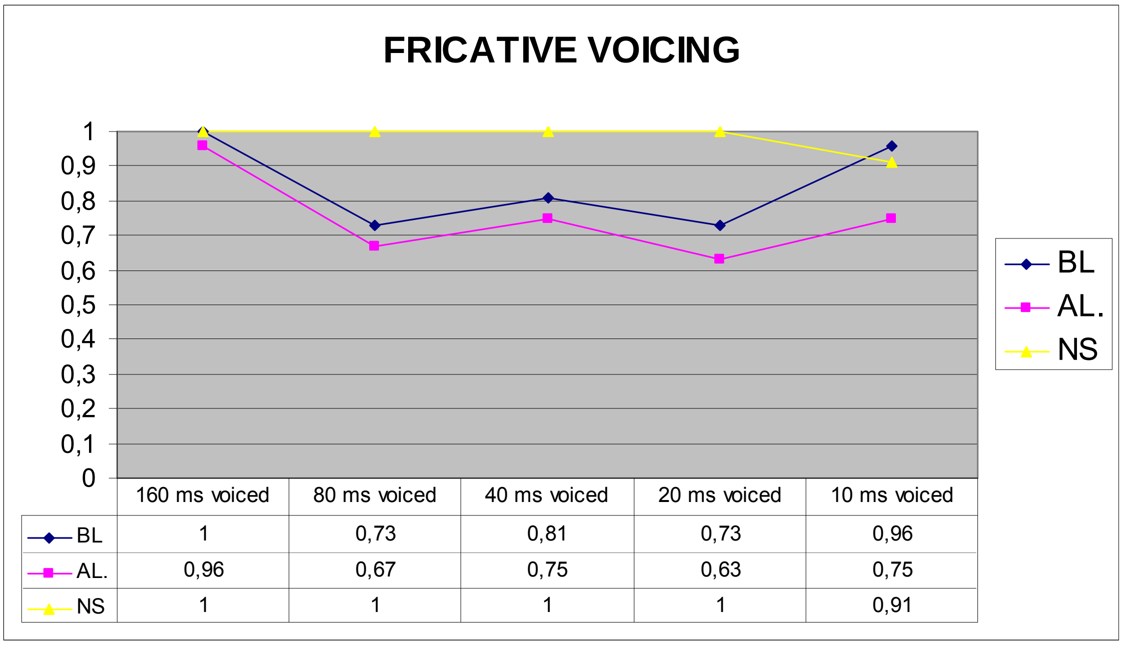
As predicted by the interlanguage hypothesis, the ALs have a perceptual pattern that mingles the patterns typical for their L1 and L2. It is interesting to note that the ALs do not reach a complete shift into a voiced percept, even at very low VOT values, which is the case for the NSs. However, unlike the BLs, they have a certain categorisation peak, even though it is 30 ms lower than the one observed for the NSs and is not so rapid; it straggles two VOT values (20 ms and 10 ms).

A fairly consistent decrease in the voiceless judgments along the reduced VOT values reported by the BLs may mean that this perceptual feature is learnt fairly rapidly. The fact that almost 80% of the BL subjects recognised the 10ms VOT point as voiced cannot be disregarded, taking into consideration the fact that this VOT value lies in the voiceless region in Polish. Although the BLs do not have a sharp category boundary, they might have learnt, to a certain degree, to recognise low VOT values as belonging to the voiced category.

3.1.2. Partially devoiced initial fricatives

Hypothesis: Polish learners will not match native speakers in recognising partially devoiced initial fricatives as voiced

The hypothesis that Polish learners of English may have problems with recognising English partially devoiced initial fricatives as voiced has been confirmed in the obtained results, however the confusion frequency is not as severe as might be expected.



	160 ms voiced	80 ms voiced	40 ms voiced	20 ms voiced	10 ms voiced
BL	1	0,73	0,81	0,73	0,96
AL.	0,96	0,67	0,75	0,63	0,75
NS	1	1	1	1	0,91

Figure 3.2. Recognition of an initial sound in *zeef* as /z/ across varying degrees of devoicing by Beginner Learners, Advanced Learners, and Native Speakers

Figure 3.2. shows a strikingly similar pattern of the perception of a partially devoiced initial fricative /z/ in the two Polish groups, however, it must be emphasised that while the stimulus effect was highly significant for the BL group (Q=14.667, p=0.005**), the results in the AL group did not meet the criteria of significance (Q=8.444, p=0.077). Despite this reservation, one can observe a lack of consistency evident in performance of the two Polish groups. The segment with 80 ms of a voiceless period caused a decrease in voiced judgements by 27% and 29% in the BL and AL groups respectively. However, a voiceless period of 120 ms brought about an unexpected increase in voiced reports. The segment with 140 ms of voicelessness again made the Polish subjects decrease their voiced judgments but, quite surprisingly, the most devoiced /z/ caused a sudden and unexpected increase in voiced judgments.

Even if one is very careful with the results obtained for the AL group, since for this group the null hypothesis could not be rejected, one must notice a lack of consistency in the identification of partially devoiced /z/ by the Polish subjects. The stimuli with intermediate devoicing values, i.e. 80 ms voiced, 40 ms voiced, and 20 ms voiced, were the source of confusion for the Polish subjects in their voicing judgments, with the most devoiced element, 10 ms voiced, having an extremely high rate of voiced reports. It is difficult to speculate why this might be the case. The more devoiced the segment, the more voiceless identifications it should enjoy. The increase in voiced reports for the most devoiced segment is evident not only in the BL but also in the AL group, even if one provides for the fact the stimulus effect in the latter group is not significant.

As expected, the NSs demonstrated very consistent immunity to the effect of partial devoicing. The stimuli ranging from 160 ms to 20 ms of a voicing period were all identified as voiced. Only the most devoiced element, 10 ms of a voicing period, caused a slight decrease in voiced reports by 9%.

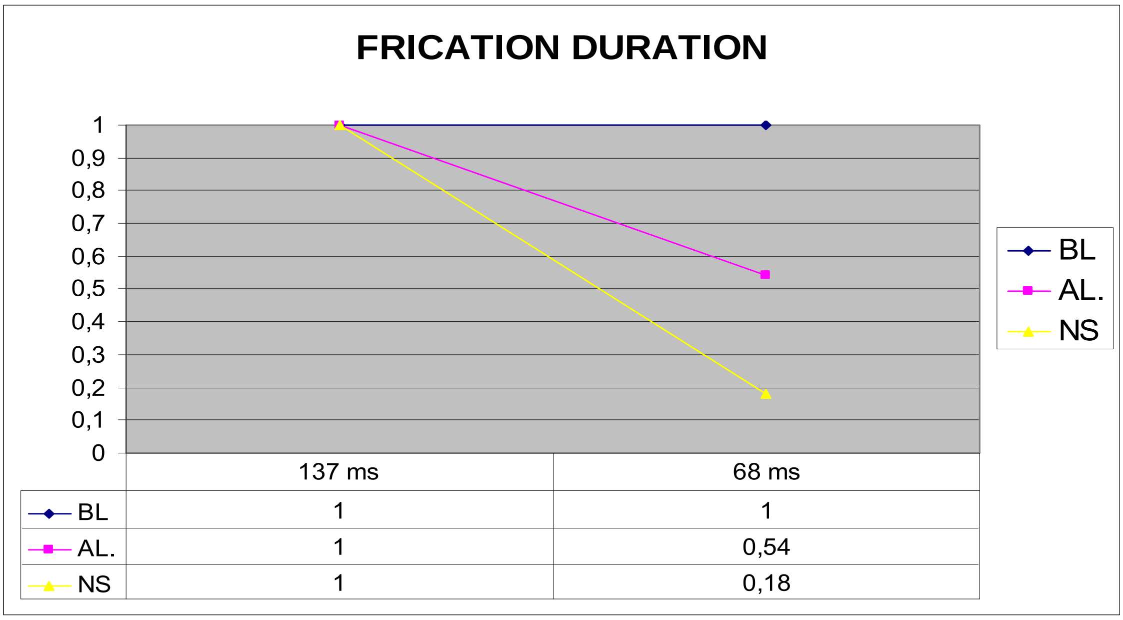
As evident in the results, partial devoicing of English initial fricatives is not as problematic for Polish listeners as the articulatory cross-linguistic comparison might suggest. Partial devoicing did not result in a drastic shift from voiced to voiceless judgments among the Polish subjects. Moreover, the most devoiced, 10 ms voiced, stimulus attained similar recognitions among the BLs, ALs, and Native Speakers.

Although it is difficult to compare the BL and AL performance due to the fact that the stimulus effect was not significant in the AL group, the identification pattern suggests that Advance Learners do not approximate Native Speakers' performance in the identification of partially devoiced initial fricatives. The regularity found for the ALs is very similar to the one observed for the BLs. Unlike the NSs, the ALs are not consistent in identifying devoiced segments as voiced, but rather follow a chaotic pattern observed for the BL group.

3.1.3. Frication duration in initial fricatives

Hypothesis: Polish learners will not match native speakers in reading reduced frication noise in initial fricatives as a cue to the voiced category

The hypothesis has been confirmed. The shortening of an initial fricative was a source of substantial cross-linguistic differences in the three tested groups.



	137 ms	68 ms
BL	1	1
AL.	1	0,54
NS	1	0,18

Figure 3.3. Recognition of an initial sound in *foss* as /f/ across varying frication duration by Beginner Learners, Advanced Learners, and Native Speakers

The BLs did not find the reduced frication duration to be a cue for the voiced status of the fricative. All subjects in this group identified the 68 ms stimulus as voiceless.

The identification pattern among the ALs indicated a 46% change in the voicing reports. Almost half of the subjects in this group reported hearing a voiced segment for the 68 ms stimulus. However, the null hypothesis for this effect could not be rejected since the comparison did not meet the criteria of significance (McNemar *Chi*-square=2.7, p=0.1).

The group of NSs demonstrated a strong regularity in recognising shortened frication duration as a cue to the voiced category. The subjects in this group perceived the 68 ms stimulus as voiced 82% of the time with a statistically significant stimulus effect (McNemar *Chi*-square=4.92, p=0.027*).

The comparison of identification patterns observed for the three groups indicates that the ALs appear to be in an intermediary stage between Native Speakers of English and Polish Beginner Learners. Even though they do not match the performance of the NSs, the ALs, unlike the BLs, react to the shortening of the frication duration and change their voicing judgments. The strength of this effect, however, is difficult to estimate since the ALs' results did not stand the test of statistical significance.

3.1.4. Duration of the release burst in initial affricates

Hypothesis: Polish learners will not match native speakers in categorising the reduced release burst in initial affricates as a cue to the voiced category

The hypothesis has only been partially confirmed. The duration of the release burst was found to be a significant factor determining the voicing contrast in initial affricates for all three tested groups.

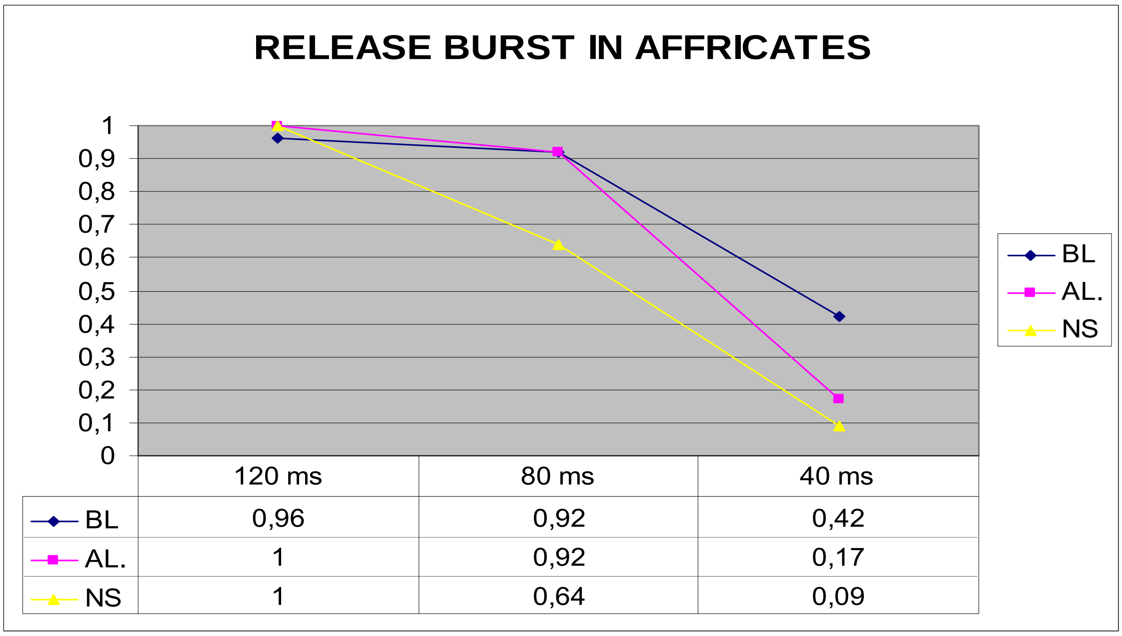


Figure 3.4. Recognition of an initial sound in *cheeth* as voiceless /t♠/ across varying duration of the release burst by Beginner Learners, Advanced Learners, and Native Speakers

120ms	80ms	40ms
-	p=0.044*	p=0.558

Table 3.3. Comparison of recognition of an initial sound in *cheef* as voiceless /t♠/ across

varying duration of the release burst between Advanced Learners and Native Speakers. *V*-square.

120ms	80ms	40ms
p=0.337	p=0.934	p=0.05*

Table 3.4. Comparison of recognition of an initial sound in *cheef* as voiceless /t♠/ across

varying duration of the release burst between Beginner Learners and Advanced Learners. *V*-square.

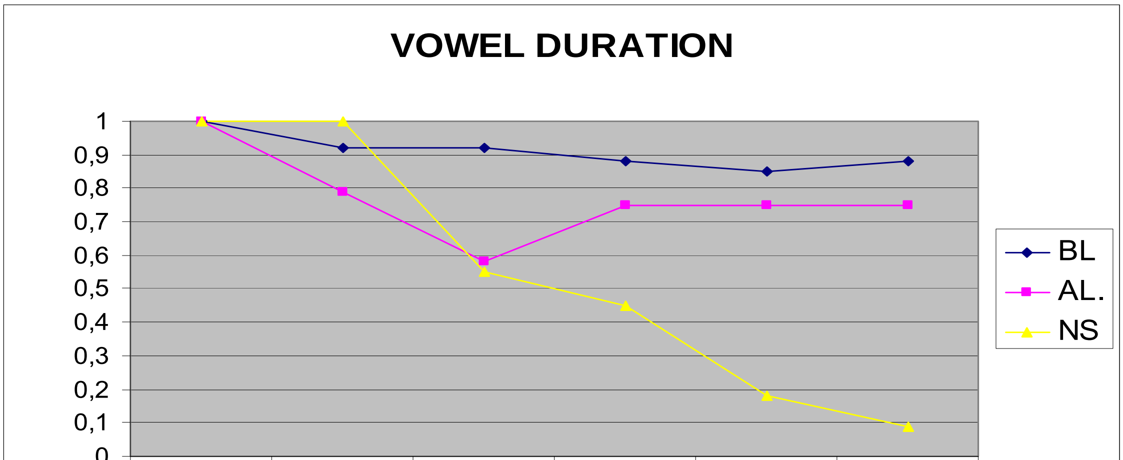
Performance graphs in Figure 3.4. indicate that, again, the ALs' performance goes in-between the regularities found for the BLs and NSs. For the 80 ms stimulus, both Polish groups reported hearing a voiceless segment 92% of the time, whereas the NSs decreased their voiceless reports to 64%. When presented with the 40 ms stimulus, the ALs approximated the NSs in their voicing judgments by reporting hearing a voiceless affricate only 17% and 9% of the time respectively. At the same time, the same stimulus was perceived as voiceless by only half of the BL subjects.

The data point to the fact that the duration of the release burst in affricates plays a significant role in cuing the voicing contrast. This feature appears to be another cross-linguistic Polish-English perceptual learning task for Polish learners of English. It is interesting to note higher sensitivity of Native Speakers to shorter reduction in the release durations. The reduction by 40 ms, to the release length of 80 ms, resulted in almost 40% decrease in voiceless recognitions for the NSs. The same value was categorised by both Polish groups as voiceless 92% of the time. Only when listening to the 80 ms shorter release burst, did the voiceless reports fall more conspicuously in the Polish groups, albeit never reaching the NSs' performance.

3.1.5. Vowel duration

Hypothesis: Polish learners of English will not match native speakers in reading increased vowel duration as a cue to the voiced category of a following stop

The hypothesis has been confirmed. The data for the vowel duration as a cue to the final voicing contrast show that this temporal feature is among the most difficult to learn for Polish learners of English.



	142 ms	172 ms	202 ms	232 ms	262 ms	292 ms
BL	1	0,92	0,92	0,88	0,85	0,88
AL.	1	0,79	0,58	0,75	0,75	0,75
NS	1	1	0,55	0,45	0,18	0,09

Figure 3.5. Recognition of a final sound in *theep* as voiceless /p/ across varying vowel duration by Beginner Learners, Advance Learners, and Native Speakers

142ms	172ms	202ms	232ms	262ms	292ms
-	p=0.1	p=0.84	p=0.09	p=0.00**	p=0.00**

Table 3.5. Comparison of recognition of a final sound in *theep* as voiceless /p/ across varying vowel duration between Advanced Learners and Native Speakers. *V*-square.

Graphic representations of regularities found for the vowel duration as a cue to the voicing contrast of the final stop indicate that Polish learners do not make use of this feature in any consistent way. Beginner Learners were completely unaware of increasing durational values of the vowel. The same proportion of voiceless judgments is maintained across increasing vowel durations. The stimulus effect was obviously statistically insignificant.

Although the AL group demonstrated a consistent decrease in voiceless reports for the first two, 172 ms and 202 ms, vowel duration values, the 232 ms stimulus caused an unexpected reversal of this tendency. Even longer durational values, namely 262 ms and 292 ms, did not bring about a decrease in voiceless reports, but rather obtained a levelled 75% voiceless recognition rate.

The NSs demonstrated a consistent, even if not complete, reliance on the vowel duration as a cue to the voicing of the following stop. The voiceless identifications decrease regularly along the increasing vowel duration. The longest, 292 ms, stimulus caused an almost complete shift from the voiceless to voiced stop, where /b/ was reported more than 90% of the time.

The results show that vowel duration poses a learning challenge for speakers of Polish. That the recognition of the vowel durations as signalling the final voicing contrast does not improve with learning time is demonstrated by the results obtained for Advanced Learners. In our data, the ALs reported hearing more voiced stops for the 202 ms than for 292 ms vowel durations. How effective a cue of vowel duration can be for Native Speakers is best shown by their identification performance. Each stimulus with an added durational value resulted in a subsequent increase in voiced reports.

The analysis of our data also corroborates the findings by Keating (1980), and Slowiaczek and Szymanska (1989) that vowel duration differences are not detected by Polish listeners, even though they might have a 10% production difference in vowels preceding phonologically voiced and voiceless obstruents (Slowiaczek and Dinnsen 1985). The Polish listeners in our study could not even read a 150 ms difference in presented vowel durations. The language-specific inability of Polish speakers to detect vowel durations as signalling the voicing contrast appears to be so deep-ingrained in their native perceptual system that even the AL subjects, fluent speakers of English, are not able to internalise this L2 perceptual rule.

3.1.6. Closure duration in stops

Hypothesis: Polish learners will not match native speakers in recognising decreasing closure duration as a cue to the voiced category of a final stop

The hypothesis has only been partially confirmed. Closure duration of the stop in a final position did not appear to be a powerful cue in determining the voicing contrast in any tested group.

Figure 3.6. Recognition of a final sound in *thock* as voiceless /k/ across varying closure duration by Beginner Learners, Advanced Learners, and Native Speakers

Although the stimulus effect was statistically significant, the BLs showed an inconsistent identification pattern with interspersing falls and rises along the decreasing closure duration. The recognition of a voiceless segment oscillated around 88% for the stimuli ranging from 25 ms to 50 ms. Only the last stimulus, with the shortest 25 ms closure duration, brought about a conspicuous decrease in voiceless judgments and was perceived as voiceless 73% of the time.

The group of ALs was not only far from consistent in their recognition of the closure duration as a cue to the voicing contrast, but also their reaction to the stimulus effect was not statistically significant (p=0.247). Although the group demonstrated a mild decrease in voiceless reports for the 50 ms stimulus, the next, shorter, 25 ms stimulus did not cause a further decline in voiceless judgment – both the 50 ms and 25 ms stimuli were recognised as voiceless 75% of the time.

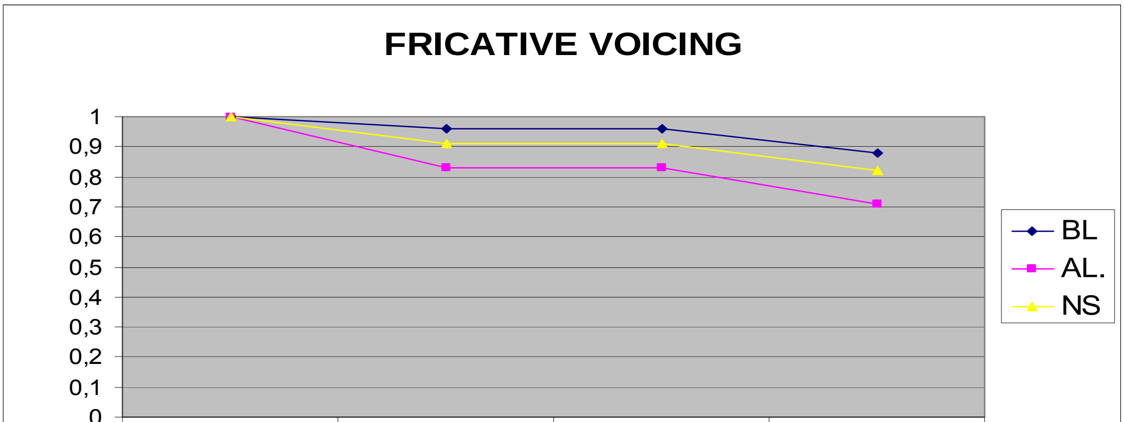
The recognition pattern observed for the NS group is characterised by a consistent and statistically significant shift from voiceless to voiced responses along the decreasing closure duration. However, it is interesting to note that even the stimulus with the shortest closure duration did not effect a complete change in voicing reports. The stimuli 75 ms and 50 ms caused a regular fall in voiceless judgments by 9% each. The shortest, 25 ms stimulus, brought about the most substantial decline and was perceived as voiceless 45% of the time.

The comparison of the performance of all three groups reveals that both Polish groups are characterised by the same lack of effectiveness in recognising the closure duration as a cue to the voicing contrast of a final plosive. Quite surprisingly, although the performance in the NS group is marked by a gradual decline in voiceless judgments along the decreasing closure values, they never reach a complete shift into a voiced category. The fact that Polish does not contrast perceptually voiced and voiceless stops in absolute word-final positions seems to be responsible for poor performance among the Polish subjects in allotting the closure duration to an appropriate voicing category. That this feature is not easy to learn is demonstrated by the performance of the ALs, whose identification pattern is very similar to the one observed for the BL subjects and characterised by a lack of regularity. As in the case of vowel duration, the closure duration seems to be another temporal parameter where the AL performance does not approximate the NS performance. However, it must be noted that even Native Speakers did not read categorically reduced closure durations. The inter-group comparison between the Polish groups and Native Speakers did not reveal any significant differences at any of the decreasing steps. Nevertheless, an observable difference is manifested in recognition consistency. Whereas the Polish groups report inconsistent rises and falls in voicing reports along decreasing closure durations, Native Speakers steadily change their judgements into a voiced category with decreasing closure durations.

3.1.7. Partially devoiced final fricatives

Hypothesis: Polish learners will not match native speakers in recognising partially devoiced final fricatives as voiced

The hypothesis has not been confirmed. The data obtained for the three tested groups indicate that partial devoicing of final fricatives has a minor influence on the perception of its voicing status.



	215 ms voiced	165 ms voiced	115 ms voiced	65 ms voiced
BL	1	0,96	0,96	0,88
AL.	1	0,83	0,83	0,71
NS	1	0,91	0,91	0,82

Figure 3.7. Recognition of a final sound in *heez* as voiced /z/ across varying degrees of devoicing by Beginner Learners, Advanced Learners, and Native Speakers

The juxtaposed performance results show a very regular identification pattern in all tested groups. The stimuli, 165 ms voiced and 115 ms voiced, effected a slight decrease and subsequent levelling in voiced judgments. The most devoiced, 65 ms voiced, stimulus caused a further, albeit far from categorical, decline in voiced reports.

The BL group appears to be a group that is the least sensitive to the devoicing effect. This fact seems to be, again, motivated by a phenomenon of final obstruent devoicing in Polish. In their L1, Polish subjects are not accustomed to find any voicing period in absolute-final fricatives. Any voicing period, even in strongly devoiced English final fricatives, appears to be a sufficient cue to recognise them as voiced. This explanation seems to be supported by the fact that even the NSs, who are well acquainted with a devoicing effect, identified more voiceless segments along progressing devoicing.

The performance of the AL group points to the fact that this group is the least immune to the partial devoicing phenomenon. Unlike the BL group, the ALs show a statistically significant shift from voiced to voiceless judgments with increasing devoicing. It seems to be a very rare case when the learning process results in a disadvantaging effect. The rationale for this regularity might be sought in a certain re-programming of perceptual sensitivity from L1 to L2. Initially, Polish learners of English are effective in hearing any voiced periods in English partially devoiced fricatives, owing to the fact that their L1 fricatives have very little or mostly no voicing at all in absolute-final positions. With time, learners seem to observe that, unlike in Polish, English final fricatives have remnants of periodicity in final portions of the frication. As a result, they lose their original sensitivity and become hypercorrect in that they expect more voiced portions in order to classify a final fricative as voiced.

The fact that the NSs react, to a certain degree, to the devoicing process may suggest that this articulatory feature may have some perceptual costs. The process of final devoicing results from lowering articulatory effort, since the speaker will not need to sustain the vocal cord vibration throughout the whole frication period. In a no-context perception task, the NS listeners' tolerance of devoicing seems to be limited. Of course, the devoicing does not cause a categorical shift from a voiced to voiceless segment, however it may reduce the effectiveness of correct identification.

3.1.8. Closure duration in affricates

Hypothesis: Polish learners of English will not match native speakers in reading reduced closure duration in final affricates as a cue to the voiced category

The hypothesis has only been partially confirmed. The shortening of the closure duration of a final voiced affricate resulted in a decrease in voiced judgements, however the effect was the most powerful and statistically significant only for the NS group.

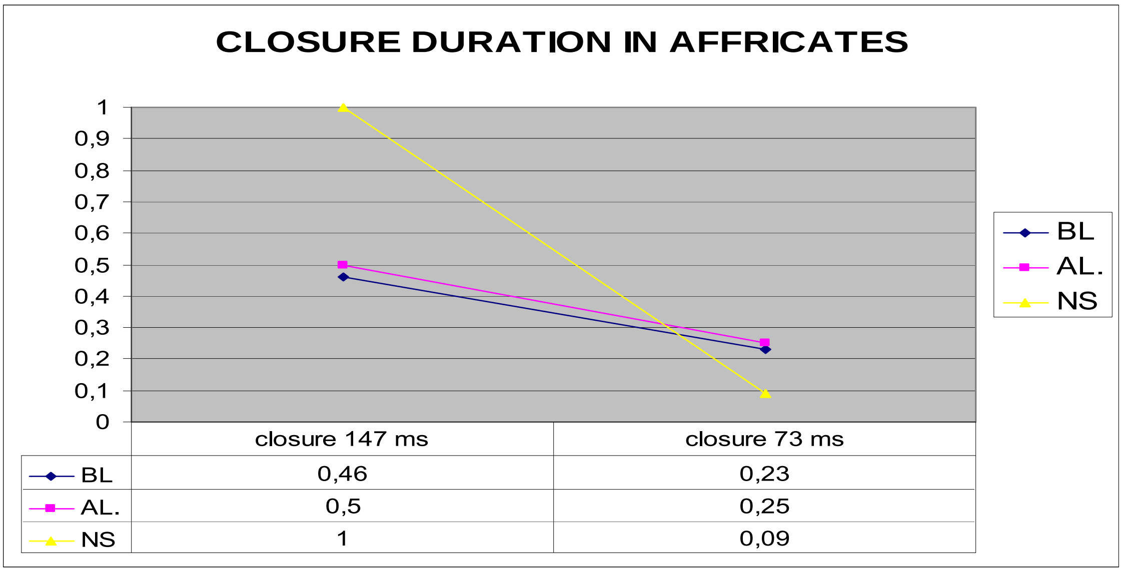


Figure 3.8. Recognition of a final sound in *heedge* as voiced /d|/ across varying closure duration by Beginner Learners, Advanced Learners, and Native Speakers

The Polish groups, both the BLs and ALs, demonstrated an almost identical recognition pattern of the reduced closure duration in final voiced affricates. It is interesting to note an unexpectedly low rate of correct identifications of the original 147 ms stimulus. Only 46% of the BL subjects and 50% of the AL subjects correctly recognised this stimulus as ending with a voiced affricate. A post-test acoustic reanalysis of the original syllable *heedge* /hi:d|/ revealed that the final affricate had a partially devoiced release burst, which seems to have contributed to a low rate of correct responses in the Polish groups. Both the BLs and ALs reacted to the shortening of the closure, however the effect was not statistically significant, largely due to misidentification of the original 147 ms stimulus.

The group of NSs showed a powerful and highly significant stimulus effect of the shortened affricate closure. All the NS subjects recognised the first stimulus as a voiced affricate and, when presented with the 73 ms stimulus, the voiced identification declined to only 9%. As pointed out in Chapter 2, Section 7, affricates are severely underresearched in the phonetic literature. The proposal that the implementation of voicing contrast in affricates is similar to the one observed for stops does not seem to be warranted, at least in final positions, in the light of our data. Voiced and not voiceless stops are characterised by shorter closure durations in production and this feature has also been demonstrated to be a significant cue in perception (see Chapter 2, Section 5).

However, our results show that it is a voiceless category that is cued by shorter closure durations in affricates. The reduction of the closure duration in a final affricate by half caused a decrease in voiced judgments in all the three tested groups, however the most powerful, categorical, shift from voiced to voiceless identifications was observed for the NSs. We have no acquaintance with any research on this problem, so it is impossible to confront the obtained results with other studies. If, however, this finding is confirmed by other independent studies in the future, it may shed some new light on the implementation of the voicing contrast in affricates.

The developmental analysis reveals that the ALs precisely emulate the recognition patter found for the BLs. Unlike the NSs, the ALs had problems with correct recognition of the original 147 ms stimulus and thus showed a substantially lower decrease in voiced reports for the 73 ms stimulus compared to the NS group.

3.2. Transfer from L2 to L1

3.2.1. 0 ms VOT

Hypothesis: 0 ms VOT value will be perceived as voiced by Advanced Learners and as voiceless by Beginner Learners

The hypothesis has not been confirmed. The idea underlying this experiment was an expectation that, if the level of fluency in English was a factor that could influence L1 Polish perception, there would be significant differences between the BLs and ALs in categorising the 0 ms VOT value in Polish. The 0 ms VOT value has been demonstrated to be a perceptual boundary separating a voiced and voiceless category in Polish initial stops (see Part 2, Section 3.3.).

However, the L2-L1 effect was not found in the present study. Not only did the difference between the two groups lack statistical significance, but also these were the BLs, rather than ALs, that showed more voiced identifications of the 0 ms VOT value. It obviously runs counter to the expectations in which we hypothesised that it would be the AL group that would be more prone to recognise the 0 ms VOT value as voiced due to the fact that English is consistent in classifying values up to around +30 ms VOT as voiced. Although the difference between the BLs and ALs is not significant, it is a puzzling finding that the BLs were more readily reporting hearing voiced stops for the 0 ms VOT value. Further studies seem to be necessary to confirm or negate this observation and, if this tendency is again demonstrated, to provide a convincing explanation.

3.2.2. Voicing lead versus voicing lag

Hypothesis: When confronted with a hybrid initial stop comprising both voicing lead and voicing lag, Beginner Learners will attend more readily to the voicing lead, whereas Advanced Learners will attend to the voicing lag

The hypothesis has not been confirmed. No between-group difference was found in the perception of an initial hybrid segment /g/ with -64 ms voicing lead and +75 ms voicing lag. The expectation was that, if an L2 English perceptual pattern interfered with L1 Polish perceptual system, the AL subjects would ignore the voicing lead period and concentrate on the voicing lag, thus recognising the sound as voiceless. It was not the case. Both groups were consistent in reporting hearing a voiced segment. It turns out that the voicing lead is a sufficiently powerful cue to determine the voicing contrast in Polish stops and cannot be overridden by the voicing lag, even among fluent speakers of English.

3.2.3. Frication duration in initial fricatives

Hypothesis: Advanced Learners will be more sensitive to the shortening of frication noise in Polish fricatives as a cue to the voiced category

The hypothesis has not been confirmed. The results show that neither the BLs nor ALs react to the shortening of frication duration as a cue to the voicing contrast in Polish fricatives. It is of no surprise in the case of the BLs, who were also insensitive to this feature in English (see Part 3, Section 3.1.3.). On the other hand, the ALs showed an almost 50% shift from voiceless to voiced judgments when the frication duration was shortened in English, which means that they are able to recognise this temporal parameter as an indicator of the voicing contrast in their L2. However, as the results suggest, Polish Advanced Learners of English do not transfer this feature into Polish. The temporal parameter that the ALs have learnt and incorporated into their L2 perceptual system does not appear to be activated in native perception. It probably results from the fact that Polish has either completely voiceless or fully voiced fricatives word-initially, quite unlike English, which may, partially or even completely, devoice its initial fricatives. When listening to Polish stimuli, the ALs seem to expect a voicing period in order to classify a fricative as voiced and do not attend to temporal parameters of frication noise. On the other hand, when listening to English, their L2 perceptual knowledge seems to inform them that they should rather look for temporal differences than expect a voicing period in English fricatives.

3.2.4. Duration of the release burst in initial affricates

Hypothesis: Advanced Learners will be more sensitive to the shortening of the release burst as a cue the voiced category in Polish initial affricates

The hypothesis has not been confirmed. Although there were no significant differences in the performance of the two Polish groups, both the BLs and ALs substantially reacted to the shortened release burst in the voiceless affricate. The sound /t♠/ with the reduced release burst to 22 ms was recognised as voiced by the BLs and ALs 62% and 67% of the time respectively. It follows the pattern obtained for the English stimuli where shortening of the release burst made both Polish groups significantly decrease their voiceless reports. The fact that the shortening of the release period in Polish affricates should cause a shift from a voiceless to voiced category makes a surprising finding regarding the common belief that Polish initial affricates are contrasted by presence or absence of periodicity (see Chapter 2, Section 7). There was obviously neither a voicing lead nor voicing in the closure phase in the presented stimulus, but still it was identified as voiced by most subjects who seem to have responded only to a temporal cue of the shortened release burst. It supports a belief that affricates do not follow a voicing contrast implementation observed for stops and call for a separate implementation mechanism.

As already mentioned, there were no significant differences in the performance of the BLs and ALs. It mostly results from the fact that, as shown by performance of the BL group, Polish learners come well-equipped with sensitivity to the durations of the release burst in affricates when beginning to learn English. This sensitivity seems to be positively transferred to, and accommodated in, the L2 perceptual system.

4. Pedagogical implications

Production has always been the main concern of phonetic components of all approaches to teaching English. It is of no surprise – teaching models aim at providing learners with comprehensible output. Perception, on the other hand, is mostly submerged in a contextual crust, in that the learners listen to authentic recordings and answer questions concerning a semantic load. The present data call for a debate on whether a certain purely phonetic perception component would not enrich an English teaching course. Of course, since the curriculum requirements concentrate mostly on skills such as reading, writing, speaking, or listening comprehension, any incorporation of purely phonetic perception training would necessarily have to be limited in size. Yet, we believe that it might constitute a valuable enrichment.

The present analysis of the obtained data allows us to denote potential perceptual problems for Poles learning English. In initial positions, it is the location of English-like VOT boundaries, which is a prerequisite for distinguishing voiced from voiceless stops. Even if, as has been demonstrated in Section 2.1.1, Polish Advanced Learners approximate such a boundary it is positioned at different values from the one typical for Native Speakers.

Partial devoicing of initial fricatives causes confusion among both Beginner Learners and Advanced Learners. Polish learners unnecessarily react to changes in a voicing period, as demonstrated in Section 2.1.2, hence confusing partially devoiced fricatives for voiceless ones. Native Speakers are characterised by consistent immunity to reductions in a voicing period in partially devoiced fricatives.

The shortening of frication noise in fricatives shifts the category from voiceless to voiced in English. According to the results in Section 2.1.3, Beginner Learners must face this feature as a novelty to learn and effectively apply. That this is a manageable task is demonstrated by the performance of Advanced Learners, who approximate the perceptual pattern of Native Speakers in shifting from voiceless to voiced categories.

The duration of the release burst as a cue to the voicing contrast in initial affricates need not be practised as a new or difficult feature. The results in Section 2.1.4 suggest that speakers of Polish come well-acquainted with this temporal parameter which is also present in their L1.

Vowel duration as a temporal parameter cueing the voicing contrast in final obstruents should be paid significant heed. Since Polish neutralises voiced obstruents in absolute word-final positions, Polish learners must learn a new parameter which is perceptually absent in their L1. That this process is extremely difficult to successfully accomplish is best manifested in the performance of Advanced Learners, who perform much worse than Native Speakers in recognising vowel durations as a cue to the voicing contrast in following stops (see Section 2.1.5.). In effect, their performance is comparable to the one observed for Beginner Learners, in that they do not read increasing durational values of a vowel in any consistent way (Section 2.1.5.).

Closure duration in stops is another problematic feature resulting form the fact that Polish neutralises absolute word-final obstruents. Polish learners are not able to detect differences in closure duration in identifying voiced from voiceless stops. However, as shown in the data in Section 2.1.6, Native Speakers seem to rely on this feature only to a limited extent word-finally. Therefore, there seems to be no need to practise this aspect separately, but rather as a subsidiary and compensating parameter to changing vowel durations.

Final partial devoicing of English fricatives poses no problem for Polish learners. As manifested in the results in Section 2.1.7, Polish learners, like Native Speakers, are insensitive to a decreasing voicing period and correctly identify partially devoiced final fricatives as voiced.

Although Polish learners follow the Native Speakers' pattern in identifying shortened closure duration in affricates as a cue the voiceless status, there is a surprisingly low correct recognition rate, both among Beginner and Advanced Learners, of voiced fricatives (see Section 2.1.8.). Partial devoicing in English affricates applies a different implementation strategy than the one observed for fricatives, i.e. devoiced frication noise versus devoiced release burst followed by frication. Although Polish learners are perceptually immune to partial devoicing in final fricatives, they need to learn how to correctly handle partial devoicing in final affricates.

Bearing all the potential cross-linguistic difficulties in learning English in mind, we believe that the inclusion of acoustically manipulated tokens can be an enriching element in a perception training component. Learners on different levels would be given an opportunity to distinguish between presented minimal pairs and learn to ascribe given temporal and acoustic parameters to an appropriate category typical for English. Such learning without a semantic burden would allow learners to concentrate on selected features in a conscious mode. Further listening comprehension exercises would afford a chance to apply learnt skills in a natural, contextualised scenario. But then, however, learners would already know what features they should seek to detect.

Concluding remarks

Speech perception has rightly found its place in a linguistic and psycholinguistic research mode. The ability of humans to perceive, recognise, and process acoustic signals is a prerequisite for any verbal communication. Infants come equipped to acquire any of hundreds of different sounds held in the inventories of the world's languages. Within the first years of life, they learn to match acoustic impressions with appropriate segments, even if there is very often no one-to-one mapping between acoustic properties of the signal and a phoneme.

However, early global sensitivity of infants is subsequently narrowed down to language-specific parameters of their ambient language. Languages exploit different features and parameters in forming their segmental stock. This leads to a situation in which original infants' sensitivity is suppressed and only parameters found in their ambient language are reinforced while for others their sensitivity is quickly fading away.

It is of no surprise then that every L2 learner must face the task of perceptual tuning into new phonetic contrasts. In early stages of a learning process, the learners, circumscribed by their L1 perceptual map, filter and warp L2 categories, which necessarily leads to impaired recognition. In order to be successful, the L2 learners must readjust their perceptual device which has been shaped by L1 experiences. Whether it is possible for L2 learners to attain native-like performance is a still debated issue. The Speech Learning Model claims that L2 learners filter out phonetic features that are used to distinguish contrasts in L2, but are absent in their L1, and even when proficient learners create L2 categories, they will never perform native-like due to constant interference from L1. On the other hand, the Second Language Linguistic Perception Model hypothesises that L2 learners are equipped with the same learning mechanisms they used in acquiring L1 and – depending on the cross-linguistic scenario – the learners will successfully create totally new L2 categories, or reduce or readjust pre-existing L1 categories.

The voicing contrast has been found to be among the most intricate contrasting devices used in languages. Early beliefs that it is implemented by the presence or absence of the vocal cord vibration proved to be essentially oversimplified. Moreover, cross-linguistic comparisons have demonstrated that a superficially universal phonological division into voiced and voiceless applies variegated fine-grained temporal and acoustic parameters in different languages. As demonstrated in Part 2, English and Polish make a good source of contrastive analysis of the voicing contrast. On the VOT continuum, English contrasts short lag versus long lag VOT values, whereas Polish locates its VOT boundary between the voicing lead and short lag. English is reported to be heavily dependent on the preceding vowel duration – it significantly reduces the length of the vowel preceding voiceless obstruents. Polish, on the other hand, belongs to word-final neutralising languages and, although it has been found to incompletely neutralise the voicing contrast word-finally and to vary, to a certain extent, the vowel duration, perception studies have demonstrated that any vowel duration in Polish is perceptually suppressed. English has shorter closure duration in voiced than in voiceless stops – the parameter that has been demonstrated to be a sufficient cue to the perception of the voicing contrast word-medially. In Polish, although closure duration differences have been observed as a result of the non-complete word-final neutralising process, its potential to cue the voicing contrast has not been confirmed in perception studies. The implementation of the voicing distinction in English fricatives is realised both temporally and acoustically. Shorter frication noise in fricatives obtains voiced percepts for speakers of English. Moreover, English fricatives are known to partially devoice both word-initially and word-finally. Polish fricatives are considered to be fully voiced or voiceless word-initially and only voiceless word-finally in absolute coda, i.e. when not followed by another voiced segment.

The experiments presented in Part 3 show a developmental view of the perceptual learning process that Polish learners of English must face. Although some parameters appear to be relatively easy, in that the learners already come well-equipped to detect them, most of them, however, are difficult, and even Advanced Learners do not match the patterns found for Native Speakers.

Although sensitivity to an English boundary within the positive VOT values appears to be learnt fairly rapidly, as demonstrated by Beginner Learners' performance (see Section 2.1.1.2.), and even though Advanced Learners finally learn to divide this continuum almost categorically (see Section 2.1.1.3.), Native Speakers place their categorisation peak at higher VOT values (see Section 2.1.1.4.). Moreover, unlike Native Speakers, Polish learners never reach a complete shift into a voiced category, even for the lowest VOT values.

Both Beginner Learners and Advanced learners show a chaotic pattern in reacting to partial devoicing of initial English fricatives (see Section 2.1.2.). On the contrary, Native Speakers are characterised by consistent insensitivity to a decreasing voicing period until almost complete devoicing. Learning process does not seem to be a contributing factor here, as demonstrated by a similar performance pattern found for both Beginner Learners and Advanced Learners.

The analysis of reactions to the shortening of frication noise in fricatives reveals that, although at the beginning of a learning process Polish learners are not able to detect this parameter, it is mastered fairly effectively with time, as demonstrated by converging recognition patterns observed for both Advanced Learners and Native Speakers in Section 2.1.3.

The data obtained for the duration of the release burst and subsequent frication noise in affricates point to the fact that Polish learners come already equipped with sensitivity to this parameter (see Section 2.1.4.). Although slight shortening leads to a weaker shift into a voiced category among Polish learners than among Native Speakers, for longer reductions Polish subjects perform similarly to Native Speakers.

The parameters that Polish listeners find most difficult to adjust to in perception of English are the vowel duration and closure duration (see Sections 2.1.5. and 2.1.6.). It seems to result from the fact that they are the parameters which are suppressed perceptually in Polish by the word-final neutralising rule. Although some speakers may retain residues of temporal variability in the vowel and closure duration in Polish (see Part 2, Section 4.1. and 5.1.), the auditory experiments have demonstrated that they are not detected perceptually. The difficulty of those parameters, as an L2 perceptual learning task, is best manifested by poor performance not only among Beginner Learners but also Advanced Learners (see Sections 2.1.5. and 2.1.6.).

Partial devoicing of fricatives in final positions does not seem to be a learning challenge. Similar patterns found for both Polish learners and Native Speakers (see Section 2.1.7.) suggest that Poles will correctly recognise English partially devoiced fricatives in final positions.

The fact that shorter closure durations in final voiced affricates yielded voiceless judgments (see Section 2.1.8.) is an interesting regularity that calls for further studies on the voicing implementation in affricates. It casts some doubt on claims that affricates use the same devices as stops in signalling the voicing contrast. As demonstrated by the results in Section 2.1.8, although Polish learners show only a slight shift towards a voiceless category, Native Speakers are characterised by strong sensitivity to the shortening effect. Further studies are needed, however, to explain this phenomenon.

Finally, we have found no backwards transfer from English into Polish. The tests comparing the perception of Polish stimuli for the 0 ms VOT value (see Section 2.2.1.), voicing lead versus voicing lag (see Section 2.2.2.), frication duration (see Section 2.2.3.), and the release burst in initial affricates (see Section 2.2.4.) did not reveal any significant differences between Beginner and Advanced Learners. Of course, these results do not rule out the possibility that such a transfer may take place. The subjects in the Advanced Learner group, although fluent speakers of English, were all living in Poland at the time of experiments and spoke Polish outside university on a regular basis. It would be necessary to apply the same tests to fluent speakers of English living in an English speaking community, and using English and Polish at least in the same proportions.

References

Abberton, E. (1972). Some laryngographic data for Korean stops. *Journal of the International Phonetic Association* 2: 67-78.

Abercrombie, D. (1967). *Elements of General Phonetics*. Chicago: Aldine Atherton.

Abramson, A. S. (1977). Laryngeal timing in consonant distinctions. *Phonetica* 34: 295-303.

Abramson, A. S. (2000). The perception of voicing distinctions. In: Burnham, D., Luksaneeyanawin, S., Davis, C., Lafourcade, M. (eds.) *Interdisciplinary Approaches to Language Processing*. Bangkok: NECTEC: 25-31.

Abramson, A. S., Lisker, L. (1965). Voice onset time in stop consonants: Acoustic analysis and synthesis. *Proceedings of the 5th International Congress of Acoustics*. Liege: Thone: A51.

Abramson, A. S., Lisker, L. (1967). Discrimination along the voicing continuum: Cross-language tests. *Status Report on Speech Research*. Haskins Laboratories: 17-22.

Abramson, A. S., Lisker, L. (1970). Discriminability along the voicing continuum: Cross language tests. *Proceedings of the Sixth International Congress of Phonetic Sciences, Prague 1967*. Prague: Academia: 569-573.

Abramson, A. S., Lisker, L. (1973). Voice-timing perception in Spanish word-initial stops. *Journal of Phonetics* 1: 1-8.

Abramson, A. S., Lisker, L. (1985). Relative power of cues: F0 shift versus voice timing. In: Fromkin, V. (ed.) *Phonetic Linguistics: Essays in Honor of Peter Ladefoged*. Orlando: Academic Press: 25-33.

Abutalebi, J., Cappa, S. F., Perani, D. (2001). The bilingual brain as revealed by functional neuroimaging. *Bilingualism: Language and Cognition* 4: 169-178.

Abutalebi, J., Keim, R., Brambati, S. M., Tettamanti, M., Cappa, S. F., De Bleser, R., Perani, D. (2006). Late acquisition of literacy in native language. *Human Brain Mapping* 28: 19-33.

Allen, G. D. (1985). How the young French child avoids the pre-voicing problem for word-initial voiced stops. *Journal of Child Language* 12: 37-46.

Allen, W. S. (1953). *Phonetics in Ancient India*. London: Oxford University Press.

Anderson, J. R. (1990). *Cognitive Psychology and its Implications*. New York: Freeman.

Appelbaum, I. (1996). The lack of invariance problem and the goal of speech perception. *ICSLP-1996*: 1541-1544.

Arabski, J. (1968). A linguistic analysis of English composition errors made by Polish students. *SAP* 1: 71-89.

Arabski, J. (1979a). *Errors as Indications of the Development of Interlanguage*. Katowice: Wydawnictwo Uniwersytetu Śląsiego.

Arabski, J. (1979b). Contrastive studies and intelanguage. *Papers and Studies in Contrastive Linguistics* 10: 135-143.

Arabski, J. (1987). *Wymowa Amerykańska*. Warszawa: Państwowe Wydawnictwo Naukowe.

Arabski, J. (1997). *Przyswajanie Języka Obcego i Pamięć Werbalna*. Katowice: Śląsk.

Arabski, J. (2006). Language transfer in language learning and language contact. In: Arabski, J. (ed.) *Cross-linguistic Influences in the Second Language Lexicon*. Clevedon: Multilingual Matters.

Arabski, J. (2007). General trends in language transfer studies. In:. Arabski, J. (ed.) *Challenging Tasks for Psycholinguistics. Proceedings of the 7th Congress of International Society of Applied Psycholinguistics*. Katowice: Wydawnictwo Uniwersytetu Śląskiego.

Archibald, J., Young-Scholten, M. (2003). The second language segment revisited. *Second Language Research* 19: 163-167.

Aslin, R. N., Pisoni, D. B., Hennessy, B. L., Perrey, A. V. (1981). Discrimination of voice onset time by human infants: New findings and implications for the effect of early experience. *Child Development* 52: 1135-1145.

Baer, T. (1975). *Investigation of Phonation Using Excised Larynxes*. Ph. D. Dissertation, MIT.

Bailey, P. J., Summerfield, Q. (1980). Information in speech: Observations on the perception of [s]-stop clusters. *Journal of Experimental Psychology: Human Perception and Performance* 6: 536-563.

Baker, S., Idsardi, W., Golinkoff, R. M., Petitto, L. A. (2005). The perception of handshapes in American Sign Language. *Memory and Cognition* 33: 887-904.

Balise, R., Diehl, R. (1994). Some distributional facts about fricatives and a perceptual explanation. *Phonetica* 51: 99-110.

Bałutowa, S. (1974). *Wymowa Angielska dla Wszystkich.* Warszawa: Wiedza Powszechna.

Barry, W. J. (1979). Complex encoding in word-final voiced and voiceless stops. *Phonetica* 36: 361-372.

Belasco, S. (1953). The influence of articulation of consonants on vowel duration. *Journal of the Acoustical Society of America* 25: 1015-1016.

Bell, A. M. (1867). *Visible speech*. London: Simpkin, Marshall & Co.

Bell-Berti, F. (1975). Control of pharyngeal cavity size for English voiced and voiceless stops. *Journal of the Acoustical Society of America* 57: 456-461.

Bent, T., Bradlow, A. (2003). The interlanguage speech intelligibility benefit. *Journal of the Acoustical Society of America* 114: 1600-1610.

Benzeghiba, M., De Mori, R., Deroo, O., Dupont, S., Erbes, T., Jouvet, D., Fissore, L., Laface, P., Mertins, A., Ris, C., Rose, R., Tyagi, V., Wellekens, C. (2007). Automatic speech recognition and speech variability: A review. *Speech Communication* 49: 763-786.

Best, C. (1995). A direct realistic perspective on cross-language speech perception. In: Strange, W. (ed.) *Speech Perception and Linguistic Experience: Issues in Cross-language Research*. Timonium: York Press: 171-206.

Best, C., McRoberts, G., Goodell, E. (2001). Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener's native phonological system. *Journal of the Acoustical Society of America* 109: 775-794.

Best, C. T., McRoberts, G. W., Sithole, N. M. (1988). Examination of perceptual reorganisation for nonnative speech contrasts: Zulu click discrimination by English speaking adults and infants. *Journal of Experimental Psychology: Human Perception and Performance* 14: 345-360.

Best, C. T., Strange, W. (1992). Effects of phonological and phonetic factors on cross-language perception of approximants. *Journal of Phonetics* 20: 305-331.

Boersma, P. (1998). *Functional Phonology*. Ph. D. Dissertation, University of Amsterdam. The Hague: Holland Academic Graphics.

Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glot International* 10: 341-345.

Boersma, P., Escudero, P., Hayes, R. (2003). Learning abstract phonological from auditory phonetic categories: An integrated model for the acquisition of language-specific sound categories. In: Sole, M. J., Recansens, D., Romero, J. (eds.) *Proceedings of the 15th International Congress of Phonetic Sciences*. Barcelona: Casual Productions: 1013-1016.

Boersma, P., Weenink, D. (2007). Praat: Doing phonetics by computer. (Version 4.6.18) [Computer program]. Retrieved August 31, 2007, from http://www.praat.org.

Bond, Z. S., Moore, T. J. (1994). A note on the acoustic-phonetic characteristics of inadvertently clear speech. *Speech Communication* 14: 324-337.

Bradlow, A. R., Pisoni, D. B., Arkahane-Yamada, R., Tohkura, Y. (1997). Training Japanese listeners to identify English /r/ and /l/: Some effects of perceptual learning on speech production. *Journal of the Acoustical Society of America* 101: 2299-2310.

Bregman, A. S. (1990). *Auditory scene analysis: The perceptual organisation of sound*. Cambridge: MIT Press.

Broersma, M. (2005). Perception of familiar contrasts in unfamiliar positions. *Journal of the Acoustical Society of America* 117: 3890-3901.

Browman, C., Goldstein, L. (1986). Towards an articulatory phonology. *Phonology Yearbook* 3: 219-252.

Browman, C., Goldstein, L. (1989). Articulatory gestures as phonological units. *Phonology* 6: 201-251.

Browman, C., Goldstein, L. (1992). Articulatory phonology: An overview. *Phonetica* 49: 155-180.

Browman, C., Goldstein, L. (1995). Dynamics and articulatory phonology. In: Port, R., Van Gelder, T. (eds.) *Mind as Motion: Explorations in the Dynamics of Cognition*. Cambridge: MIT Press: 175-193.

Brown, C. (1998). The role of the L1 grammar in the L2 acquisition of segmental structure. *Second Language Research* 14: 136-193.

Brown, C. (2000). The interrelation between speech perception and phonological acquisition from infant to adult. In: Archibald, J. (ed.) *Second Language Acquisition and Linguistic Theory*. Oxford: Blackwell.

Brown, C., Matthews, J. (1997). The role of feature geometry in the development of phonemic contrasts. In: Hannahs, S. J., Young-Scholten, M. (eds.) *Focus on Phonological Acquisition*. Amsterdam: John Benjamins Publishing Company.

Byrd, D. (1992). Preliminary results on speaker-dependent variation in the TIMIT database. *Journal of the Acoustical Society of America* 92: 593-596.

Byrd, D. (1993). 54,000 American stops. *UCLA Working Papers in Phonetics* 83: 97-115.

Byrd, D. (1994). Relations of sex and dialect to reduction. *Speech Communication* 15: 39-54.

Caramazza, A., Costa, A., Miozzo, M., Yanchao, B. (2001). The specific-word frequency effect: Implications for the representation of homophones in speech production. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 27: 1430-1450.

Caramazza, A., Yeni-Komshian, G., Zurif, E., Carbone, E. (1973). The acquisition of a new phonological contrast: The case of stop consonants in French-English bilinguals. *Journal of the Acoustical Society of America* 54: 421-428.

Chang, S. S., Plauché, M. C., Ohala, J. J. (2001). Markedness and consonant confusion asymmetries. In: Hume, E., Johnson, K. (eds.) *The Role of Speech Perception in Phonology*. New York: Academic Press.

Charles-Luce, J. (1993). The effects of semantic context on voicing neutralization. *Phonetica* 50: 28-43.

Charles-Luce, J., Dinnsen, D. A. (1987). A reanalysis of Catalan voicing. *Journal of Phonetics* 15: 187-190.

Chen, M. (1970). Vowel length variation as a function of the voicing of consonant environment. *Phonetica* 22: 129-159.

Cho, T., Jun, S. A., Ladefoged, P. (2002). Acoustic and aerodynamic correlates to Korean stops and fricatives. *Journal of Phonetics* 30: 193-229.

Cho, T., Ladefoged, P. (1997). Variations and universals in VOT: Evidence from 17 endangered languages. *UCLA Working Papers in Phonetics* 94: 18-40.

Cho, T., Ladefoged, P. (1999). Variations and universals in VOT: Evidence from 18 languages. *Journal of Phonetics* 27: 207-222.

Cho, T., McQueen, J. M. (2006). Phonological versus phonetic cues in native and non-native listening: Korean and Dutch listeners' perception of Dutch and English consonants. *Journal of the Acoustical Society of America* 119: 3085-3096.

Chomsky, N. (1957). *Linguistic Structures*. The Hague: Mouton.

Chomsky, N. (1980). *Rules and Representations*. Oxford: Blackwell.

Chomsky, N., Halle, M. (1968). *The Sound Pattern of English*. New York: Harper & Row.

Clarke, C. M., Garrett, M. F. (2004). Rapid adaptation to foreign-accented English. *Journal of the Acoustical Society of America* 116: 3647-3658.

Clarke, C. M., Luce, P. A. (2005). Perceptual adaptation to speaker characteristics: VOT boundaries in stop voicing categorisation. *PSP2005*: 23-26.

Cole, J., Kim, H., Choi, H., Hasegawa-Johnson, M. (2007). Prosodic effects on acoustic cues to stop voicing and place of articulation. Evidence from Radio News speech. *Journal of Phonetics* 35: 180-209.

Cole, R. A., Cooper, W. E. (1975). Perception of voicing in English affricates and fricatives. *Journal of the Acoustical Society of America* 58: 1280-1287.

Cooper, A. (1991). *An Articulatory Account of Aspiration in English*. Ph. D. Dissertation, Yale University.

Cooper, F. S., Delattre, P. C., Liberman, A. M., Borst, J. M., Gerstman L. J. (1952). Some experiments on the perception of synthetic speech sounds. *Journal of the Acoustical Society of America* 24: 597-606.

Crystal, T., House, A. (1988a). Segmental durations in connected-speech signals: Current results. *Journal of the Acoustical Society of America* 83: 1553-1573.

Crystal, T., House, A. (1988b). Segmental durations in connected-speech signals: Syllabic stress. *Journal of the Acoustical Society of America* 83: 1574-1585.

Cutler, A., Boersma, M. (2005). Phonetic precision in listening. In: Hardcastle, W., Beck, J. (eds.) *A Figure of Speech*. Mahwah: Erlbaum: 63-91.

Cutler, A., Clifton, Ch. (1999). Comprehending spoken language: A blueprint of the listener. In: Brown, C., Hagoort, P. (eds.) *The Neurocognition of Language*. Oxford, New York: Oxford University Press.

Dahan, D., Gaskell, M. G. (2007). The temporal dynamics of ambiguity resolution: Evidence from spoken-word recognition. *Journal of Memory and Language* 57: 483-501.

Damper, R. I. (2000). Ontogenetic versus phylogenetic learning of the emergence of phonetic categories. *Proceedings of the 3rd International Workshop on the Evolution of Language*. Paris: France: 55-58.

Damper, R. I., Harnad, S. R. (2000). Neural network models of categorical perception. *Perception and Psychophysics* 62: 843-867.

Davidsen-Nielsen, N. (1969). English stops after /s/. *English Studies* 50: 321-339.

Davis, S., Summers, W. V. (1989). Vowel length and closure duration in word-medial VC sequences. *Journal of Phonetics* 17: 339-353.

De Boer, B. (1999). Investigating the emergence of speech sounds. In: Dean, T. (ed.) *IJCAI99*. San Francisco: Morgan Kaufmann Publishers: 364-369.

De Jong, K. (1991). An articulatory study of consonant-induced vowel duration changes in English. *Phonetica* 48: 1-17.

De Jong, K. (1995). The supraglottal articulation of prominence in English. *Journal of the Acoustical Society of America* 97: 491-504.

De Jong, K. (2004). Stress, lexical focus, and segmental focus in English: Patterns of variation in vowel duration. *Journal of Phonetics* 32: 493-516.

DeKeyser, R. (2000). The robustness of critical period effects in second language acquisition. *Studies in Second Language Acquisition* 22: 499-534.

Delattre, P. C., Liberman, A. M., Cooper, F. S. (1955). Acoustic loci and transitional cues for consonants. *Journal of the Acoustical Society of America* 27: 769-773.

Delattre, P. C., Liberman, A. M., Cooper, F. S. (1964). Formant transitions and loci as acoustic correlates of place of articulation in American fricatives. *Studia Linguistica* 18: 104-121.

Delattre, P. C., Liberman, A. M., Cooper, F. S., Gerstman, L. J. (1952). An experimental study of the acoustic determinants of vowel color. *Word* 8: 195-210.

Dell, G. S. (1990). Effects of frequency and vocabulary type on phonological speech errors. *Language and Cognitive Speech Processes* 5: 313-349.

Dell, G. S., Gordon, J. K. (2003). Neighbors in the lexicon: Friends or foes? In: Schiller, N. O., Meyer, A. S. (eds.) *Phonetics and Phonology in Language Comprehension and Production: Differences and Similarities.* Berlin, New York: Mouton de Gruyter: 9-37.

Denes, P. B. (1955). Effect of duration on the perception of voicing. *Journal of the Acoustical Society of America* 27: 761-764.

Denes, P. B. (1963). On the statistics of spoken English. *Journal of the Acoustical Society of America* 35: 892-904.

Denes, P. B., Pinson, E. N. (1963). *The Speech Chain: The Physics and Biology of Spoken Language*. Baltimore: Waverly Press.

Diehl, R. L. (1991). The role of phonetics within the study of language. *Phonetica* 48: 120-134.

Diehl, R. L., Kluender, K. R. (1989). On the objects of speech perception. *Ecological Psychology* 1: 121-144.

Diehl, R. L., Kluender, K. R, Walsh, M. A. (1990). Some auditory bases of speech perception and production. In: Ainsworth, W. A. (ed.) *Advances in Speech Hearing and Language Processing*. London: JAI Press: 243-268.

Diehl, R. L., Lotto, A. J., Holt, L. L. (2004). Speech perception. *Annual Review of Psychology* 55: 149-179.

Diehl, R. L., Molis, M. R. (1995). Effects of fundamental frequency on medial [voice] judgements. *Phonetica* 52: 188-195.

Diehl, R. L., Walsh, M. A., Kluender, K. R. (1991). On the interpretability of speech/nonspeech comparisons: A reply to Fowler. *Journal of the Acoustical Society of America* 89: 2905-2909.

Dijkstra, T. (2007). The multilingual lexicon. In: Gaskel, M. G. (ed.) *The Oxford Handbook of Psycholinguistics*. Oxford: Oxford University Press: 251-265.

Dinnsen, D. A. (1985). A re-examination of phonological neutralization. *Journal of Linguistics* 21: 265-279.

Dinnsen, D. A., Charles-Luce, J. (1984). Phonological neutralisation, phonetic implementation and individual differences. *Journal of Phonetics* 12: 46-60.

Dinnsen, D. A., Garcia-Zamor, M. (1971). The three degrees of vowel length in German. *Papers in Linguistics* 4: 111-126.

Docherty, G. (1992). *The Timing of Voicing in British English Obstruents*. Dordrecht: Foris.

Dooling, R. J., Okanoya, K., Brown, S. D. (1989). Speech perception by budgerigars (*Melopsittacus undulatus*): The voiced-voiceless distinction. *Perception and Psychophysics* 46: 65-71.

Duffy, S. A., Pisoni, D. B. (1992). Comprehension of synthetic speech produced by rule: A review and theoretical interpretation. *Language and Speech* 35: 351-389.

Ehret, G. (1987). Categorical perception of sound signals: Facts and hypotheses from animal studies. In: Harnad, S. (ed.) *Categorical Perception: The Groundwork of Perception*. Cambridge: Cambridge University Press: 301-331.

Eilers, R. E., Oller, D. K., Benito-Garcia, C. R. (1984). The acquisition of voicing contrasts in Spanish and English learning infants and children: A longtitudinal study. *Journal of Child Language* 11: 313-336.

Eilers, R. E., Wilson, W. R., Moore, J. M. (1977). Developmental changes in speech discrimination in infants. *Journal of Speech and Hearing Research* 18: 158-167.

Eilers, R. E., Wilson, W. R., Moore, J. M. (1979). Speech discrimination in the language-innocent and the language-wise: A study in the perception of voice onset time. *Journal of Child Language* 6: 1-18.

Eimas, P. D. (1974). Auditory and linguistic processing of cues for place of articulation by infants. *Perception and Psychophysics* 16: 564-570.

Eimas, P. D. (1975). Speech perception in early infancy. In: Cohen, L. B., Salapatek, P. (eds.) *Infant Perception: From Sensation to Cognition, Vol. II*. New York: Academic Press: 193-231.

Eimas, P. D. (1985). The perception of speech in early infancy. *Scientific American* 252: 46-52.

Eimas, P. D. (1991). Comment: Some effects of language acquisition on speech perception. In: Mattingly, I. G., Studdert-Kennedy, M. (eds.) *Modularity and the Motor Theory of Speech Perception*. Hillsdale: Erlbaum: 111-116.

Eimas, P. D., Miller, J. L. (1977). Discrimination of the information for manner of articulation by young infants. *Infant Behavior and Development* 3: 247-252.

Eimas, P. D., Miller, J. L., Juszczyk, P. W. (1987). On infant speech perception and the acquisition of language. In: Harnad, S. (ed.) *Categorical Perception*. Cambridge: Cambridge University Press: 161-195.

Eimas, P. D., Siqueland, E. R., Juszczyk, P., Vigorito, J. (1971). Speech perception in infants. *Science* 171: 303-306.

Elman, J., McClelland, J. (1988). Cognitive penetration of the mechanisms of perception: Compensation for coarticulation of lexically restored phonemes. *Journal of Memory and Language* 27: 143-165.

Emmorey, K. (2007). The psycholinguistics of signed and spoken languages: How biology affects processing. In: Gaskell, M. G. (ed.) *The Oxford Handbook of Psycholinguistics*. Oxford: Oxford University Press: 703-721.

Emmorey, K., McCullough, S., Brentari, D. (2003). Categorical perception in American Sign Language. *Language and Cognitive Processes* 18: 21-45.

Epstein, M. A. (2000). All the sounds of all the world's languages. *UCLA Working Papers in Phonetics*.

Escudero, P. (2005). *Linguistic Perception and Second Language Acquisition*. Ph. D. Dissertation, Utrecht University. LOT Dissertation Series 113.

Escudero, P. (2006). Second language phonology: The role of perception. In: Pennington, M. (ed.) *Phonology in Context*. UK: Macmillan.

Escudero, P., Boersma, P. (2003). Modelling the perceptual development of phonological contrasts with Optimality Theory and the Gradual Learning Algorithm. In: Arunachalam, S., Kaiser, E., Williams, A. (eds.) *Proceedings of the 25th Annual Penn Linguistics Colloquium. Penn Working Papers in Linguistics* 8: 71-85.

Escudero, P., Boersma, P. (2004). Bridging the gap between L2 speech perception research and phonological theory. *Studies in Second Language Acquisition*. 26: 551-585.

Fischer-J↓rgensen, E. (1954). Acoustic analysis of stop consonants. *Miscellanea Phonetica* 2: 42-59.

Flavell, J. H. (1985). *Cognitive Development*. Englewood Cliffs: Prentice-Hall.

Flege, J. E. (1982). Laryngeal timing and phonation onset in utterance-initial English stops. *Journal of Phonetics* 10: 177-192.

Flege, J. E. (1987). The production of “new” and “similar” phones in a foreign language: Evidence for the effect of equivalence classification. *Journal of Phonetics* 15: 47-65.

Flege, J. E. (1988). The production and perception of speech sounds in a foreign language. In: Winitz, H. (ed.) *Human Communication and Its Disorders, A Review 1988*. Norwood: Ablex: 224-401.

Flege, J. E. (1991). Age of learning affects the authenticity of voice onset time (VOT) in stop consonants produced in second language. *Journal of the Acoustical Society of America* 89: 395-411.

Flege, J. E. (1992). Speech learning in a second language. In: Ferguson, Ch., Menn, L., Stoel-Gammon, C. (eds.) *Phonological Development: Models, Research, and Implications*. Timonium: York Press: 565-604.

Flege, J. E. (1995). Second-language speech learning: Theory, findings, and problems. In: Strange, W. (ed.) *Speech Perception and Linguistic Experience: Issues in Cross-language Research*. Timonium: York Press: 220-273.

Flege, J. E. (1999). Age of learning and second-language speech. In: Birdsong, D. (ed.) *New Perspectives on the Critical Period Hypothesis for Second Language Acquisition*. Hillsdale: Erlbaum: 101-132.

Flege, J. E. (2002). Interactions between the native and second-language phonetic systems. In: Burmeister, P., Piske, T., Rohde, A. (eds.) *An Integrated View of Language Development: Papers in Honor of Henning Wode*. Trier: Wissenschaftlicher Verlag: 217-244.

Flege, J. E. (2003). Assessing constraints on second-language segmental production and perception. In: Schiller, N. O., Meyer, A. S. (eds.) *Phonetics and Phonology in Language Comprehension and Production: Differences and Similarities*. Berlin: Mouton de Gruyter: 319-355.

Flege, J. E., Eefting, W. (1987). Production and perception of English stops by native Spanish speakers. *Journal of Phonetics* 15: 67-83.

Flege, J. E., Hillenbrand, J. (1986). Differential use of temporal cues to the /s/-/z/ contrast by native and non-native speakers of English. *Journal of the Acoustical Society of America* 79: 508-517.

Flege, J. E., MacKay, I. R. A. (2004). Perceiving vowels in a second language. *Studies in Second Language Acquisition* 5: 1-18.

Flege, J. E., MacKay, I. R. A., Meador, D. (1999). Native Italian speakers' production and perception of English vowels. *Journal of the Acoustical Society of America* 106: 2973-2987.

Flege, J. E., Schirru, C., MacKay, I. R. A. (2003). Interaction between the native and second language. *Speech Communication* 40: 467-491.

Flemming, E. S. (2002). Auditory representations in phonology. In: Horn, L. (ed.) *Outstanding Dissertations in Linguistics*. New York, London: Routledge.

Foulke, E., Sticht, T. (1969). Review of research on the intelligibility and comprehension of accelerated speech. *Psychological Bulletin* 72: 50-62.

Fourakis, M., Iverson, G. (1984). On the "incomplete neutralisation" of German finals. *Phonetica* 41: 140-149.

Fowler, C. A. (1981). Production and perception of coarticulation among stressed and unstressed vowels. *Journal of Speech and Hearing Research* 24: 127-139.

Fowler, C. A. (1984). Segmentation of coarticulated speech in perception. *Perception Psychophysiology* 36: 359-368.

Fowler, C. A. (1990). Sound-producing sources as objects of perception: Rate normalization and nonspeech perception. *Journal of the Acoustical Society of America* 88: 1236-1249.

Fowler, C. A. (1991). Auditory perception is not special: We see the world, we feel the world, we hear the world. *Journal of the Acoustical Society of America* 89: 2910-2915.

Fowler, C. A. (1992). Vowel duration and closure duration in voiced and unvoiced stops: There are no contrast effects here. *Journal of Phonetics* 20: 143-165.

Fowler, C. A. (1994). Speech perception: direct realist theory. In: Asher, R. E. (ed) *The Encyclopedia of Language and Linguistics*. Oxford: Pergamon: 4199-4203.

Fowler, C. A. (1996). Listeners do hear sounds, not tongues. *Journal of the Acoustical Society of America* 99: 1730-1741.

Fowler, C. A. (2003). Speech production and perception. In: Healy, A., Proctor, R. (eds.) *Handbook of Psychology, Vol. 4: Experimental Psychology*. New York: John Wiley & Sons: 237-266.

Fowler, C. A, Galantucci, B. (2005). The relation of speech perception and speech production. In: Pisoni, D., Remez, R. (eds.) *The Handbook of Speech Perception*. Oxford: Blackwell Publishing: 633-652.

Francis, A., Nusbaum, H. (2002). Selective attention and the acquisition of new phonetic categories. *Journal of Experimental Psychology: Human Perception and Performance* 28: 349-366.

Frieda, E. M., Walley, A. C., Flege, J. E., Sloane, M. E. (1999). Adults' perception of native and non-native vowels: Implications for the perceptual magnet effect. *Perception and Psychophysics* 61: 561-577.

Fromkin, V. A. (1966). Neuromuscular specification of linguistic units. *Language and Speech* 9: 170-199.

Fry, D. B. (2001). *The Physics of Speech*. Cambridge: Cambridge University Press.

Fry, D. B., Abramson, A. S., Eimas, P. D., Liberman, A. M. (1962). The identification and discrimination of synthetic vowels. *Language and Speech* 1: 35-58.

Fujimura, O. (1971). Remarks on stop consonants synthesis experiments and acoustic cues. In: Hammerich, L. L., Jakobson, R., Zwirner, E. (eds.) *Form and Substance: Phonetic and Linguistic Papers Presented to Eli Fischer-Jorgensen*. Copenhagen: 221-232.

Fujimura, O., Tatsumi, I. F., Kagaya, R. (1973). Computational processing of palatographic patterns. *Journal of Phonetics* 1: 47-54.

Gandour, J., Pett, S. H., Dardarananda, R., Dechongkit, S., Mukngoen, S. (1986). The acquisition of the voicing contrast in Thai: A study of voice onset time in word-initial stop consonants. *Journal of Child Language* 13: 561-572.

Ganong, W. F. (1980). Phonetic categorization in auditory word perception. *Journal of Experimental Psychology: Human Perception and Performance* 6: 110-125.

Gardner, B. T., Gardner, R. A. (1994). Development of phrases in the utterances of children and cross-fostered chimpanzees. In: Gardner, R. A., Gardner, B. T., Chiarelli, B., Plooj, R. (eds.) *The Ethological Roots of Culture*. Dordrecht: Kluwer: 223-225.

Gardner, R. A., Gardner, B. T. (1969). Teaching sign language to a chimpanzee. *Science* 165: 664-667.

Gardner, R. A., Gardner, B. T. (1971). Two-way communication with an infant chimpanzee. In: Schrier, A., Stollnitz, F. (eds.) *Behavior of Nonhuman Primates* Vol. 4. New York: Academic Press.

Gay, T. (1978). Effect of speaking rate on vowel formant movements. *Journal of the Acoustical Society of America* 63: 223-230.

Gerosa, M., Giuliani, D., Brugnara, F. (2007). Acoustic variability and automatic recognition of children's speech. *Speech Communication* 49: 847-860.

Gibson, J. J. (1966). *The Senses Considered as Perceptual Systems*. Boston: Houghton Mifflin.

Gibson, J. J. (1979). *The Ecological Approach to Visual Perception*. Boston: Houghton Mifflin.

Gimson, A. (1994). *Gimson's Pronunciation of English*. 5th ed. rev. by Cruttenden, A. London: Arnold.

Gonet, W. (2001). Obstruent voicing in English and Polish: A pedagogical perspective. *International Journal of English Studies* 1: 73-92.

Goodman, J. C., Nusbaum, H. C. (1994). *The Development of Speech Perception: The Transition from Speech Sounds to Spoken Words*. Cambridge: MIT Press.

Goodsitt, J. V., Morgan, J. L., Kuhl, P. K. (1993). Perceptual strategies in prelingual speech segmentation. *Journal of Child Language* 20: 229-252.

Gottfried, T, Beddor, P. (1988). Perception of temporal and spectral information in French vowels. *Language and Speech* 31: 57-75.

Griffin, Z. M., Bock, J. K. (1998). Constraint, word frequency, and the relationship between lexical processing levels in spoken word production. *Journal of Memory and Language* 38: 313-338.

Grosjean, F. (1989). Neurolinguists, beware! The bilingual is not two monolinguals in one person. *Brain and Language* 36: 3-15.

Grosjean, F. (2000). The bilingual's language modes. In: Nicol, J. (ed.) *One Mind, Two Languages: Bilingual Language Processing*. Oxford: Blackwell: 1-22.

Haggard, M. P., Ambler, S., Callow, M. (1970). Pitch as a voicing cue. *Journal of the Acoustical Society of America* 47: 613-617.

Halle, M., Hughes, G. W., Radley, J. P. (1957). Acoustic properties of stop consonants. *Journal of the Acoustical Society of America* 29: 107-116.

Hammanberg, B. (1997). Conditions on transfer in phonology. In: James, A., Leather, J. (eds.) *Second Language Speech: Structure and Process*. Berlin: Mouton de Gruyter: 161-180.

Han, N. (1998). *A Comparative Acoustic Study of Korean by Native Korean Children and Korean-American Children*. M. A. Thesis, UCLA.

Hardcastle, W. J. (1973). Some observations on the Tense-Lax distinction in initial stops in Korean. *Journal of Phonetics* 1: 263-271.

Harris, K. S., Lysaught, G. F., Schvey, M. M. (1965). Some aspects of the production of oral and nasal labial stops. *Language and Speech* 8: 135-147.

Harwood, K. (1955). Listenability and rate of presentation. *Speech Monographs* 22: 57-59.

Hawkins, S. (1999). Looking for invariant correlates of linguistic units: Two classical theories of speech perception. In: Picket, J. M. (ed.) *The Acoustics of Speech Communication: Fundamentals, Speech Perception Theory, and Technology*. Boston: Allyn and Bacon: 198-231.

Hawkins, S. (2003). Roles and representations of systematic fine phonetic detail in speech understanding. *Journal of Phonetics* 31: 373-406.

Hazan, V., Barrett, S. (1999). The development of phoneme categorisation in children aged 6 to 12 years. In: Ohala, J., Hasegawa, Y., Ohala, M., Granville, D., Bailey, A. (eds.) *Proceedings of the XIVth International Congress of Phonetics Sciences*. Berkeley: Department of Linguistics: 2493-2496.

Healy, A. F., Levitt, A. G. (1980). Accessibility of the voicing distinction for learning phonological rules. *Memory and Cognition* 8: 107-114.

Heffner, R. M. S. (1941). Notes on the length of vowels. *American Speech* 16: 204-207.

Heffner, R. M. S. (1964). *General Phonetics*. Madison: University of Wisconsin Press.

Henton, C., Ladefoged, P., Maddieson, I. (1992). Stops in the world's languages. *Phonetica* 49: 65-101.

Hillenbrand, J., Ingrisano, D. R., Smith, B. L., Flege, J. E. (1984). Perception of the voiced-voiceless contrast in syllable-final stops. *Journal of the Acoustical Society of America* 1984: 18-26.

Hirose, H., Gay, T. (1972). The activity of the intrinsic laryngeal muscles in voicing control: An electromyographic study. *Phonetica* 25: 140-164.

Hockett, C. (1955). *A Manual of Phonetics*. Bloomington: Indiana University Press.

Hogan, J. T., Rozsypal, A. J. (1980). Evaluation of vowel duration as a cue for the voicing distinction in the following word-final consonant. *Journal of the Acoustical Society of America* 67: 1764-1771.

Holmberg, T. L., Morgan, K. A., Kuhl, P. K. (1977). Speech perception in early infancy: Discrimination of fricative consonants. *The 94th Meeting of the Acoustical Society of America.* Miami Beach, Florida.

Holt, L. L., Lotto, A. J., Kluender, K. R. (2001). Influence of fundamental frequency on stop-consonant voicing perception: A case of learned covariation or auditory enhancement? *Journal of the Acoustical Society of America* 109: 764-774.

House, A. S., Fairbanks, G. (1953). The influence of consonant environment upon the secondary acoustical characteristics of vowels. *Journal of the Acoustical Society of America* 25: 105-113.

House, A. S., Stevens, K. N. (1956). Along studies of the nasalization of vowels. *Journal of Speech and Hearing Disorders* 21: 218-232.

Householder, F. W. (1956). Unreleased PTK in American English. In: Halle, Lunt, McLean, van Schooneveld (eds.) *For Roman Jakobson*. The Hague: Mouton: 235-244.

Howell, D. C. (1999). *Fundamental Statistics for Behavioural Sciences*, 4th ed. Pacific Grove: Duxbury Press.

Hughes, A., Trudgill, P., Watt, D. (2005). *English Accents and Dialects: An Introduction to Social and Regional Varieties of English in the British Isles, 4th edition*. London: Hodder Arnold.

Hume, E., Johnson, K. (2001). A model of the interplay of speech perception and phonology. In: Hume, E., Johnson, K. (eds.) *The Role of Speech Perception in Phonology*. New York: Academic Press.

Hyltenstam, K., Abrahamsson, N. (2003). Maturational constraints in SLA. In: Doughty, C., Long, M. (eds.) *Handbook of Second Language Acquisition*. Cambridge: Blackwell: 104-129.

Imsri, P. (2002). *The Perception of English Stop Consonants by Thai children and Adults.* Ph. D. Dissertation, University of Delaware.

Indefrey, P. (2007). Brain-imaging studies of language production. In: Gaskel, M. G. (ed.) *The Oxford Handbook of Psycholinguistics*. Oxford: Oxford University Press: 547-564.

Iverson, P., Kuhl, P. K., Akahne-Yamada, R., Diesch, E., Tohkura, Y., Kettermann, A., Siebert, C. (2001). A perceptual interference account of acquisition difficulties for nonnative phonemes. *Speech, Hearing and Language: Work in Progress* 13: 106-118.

Jacob, J. (1977). Evolution and tinkering. *Science* 196: 1161-1166.

Jakobson, R., Fant, G., Halle, M. (1951). *Preliminaries to Speech Analysis*. Cambridge: MIT Press.

Jakobson, R., Halle, M. (1962). Tenseness and laxness. In: Jakobson, R. (ed.) *Selected Writings I: Phonological Studies*. The Hague: Mouton: 550-555.

Jakobson, R. Waugh, L. (1979). *The sound shape of language*. Bloomington: Harvester Press.

Jamieson, D. G., Morosan, D. E. (1986). Training non-native speech contrasts in adults: Acquisition of the English /☟/ - /❄/ contrasts by francophones. *Perception and Psychophysics* 40: 205-215.

Janker, P. M., Piroth, H. G. (1999). On the perception of voicing in word-final stops in German. *Proceedings of the 14th International Congress on Phonetic Sciences (ICPhS)*: 2219-2222.

Jansen, W. (2004). *Laryngeal Contrast and Phonetic Voicing: A laboratory Phonology Approach to English, Hungarian, and Dutch*. Ph. D. Dissertation, University of Groningen.

Jassem, W. (1973). *Podręcznik Wymowy Anielskiej*. Warszawa: Państwowe Wydawnictwo Naukowe.

Jassem, W. (1974). *Fonetyka Języka Angielskiego*. Warszawa: Państwowe Wydawnictwo Naukowe.

Jassem, W. (1983). *The Phonology of Modern English*. Warszawa: Państwowe Wydawnictwo Naukowe.

Jassem, W., Richter, L. (1989). Neutralization of voicing in Polish obstruents. *Journal of Phonetics* 17: 317-325.

Jescheniak, J. D., Levelt, W. J. M. (1994). Word frequency effects in speech production: Retrieval of syntactic information and of phonological form. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 20: 824-843.

Jessen, M. (1998). *Phonetics and Phonology of Tense and Lax obstruents in German.* Amsterdam: John Benjamins.

Johnson, C. (2000). Children’s phoneme identification in reverberation and noise. *Journal of Speech, Language, and Hearing Research* 43: 129-143.

Jones, C., Berry, L., Stevens, C. (2007). Synthesised speech intelligibility and persuasion: Speech rate and non-native listeners. *Computer Speech and Language* 21: 641-651.

Jones, D. (1950). *The Phoneme: Its Nature and Use*. Cambridge: Cambridge University Press.

Jones, D. (1956). *An Outline of English Phonetics.* 8th ed. Cambridge: Heffer.

Jurafsky, D., Martin, J. H. (2000). *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition.* New Jersey: Prentice Hall.

Jusczyk, P. W. (1982). Auditory versus phonetic coding of signals during infancy. In: Mehler, J., Garrett, M., Walker, E. C. T. (eds.) *Perspectives in Mental Representation: Experimental and Theoretical Studies of Cognitive Processes and Capacities*. Hillsdale: Erlbaum: 361-387.

Jusczyk, P. W. (1993). From general to language-specific capacities: The WRAPSA model of how speech perception develops. *Journal of Phonetics* 21: 3-28.

Jusczyk, P. W. (1997). *The Discovery of Spoken Language*. Cambridge: MIT Press.

Jusczyk, P. W., Bertoncini, J., Bijeljac-Babic, R., Kennedy, L. J., Mehler, J. (1990). The role of attention in speech perception by young infants. *Cognitive Development* 5: 265-286.

Jusczyk, P. W., Copan, H. C., Thompson, E. J. (1978). Perception by two-month olds of glide contrasts in multisyllabic utterances. *Perception and Psychophysics* 24: 515-520.

Jusczyk, P. W., Rosner, B. S., Reed, M. A., Kennedy, L. J. (1989). Could temporal order differences underlie 2-month-olds’ discrimination of English voicing contrasts? *Journal of the Acoustical Society of America* 90: 83-96.

Keating, P. A. (1980). *A Phonetic Study of a Voicing Contrast in Polish*. Ph. D. Dissertation, Brown University.

Keating, P. A. (1984). Phonetic and phonological representation of stop consonant voicing. *Language* 60: 286-319.

Keating, P. A. (1985). Universal phonetics and the organization of grammars. In: Fromkin, V. (ed.) *Phonetic Linguistics: Essays in Honor of Peter Ladefoged*. Orlando: Academic Press: 115-132.

Keating, P. A., Linker, W., Huffman, M. (1983). Patterns in allophone distribution for voiced and voiceless stops. *Journal of Phonetics* 11: 277-290.

Keating, P. A., Mikoś, M. J., Ganong III, W. F. (1981). A cross-language study of range of voice onset time in the perception of initial stop voicing. *Journal of the Acoustical Society of America* 70: 1261-1271.

Kessinger, R. H., Blumstein, S. E. (1997). Effects of speaking rate on voice-onset time in Thai, French, and English. *Journal of Phonetics* 25: 143-168.

Kessinger, R. H., Blumstein, S. E. (1998). Effects of speaking rate on voice-onset and vowel production: Some implications for perception studies. *Journal of Phonetics* 26: 117-128.

Kewley-Port, D. K., Preston, M. S. (1974). Early apical stop production: A voice onset time analysis. *Journal of Phonetics* 2: 195-210.

Kijak, A., Rojczyk, A. (in press). CV as the only syllable type. *Proceedings from PASE 2007, Szczyrk, Poland.*

Kim, K. H., Relkin, N. R., Lee, K. M., Hirsch, J. (1997). Distinct cortical areas associated with native and second languages. *Nature* 388: 171-174.

Kingston, J. (1986). Are F0 differences after stops deliberate or accidental? *Journal of the Acoustical Society of America* 79: S27.

Kingston, J., Diehl, R. L. (1994). Phonetic knowledge. *Language* 70: 419-454.

Kingston, J., Diehl, R. L. (1995). Intermediate properties in the perception of distinctive feature values. In: Connel, B., Arvaniti, A. (eds.) *Papers in Laboratory Phonology IV: Phonology and Phonetic Evidence*. Cambridge: Cambridge University Press: 7-27.

Klatt, D. H. (1973). Interaction between two factors that influence vowel duration. *Journal of the Acoustical Society of America* 54: 1102-1104.

Klatt, D. H. (1975). Voice-onset time, frication, and aspiration in word-initial consonant clusters. *Journal of Speech and Hearing Research* 18: 687-703.

Kluender, K. R. (1991). Effects of first formant onset properties on voicing judgements result from processes not specific to humans. *Journal of the Acoustical Society of America* 90: 83-96.

Kluender, K. R. (1994). Speech perception as a tractable problem in cognitive science. In: Gernsbacher, M. A. (ed.) *Handbook of Psycholinguistics*. San Diego: Academic Press: 173-217.

Kluender, K. R., Coady, J. A., Kiefte, M. (2003). Sensitivity to change in perception of speech. *Speech Communication* 41: 59-69.

Kluender, K. R., Diehl, R. L., Killeen, P. R. (1987). Japanese quail can learn phonetic categories. *Science* 237: 1195-1197.

Kluender, K. R., Diehl, R. L., Wright, B. A. (1988). Vowel-length differences before voiced and voiceless consonants: An auditory explanation. *Journal of Phonetics* 16: 153-169.

Kluender, K. R., Lotto, A. J. (1994). Effects of first formant onset frequency on [-voice] judgments result from auditory processes not specific to humans. *Journal of the Acoustical Society of America* 95: 1044-1052.

Kohler, K. J. (1979). Dimensions in the perception of fortis and lenis plosives. *Phonetica* 36: 332-343.

Kohler, K. J. (1982). F0 in the production of lenis and fortis plosives. *Phonetica* 39: 199-218.

Kohler, K. J. (1984). Phonetic explanation in phonology. The feature fortis lenis. *Phonetica* 41: 150-174.

Kopczyński, A. (1977). *Polish and American English Consonant Phonemes: A Contrastive Study*. Warszawa: Państwowe Wydawnictwo Naukowe.

Kopkalli, H. (1993). *A Phonetic and Phonological Analysis of Final Devoicing in Turkish*. Ph. D. Dissertation, University of Michigan.

Krakow, R. A., Beddor, P. S., Goldstein, L. M., Fowler, C. A. (1988). Coarticulatory influences on the perceived height of nasal vowels. *Journal of the Acoustical Society of America* 83: 1146-1158.

Krause, S. E. (1982). Developmental use of vowel duration as a cue to postvocalic consonant voicing. *Journal of Speech and Hearing Research* 25: 388-393.

Kuhl, P. K. (1987). Perception of speech and sound in early infancy. In: Salapatek, P., Cohen, L. (eds.) *Handbook of Infant Perception, Vol. 2*. New York: Academic Press: 257-381.

Kuhl, P. K. (1991). Human adults and human infants show a "perceptual magnet effect" for the prototypes of speech categories, monkeys do not. *Perception and Psychophysics* 50: 93-107.

Kuhl, P. K. (1993). Innate predispositions and the effects of experience in speech perception: The native language magnet theory. In: de Boysson-Bardies, B., de

Schonen, S., Juszczyk, P., MacNeilage, P., Morton, J. (eds.) *Speech and Face Processing in the First Year of Life*. Dordrecht: Kluwer Academy: 259-274.

Kuhl, P. K. (1998). The development of speech and language. In: Carew, T. C., Menzel, R., Shatz, C. J. (eds.) *Mechanistic Relationships between Development and Learning*. New York: Wiley: 53-73.

Kuhl, P. K. (2000a). A new view of language acquisition. *Proceedings of the National Academy of Sciences* 97: 11850-11857.

Kuhl, P. K. (2000b). Language, mind, and the brain: Experience alters perception. In: Gazzaniga, M. S. (ed.) *The New Cognitive Neurosciences*. Cambridge: MIT Press: 99-115.

Kuhl, P. K., Miller, J. D. (1975). Speech perception by the chinchilla: Voiced-voiceless distinction in alveolar plosive consonants. *Science* 190: 69-72.

Kuhl, P. K., Miller, J. D. (1978). Speech perception by the chinchilla: Identification functions for synthetic VOT stimuli. *Journal of the Acoustical Society of America* 63: 905-917.

Kuhl, P. K., Williams, K., Lacerda, F., Stevens, K., Lindblom, B. (1992). Linguistic experience alters phonetic perception in infants six months of age. *Science* 255: 606-608.

Kuzla, C., Cho, T., Ernestus, M. (2007). Prosodic strengthening of German fricatives in duration and assimilatory devoicing. *Journal of Phonetics* 35: 301-320.

Ladefoged, P. (1971). *Preliminaries to Linguistic Phonetics*. Chicago: University of Chicago Press.

Ladefoged, P. (2000). *A Course in Phonetics,* 4th ed. Fort Worth: Harcourt College Publishers.

Ladefoged, P, Cho, T. (2000). Linking linguistic contrasts to reality: The case of VOT. *UCLA Working Papers on Phonetics* 98: 1-9.

Ladefoged, P., Maddieson, I. (1996). *Sounds of the World's Languages*. London: Blackwell.

Lado, R. (1957). Linguistics across cultures. *Language Transfer and Language Learning*. Rowley: Newbury House: 21-32.

Lasky, R. E., Syrdal-Lasky, A., Klein, R. E. (1975). VOT discrimination by four to six and a half month old infants from Spanish environments. *Journal of Experimental Child Psychology* 20: 215-225.

Leanderson, R., Lindblom, B. E. F. (1972). Muscle activation for labial speech gestures. *Acta Otolaryngologica* 73: 362-373.

Leather, J., James, A. (1996). The acquisition of second language speech. In: Ritchie, W., Bhatia, T. (eds.) *Handbook of Second Language Acquisition*. San Diego: Academic Press: 269-316.

Lehiste, I. (1970). *Suprasegmentals*. Cambridge: MIT Press.

Lehiste, I., Peterson, G. E. (1961). Some basic considerations in the analysis of intonation. *Journal of the Acoustical Society of America* 33: 419-425

Lenneberg, E. (1967). *Biological Foundations of Language*. New York: Wiley.

Liberman, A. M. (1957). Some results of research on speech perception. *Journal of the Acoustical Society of America* 29: 117-123.

Liberman, A. M. (1996). Introduction: Some assumptions about speech and how they changed. In: Liberman, A. M. (ed.) *Speech: A Special Code*. Cambridge: MIT Press.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P. Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review* 74: 431-461.

Liberman, A. M., Delattre, P. C., Cooper, F. S. (1952). The role of selected stimulus-variables in the perception of the unvoiced stop consonants. *The American Journal of Psychology* 65: 497-516.

Liberman, A. M., Delattre, P. C., Cooper, F. S. (1954). The role of consonant-vowel transitions in the stop and nasal consonants. *Psychological Monographs* 68: 1-13.

Liberman, A. M., Delattre, P. C., Cooper, F. S. (1958). Some cues for the distinction between voiced and voiceless stops in initial position. *Language and Speech* 1: 153-167.

Liberman, A. M., Delattre, P. C., Gerstman, L. J., Cooper, F. S. (1956). Tempo of frequency change as a cue for distinguishing classes speech sounds. *Journal of Experimental Psychology* 52: 127-137.

Liberman, A. M., Fitch, H. L., Halwes, T., Erickson, D. M. (1980). Perceptual equivalence of two acoustic cues for stop-consonant manner. *Perception and Psychophysics* 27: 343-350.

Liberman, A. M., Harris, K. S., Eimas, P., Lisker, L., Bastian, J. (1961a). An effect of learning on speech perception: the discrimination of durations of silence with and without phonemic significance. *Language and Speech* 4: 175-195.

Liberman, A.M., Harris, K. S., Kinney, J. A., Lane, H. (1961b). The discrimination of relative onset-time of the components of certain speech and nonspeech patterns. *Journal of Experimental Psychology* 61: 379-388.

Liberman, A. M., Mattingly, I. (1985). The motor theory revised. *Cognition* 21: 1-36.

Lieberman, P. (2006). *Toward an Evolutionary Biology of Language.* Cambridge MA, London: The Belknap Press of Harvard University Press.

Lieberman, P., Blumstein, S. E. (2002). *Speech Physiology, Speech Perception, and Acoustic Phonetics*. Cambridge: Cambridge University Press.

Lim, B. J., Jong, K. D., Nagao, K. (2001). Cross-language perception of syllable affiliation: Effects of voicing and language background. *Berkley Linguistic Society* 27: 43-67.

Lisker, L. (1957). Closure duration and the intervocalic voiced-voiceless distinction in English. *Language* 33: 42-49.

Lisker, L. (1970). Supraglottal air pressure in the production of English stops. *Language and Speech* 13: 215-230.

Lisker, L. (1978a). In qualified defense of VOT. *Language and Speech* 21: 375-383.

Lisker, L. (1978b). Segment duration, voicing and the syllable. In: Hooper, B. (ed.) *Syllables and Segments*. Amsterdam: North-Holland: 133-140.

Lisker, L. (1981). On generalising the *Rabid-Rapid* distinction based on silent gap duration. *Haskins Laboratories Status Report on Speech Research* SR-65: 251-259.

Lisker, L. (1984). How is the aspiration of English /p, t, k/ "predictable"?. *Language and Speech* 27: 391-394.

Lisker, L. (1986). "Voicing" in English: A catalogue of acoustic features signalling /b/ versus /p/ in trochees. *Language and Speech* 29: 3-11.

Lisker, L. (2002). The voiceless unaspirated stops in English. In: Nevin, B. E. (ed.) *The Legacy of Zellig Harris*. John Benjamins Publishing Co.: 233-240.

Lisker, L., Abramson, A. S. (1964). A cross language study of voicing in initial stops: Acoustic measurements. *Word* 20: 384-422.

Lisker, L., Abramson, A. S. (1965). Stop categorisation and voice onset time. *Proceedings of the 5th International Congress of Phonetic Sciences*. Basel: Karger: 389-391.

Lisker, L., Abramson, A. S. (1967). Some effects of context on voice onset time in English stops. *Language and Speech* 10: 1-28.

Lisker, L., Abramson, A. S. (1970). The voicing dimension: Some experiments in comparative phonetics. *Proceedings of the Sixth International Congress of Phonetic Sciences, Prague 1967*. Prague: Academia: 563-567.

Lisker, L., Abramson, A. S. (1971). Distinctive features and laryngeal control. *Language 47*: 767-785.

Lisker, L., Liberman, A. M., Erickson, D. M., Dechovitz, D., Mandler, R. (1977). On pushing the voice-onset-time (VOT) boundary about. *Language and Speech* 20: 209-216.

Lively, S. E., Pisoni, D. B. (1997). On prototypes and phonetic categories: A critical assessment the perceptual magnet effect in speech perception. *Journal of Experimental Psychology* 23: 1665-1679.

Long, M. (1990). Maturational constraints on language development. *Studies in Second Language Acquisition* 12: 251-285.

Lotto, A. J., Kluender, K. R. (2002). Synchrony capture hypothesis fails to account for effects of amplitude on voicing perception. *Journal of the Acoustical Society of America* 111: 1056-1062.

Lotto, A. J., Kluender, K. R., Holt, L. L. (1997a). Animal models of speech perception phenomena. In: Singer, K., Eggert, R., Anderson, G. (eds.) *Chicago Linguistic Society, Vol. 33*. Chicago: Chicago Linguistic Society: 357-367.

Lotto, A. J., Kluender, K. R., Holt, L. L. (1997b). Perceptual compensation for coarticulation by Japanese quail. *Journal of the Acoustical Society of America* 102: 1134-1140.

Lotto, A. J., Kluender, K. R., Holt, L. L. (1998). The perceptual magnet effect depolarised. *Journal of the Acoustical Society of America* 103: 3648-3655.

Lotz, J., Abramson, A. S., Gerstman, L. J., Ingemann, F., Nemser, W. J. (1960). The perception of English stops by speakers of English, Spanish, Hungarian and Thai: A tape-cutting experiment. *Language and Speech* 3: 71-77.

Lubker, J. F. (1968). An electromyographic-cinefluorographic investigation of velar function during normal speech production. *Cleft Palate Journal* 5. 1-18.

Lubker, J. F., Parris, P. J. (1970). Simultaneous measurements of intraoral pressure, force of labial contact, and labial electromyographic activity during production of the stop consonant cognates /p/ and /b/. *Journal of the Acoustical Society of America* 47: 625-633.

Luce, P. A., Charles-Luce, J. (1985). Contextual effects on vowel duration, closure duration and the vowel/consonant ratio in speech production. *Journal of the Acoustical Society of America* 78: 1949-1957.

Luce, P. A., Pisoni, D. B. (1998). Recognizing spoken words: The Neighborhood Activation Model. *Ear and Hearing* 20: 19-36.

MacKay, I. R. A., Flege, J. E., Piske, T. Schirru C. (2001). Category restructuring during second-language (L2) speech acquisition. *Journal of the Acoustical Society of America* 110: 516-528.

Macken, M. A., Barton, D. (1980). The acquisition of the voicing contrast in Spanish: A phonetic and phonological study of word-initial stop consonants. *Journal of Child Language* 7: 433-458.

MacNeilage, P. (1972). Speech physiology. In: Gilbert, J. (ed.) *Speech and Cortical Functioning*. New York: Academic Press: 1-72.

MacWhinney, B. (1998). Models of the emergence of language. *Annual Reviews of Psychology* 49: 199-227.

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

Maddieson, I. (1997). Phonetic universals. In: Hardcastle, W. J., Laver, J. (eds.) *The Handbook of Phonetic Sciences*. Oxford: Blackwell: 619-639.

Magloire, J., Green, K. (1999). A cross-language comparison of speaking rate effects on the production of Voice Onset Time in English and Spanish. *Phonetica* 56: 158-185.

Major, R. C. (2001). *Foreign Accent*. Mahwah: Lawrence Erlbaum Associates.

Major, R. C., (2002a). The ontogeny and phylogeny of second language phonology. In: Leather, J., James, A. (eds.) *New Sounds 2000: Proceedings of the 4th International Symposium on the Acquisition of Second-Language Speech*. Klagenfurt: University of Klagenfurt: 223-235.

Major, R. C. (2002b). The phonology of the L2 user. In: Cook, V. (ed.) *Portraits of the L2 User*. Clevedon: Multilingual Matters: 65-92.

Malécot, A. (1958). The role of releases in the identification of released final stops. *Language* 34: 370-380.

Malécot, A. (1966). Mechanical pressure as an index of 'force of articulation'. *Phonetica* 14: 169-180.

Malécot, A. (1970). The lenis-fortis opposition: Its physiological parameters. *Journal of the Acoustical Society of America* 47: 1588-1592.

Mani, N., Plunkett, K. (2007). Phonological specificity of vowels and consonants in early lexical representations. *Journal of Memory and Language* 57: 252-272.

Martin, N., Saffran, E. M., Dell, G. S. (1996). Recovery in deep dysphasia: Evidence for a relationship between auditory-verbal STM capacity and lexical errors in repetition. *Brain and Language* 52: 83-113.

Massaro, D. W. (1994). Psychological aspects of speech perception. In: Gernsbacher, M. A. (ed.) *Handbook of Psycholinguistics*. San Diego: Academic Press: 173-217.

McCarthy, J., Prince, A. (1993). *Prosodic morphology*. Unpublished manuscript, University of Massachusetts, Amherst..

Meuter, R. F. I., Allport, A. (1999). Bilingual language switching in naming: Assymetrical costs of language selection. *Journal of Memory and Language* 40: 25-40.

Michaels, D. (1974). Some replacements and phonological systems. *Linguistics* 126: 69-81.

Mikoś, M. J., Keating, P. A., Moslin, B. J. (1978). The perception of voice onset time in Polish. *Journal of the Acoustical Society of America* (S1): 63, S19.

Miller, C., Morse, P., Dorman, M. (1977). Cardiac indices of infant speech perception: Orienting and burst discrimination. *Quarterly Journal of Experimental Psychology* 29: 533-545.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits in our capacity for processing information. *Psychological Review* 63: 81-97.

Miller, G. A. (1981). *Language and Speech*. San Francisco: Freeman.

Miller, J. D., Wier, C. C., Pastore, R. E., Kelly, W. J., Dooling, R. J. (1976). Discrimination and labelling of noise-buzz sequences with varying noise-lead times: An example of categorical perception. *Journal of the Acoustical Society of America* 60: 410-417.

Miller, J. L., Eimas, P. D. (1983). Studies on the categorization of speech by infants. *Cognition* 12: 135-165.

Miller, J. L., Green, K. P., Reeves, A. (1986). Speaking rate and segments: A look at the relation between speech production and speech perception for the voicing contrast. *Phonetica* 43: 104-115.

Miller, J. L., Volaitis, L. E. (1989). Effects of speaking rate on the perceptual structure of a phonetic category. *Perception and Psychophysics* 46: 505-512.

Moffitt, A. R. (1971). Consonant cue perception by twenty to twenty-four week old infants. *Child Development* 42: 717-731.

Moore, B. C. J. (1997). *An Introduction to the Psychology of Hearing*. Sand Diego: Academic.

Moore, K. M. (2007). Spoken language processing by machine. In: Gaskell, M. G. (ed.) *The Oxford Handbook of Psycholinguistics*. Oxford: Oxford University Press: 724-738.

Morse, P. A. (1972). The discrimination of speech and nonspeech stimuli in early infancy. *Journal of Experimental Child Psychology* 14: 477-492.

Morton, J., Long, J. (1976). Effects of word transition probability on phoneme identification. *Journal of Verbal Learning and Verbal Behavior* 15: 43-52.

Moslin, B. J. (1978). *The Role of Phonetic Input in the Child's Acquisition of the Voiced-Voiceless Contrast in English Stops: A Voice Onset Time Analysis*. Ph. D. Dissertation, Brown University.

Moulines, E., Charpentier, F. (1990). Pitch-synchronous waveform processing techniques for text-to-speech synthesis using diphones. *Speech Communication* 9: 453-467.

Moulines, E., Verhelst, W. (1995). Time-domain and frequency-domain techniques for prosodic modification of speech. In: Kleijn, W., Paliwal, K. (eds.) *Speech Coding and Synthesis*. Amsterdam: Elsevier: 519-555.

Munro, M. (1993). Productions of English vowels by native speakers of Arabic: Acoustic measurements and accentedness ratings. *Language and Speech* 36: 39-66.

Nearey, T. M. (1995). A double-weak view of trading relations: Comments on Kingston and Diehl. In: Connell, B., Arvaniti, A. (eds.) *Phonology and Phonetic Evidence: Papers in Laboratory Phonology IV*. Cambridge: Cambridge University Press: 28-39.

Nearey, T. M., Rochet, B. I. (1994). Effects of place of articulation and vowel context on VOT production and perception in French and English stops. *Journal of the International Phonetic Association* 24: 1-19.

Nelson, H. (1948). The effect of variation of rate on the recall by radio listeners of "straight" newscasts. *Speech Monographs* 15: 173-180.

Netsell, R. W. (1969). Subglottal and intraoral air pressures during the intervocalic contrasts of /t/ and /d/. *Phonetica* 20: 68-73.

Nowocień, P. (2000). *Voicing of Polish and English Initial Lenis Obstruents in Isolated Words.* M. A. Thesis, University of Marie Curie Sklodowska.

Oakes, M. (1998). *Statistics for Corpus Linguistics*. Edinburgh: Edinburgh University Press.

Ohala, J. (1974). Experimental historical phonology. In: Anderson, J. M., Jones, C. (eds.) *Historical Linguistics: II. Theory and Description in Phonology*. Amsterdam: North-Holland: 353-389.

Ohala, J. (1981). The listener as a source of sound change. In: Masek, C., Hendrick, R., Miller, R., Miller, M. (eds.) *Papers from the Parasession on Language and Bahavior*. Chicago: Chicago Linguistics Society: 178-203.

Ohala, J. (1986). Against the direct realist view of speech perception. *Journal of Phonetics* 14: 75-82.

Ohala, J., Hiki, S., Hubler, S., Harshman, R. (1968). Transducing jaw and lip movements in speech. *Journal of the Acoustical Society of America* 45: 324.

Ohde, R. (1984). Fundamental frequency as an acoustic correlate of stop consonant voicing. *Journal of the Acoustical Society of America* 75: 224-230.

O'Kane, D. (1978). Manner of vowel termination as a perceptual cue to the voicing of post-vocalic stop consonants. *Journal of Phonetics* 6: 311-318.

Oldfield, R. C., Wingfield, A. (1965). Response latencies in naming objects. *Quarterly Journal of Experimental Psychology* 17: 273-281.

Ostaszewska, D., Tambor, J. (2000). *Fonetyka i Fonologia Współczesnego Języka Polskiego.* Warszawa: Państwowe Wydawnictwo Naukowe.

Parker, E. M. (1988). Auditory constraints on the perception of stop voicing: The influence of lower-tone frequency on judgements of tone-onset simulatenity. *Journal of the Acoustical Society of America* 83: 1579-1607.

Parker, F. (1974). The coarticulation of vowels and consonants. *Journal of Phonetics* 2: 211-221.

Parker, F., Walsh, T. (1981). Voicing cues as a function of the tense/lax distinction in vowels. *Journal of Phonetics* 9: 353-358.

Paus, T., Perry, D., Zatorre, R., Worsley, K., Evans, A. (1996). Modulation of cerebral blood flow in the human auditory cortex during speech: Role of motor-to-sensory discharges. *European Journal of Neuroscience* 8: 2236-2246.

Penfield, W., Roberts, L. (1959). *Speech and Brain Mechanisms*. Princeton: Princeton University Press.

Peterson, G. E., Lehiste I. (1960). Duration of syllable nuclei in English. *Journal of the Acoustical Society of America* 32: 693-703.

Pickett, J. M. (1999). *The Acoustics of Speech Communication: Fundamentals, Speech Perception Theory, and Technology*. Boston: Allyn and Bacon.

Picheny, M. A., Durlach, N. I., Braida, L. D. (1985). Speaking clearly for the hard of hearing. I: Intelligibility differences between clear and conversational speech. *Journal of Speech and Language Hearing Research* 28: 96-103.

Picheny, M. A., Durlach, N. I., Braida, L. D. (1986). Speaking clearly for the hard of hearing. II: Acoustic characteristics of clear and conversational speech. *Journal of Speech and Language Hearing Research* 29: 434-446.

Pike, K. (1943). *Phonetics*. Ann Arbor: University of Michigan Press.

Pirello, K., Blumstein, S. E., Kurowski, K. (1997). The characteristics of voicing in syllable-initial fricatives in American English. *Journal of the Acoustical Society of America* 101: 3754-3765.

Piske,. T. I., MacKay, I., Flege, J. (2001). Factors affecting degree of foreign accent in an L2: A review. *Journal of Phonetics* 29: 191-215.

Pisoni, D. B. (1973). Auditory and phonetic memory codes in the discrimination of consonants and vowels. *Perception and Psychophysics* 13: 253-260.

Pisoni, D. B. (1977). Identification and discrimination of the relative onset of time of two component tones: Implications for voicing perception in stops. *Journal of Acoustical Society of America* 61: 1352-1361.

Pisoni, D. B., Lively, S. E., Logan, J. S. (1994). Perceptual learning of nonnative speech contrasts: Implications for theories of speech perception. In: Goodman, J. C., Nusbaum, H. C. (eds.) *The Development of Speech Perception: The Transition from Speech Sounds to Spoken Words.* Cambridge: MIT Press: 121-166.

Polański, K. (2003). Stare i nowe problemy w badaniach nad dzwiękami mowy. *Doctor Honoris Causa Universatis Opoliensis*. Opole: Wydawnictwo Uniwersytetu Opolskiego.

Polivanov, E. D. (1931). La perception des sons d'une langue etrangere. *Travaux du Cercle Linguistique de Prague* 4: 79-96.

Polka, L. (1991). Cross-language speech perception in adults: Phonemic, phonetic, and acoustic contributions. *Journal of the Acoustical Society of America* 89: 2961-2977.

Polka, L. (1992). Characterizing the influence of native language experience on adult speech perception. *Perception and Psychophysics* 52: 37-52.

Port, R. F. (1979). The influence of tempo on stop closure duration as a cue for voicing and place. *Journal of Phonetics* 7: 45-56.

Port, R. F. (1980). Linguistic timing factors in combination. *Journal of the Acoustical Society of America* 69: 262-274.

Port, R. F., Crawford, P. (1989). Incomplete neutralization and pragmatics in German. *Journal of Phonetics* 17: 257-282.

Port, R. F., Dalby, J. (1982). Consonant/vowel ratio as a cue for voicing in English. *Perception and Psychophysics* 32: 141-152.

Port, R. F., Mitleb, F. M., O'Dell, M. (1981). Neutralization of obstruent voicing in German is incomplete. *Journal of the Acoustical Society of America* 70: S10.

Port, R. F., O'Dell, M. (1985). Neutralization of syllable-final voicing in German. *Journal of Phonetics* 13: 455-471.

Prince, A., Smolensky, P. (1993). *Optimality Theory: Constraint Interaction in Generative Grammar*. Rutgers University Center from Cognitive Science Technical Report 2.

Quené, H. (2007). On the just noticeable difference for tempo in speech. *Journal of Phonetics* 35: 353-362.

Raphael, L. J. (1971). *Vowel Duration as a Cue to the Perceptual Separation of Cognate Sounds in American English*. Ph. D. Dissertation, City University of New York.

Raphael, L. J. (1972). Preceding vowel duration as a cue to the perception of the voicing characteristics of word-final consonants in English. *Journal of the Acoustical Society of America* 51: 1296-1303.

Raphael, L. J. (1975). The physiological control of durational differences between vowels preceding voiced and voiceless consonants in English. *Journal of Phonetics* 3: 25-33.

Raphael, L. J. (1981). Durations and contexts as cues to word-final cognate opposition in English. *Phonetica* 69: 1462-1464.

Raphael, L. J., Dorman, M. F., Freeman, F., Tobin, C. (1975). Vowel and nasal duration as cues to voicing in word-final stop consonants: Spectrographic and perceptual studies. *Journal of Speech and Hearing Research* 18: 380-400.

Raphael, L. J., Dorman, M. F., Geffner, D. (1980). Voicing-conditioned durational differences in vowels and consonants in the speech of three- and four-year old children. *Journal of Phonetics* 8: 335-341.

Raphael, L. J., Dorman, M. F., Liberman, A. M. (1980). On defining the vowel duration that cues voicing in final position. *Language and Speech* 23: 297-308.

Raphael, L. J., Tobin, Y., Faber, A., Most, T., Kollia, H. B., Milstein, D. (1995). Intermediate values of voice onset time. In: Bell-Berti, F., Raphael, L. J. (eds.) *Producing Speech: Contemporary Issues. For Katherine Safford Harris*. New York: AIP Press.

Reeds, J. A., Wang, W. S. W. (1961). The perception of stops after s. *Phonetica* 6: 78-81.

Remez, R. E. (1994). On the perception of speech. In: Gernsbacher, M. A. (ed.) *Handbook of Psycholinguistics*. San Diego: Academic Press: 145-171.

Repp, B. H. (1984). Categorical perception: Issues, methods, findings. In: Lass, N. J. (ed.) *Speech and Language: Advances in Basic Research and Practice, Vol. 10*. New York: Academic Press: 243-335.

Repp, B. H., Williams, D. R. (1985). Influence of following context on perception of the voiced-voiceless distinction in syllable final stop consonants. *Journal of the Acoustical Society of America* 78: 445-457.

Reszkiewicz, A. (1981). *Correct your English Pronunciation*. Warszawa: Państwowe Wydawnictwo Naukowe.

Riney, T. J., Takagi, N., Ota, K., Uchida, Y. (2007). The intermediate degree of VOT in Japanese initial voiceless stops. *Journal of Phonetics* 35: 439-443.

Roach, P. (2003). *English Phonetics and Phonology*: *A Practical Course*, *Third Edition*. Cambridge: Cambridge University Press.

Rojczyk, A. (2006). Categorisation of speech sounds in auditory perception. *Linguistica Silesiana* 27: 35-44.

Rojczyk, A. (forthcoming). Priming effect on vowel duration in Polish: Data from Polish advanced learners of English. *Proceedings from the XIXth International Conference on Foreign/Second Language Acquisition*, 17-19 May 2007, Szczyrk, Poland.

Rosch, E. H. (1978). Principles of categorization. In: Rosch, E. H., Lloyd, B. (eds.) *Cognition and Categorization*. Hillsdale: Erlbaum.

Rosch, E. H., Mervis, C. B., Gray, W. D., Johnson, D., Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology* 8: 382-439.

Rositzke, H. (1939). Vowel length in general American speech. *Language* 15: 99-109.

Rousselot, P. J. (1924). *Principes de Phonetique Experimentale*. Paris: Didier.

Saffran, J. R., Aslin, R. N., Newport, E. L. (1996). Statistical learning by 8-month old infants. *Science* 274: 1926-1928.

Sampson, G. R. (1989). Language acquisition: Growth or learning? *Philosophical Papers* 18: 203-240.

Samuel, A. (1981). Phonemic restoration: Insights for a new methodology. *Journal of Experimental Psychology: General* 110: 474-494.

Samuel, A. (1996). Does lexical information influence the perceptual restoration of phonemes? *Journal of Experimental Psychology: General* 125: 28-51.

Samuel, A. (2000). Some empirical tests on Merge's architecture. In: Cutler, A., McQueen, J., Zondervan, R. (eds.) *Proceedings of the Workshop on Spoken Word Access Processes*. Nijmegen: Max Planck Institute for Psycholinguistics: 51-54.

Sawashima, M. (1968). Movements of the larynx in the articulation of Japanese consonants. *Annual Bulletin of the Research Institute of Logopedics and Phoniatrics* 2: 11-20.

Sawashima, M., Miyazaki, S. (1973). Glottal opening for Japanese voiceless consonants. *Annual Bulletin of the Research Institute of Logopedics and Phoniatrics* 7: 1-10.

Sawicka, I. (1995). Fonologia. In: Wróbel, H. (ed.) *Gramatyka Współczesna Języka Polskiego: Fonetyka i Fonologia*. Kraków: Państwowe Wydawnictwo Naukowe.

Sawusch, J. R., Pisoni, D. B. (1974). On the identification of place and voicing features in synthetic stop consonants. *Journal of Phonetics* 2: 181-194.

Scharenborg, O., Wan, V., Moore, R. K. (2007). Towards capturing fine phonetic variation in speech using articulatory features. *Speech Communication* 49: 811-826.

Scholfield, P. (1991). Statistics in linguistics. *Annual Review of Anthropology* 20: 377-393.

Schwartz, B., Sprouse, R. (1996). L2 cognitive states and the Full Transfer / Full Access model. *Second Language Research* 12: 40-72.

Scobbie, J. M., Gibbon, F., Hardcastle, W. J., Fletcher, P. (1996). Covert contrast as a stage in the acquisition of phonetics and phonology. *QMC Working Papers in Speech and Language Sciences, Vol 1*: 43-62.

Scovel, T. (1988). *A Time to Speak: A Psycholinguistic Inquiry into the Critical Period for Human Speech*. Cambridge: Newbury House.

Sebastián-Gallés, N., Kroll, J. F. (2003). Phonology in bilingual processing: Acquisition, perception, and production. In: Schiller, N. O., Meyer, A. S. (eds.)

Phonetics and Phonology in Language Comprehension and Production: Differences and Similarities. Berlin: Mouton de Gruyter: 279-317.

Selinker, L. (1972). Interlanguage. *International Review of Applied Linguistics* 10: 209-231.

Serniclaes, W. (1987). *Etude Experimentale de la Perception du Trait de Voisement des Occlusives du Francais*. Ph. D. Dissertation, Universite Libre de Bruxelles.

Serniclaes, W. (2005). On the invariance of speech percepts. *ZAS Papers in Linguistics* 40: 177-194.

Shankweiler, D. P., Strange, W., Verbrugge, R. R. (1977). Speech and the problem of perceptual constancy. In: Shaw, R., Bransford, J. (eds.) *Perceiving, Acting, and Knowing: Towards an Ecological Psychology*. Hillsdale: Erlbaum.

Shimizu, K. (1996). *A Cross-language Study of Voicing Contrasts of Stop Consonants in Six Asian Languages*. Tokyo: Seibido.

Shinn, P., Blumstein, S. E. (1984). On the role of the amplitude envelope for the perception of /b/ and /w/. *Journal of the Acoustical Society of America* 75: 1243-1252.

Sholes, G. (1959). Synthesis of final /z/ without voicing. *Journal of the Acoustical Society of America* 31: 1568.

Slis, I. H. (1975). Consequences of articulatory effort on articulatory training. In: Fant, G., Tatham, M. (eds.) *Auditory Analysis and the Perception of Speech*. New York: Academic: 397-411.

Slowiaczek, L. M., Dinnsen, D. A. (1985). On the neutralizing status of Polish word-final devoicing. *Journal of Phonetics* 13: 325-341.

Slowiaczek, L. M., Szymanska, H. J. (1989). Perception of word-final devoicing in Polish. *Journal of Phonetics* 17: 205-212.

Smit, A. B., Bernthal, J. E. (1983). Voicing contrasts and their phonological implications in the speech of articulation-disordered children. *Journal of Speech and Hearing Research* 26: 486-500.

Smith, B. (1978). Temporal aspects of English speech production: A developmental perspective. *Journal of Phonetics* 6: 37-67.

Smith, B., Bradlow, A., Bent, T. (2003). Production and perception of temporal contrasts in foreign-accented English. *Proceedings of the 15th International Congress of Phonetic Sciences, Barcelona 2003*: 519-522.

Smith, B., Westbury, J. R. (1975). Temporal control of voicing during occlusion. Paper presented at the 89th meeting of the Acoustical Society of America.

Snoeren, N. D., Halle, P. A., Segui, J. (2006). A voice for the voiceless: Production and perception of assimilated stops in French. *Journal of Phonetics* 34: 241-268.

Sobkowiak, W. (2001). *English Phonetics for Poles: A Resource Book for Learners and Teachers*. Poznań: Wydawnictwo Poznańskie.

Solé, M. J. (2007). Controlled and mechanical properties of speech: A review of the literature. In: Solé, M. J., Beddor, P. S., Ohala, M. (eds.) *Experimental Approaches to Phonology*. Oxford: Oxford University Press: 302-321.

Soloman, R. L., Postman, L. (1952). Frequency of usage as a determinant of recognition thresholds for words. *Journal of Experimental Psychology* 43: 195-201.

Sommerstein, A. H. (1977). *Modern Phonology*. Baltimore: University Park Press.

Stevens, K. N. (1977). Physics of laryngeal behavior and larynx models. *Phonetica* 34: 264-279.

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

Stevens, K. N., Blumstein, S. E., Glicksman, L., Burton, M., Kurowski, K. (1992). Acoustic and perceptual characteristics of voicing in fricatives and fricative clusters. *Journal of the Acoustical Society of America* 91: 2979-3000.

Stevens, K. N., Fant, G., Hawkins, S. (1987). Some acoustical and perceptual correlates of nasal vowels. In: Channon, R., Shockey, L. (eds.) *In Honor of Ilse Lihiste*. Paris: Dordrecht: 241-254.

Stevens, K. N., Keyser, S. J., Kawasaki, H. (1986). Toward a phonetic and phonological theory of redundant features. In: Perkell, J. S., Klatt, D. H. (eds.) *Invariance and Variability in Speech Processes*. Hillsdale: Erlbaum: 426-429.

Stevens, L., Klatt, D. H. (1974). Role of formant transitions in the voiced-voiceless distinction for stops. *Journal of the Acoustical Society of America* 55: 653-659.

Strain, E., Patterson, K., Seidenberg, M. S. (1995). Semantic effects in single-word naming. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 21: 1140-1154.

Strange, W. (1995). Cross-language study of speech perception: A historical review. In: Strange, W. (ed.) *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*. Timonium: York Press: 3-45.

Strange, W., Verbrugge, R. R., Shankweiler, D. P., Edman, T. R. (1976). Consonantal environment specifies vowel identity. *Journal of the Acoustical Society of America* 60: 213-221.

Streeter, L. A. (1976). Language perception of two-month old infants shows effects of both innate mechanisms and experience. *Nature* 259: 39-41.

Strevens, P. (1960). Spectra of fricative noise in human speech. *Language and Speech* 3: 32.

Summerfield, A. Q. (1975a). Aerodynamics vs. mechanics in the control of voicing onset in consonant-vowel syllables. *Speech Perception* 4: 61-72.

Summerfield, A. Q. (1975b). How a full account of segmental perception depends on prosody and vice versa. In: Cohen, A., Nooteboom, S. G. (eds.) *Structure and Process in Speech Perception*. New York: Springer-Verlag.

Summerfield, A. Q. (1981). Articulatory rate and perceptual constancy in phonetic perception. *Journal of Experimental Psychology: Human Perception and Performance* 7: 1074-1095.

Summerfield, A. Q., Haggard, M. (1977). On the dissociation of spectral and temporal cues to the voicing distinction in initial stop consonants. *Journal of the Acoustical Society of America* 62: 435-448.

Summers, W. V. (1987). Effects of stress and final-consonant voicing on vowel production: Articulatory and acoustic analyses. *Journal of the Acoustical Society of America* 82: 847-863.

Sussman, J. E., Lauckner-Morano, V. J. (1995). Further tests on the “perceptual magnet effect” in the perception of [i]: Identification and change-no-change discrimination. *Journal of the Acoustical Society of America* 97: 539-552.

Sweet, H. (1877). *A Handbook of Phonetics*. Oxford: Clarendon Press.

Tatham, M. A. A., Morton, K. (1969). Some electromyographic data towards a model of speech production. *Language and Speech* 12: 39-53.

Tatham, M. A. A., Morton, K. (1973). Electromyographic and intraoral air pressure studies of bilabial stops. *Language and Speech* 16: 336-350.

Taylor, J. R. (1991). *Linguistic Categorization: Prototypes in Linguistic Theory*. Oxford: Clarendon Press.

Tieszen, B. (1997). *Final Stop Devoicing in Polish: An Acoustic and Historical Account for Incomplete Neutralization*. Ph. D. Dissertation, University of Wisconsin-Madison.

Trubetzkoy, N. S. (1929). Zur algemeinen theorie der phonologischen vokalsysteme. *Travaux du Cercle Linguistique de Prague* 1: 39-67.

Trubetzkoy, N. S. (1969). *Principles of Phonology*. Berkley: University of California Press.

Tsukada, K., Birdsong, D., Mack, M., Sung, H., Bialystok, E., Flege, J. (2004). Release bursts in English word-final voiceless stops produced by native English and Korean adults and children. *Phonetica* 61: 67-83.

Umeda, N. (1977). Consonant duration in American English. *Journal of the Acoustical Society of America* 74: 715-725.

Umeda, N. (1981). Influence of segmental factors on fundamental frequency in fluent speech. *Journal of the Acoustical Society of America* 70: 350-355.

Van den Berg, J. (1958). Myoelastic theory of voice production. *JSHR* 1: 227-244.

Ventura, P., Kolinsky, R., Fernandes, S., Querido, L., Morais, J. (2007). Lexical restructuring in the absence of literacy. *Cognition* 105: 334-361.

Ventura, P., Morais, J., Kolinsky, R. (2007). The development of the orthographic consistency effect in speech recognition: From sublexical to lexical development. *Cognition* 105: 547-576.

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

Walley, A., Flege, J. (2000). Effects of lexical status on children's and adults' perception of native and non-native vowels. *Journal of Phonetics* 27: 307-332.

Walsh, M. A., Diehl, R. L. (1991). Formant transition and amplitude rise times as cues to the stop/glide distinction. *Quarterly Journal of Experimental Psychology* 43: 603-620.

Wang, W. S. Y. (1959). Transition and release as perceptual cues for final plosives. *Journal of Speech and Hearing Research* 2: 66-73.

Waniek-Klimczak, E. (1993). *The Sociolinguistic Study of Phonetic Parameters in "Polish English": A Case Study on Phonetic Implementation of Voicing in English Spoken by Polish Immigrants to Britain*. Ph. D. Dissertation, University of Łódź.

Waniek-Klimczak, E. (1996). The strategies of voice implementation in English as the second language: A case study. *Studia Phonetica Posnaniensia* 5: 101-140.

Waniek-Klimczak, E. (2005). *Temporal Parameters in Second Language Speech: An Applied Linguistic Phonetics Approach*. Łódź: Wydawnictwo Uniwersytetu Łódźkiego.

Warner, N., Good, E., Jongman, A., Sereno, J. (2006). Orthographic vs. morphological incomplete neutralization effects. *Journal of Phonetics* 34: 285-293.

Warner, N., Jongman, A., Sereno, J., Kemps, R. (2004). Incomplete neutralization and other sub-phonemic durational differences in production and perception: Evidence from Dutch. *Journal of Phonetics* 32: 251-276.

Warren, D. W., Hall, D. J. (1973). Glottal activity and intraoral pressure during stop consonant productions. *Folia Phoniatrica* 25: 121-129.

Warren, R. M. (1970). Perceptual restoration of missing speech sounds. *Science* 167: 392-393.

Waters, R., Wilson, W. (1976). Speech perception by rhesus monkeys: The voicing distinction in synthesised labial and velar stops consonants. *Perception and Psychophysics* 19: 285-289.

Weismer, G. (1979). Sensitivity of voice onset measures to certain segmental features in speech production. *Journal of Phonetics* 7: 194-204.

Weismer, G., Dinnsen, D. A., Elbert, M. (1981). A study of the voicing distinction associated with omitted word-final stops. *Journal of Speech and Hearing Disorders* 46: 320-327.

Wells, J. (1982). *Accents of English, Vol. 2: The British Isles*. Cambridge: Cambridge University Press.

Werker, J. F., Lalonde, C. E. (1988). Cross-language speech perception: Initial capabilities and developmental stage. *Developmental Psychology* 24: 672-638.

Werker, J. F., Pegg, J. E. (1992). Infant speech perception and phonological acquisition. In: Ferguson, C. A., Menn, L., Stoel-Gammon, C. (eds.) *Phonological Development: Models, Research, Implications*. York: Timonium: 285-311.

Werker, J. F., Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganisation during the first year of life. *Infant Behavior and Development* 7: 49-63.

Werker, J. F., Yeung, H. H. (2005). Infant speech perception bootstraps word learning. *TRENDS in Cognitive Sciences* 9:519-527.

Westbury, J. R., Keating, P. A. (1980). A model of stop consonant voicing and a theory of markedness. Paper presented at the Linguistic Society of America.

Westbury, J. R., Keating, P. A. (1986). On the naturalness of stop consonant voicing. *Journal of Linguistics* 22: 145-146.

Whalen, D. H., Levitt, A. G., Goldstein, L. M. (2007). VOT in the babbling of French- and English-learning infants. *Journal of Phonetics* 35: 341-352.

Wierzchowska, B. (1967). *Opis Fonetyczny Języka Polskiego*. Warszawa: PZWS.

Wierzchowska, B. (1980). *Fonetyka i Fonologia Języka Polskiego.* Wrocław: Ossolineum.

Williams, L. (1980). Phonetic variation as a function of second-language learning. In: Yeni-Komishian, G., Kavanagh, J., Ferguson, C. (eds.) *Child Phonology, Vol. 2*. New York: Academic Press: 185-216.

Wilson, S. M. (2003). A phonetic study of voiced, voiceless and alternating stops in Turkish. *CRL Newsletter* 15: 1-13.

Wright, J. T. (1986). The behaviour of nasalized vowels in the perceptual vowel space. In: Ohala, J. J., Jaeger, J. J. (eds.) *Experimental Phonology*. Orlando: Academic Press: 56-112.

Wright, R. (2001). Perceptual cues in contrast maintenance. In: Hume, E., Johnson, K. (eds.) *The Role of Speech Perception in Phonology*. New York: Academic Press: 251-277.

Zlatin, M. A. (1974). Voicing contrast: Perceptual and productive voice onset time characteristics of adults. *Journal of the Acoustical Society of America* 56: 981-994.

Zue, V. W. (1976). *Acoustic Characteristics of Stop Consonants: A Controlled Study (Technical Report 523)*. Lexington: Lincoln Laboratory, MIT.

Streszczenie

Summary in Polish

Percepcja angielskich i polskich spółgłosek właściwych

Praca niniejsza koncentruje się na kontraście dźwięczna-bezdźwięczna w percepcji angielskich i polskich spółgłosek właściwych. Metodologia badań oparta została na manipulacji akustycznej parametrów temporalnych i spektralnych, które biorą udział w implementacji kontrastu dźwięczności w badanych językach. Porównane zastałych trzy grupy badanych – początkujący uczący się języka angielskiego, zaawansowani użytkownicy języka angielskiego, oraz rodowici mówcy języka angielskiego. Praca składa się z dwóch części teoretycznych, ilustrujących problematykę i kontrastujących strategie implementacji kontrastu dźwięczności w badanych językach, oraz części badawczej, prezentującej zastosowaną metodologię badań oraz analizę wyników.

Część pierwsza porusza problem roli percepcji mowy w badaniach językoznawczych. Dotyka takich aspektów jak brak bezpośredniej relacji między sygnałem dźwiękowym a kategorią fonologiczną, wyjątkowa plastyczność i zdolność adaptacyjna ludzkiej percepcji mowy, oraz referuje propozycje dotyczące kompleksowego opisu działania ludzkiej percepcji mowy. W kolejnych podrozdziałach praca omawia percepcję w kontekście kontaktu językowego, a więc rozróżnianie kontrastów akustycznych występujących w języku obcym, ale nieobecnych w języku pierwszym. Zostają również zrecenzowane modele, które taki proces opisują, jak i hipotezy opisujące potencjalny sukces w opanowaniu efektywnej percepcji kontrastów percepcyjnych występujących w języku obcym.

Część druga koncentruje się na różnicach temporalnych i akustycznych w implementacji dźwięczności w języku angielskim i polskim. Opisane zostają aspekty takie jak; Voice Onset Time, długość samogłoski, długość zwarcia, długość frykcji, ubezdźwięcznienie, długość wybuchu.

Cześć trzecia, badawcza, prezentuje materiał poddany badaniu, metodologię manipulacji materiału, oraz charakterystykę grup. Hipotezy oparte na założeniach teoretycznych są następnie weryfikowane przy pomocy otrzymanych wyników. Część końcowa omawia problemy percepcyjne, jakie spotykają Polaków uczących się języka angielskiego oraz wyciąga wnioski pedagogiczne.

Appendix 1: Forced choice format for Native Speakers

ENGLISH

1. **(k / g)**eef
2. **(s / z)**eef
3. **(k / g)**eef
4. **(ch / j)**eeth
5. **(s / z)**eef
6. **(k / g)**eef
7. **(s / z)**eef
8. **(ch / j)** eeth
9. **(k / g)**eef
10. **(s / z)**eef
11. **(f / v)**os
12. **(k / g)**eef
13. **(s / z)**eef
14. **(k / g)**eef
15. **(ch / j)** eeth
16. **(s / z)**eef
17. **(k / g)**eef
18. **(f / v)**os
19. **(k / g)**eef
20. thee**(p / b)**
21. tho**(k / g)**
22. hee**(tch / dge)**
23. thee**(p / b)**
24. hee**(s / z)**
25. tho**(k / g)**
26. thee**(p / b)**
27. hee**(s / z)**
28. hee**(tch / dge)**
29. thee**(p / b)**
30. tho**(k / g)**
31. hee**(s / z)**
32. thee**(p / b)**
33. tho**(k / g)**
34. hee**(s / z)**
35. thee**(p / b)**
36. tho**(k / g)**
37. hee**(tch / dge)**
38. hee**(s / z)**

Appendix 2: Forced choice format for Beginner Learners

ENGLISH

1. **(k / g)**eef
2. **(s / z)**eef
3. **(k / g)**eef
4. **(cz / dż)**eeth
5. **(s / z)**eef
6. **(k / g)**eef
7. **(s / z)**eef
8. **(cz / dż)**eeth
9. **(k / g)**eef
10. **(s / z)**eef
11. **(f / w)**os
12. **(k / g)**eef
13. **(s / z)**eef
14. **(k / g)**eef
15. **(cz / dż)**eeth
16. **(s / z)**eef
17. **(k / g)**eef
18. **(f / w)**os
19. **(k / g)**eef
20. thee**(p / b)**
21. tho**(k / g)**
22. hee**(cz / dż)**
23. thee**(p / b)**
24. hee**(s / z)**
25. tho**(k / g)**
26. thee**(p / b)**
27. hee**(s / z)**
28. hee**(cz / dź)**
29. thee**(p / b)**
30. tho**(k / g)**
31. hee**(s / z)**
32. thee**(p / b)**
33. tho**(k / g)**
34. hee**(s / z)**
35. thee**(p / b)**
36. tho**(k / g)**
37. hee**(cz / dż)**
38. hee**(s / z)**

POLSKI

39. **(k / g)**ir
40. **(p / b)**ir
41. **(f / w)**os
42. **(cz / dż)** ir

Appendix 3: Forced choice format for Advanced Learners

ENGLISH

1. **(k / g)**eef
2. **(s / z)**eef
3. **(k / g)**eef
4. **(tB / d|)**eeth
5. **(s / z)**eef
6. **(k / g)**eef
7. **(s / z)**eef
8. **(tB / d|)** eeth
9. **(k / g)**eef
10. **(s / z)**eef
11. **(f / v)**os
12. **(k / g)**eef
13. **(s / z)**eef
14. **(k / g)**eef
15. **(tB / d|)** eeth
16. **(s / z)**eef
17. **(k / g)**eef
18. **(f / v)**os
19. **(k / g)**eef
20. thee**(p / b)**
21. tho**(k / g)**
22. hee**(tB / d|)**
23. thee**(p / b)**
24. hee**(s / z)**
25. tho**(k / g)**
26. thee**(p / b)**
27. hee**(s / z)**
28. hee**(tB / d|)**
29. thee**(p / b)**
30. tho**(k / g)**
31. hee**(s / z)**
32. thee**(p / b)**
33. tho**(k / g)**
34. hee**(s / z)**
35. thee**(p / b)**
36. tho**(k / g)**
37. hee**(tB / d|)**
38. hee**(s / z)**

POLSKI

39. **(k / g)**ir
40. **(p / b)**ir
41. **(f / v)**os
42. **(tB / d|)**ir