Interference in Early Acquisition Dutch-English Bilinguals: A Phonetic Examination of Voice Onset Time in Dutch and English Bilabial Plosives

Student: Mary Bless
Student ID: 10619100
Supervisor: Silke Hamann
MA General Linguistics
Date: 5 July 2015

Table of Contents

1. Introduction

There is much interest and debate about how language is acquired in bilingual children. There are several different factors and levels of language which play into this debate, but this paper will seek to look at very small, measurable phonetic differences. This paper will examine the phonetic realizations of bilabial plosives, specifically Dutch [p] and [b] and English [p] and [b] in adult bilingual speakers who were raised speaking both languages, all before the age of 6. While [p] has the same phonological definition in all languages of being a voiceless bilabial plosive, a sound produced by creating a small burst of air with the parting of the lips, subtle characteristics of the sound are unique to different languages. Voicing is a feature used in phonetics to describe the presence of low-frequency periodicity produced by the vocal folds during a sound. Voicing alone is a way that sounds produced in the same location can be distinguished from each other. Voice Onset Time (VOT), the duration between a burst and the beginning of voicing in a plosive, and aspiration, a substantial puff of air as a result of the burst which can be measured by positive VOT, are examples of those subtle distinctions and they are measurable aspects that can be compared between speakers.

If the Dutch and English [p] and [b] sounds are listed in order of which sounds have VOT starting first on a real-time continuum, beginning with the longest negative VOT and ending with the longest positive VOT, the hierarchy is as follows: Dutch [b], English [b], Dutch [p], English [p]. The voicing begins earlier in both plosives in Dutch, going so far as to have voicing beginning even relatively long before the [b] is even released in a burst. According to Lisker and Abramson, Dutch /p/ can have an average VOT of 10 milliseconds and English /b/ can have an average VOT of 1 millisecond (Lisker and Abramson 1964). The question of whether Dutch-English early acquisition bilinguals will have completely separate realizations of these two sounds and whether they will have equally separate categories for the other sounds mentioned will be explored here.

This is the phonetic level, and the experiment discussed in this paper is aimed towards understanding more about the phonetic and phonological inter-workings between the two languages in the bilingual brain. Phonetic expressions are often used to determine the underlying phonological systems in place in the speaker's mind. If a bilingual person pronounces [p] differently in each language, do they process two different underlying forms of that consonant in their two languages, or are a long VOT [p] and short VOT [p] both simply /p/? This question can be answered in different ways. There will be many different realizations of [p] that come from one speaker speaking one language, so phonetically, it may appear that the underlying form /p/ may vary in acoustical measurement, but maintain the meaning it holds. However, the phonetic realizations of [p] by one speaker in one language depend also on the sound's interaction with surrounding sounds, which is dictated by the overall phonological system of the language. Then, a speaker may have one [p] in their first language (L1) and that sound exerts influence and has influence exerted upon it by the other sounds present in that language. The other [p] will undergo the same process in the speaker's second language (L2). For instance, if a speaker's L1 does not allow /p/ to precede /s/, then certain phonetic realizations of /p/ will not occur in that language. If that speaker's L2 allows that consonant sequence, then variants of /p/ will exist in the speaker's L2 that do not exist in their L1. Since the two /p/s cannot interact with surrounding sounds in the same way, are they stored and processed in the brain as the different sounds and different meanings? In dual first language bilinguals, this phonological distinction is interesting to study since these people had the opportunity to possibly develop two separate categories at the same time, unlike those who learn a second language late enough in life for first language interference to be immediately obvious.

One issue with bilingual studies is the different definitions of what constitutes a bilingual speaker. Du (2010) refers to a definition of bilingual first language acquisition put forth by Mclaughlin (1984), which holds that these speakers learned both languages before the age of 3. Du utilizes the definition of speakers that learned both languages from birth. In the present study, the previous definitions of bilingual, first language learner were taken into consideration when deciding how to define dual first language bilinguals, but there were limitations to the available participants. Bilinguals who considered themselves native speakers of both Dutch and English and learned both languages before the age of 6 with unceasing sources of both languages were used in the current study and were considered to fall under this dual native language bilingual definition. This definition may be broader than some, and the issue of defining bilinguals, especially dual first language bilinguals is discussed further in the Theoretical Background section (2.2). Another distinction to be explored in this paper is the effect of native speaking parents and environment.

For this paper, an experiment was conducted to determine the average VOT of Dutch and English [p] and [b] in dual native language bilinguals. Dutch-English bilinguals were recorded in a studio reciting a passage in both languages and they also filled out a questionnaire to explain their individual linguistic backgrounds. All of the bilabial plosives in the recordings which adhered to standard criteria, outlined in the Data and Methods section, were measured for VOT duration, recorded, and analyzed for average VOT lengths. These averages were then used to compare the speaker's productions to former averages found in relevant literature as well as to other speakers in the study. The goal was to understand whether bilinguals actually realize these sounds differently in each language or not, and whether they pronounce them the way monolingual speakers of each language do. Monolinguals were not used in this study for comparison; however, monolingual VOT averages from previous studies were used. The hypothesis for this experiment is that the English VOT productions will exhibit signs of interference, since the participants currently live in a Dutch environment and are exposed to many Dutch-accented realizations of English. The majority of Dutch people speak English and English is used quite regularly in the Netherlands. How input leads to interference that can affect speech in bilinguals is discussed in the Theoretical background section.

The current study also aimed to learn whether Dutch-English bilinguals in fact use four separate categories for voiced and voiceless bilabial plosives across the two languages. Alternatives to this scenario and exceptions are explored, as there are several factors at play that influence bilingual acquisition, especially dual first language acquisition. Along with acquisition, the speakers' maintenance of both languages must be addressed, and how their environment may change the way they speak. It has been proven in a previous study that speakers with different accents in a language may pick up features of their language partners' speech (Miller, 2010). It is possible that even native speakers of a language exhibit small, sometimes even barely perceptible, changes to their accent. In the Netherlands, most of the population are Dutch-native speakers who speak English as a second language, and therefore usually with an accent. The environment is a possible influence on the participants in this research, which warranted a survey of background information from the participants. The presence of both languages in the environment of the bilinguals studied here could influence either of their languages, both, or neither.

While there are many differences between Dutch and English phonetically, these particular sounds are interesting to compare because in Dutch, the [p] has been recorded as having such a short VOT that it is similar to the English [b]. VOT is an important distinction especially because voicing in one consonant with very broad use can also affect the surrounding sounds and the overall phonology of the language. To be clear, the values sought and analyzed statistically were the voice onset times of phonetic realizations of Dutch and English [p] and [b] laid forth in a strategically organized word list. These values were important to look at on an individual level as well as a group level. The data and methods section provides mainly group statistics and findings and some individual speaker tendencies are discussed, especially exceptions to the main findings.

Similar studies conducted in the past were consulted for methods of analyzation and elicitation of samples in the current experiment, including research by Lisker and Abramson (1964) and Caramazza (1973). These also provided average VOT measurements which were helpful for comparison to the data collected in the current study. These sources were consulted often throughout the paper and their points are very relevant in the discussion of the acoustical aspects of these stop consonants and measurements thereof.

As an outline of what will follow, next is the Theoretical Background section (2), which will include definitions of technical terms that will be used throughout the paper and a brief overview of voice onset time and how it functions. A brief review of glottal operations is also outlined in the discussion section, in regards to the physical mechanics behind the production of these sounds. Information about the acoustic properties of the target phones that are examined in the experiment will be given with supporting literature, including previous studies of stop consonants and VOT. The Lisker and Abramson study mentioned above as well as a study by Caramazza et al. which held experiments to examine VOT in French-English bilinguals in Montreal are among sources that have previously contributed to the specific field the current experiment deals with. Finally, this section will cover the topic of bilinguals, specifically first language bilinguals, and language contact phenomena, since here it is hypothesized that these Dutch-English bilinguals tested in Amsterdam will show interference from Dutch onto English productions of bilabial plosives. The discussion of bilinguals will also include descriptions of different models that describe the speech perception and production processes to an extent.

Following the Theoretical Background is the Data and Methods chapter (3), which will provide information about the subjects of the experiment, details of how the experiment was carried out and the rationale for the methodological practices used for data collection. Also discussed in this section are the procedures for data analysis, including the measurement conventions employed specifically in the Praat software (Boersma 2013), and how the data was divided into sets for comparison. Here the reader will find how the data set evaluations were based on the target sounds' surrounding sounds rather than word position and an explanation for why the type of speech sample elicited, a recited passage, warrants such treatment of the data.

The Results chapter encompasses the outcomes of the experiment, namely the VOT averages across categories for each speaker. The averages of individual speakers are compared to each other taking into account personal background information, such as whether a speaker learned both languages simultaneously or sequentially. Also taken into account are whether speakers' VOT averages are evenly distributed in voiced and voiceless modes, and whether each speaker's VOT averages adhere to previously recorded averages in each language or if the speakerse lean toward one language or the other. Here the direction of interference is also discussed, especially in terms of the data.

Finally, the discussion section will align the findings of the experiment with the current literature. The limitations of this experiment will be discussed in light of which ways the experiment could be extended to be more conclusive or useful in understanding bilingual first language acquisition. The findings of the experiment will be reviewed and the interpretation of an interference effect will be determined and tied to the hypothesis stated here, that Dutch-English bilinguals residing in a Dutch-dominant environment will show some interference in their production of English phones, voiced and voiceless bilabial plosives [b] and [p].

2. Theoretical Background

2.1 Voice Onset Time and Bilabial Plosives

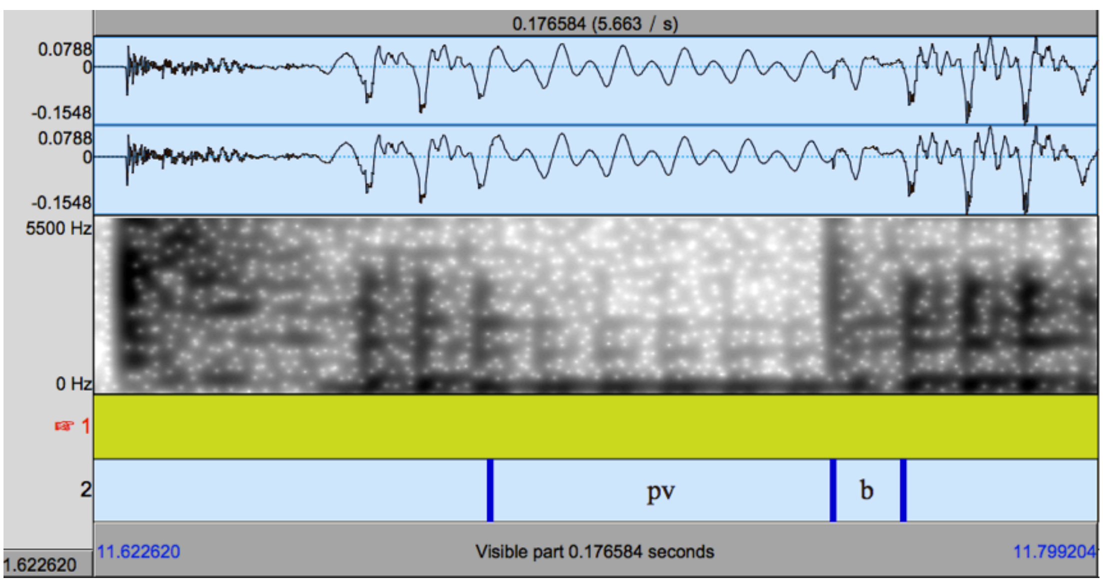
Beginning this chapter, is an overview of the main variable being analyzed in the experiment, Voice Onset Time (VOT). Following this small section will be a review of literature on bilingual first language acquisition and the stop consonants to be examined in this research, produced by bilinguals, the majority of which are, by many standards native speakers of both languages.

First, a clarification of the terms in use in this paper and some of the inter-workings involved in speech production as well as how those mechanics contribute to constructing current understandings of the bilingual mind and language processes.

The voice onset time (VOT) is a feature of a phone, or single expression of a phoneme. A phoneme carries phonological meaning while the term ‘phone’ refers to the phonetic segment that is measurable and recordable. Unlike phonemes, which are part of a more abstract, theoretical structure, linguists are able to obtain recordings of phones, speech sounds. These can be analyzed for a myriad of characteristics from the recorded speech signal to physical attributes of the speaker who produced them. That structure is based on observations of phones, individually and of course as part of larger signals.

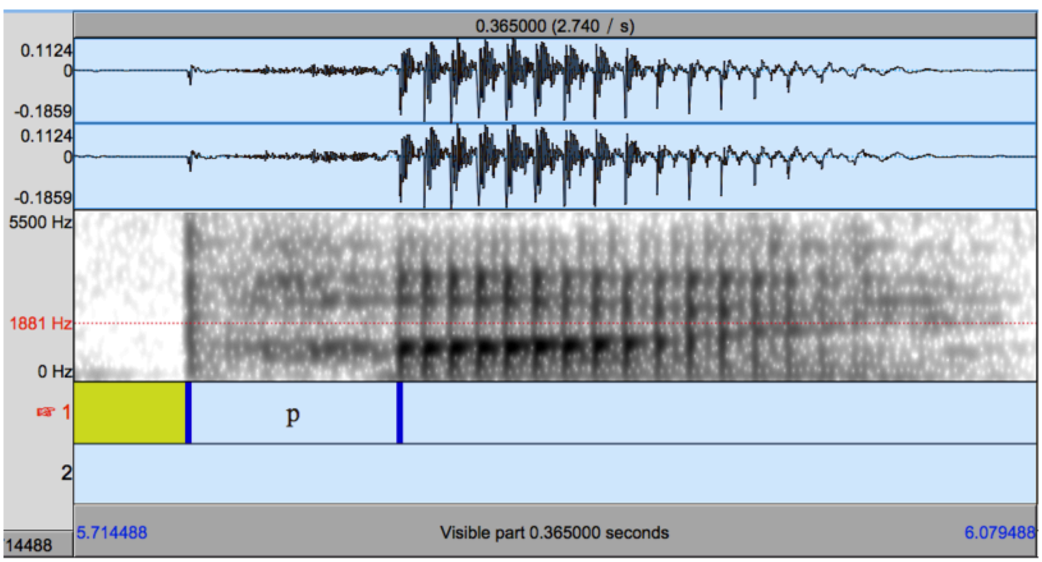
As far as what VOT physically represents, Caramazza et al (1973) give this concise, technical definition: “VOT stands for the temporal relation between changes in the glottal aperture and the supraglottal gestures; and acoustically, it is realized as the timing difference between the release of the stop occlusion and the onset of quasi-periodical laryngeal vibrations” (Caramazza et al 1973: 421). Lisker and Abramson (1964) describe their process of measuring VOT, stating, “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 were the pattern shows an abrupt change in overall spectrum“ (p. 389). The glottis is the opening between the vocal folds, located between the lungs and the mouth. When air is pushed out of the lungs through the glottis with enough force to make the vocal folds vibrate, a sound signal is produced. The anatomical, non-acoustical definition states that VOT refers to the actual time between actions of the glottis and the actions taking place in the mouth, configurations of the tongue and/or lips, to mark the speech signal. The acoustical definition uses terms which refer to specific parts of the speech signal visible in a waveform of a stop consonant. As illustrated in Figure 1, taken from a Dutch sample by one of the participants in the current study, the "release of the stop" is visible in the spectrogram as the column of energy at the beginning of the section labeled 'b'. The frequencies are indicated on the y axis of the spectrogram, ranging from 0 - 5,500 Hz. The reason the release, or burst, shows up as a column of energy is because the signal at that moment is an emission of noisiness with energy at all observable frequencies. This corresponds to a small vertical dash extending shortly down from the waveform at the location of the column of energy.

Figure 1: Example Voiced Bilabial Plosive with Prevoicing Period



The "quasi-periodical laryngeal vibrations" are visible in the waveform as the repeating dual-peaked wave which occurs five times above the section labeled 'pv'. That section of the waveform is an example of prevoicing, because the voicing begins before the release of the burst. The two-hump wave is distinct from the preceding wave that has four peaks and occurs for two cycles, so it is confirmed that the two-hump wave is prevoicing rather than the expression of the vowel in the preceding word, since it should match the four-peak wave in that case. The two-hump wave is a quite common form of prevoicing, but the height of the humps can vary from speaker to speaker, as well as the shape in general. This prevoicing period occurs when the mouth is still closed, lips together, before the burst, which must be distinguished from preceding voiced segments when it comes to measuring the prevoicing of the voiced [b]. The prevoicing signal is also visible in the spectrogram running through the middle of Figure 1, as the dark line at the bottom of the spectrogram window. This dark area is the visualization of low frequency energy, or voicing. One can contrast this voicing with the voiceless period at the beginning of the window, where there is no dark area at the bottom of the window. That is because a [t] is produced there, so there is energy at the higher frequencies, which indicates noise, and often aspiration, in the signal.

Figure 2: Example of Voiceless Bilabial Plosive



Illustrated in Figure 2 is an example of VOT that begins after the release of the stop consonant. English /p/ is shown here, this picture also taken from a recording made for this study. Visible at the beginning of the ‘p’ section is the same type of column of energy discussed in the explanation of the /b/, as well as the abrupt, vertical drop in the corresponding location in the waveform. This is where a release of the lips occurs along with a puff of air and the voicing does not start until the periodic wave, or wave which repeats in the same pattern. In the waveform, the aspiration of the [p], following the burst, is recognizable by the lengthy segment of noisiness. The pictured excerpts also serve to show how the VOT was measured in the current research. The burst in both cases of [p] and [b] was always marked at the precise place of the drop in the waveform. The voicing was consistently marked at the first ascending slope of the first periodic wave, regardless of whether the voicing began before or after the burst, since the VOT may be positive or negative in relation to the stop consonant release.

What are some characteristics of the sounds /p/ and /b/?

In order to make a fair assessment of the English VOT values, it is important to take into consideration some phonetic features of the bilabial stop consonants as expressed in English. Yavas delineates /p/ and /b/ as having fluctuating voicing properties based on their position in a word (Yavas 2006). While /p/ tends to be more constant since it is always voiceless in any position, /b/ can vary in voicing, but importantly, voiced /b/ occurs consistently in intervocalic positions. Yavas states further that in English, /b/ typically has little to no voicing in initial position, but to clarify, devoicing is much more prominent in final position than initial. This provides support for the sought after differences between Dutch and English bilabial stops investigated in the experiment. To easier classify the differences between /p/ and /b/ it is also possible to apply the terms fortis and lenis. Fortis stops like /p/ are "pronounced with more muscular energy (force), higher intra-oral pressure, and a stronger breath effort than their lenis counterparts," (Yavas 2006: 58) like /b/ . Many early sources also used the terms fortis and lenis, as discussed by Lisker and Abramson (1963), but these terms are not as specific as VOT measurements when it comes to understanding voicing. However, in the case of [p], the aspiration following the burst can be longer based on this component of articulatory force. Articulatory force, voicing, and aspiration are the three components of stop consonants discussed in Lisker and Abramson (1963). It is imperative to recognize that these factors work in tandem, meaning the stronger the air current exiting the lips to create the burst, the longer the aspiration period, and in turn, the longer the VOT in milliseconds. As the surrounding sounds of the stop consonants in the experiment are also involved in the analysis, some tendencies outlined by Yavas are also important. When /p/ or /b/ follow an /s/, the /p/ will have a fortis production and /b/ a lenis production, and there is no aspiration after /sp/, meaning the VOT will be very short. Furthermore, another characteristic which helps to distinguish between /p/ and /b/ is the aspiration present after the voiceless variety of bilabial stops. Aspiration length and strength can be based on syllable stress, which varies from word to word and language to language. Plag et al. state “the phonetic correlates of stress in English are usually given as pitch, intensity, duration and vowel quality, with stressed syllables tending to have higher pitch, higher intensity and longer duration” (Plag et al., 2011). Moreover, in words with an initial syllable stress, the bilabial plosive may be pronounced more strongly with a more intense air stream and consequently a longer VOT in the case of voiceless bilabial plosives. Yavas adds that when two stop consonants occur near each other, the first stop consonant does not release before the second stop consonant is produced.

There are also some differences in the orthography of the language which should be addressed. What makes this case interesting is that similar sounds are captured by the letters p and b, but those sounds typically have different characteristics in the two languages, specifically in the VOT feature. Dutch has a shallow orthography, or an orthography in which the symbols closely reflect pronunciation. English on the other hand has a deep orthography, wherein the pronunciation of words is less correlated with the spelling conventions of the language. There is no orthographical evidence that the sounds should be produced any differently, even in Dutch, wherein other nuances of pronunciation are more prominent in the written form than in English. Since the participants were in fact reading the texts aloud, the orthography could have had some minor influence on the reader, or the style of pronunciation in general could be considered more relaxed in English, as the orthography is less correspondent to spoken forms.

Lisker and Abramson are responsible for very in-depth and typological discussion of stop consonants /b d g/ and /p t k/ with references to several different languages. They refer to three different features that are exhibited in the production of /p/ and /b/ in English, voicing, aspiration and articulatory force (p. 385). Their 1964 study aimed to understand voicing and voicing onset distinctions across a handful of languages that were reported to have distinct VOT for the same phones. English /p/ and its allophones are still represented by the overarching phoneme /p/ and share base identifying characteristics with /p/ and its allophones in Dutch. The other features of the sounds make them actually variants of each other rather than simply being /p/s across the board. Both English and Dutch are classified as two-phoneme languages in the study, indicating that it is appropriate to analyze only the use of /p/ and /b/ in the current study's analysis of English and Dutch, as opposed to the possibility of three- and four-phoneme languages like Korean and Hindi, respectively.

2.2 Bilingualism

What types of bilingualism will be addressed here?

Bilingualism is a general term for those who speak two languages, but many other characteristics may become tacked on to this term to be more specific about the type of bilinguals at hand. The term bilingual bears the implication that the speaker is fluent in both languages, but does not necessarily imply that the speaker has no accent in either language. The goal in this study was to target those who are least likely to have an accent in either language, those who consider themselves native speakers, and to determine whether there is some accent present and how come.

Here, the study deals with early acquisition bilinguals, who acquired both languages at a very young age, before the end of the Critical Period. Early acquisition bilinguals is a term that also encompasses bilingual first language learners. The first group describes both simultaneous and sequential young learners while those who are considered to bilingual first language learners really implies only simultaneous learners. Simultaneous bilinguals are those who learn both languages at the same time. Sequential learners learn one language after the other, and this term can be applied to L2 learners of any age. Vihman (2002) takes a points to the issues in determining when a learner leaves the simultaneous learner category and enters into the sequential learner category. She suggests that as motor skills develop in a young child, so does her or his ability to acquire language. This could explain different people having different critical periods in terms of age. She writes, “the gradual increase in motoric skills, range of ‘vocal motor schemes’ (VMS) or phonetic patterns that the child can produce at will, depends in part on practice, that is, on recurrent production of the same schemes,” and she cites McCune & Vihman 2001. For the study at hand, this means that it is nearly impossible to pin down exactly which bilinguals had more output in either language. Although some background information was collected, it is not enough to make these assessments. This information also provides some insight on the process that may influence the development of one or two phonologies in the bilingual learner, which in turn effects the outcome in the examination of adult VOT productions of bilinguals.

Do early acquisition bilinguals develop both languages at once?

Genesee explains the Unitary Language System Hypothesis, in which BFL learners initially have one language system and then separate their two languages later on. He later explains why this is probably not the case, citing many studies which found that BFL learners show language-specific traits in each language early on. The language-specific traits they display match the traits displayed by monolingual language learners. He goes on to say that generally, bilingual children “exhibit the same rate of morphosyntactic development as monolingual children, at least in their dominant language” (Genesee 2006 p.5). Reports show that even bilingual children with specified language impairments will reflect monolingual children’s impairments in the same age groups.

Genesee also discusses language dominance, stating that bilingual people may show features of their dominant language in their weaker language, and that situation is more common than attributes of their weaker language being exhibited in their dominant language. The dominant language in bilingual children is typically the first language they learn. Instances of this cross-linguistic transfer are noted by Genesee to only take place under certain circumstances and with certain grammatical structures.

When children use one language more they tend to have a bigger vocabulary in that language. When it comes to the translation equivalents in both languages, some believe that since bilingual children have two words for the same referent, it is proof of the dual language system.

Do early acquisition bilinguals acquire two languages and phonological systems at the same time, or one language and phonological system that eventually separates?

Bilingual children frequently code-switch and mix languages early on in the acquisition phase. They are typically corrected and encouraged to complete full ideas in one language or the other. Whether this is true of their pronunciation of certain sounds is another case. When the sounds are similar enough, children may not develop two separate categories for sounds, [p] and [b] in this case, for the efficiency of storing less sounds. However, bilingual children and children exposed to multiple languages within the first year are more sensitive to distinguishing non-native speech sounds than their monolingual counterparts (Saffran et al. 2003). Does this mean that these children then have an ear for distinguishing sounds in their own languages and are more likely to do so? Besides those possibilities, it may also be the case that the Dutch [p] and [b] have phonological translations that cannot be replaced by the English [p] and [b], and visa versa.

Lihong Du provides many historical understandings of bilingual acquisition. He references the 1978 study by Volterra and Taeschner which outlines three stages of bilingual acquisition. In the first, the children make no language distinction based on whom they are speaking with and frequently mix both languages into single sentences (Du 2010). In the second stage, they still show trouble with syntax, but achieve more translative equivalents in both languages, and in the final stage, they master all of the issues described (Du 2010). Syntactical errors are the most noteworthy expression of difficulties the bilingual children face in separating their languages (Du 2010). The question here becomes, does the dual first language bilingual child acquire all the different phonemes at the same time, or does the child mix the phonemes at the initial stage when they are not differentiating the two languages? One would assume that a native speaker of two languages uses both languages correctly because they acquired those languages from birth, but does the dual exposure actually ensure that the child does not hear enough isolated realizations of a sound in either language? Another possibility is that the child acquires a phoneme in one language and uses the phoneme of the other language as a variant of the initially acquired phoneme until a later stage when two completely distinguished phonemes exist in either language. In the current experiment, the end result is under scrutiny, whether speakers indeed have two distinct categories in each language.

One explanation that Genesee gives for interference in one of the two languages of a BFL (Bilingual First Language) person is the presence of interdependent development, as opposed to autonomous development (Du 2010). A later study performed in 1980 by Redlinger and Park supported the previous results in showing high language mixing rates in the early stages of acquisition. These researchers put forth that the bilingual child initially has one language (Du 2010). This is relevant to the study at hand, showing that on a broader scale, researchers have witnessed languages initially undifferentiated by children, in the use of words and phrases. Quay (1995) found that the first and second stage of the 1975 study have no boundary between them, based on a study of a bilingual child who did produce translational equivalents from the earliest stage. Lexically (Quay 1995) and syntactically (Meisel 2000), bilingual children do not begin with a single system before differentiating between the two languages (Du 2010). This suggests that most issues analyzed are not on the phonetic level, although there have been plenty of studies conducted on the bilingual productions of specific phonemes (Caramazza 1973, Flege 1995, Lisker and Abramson 1964).

Paradis (2001) found that French-English bilingual children in Quebec appeared to have two separate phonological systems, although those systems could influence each other (Du 2010). In their experiment, they tested monolingual French and English children and also bilingual children to see which language's phonological rules the children would apply to nonsense words. Context was important in the case of the bilingual children. All children, both the French monolinguals and French-English bilinguals, were very sensitive to producing the French nonsense words, but only the English monolingual children were very sensitive to producing English nonsense words with correct phonology.

2.3 Interference

What is interference?

Defined in a 2011 article by Francois Grosjean, transfer, interference, code-switching, and borrowing are all terms that describe phenomena that can affect bilinguals. These processes, which are also termed contact phenomena, are still a bit mysterious, and they refer to different relationships between languages that become manifested in speech. Grosjean's article explores the issues with studying and capturing instances of these processes when they actually occur. He differentiates between language "transfer," which should refer to "static phenomena," and language "interference," which "are linked to processing and have to be accounted for by encoding mechanisms" (Grosjean 2011: 12). In other words, transfer between two languages should be considered when a rule, construction, or category from one language is consistently expressed in the other language. Interference, on the other hand, would describe instances of the same thing, which appear inconsistently. For example, an adult learner of English may dentalize /t/, if she does not have alveolar /t/ in her native language. Then she has transferred a phonetic category from her native language to her second language. An example of interference would be if that same hypothetical speaker were to infrequently produce an alveolar /t/ in her native language without much pattern after becoming fluent in English. Grosjean calls transfer "permanent (static)" and interference "ephemeral (dynamic)" (Grosjean 2011: 12). Another part of language contact, of the individual speaker variety, includes "the permanent extension of meanings of words due to the other language, as well as specific syntactic structures that are permanently present (e.g. the constant misuse of a preposition)" (Grosjean 2011: 15). This elucidation on interference gives a clearer picture of how accents also fall under this interference category, since in the most basic terms, interference describes the use of a bilingual's first language to express their second language. Then, the first language may be expressed in the second language as the second language will not only have a different phonetic system, but a different phonological system. That is why the present research took age of acquisition into serious consideration, since sequential language learners show some differences from their simultaneous language learning counterparts. Examining the small details of phonetic production in the speech of Dutch-English bilinguals allows us to detect the extent to which interference has taken hold, and even whether it occurs in higher degrees in correlation to aging, even in children under 6.

Under the contact phenomena label, there are transfer, interference, code-switching and borrowing, as well as other meta-forms of language contact like pidgins and creoles. Transfer and interference are explained above, but as for code-switching, it should be defined as when a multilingual person uses more than one language during an interaction. This is in contrast with borrowing, which is when a person, not necessarily multilingual, uses words from a different language. This can occur on a meta-level scale as well. These are less relevant to the present research, but their possible presence in the recordings should be addressed nonetheless, since it is possible for at least code-switching to occur on a micro-level, in pronunciation. Possible influences on the language mode which could prompt instances of language contact are produced in the Discussion section with details from the present research methods.

According to Schulpen et. al., in “Interlingual Homophone Recognition” (2003), bilingual lexical access to both languages simultaneously are accessed “via a shared acoustic-phonetic input system” (p. F1157). The acoustic-phonetic input refers to speech that the bilinguals perceive. In bilinguals, the two separate lexicons are available for access but one lexicon will be accessed more readily based on the input, while possible targets in the other language lexicon are accessed to a lesser degree. This means that bilinguals will begin to access one language lexicon more than the other based on the type of phonetic cues they recognize and which language they recall them belonging to.

2.4 Phonetic Study of Bilingualism

How should the input (speech perception) and output (speech production) be modeled and understood?

There are two models which can help to address the production and perception in speech processing. The Exemplar theory model explains how the input itself, especially the volume of input, becomes stored in the brain, and eventually expressed. The Prototype model is also an important component of processing that works with the Exemplar concept. Walsh et al. succinctly describe it in stating, “at the core of exemplar theory is the idea that the acquisition of language is significantly facilitated by repeated exposure to concrete language input” (Walsh et al. 2010). It would be impossible to measure the amount of exposure each speaker had in their lifetime to each language, but we know that the participants were in Amsterdam during the time these recordings took place. The most recent and abundant exemplars for the majority of participants were Dutch productions of [p] and [b], even in both languages. Dutch speakers of English may provide dual first language Dutch-English bilinguals with exemplars that are not representative of how native English speakers actually produce such sounds. Walsh et al go on to explain that it is a “combination of extensive storage, similarity, frequency, and recency” which compose or sculpt the “exemplar memory which is in a constant state of flux” (Walsh et al. 2010). To expand on what an exemplar memory is, it is when a category in the mind is built upon several instances of similar inputs. A chair, for instance, may come in several shapes and sizes and all these individual examples of chairs are an input, or exemplar, that forms a sort of cloud of categorization. This idea can be applied in phonetics when a speaker of a language hears the same sounds over and over again, but with slightly different characteristics each time. This is especially helpful in linguistics since its nearly impossible to produce the exact same sound every single time one would like to, since the characteristics of surrounding sounds inhibit the process of completely uniform productions of a single sound. So, a [b] will sound different if it follows a vowel than it will if it follows a consonant. In exemplar theory, the [b] productions, with the varying characteristics still become part of the same “cloud” of productions. Walsh et al. add to this idea in reporting that phonetic exemplars “are categorized on the basis of their similarity to extant stored exemplars (using a variety of metrics), into clouds of memory traces with similar traces lying close to each other while dissimilar traces are more distant” (Walsh et al. 2010). At the time of recording, the Dutch language environment could have reinforced the static productions of [p] and [b]. This seems like it could be a fair explanation in the case that the Dutch values are more in line with each other or more stable than the English.

Exemplar theory could support the results here if the English productions of [p] and [b] are more variable even when it is the same sound produced in the same word by the same speaker. There are two more dimensions of exemplar and prototype theories which are very relevant to the current research, typicality and recency. Walsh et al. clarify that the “strength or activation of an exemplar is a function of its frequency and recency" (Walsh et al. 2010). Speakers living in an environment where they may often hear English productions by non-native speakers, especially Dutch speakers, could lead to variable inputs of the same sound which are then stored as exemplars with variation.

Exemplar theory and prototype theory are two theories that are discussed in psychology as ways that the brain organizes and recognizes information. They are two different types of ways that the brain may categorize input, or perceived speech, not only in linguistics but also in other forms of sensing information. Both theories also ultimately concern output, or produced speech, as well, but here the input aspect is initially discussed since the productions of [p] and [b] by the speakers in this study were likely influenced by the input from their environment upon one of their native languages, which is different from the native language of their environment.

Prototype theory is another theorization of the way in which the brain categorizes input. This theory works with exemplar theory, although there are areas in which the two theories diverge. Prototype theory concerns the "internal structure and the psychological organization of speech categories--the 'centers' of speech categories" Kuhl (1991: 105). In prototype theory it is put forth that categories are not so much a composite of examples, but that the examples reinforce the description already assigned to the category, or the list of features which distinguishes the category. Speech input is processed and sorted for its similarity to the prototype and the "members of a category form a gradient of typicality," (Lively 1997: 1665) meaning there exists a organized structure of likeness within the prototype category. For instance, the chair category may require four chair legs and a seat to be applied to incoming data. The prototype is a bit more rigid and unchanging than the exemplar model, but the exemplar model is sometimes faulted with being too flexible and less efficient than the prototype model. In the prototype way, the brain simply completes a comparison of the current input with the stored category, whereas the exemplar model allows less strict categorizations, and in that way, perhaps less helpful. In the case of language, the effect of accent change has been observed in those engaging in conversation with people with other accents, so it is quite certain that flexibility is more common in language. Language tends to be less like other items mapped in the brain, in that the modality already allows much flexibility. Unlike the mapping of a specific concept, a language and speech perception and production are so varied and complex that prototype theory requires a more complex application.

How and why should one examine VOT in bilingual speech to recognize interference?

According to Laura Bosch and Marta Ramon-Casas, “adequate categorization in bilinguals, that is, the ability to form separate categories for similar sounds in each of the languages in their environment, appears to be a phenomenon that is linked to input properties” (2011). In a way, finding if the bilinguals in this study will have four separate categories for four separate sounds will contribute to understanding whether there is a reasonable standard for bilinguals to be held to, when they learn both languages from birth. The expectations that could be met are that the bilinguals show signs of native speech by fully separating all four categories across the two languages. This does not take into account the recent events in their lives that could influence their accent and speech at the present moment. Antoniou et al. (2011) cite the Speech Learning Model (SLM), conceived by Flege, and use a part of it to describe why studying the phonetic realizations of bilinguals is important to understanding interference. They write, “according to SLM, bilinguals will differ from monolinguals of either languages because the phonetic categories used to produce and perceive both the L1 and L2 reside in a common acoustic-phonetic space, and will inevitably influence one another” (Antoniou et al. 2011). This is a quite intuitive suggestion, that unlike the acoustic-phonetic spaces of two monolinguals from two separate languages, who produce one or a few variants of each sound, the bilingual has not only the sounds of two languages, but also the variants of each of those sounds. The bilinguals also have two phonological systems to adhere to, meaning that the sounds in question, here /p/ and /b/, must interact accordingly within each language.

2.5 Research Practices

Data Elicitation and Language Modes

An interesting detail of the Lisker and Abramson study is how they elicited speech samples. Their method was to give participants a word and then ask the participants to produce the word twice and make a sentence with the word twice. They were instructed to speak “with the fluency and naturalness of normal conversation” (389). Later they explain that this resulted in two expressions of each word, where the word spoken in a sentence was more compressed than the expression of the word spoken alone. Current research aimed to attain natural and fluid productions as well, without notifying the participants about the target sounds.

On the topic of speech sample elicitation, consideration for the language mode is quite important. Language modes are defined as “the state of activation of the bilingual’s languages and language processing mechanisms at a given point in time” (12) (Grosjean 2011). Many academics, a few cited above, believe that bilinguals are not able to access only one language at a time. Instead it is suggested that both languages are activated in processing, but the non-relevant language is activated to a lesser degree (Genessee 2006, Du 2010). The language mode then, in reflection of that detail, pertains to both the relevant language or languages that the bilingual is operating with at a given time as well as the environment giving rise to the bilingual’s use of a certain language or languages. Antoniou et al. describes the duties of the researcher in the monolingual, also called “unilingual,” mode, determining “that all contact, instructions, and feedback occurred in only one language for a given speaker” (Antoniou et al. 2011). The implementation of this structure in the experiment suffices as far as is possible to preserving a monolingual mode in an experiment.

Grosjean conceptualizes language modes as points on a continuum, where a bilingual person is in an entirely monolingual environment (of either language) or a bilingual environment. The mode depends “on such factors as interlocutor, situation, content of discourse and function of the interaction” (Grosjean 2011 12). Those who are bilingual and know other bilinguals who speak their same languages will recognize the situation Grosjean illustrates of bilinguals using constructions, words, phrases, and pronunciations from both languages freely and fluently together as the bilinguals desire, evoking the image of a kind of Spanglish, which is neither Spanish nor English nor a language, but a mix of the two languages, that is spoken for pleasure rather than for necessity. At the other end of the spectrum, the monolingual mode, takes place when a bilingual is engaging with a monolingual person and suppresses their usage of constructions and vocabulary from the language that the other speaker is unfamiliar with. Of course these situations become more complex when the interactions are between the bilingual and a bigger group of people, such as in a classroom setting. This also takes into account that the bilingual is making assumptions about the other's linguistic background.

Although, of course, bilinguals will sometimes still show signs of transfer or interference in their speech when in the monolingual mode. Grosjean makes a fair argument that it is difficult for researchers to properly account for contact phenomena and languages modes in their research especially since it is a task in and of itself to isolate the instances of transfer and the instances of interference without extended exposure to a bilingual person. He goes on to explain that the problem may also lie with the researcher, who may shift the bilingual out of a monolingual language mode if they reveal that they have knowledge of the other language, so a researcher must show to the bilingual subject that they know only one language (Grosjean 2011). This was the case in the current research, as subjects were made aware that the researcher did not speak Dutch, but was a native English speaker. Arrangements to preserve monolingual modes in the recording sessions are relayed in the Data and Methods section. Fortunately, this aspect of the research was looked after, but the speech samples elicited were not inventions of the speaker, since the aim was to isolate and analyze productions of specific sounds.

Important factors to include in the linguistic background survey?

Flege (1995) refers to the age of learning (AOL) and length of residence (LOR) as important factors in the study of bilingual speech. Both of these components have been included in the data about the participants. Flege states that if subjects in a study have fully acquired and spoken their second language for more than five years, then the countries they have lived in will probably be less important to the study (Flege 1995: 99). "In a study by Flege, Munro, and MacKay (1995), native English listeners used a continuous scale to rate sentences spoken by the native Italian subjects and a group of native English speakers. AOL accounted for 60% of the variance in the Italian subjects' foreign accent ratings, whereas LOR accounted for very little variance (< 2%)" (Flege 1995: 99). The other finding of that study was that the Italian speakers could show a detectable foreign accent even when they learned English at a very young age, so they propose that the critical period is not the best indicator of whether or not an L2 speaker will have an accent. However, the chance of a person having a detectable accent increases with the AOL and all of the subjects in the study whose AOL was past 15 years of age had accents. From the literature on bilingual first language speakers, there seem to be several factors which impact how bilinguals' languages will be shaped. These children, unlike most of their monolingual peers, receive extra input and the way that they learn a language is very different in source and manner. In most cases, bilingual children either have parents with different native languages or grow up in one linguistic environment with parents or caretakers with a contrasting linguistic background.

3. Data & Methods

3.1 Methodology

Participants: Ten Dutch people took part in this study, six females and four males.They were compensated for their time. It was a requirement that they felt they were native speakers of both Dutch and English, but it was also a requirement that the participants had started learning both languages before six years of age. They were also required to have had a stable source through which they acquired each language during their early acquisition period, such as a native speaking parent or attending school instructed in the language, but sources like books and television were not considered efficient in this regard. One speaker who came to be recorded, outside the group of ten, told the researcher that he was a native speaker of English, but in person he had a clear accent both in pronunciation and in syntax. He stated that his source of English was television, which he had been watching since four years old. He was not used in the study since television is not a source that one interacts with. Although he may have had years of input from English television, the lack of actual communication meant that he did not receive feedback and correction the way other bilinguals in the study did. This was the only speaker that had to be discarded. Specific background information, provided by the speakers themselves, is available in the Appendix (7.2).

Two groups seemed to emerge, those who had learned the languages sequentially and those who learned both languages simultaneously. However, it seemed that the subjects mostly learned both languages before they remember speaking only one language. Five speakers learned both languages from birth and five speakers learned one language from birth and the other around ages 2-4. One participant started learning English at age 6; that was the largest gap in those who acquired the languages sequentially. All participants stated that they use both languages daily, except one who speaks Dutch once per week to his parents. Some of these participants spoke Dutch and English at home with one or both parents and some had used one language at home and the other in school. All of the participants had grown up at least partially in the Netherlands and the majority live in Amsterdam currently. Some had spent time in other countries including the United States and the United Kingdom, as far as English-Speaking countries go. Some spent time in other European countries for short periods.

Caramazza et al. note the distinction between compound and coordinate bilinguals in their description of participants, “the former category referring to those subjects who had learned both languages in the same social context, the later to those who had learned their languages in different contexts” (422) (Caramazza 1973). In their research, they found no distinctions between the two groups, neither in production nor in perception, and so they analyzed all the data from both groups together. Among the participants of the current research, only one did not learn Dutch in school, and which parent or parents the participants learned each language from varied greatly, but there seemed to be no evidence that the acquisition source was divisive enough to analyze the datas in separate categories.

Method: The research was carried out in the soundproof phonetics lab at University of Amsterdam. Participants were asked to agree to the Informed Consent Form, ensuring that their personal details would be anonymized in the research. The participants were made aware that they were to sit alone in the soundproof lab during the recording and all of the participants were comfortable with the arrangement. The first part, the English recording session, was conducted in English. The participants were given a short questionnaire to answer and then asked to read through the text to be read aloud when they finished the questionnaire. Once they read through the text, they were asked to read the text aloud three times. Once this was complete, the participants were instructed to answer the next questionnaire which was in Dutch and then to read through the Dutch text and recite it three times. The instructions were reiterated in English before the Dutch portion and they were requested to only communicate in Dutch after that point. All of the participants were successful with following the instructions and staying in either the English or Dutch language mode.

As described earlier, the use of different language modes has been shown in earlier studies to affect the way speakers speak (Antoniou et al. 2011, Grosjean 2011). It was important to begin in English, as the researcher does not speak Dutch, so the participants were given the thorough instructions and explanations in English from the start. For the first three speaker sessions, a Dutch speaker assisted to give the few instructions in Dutch during the Dutch mode portion, but for the rest of the sessions, that person was unavailable. In those cases, any necessary feedback was given in English before the switching to the Dutch mode and once the Dutch mode had commenced, any necessary instructions were given in either one-word Dutch form or using hand signals. The same hand signals, such as to signify “you may begin speaking,” were used in both language modes.

Materials:

In Figure 3, are lists of target words which were in the texts that the subjects were asked to recite. It is written in a casual, conversational style with nonsense subject matter.

Figure 3: Target Words in the Experiment Text

English		**Dutch**	
[p] initial	**[b] initial**	**[p] initial**	**[b] initial**
Pair	Bake	Paar	Bakken
Paws	Bear	Plukten	Beren
Peaches	Bears	Peren	Bedienen
Pears	Because	Perziken	Berenpoten
Pie	Began	**[p] medial**	Bakken
Pie	Began	Appelen	Begonnen
Prepare	Being	Honingpot	**[b] medial**
Put	**[b] medial**	Openen	Ijsbaren
[p] medial	Able	Berenpoten	Voorbereidingen
Apples	About		Gebak
Open	Raspberries		Aardbeien
Operate	Strawberries		Frambozen
Prepare			

The original text was in English of quite simple, common wording, just a short, made-up story. The Dutch text was a close translation of the English text and both texts included words that should elicit twelve tokens of each sound per speaker per language: [p] initial, [p] medial, [b] initial, and [b] medial. For instance, there were to be at least four words beginning with [p], four words beginning with [b] and so on. Originally, the goal of the research was to compare Dutch and English /p/ and Dutch and English /b/ in word-initial and word-medial positions. As for the surrounding sounds, the Dutch and English [p] and [b] sounds had similar placements although not exactly equal proportions. Originally, this research looked at /p/ in initial and medial positions of the word, whereas the Lisker and Abramson (1964) study looked at only word initial productions. Examining the sounds in different positions aimed to account for shorter VOT in the case of mid-word productions. However, more distinction lay between intervocalic and post-consonantal sounds, so the surrounding sounds were then analyzed. The target sounds were regrouped accordingly in both languages. In the course of composing the text, some sounds occurred more often than planned and so those sounds were included in the analysis. The sounds were placed in different contexts, so while many were intervocalic, some tokens were surrounded by glides, fricatives, and so forth. The sounds were only labeled when they occurred in the words which were on the original word list. In the Lisker and Abramson (1964) study, they stated that they felt it would be misleading to present a single average value that encompassed both positive and negative VOT values. However, the data in this paper was analyzed combining the negative and positive values for the sake of coming to a workable average number to use in comparisons directly between the two languages. During the analysis stage, some tokens had to be discarded across all speakers.

3.2 Analysis

Although the terms in the word list were carefully selected to make it possible to obtain the same number of tokens for each sound, some difficulties arose. Since the participants were asked to read a full text rather than word lists, the proximity of voiced sounds caused occasional instances of prevoicing in the phonetic implementation where it might not be phonologically bound. For instance, in the phrase “they began,” the preceding [eɪ̯] may have caused an extended prevoicing period in the following [b]. However, the use of a text read aloud allowed the closest simulation of natural speech that still ensured similar numbers of tokens for all speakers. Speaker averages were all compiled the same way and the averages of the different categories maintained this uniformity.

One English sound, the medial [p] in “optimistic,” had to be discarded, as did the corresponding sound in the Dutch counterpart word, “optimistisch,” which was also discarded. The other terms, “optimistic” and “optimistisch,” were produced as plosives by all the speakers. However, the /p/ is followed by a voiceless /t/ and therefore, regardless of position, no voicing or VOT occurs as a direct, measurable characteristic of the [p]. The medial [b] in “strawberries” also elicited many approximations but there were enough actual productions to keep from discarding the word altogether. In the case of “strawberries,” participants often produced the [b] as a voiced fricative approximant. This may have been due to the lack of phonological significance of the /b/ in this position. Since, for the word “strawberries” there are no homonyms or otherwise closely pronounced English words, it may be that the speakers cinched the consonant cluster in favor of efficiency. Also, some speakers approximated several sounds, so the number of tokens varied from speaker to speaker. In Table 1, the number of approximated sounds per speaker per language is shown. It is also important to note that using the same text for all of the speakers to recite, rather than using spontaneous speech, ensured that the type of sounds recorded would be uniform. The speakers should have used similar stress patterns for words in the text, although some speakers spoke faster than others, which led to more approximations in some recordings.

The participants also had individual styles of speech. Some speakers approximated the bilabial consonants more than others. Some consistently approximated the same sounds while others tended to approximate sounds more in general, as shown in Table 1. An expanded version of Table 1 is available in the Appendix (7.3 and 7.4). Most speakers had prevoiced /b/ in both languages.

Table 1: Approximants per speaker per sound per language:
Pi - /p/ initial
Pm - /p/ medial
Bi - /b/ initial
Bm - /b/ medial

Speaker #	English				Dutch				Total per speaker
	Pi	Pm	Bi	Bm	Pi	Pm	Bi	Bm	
1	0	0	4	8	0	6	2	6	26
2	2	0	1	3	0	0	0	0	6
3	0	0	3	1	0	1	1	1	7
4	1	0	5	7	0	2	0	4	19
5	1	1	4	8	0	0	1	5	20
6	3	0	0	5	0	0	0	0	8
7	1	0	5	8	0	1	0	4	19
8	0	0	5	6	0	0	2	4	17
9	1	1	1	2	0	0	0	2	7
10	0	1	2	1	0	0	0	0	4
Total per category	9	3	30	49	0	10	6	26	

Originally, the objective of this research was to examine Dutch [p] and [b] versus English [p] and [b] produced in different positions, initial and medial. It became soon apparent, however, that exploration in this direction would eventually require corroborative evidence of contextual influence. When the data was analyzed in that light, intervocalic [b] productions versus those following consonants, the patterns became clearer. As McClelland and Elman (1986) write in “The TRACE Model of Speech Perception,” the characteristics of a sound can vary greatly depending on its context. Vowel place can play a role in the voicing of a sound, and the authors mentioned also cite “rate of speech,” “morphological and prosodic factors,” “stress contour,” individual “fundamental frequency,” and even the build of the vocal tract as possible factors in the realizations of sounds. There was certainly much variation in the productions of each speaker and in the speakers overall, but patterns did emerge when it came to those sounds in the context of being between two vowels, and those preceded by a consonant but followed by a vowel. Unfortunately, those individual factors listed were not controlled for in this study, although the uniformity of the speech sample may have been of some assistance in unifying the pitch contours of the participants.

To analyze the data, each speaker's recordings were segmented using the Praat software. The speakers' [p] and [b] realizations were segmented in separate sections using the text grid annotation tool within the software. The design of the research was in consideration of the abilities of the Praat software, since voicing is very visible in both visualizations of the speech. Speech samples can become very extensive when converted into waveform and spectrogram however, so it was important to keep the speech sample short and uniform. Using spontaneous speech would have made for a painstaking process of hunting individual sounds and cross-checking against transcriptions, whereas the short passage read aloud was able to simulate the natural fluidity of spontaneous speech while maintaining a repeating pattern that is more reliable to follow.

The [p] sounds were measured from the first disturbance, particularly the first substantial sharp drop in the waveform, until the onset of the first clear periodic wave period. The [b] sounds were measured for prevoicing, beginning at the onset of first periodic signal recognizable before the burst when present, as well as the burst period. Ultimately, the burst period data was discarded for the purposes of this research. All [b] sounds without prevoicing in both Dutch and English were given the VOT value 0.0. The values from all approximated sounds were discarded in calculations for averages. Only plosives were considered here, although the amount of approximated sounds per speaker is compared between languages. This method of determining VOT is line with the methods used by Lisker and Abramson in the study mentioned earlier. As for the prevoicing period in the current research, it was measured in the spectrogram as being present when a low frequency signal was visible and that low frequency corresponded with a periodic wave in the waveform. The exact point at which the prevoicing was marked as beginning was at the onset of the first periodic wave.

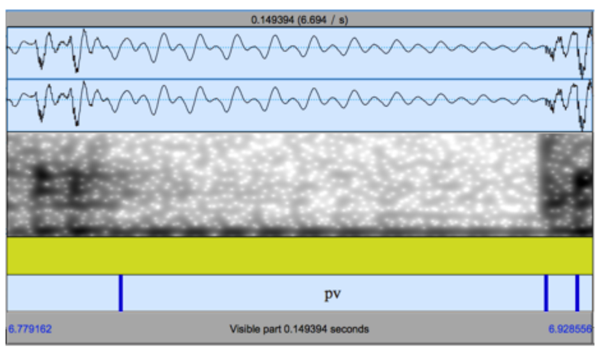
While labeling the [p] sounds, the first objective was to locate the burst of energy in the spectrogram, which appeared as a fricative, with high frequency energy and no voicing. For precision, the beginning and end of the portion were sought by matching the high frequency energy to the first corresponding, substantial interruption in the oscillogram, an abrupt, steep drop and then the onset of the first periodic wave period following. In the case of [p], this was usually the entire process necessary to determine the voice onset time.

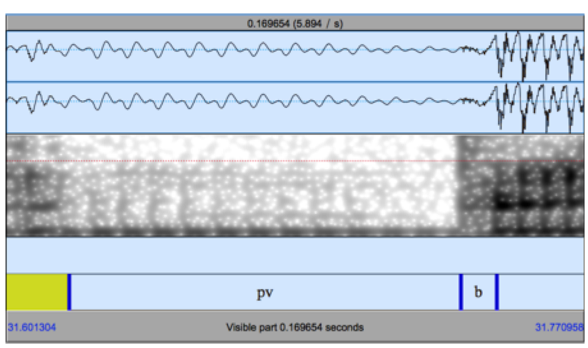
3.3 Individual Productions and Abnormalities

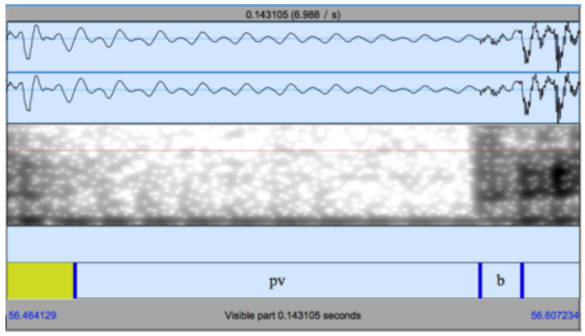
During the segmenting stage of this research, it became clear that individual differences had to be considered. In some cases, the speaker's prevoicing was very clear and consistent, while in other cases, the prevoicing could vary from word to word or from passage to passage.

Here are the three productions of the same Dutch /b/ from the same word in Figures 4-6. Speaker 6, a male, had a prevoicing period that consistently looked like this. Although, one can see that by the third iteration of the sound, the prevoicing has very slightly less pronounced first hump. In all three realizations, there are about 9-10 cycles of prevoicing before the burst.

Figures 4-6: Three Productions of Intervocalic Dutch /b/ by Speaker 6







Shown below, in Figure 7 and Figure 8-9, are the prevoicing periods of two other speakers, taken from Dutch recordings. While the prevoicing periods did not look the same for all speakers, the behavior of the prevoicing periods were similar enough to distinguish that they were comparable. The average VOT lengths tended to stay within a reasonable range of each other in each category. The sample from Speaker 4 and the two samples from Speaker 8 differ from those of Speaker 6. The two samples from Speaker 8 differ even from each other. These are just a smattering of the different prevoicing waves that were exhibited throughout the speech samples collected. While they are similar, usually composed of two-hump waves, there are clearly differences in intensity. Speaker 4 usually exhibited single-hump waves in the prevoicing period. For all of the speakers, during the segmenting phase, individual characteristics of the fundamental frequency were considered and the researcher looked for uniformity throughout each individual speaker's samples.

Figure 7: Speaker 4 Prevoicing Dutch /b/

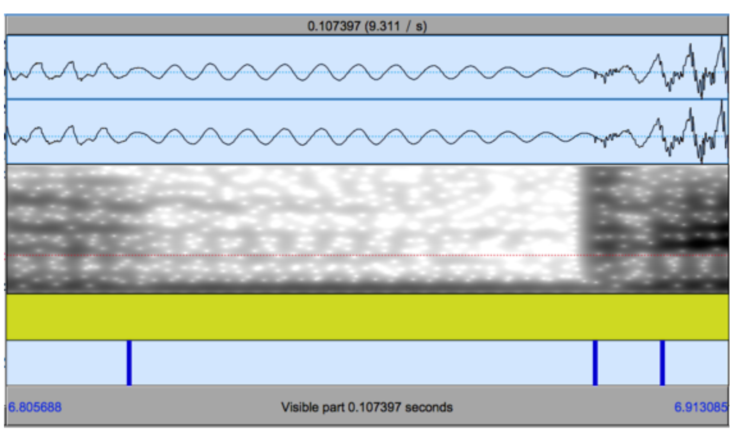
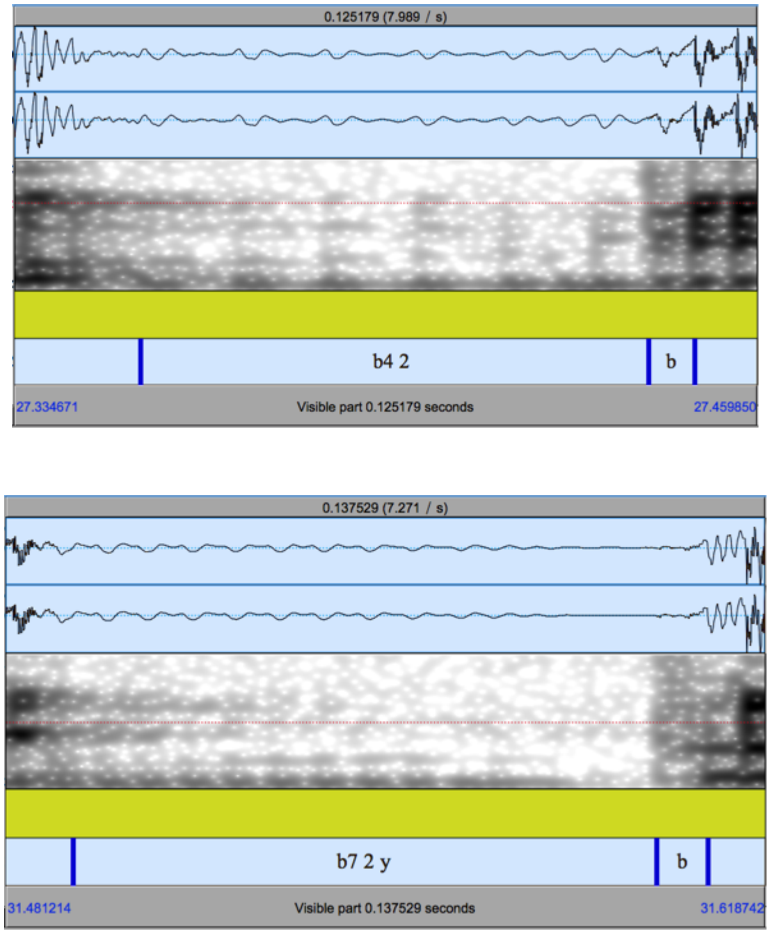
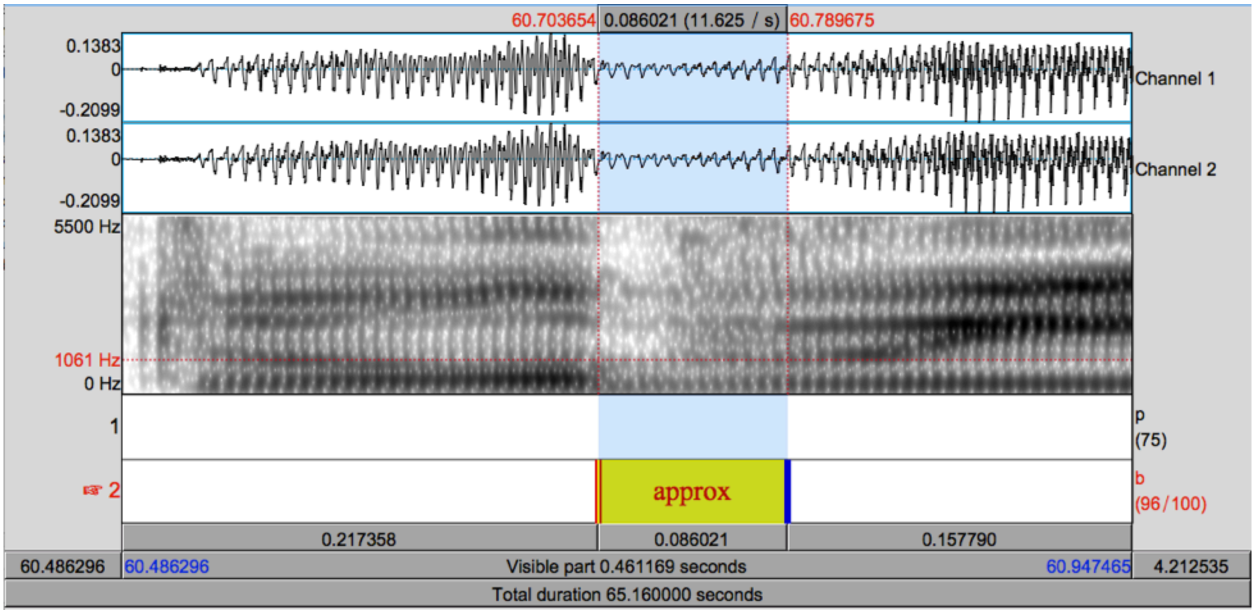


Figure 8-9: Speaker 8 Prevoicing Dutch /b/ in two different words



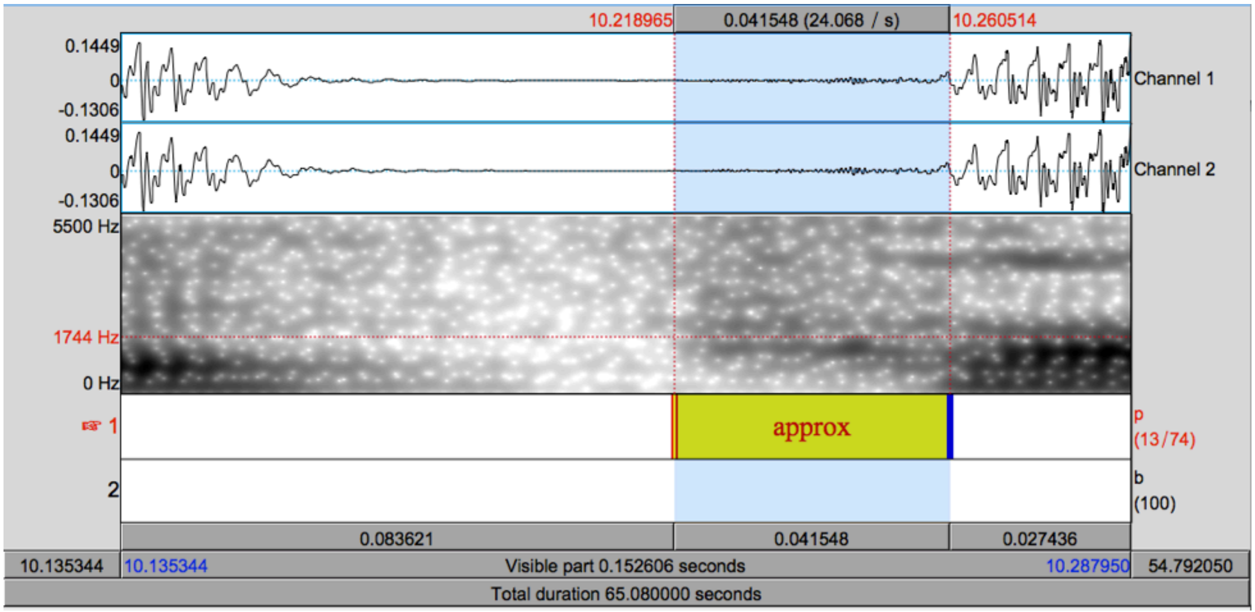
The medial [b] in English provided many issues and was in fact the most approximated sound of all the sounds studied here. In this research, the English [b] production that followed a consonant were much more likely to be approximated, meaning there was no actual burst, but only a near-closure of the signal as shown in Figure 10:

Figure 10: Approximated /b/ in Speaker 8's English production of "strawberries"



In the highlighted area, a /b/ should occur, but instead of the speaker bringing their lips together and subsequently releasing a small puff of air, the speaker brings the lips close enough together to nearly obstruct the speech signal to simulate a /b/. As the speech signal remains periodic, this type of approximant most resembles a semivowel. Most speakers produced at least one /b/ like this, especially in word medial position. This was the most common kind of approximation found throughout the speech samples and these instances were not included in the averages since there is no release to measure the relative distance of the onset time of voicing.

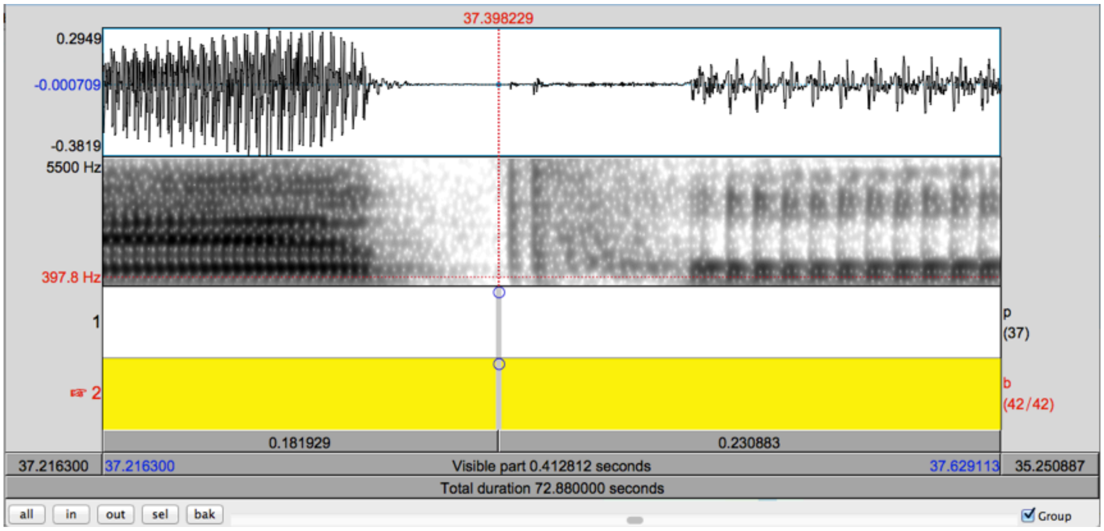
Figure 11: Approximated /p/ in Speaker 4's English production of "operate"



As for approximated /p/, the most common type exhibited was the [p] which lacked a burst. Most speakers would substitute a voiceless fricative for the plosive. These speakers managed to simulate the aspiration segment of the [p] by causing friction with air stream flowing between nearly closed lips, but not actually closing the lips so as to build enough pressure for a release of air. This type was common in English in the word initial position preceding an /r/ in the word 'prepare'.

Without a burst, it is impossible to measure VOT, so only plosives were included in the research. Instances like those pictured above were left out. A more perplexing type of expression that was most common in Dutch [p] productions were double burst [p]s as shown below in Figure 12.

Figure 12: Double Burst in Production of /p/



In many of these cases, the darkest, strongest burst was used as the release point, since other cases, the double burst might have one faint column and one dark. So the more substantial interruption in the speech signal was the preferred pinpoint. In the instance pictured, the first column is substantial enough to mark a release and would be the marking point. These instances were included in the data analyzed since they meet the criteria of being plosives.

Despite the possible implications, the physical aspects of articulation, such as glottis size/vocal fold length found through fundamental frequency, are not examined in depth here, since it would require much more equipment and analysis, as well as participant involvement. However, as far as possible, the mental aspects were considered, as each of the participants gave relevant information about her or his language acquisition background. Measurable acoustical data was measured here to make inferences about information that is not directly measurable.

In the appendix (7.3 and 7.4) are matrixes showing the amount of /b/ productions which were approximated or which measured 0 millisecond VOT, by word and by speaker. The preceding and following sounds are located next to the individual words. Speaker's individual results are listed along the bottom edge, while individual word results are located along the right hand side. The tables are provided for Dutch and

English /b/ productions only, as there were less approximations of /p/ in general and no 0 millisecond VOT productions of /p/. One can see that post-consonantal [b] productions tended to have more 0 millisecond VOT realizations. These tables also make clear that there were many more approximations in English than in Dutch. The tables are useful in demonstrating which speakers tended to approximate sounds more often.

4. Results

One of the original questions of this study was whether Dutch-English bilinguals might possibly use the same VOT in either [p] or [b] across both languages, or whether they match the standard sounds recorded previously in the language. Another possible scenario would be the use of intermediate productions in one or both languages. For example, the speaker might produce a [p] in Dutch that is closer to the English [p] than Dutch monolinguals, and so on. Data organized to examine intervocalic versus post-consonantal /p/ and /b/ tended to prove more useful than original averages of initial versus medial /p/ and /b/. Looking at the possible influence of neighboring sounds, regardless of the sound's initial or medial position in the word itself, is more defensible in some ways. Although, in Lisker and Abramson's (1964) study, they examined word position. Both aspects of a specific sound can be valuable, but in consideration of the type of flowing speech elicited, referring to surrounding sounds is more appropriate.

It is also important to note that the method for comparing the English and Dutch averages was directed by the ability to condense the data into a small enough set with few enough numbers that comparisons could be made. The average of each speaker does not reflect the individual sound in word-context averages. For instance, in all speakers, the longest [p] VOT was in pre-lateral glide position in Dutch. For consistency in the research, this distinction between intervocalic and post-consonantal sounds was applied to the analysis of [p] as well. Following are the separate data and interpretations of the /p/ VOTs and /b/ VOTs, and further on will be statistical support as well as explanations derived from literature.

4.1 /p/

The results of [p] productions, shown in Table 2, were that speakers generally showed longer VOT in the production of English intervocalic [p] as well as in the production of English post-consonantal [p]. The English intervocalic [p] VOT was 40.6 milliseconds while the Dutch intervocalic [p] VOT was 26.3 milliseconds. As for the English post-consonantal [p], the VOT was 46.2 milliseconds and the Dutch post-consonantal VOT for [p] was 16.8 milliseconds. These figures proved significant.

Table 2: Individual VOT averages (in milliseconds) of /p/ production in English and Dutch, intervocalic and post-consonantal positions.

Speaker #	Intervocalic		Consonant-preceded	
	English	Dutch	English	Dutch
1	37	27.7	36.5	20.7
2	58.9	30.8	62.9	21.7
3	64	36.7	63	19.9
4	30.4	17.2	45.6	10.4
5	25.2	26.2	25.8	19.6
6	44.6	16.5	46.8	12.6
7	30.8	35.4	39.5	13.6
8	32.3	21.2	40.3	13.5
9	28.3	25.9	39.5	21.2
10	54.4	25.4	62.5	14.7
Average	40.6	26.3	46.2	16.8
ST DEV	14	6.8	12.7	4.2

All speakers had a longer [p] in both positions in English than in Dutch. 80% of speakers had a longer post-consonantal [p] than intervocalic [p] in English. All ten speakers had a longer intervocalic [p] than post-consonantal [p] in Dutch. This contrast of [p] in different positions in either languages firmly rejects any suggestions that Dutch-English bilinguals may have combined phonetic categories for efficiency. In Dutch, the speakers made more of a distinction between the productions of the sounds in the two positions and it was very consistent. Besides Speakers 6 and 9, all of the speakers in Dutch showed a very large difference in averages of the two positions. In English however, Speakers 1, 2, 3, 5, and 6 had average VOTs for the two sounds that were within a few milliseconds of each other. In both /p/ and /b/ productions, English productions were found to show much more variation than Dutch sounds. The English /p/ productions VOT were on average longer than Dutch productions by 22 milliseconds.

The only speaker that did not produce different sounds is Speaker 5, whose categories were all closer to the Dutch average, within 5 milliseconds, in all positions in both languages. Speakers 7 and 9 also had VOT averages in the intervocalic category that were within 5 milliseconds of each other, but their post-consonantal productions were very distinct. The other half of the speakers' average VOT for intervocalic [p] was around 8-15 milliseconds shorter than post-consonantal [p]. This divide between speakers with distinct separation and minimal separation between English intervocalic and post-consonantal [p] advances the point that English [p] tends to be variable. 50% of speakers produced [p] in the two categories with around 5 milliseconds of difference, where only 20% of speakers did so in Dutch. The average VOT length difference in milliseconds intralingually and interlingually are helpful in showing general trends.

In both Dutch and English, the average VOT of [p] was closer within the language than across languages. Also, regardless of accent, source, and usage, the English [p] is always longer, in those speakers that learned both languages before 4 years of age, which were all but one speaker, Speaker 5.

4.2 /b/

As shown in Table 3, all speakers prevoiced English [b] in intervocalic positions, with a substantial VOT average. It is apparent that Dutch has generally longer prevoicing of /b/ than English in both intervocalic and post-consonantal positions. As for the VOT difference between Dutch [p] and English [b], it was originally expected that these phones would be much closer in realization, but the data revealed that they were quite distinct. The average VOT of Dutch [p] was 16.8 milliseconds in post-consonantal position and 26.3 milliseconds in intervocalic position. The average VOT of English [b] was -13.3 milliseconds in post-consonantal position and -39.2 milliseconds in intervocalic position. These are, at lowest, different by 30 milliseconds, making them quite distinct.

In this research, intervocalic sounds were contrasted with post-consonantal productions, since this would affect the prevoicing in [b] sounds. To keep the words to a reasonable length and after much practice at producing consonant clusters, speakers might show shorter prevoicing following consonants than following vowels. As shown in the tables mentioned above, (located in the Appendix 7.3), when [b] followed [t] in the English word “being,” 6 of 10 speakers had 0 millisecond VOT. Following the [z] in “raspberries,” 7 of 10 speakers had 0 millisecond VOT. Approximates, in turn, were seen more in intervocalic productions of English /b/. In Dutch (table located in Appendix 7.4), both when the [b] followed an [s] in “IJsberen” and when the [b] followed the [m] in “frambozen,” 6 out of ten speakers produced 0 millisecond VOT realizations. There are other post-consonantal productions that show no such effects in both languages, but these specific examples show that the preceding consonant affects the voicing in /b/ productions.

Table 3: Individual VOT averages (in milliseconds) of /b/ production in English and Dutch, intervocalic and post-consonantal positions.

SPEAKER #	Intervocalic		Consonant-preceded	
	English	Dutch	English	Dutch
1	-16.2	-57.0	-0.0	-12.2
2	-48.6	-70.2	-21.1	-31.5
3	-38.2	-65.3	-0.0	-10.0
4	-34.8	-73.2	-5.8	-3.6
5	-39.1	-72.1	-14.2	-19.9
6	-48.4	-89.7	-17.3	-51.9
7	-53.4	-59.6	-25.7	-37.3
8	-37.0	-78.2	-34.2	-53.5
9	-29.8	-49.3	-5.5	-18.5
10	-46.8	-58.7	-9.5	-26.8
	B	B	B	B
Averages	-39.2	-67.3	-13.3	-26.5
STDEV	10.9	11.8	11.3	15.1

4.3 Significance

Four t-tests were conducted to test the significance of the results. Normal distribution was assumed for all tests, in accordance with Kolmogorov-Smirnov test p-values for each case, and the tests were assumed to assess dependent samples. The English intervocalic [p] had a significantly longer VOT than Dutch intervocalic [p] ($p < 0.0001$). There were 258 tokens for English intervocalic [p], the largest set of tokens in the data. The average of these tokens, equal to the average of the individual averages for this category, was measured against the corresponding average for the Dutch category, which was made up of 141 tokens. The English post-consonantal [p] had a significantly longer VOT than Dutch post-consonantal [p] ($p < 0.01$). There were 89 tokens for post-consonantal English [p] and 89 tokens for post-consonantal Dutch as well. In both cases, Dutch [p] then had a significantly shorter VOT than English [p], which suggests that the Dutch [p] category is separate from the English [p] in this set of bilinguals.

The Dutch intervocalic [b] had a significantly longer VOT than English intervocalic [b] ($p < 0.0001$). There were 155 tokens for Dutch intervocalic [b] and 147 tokens for English intervocalic [b], so they were pretty evenly matched in this section. The Dutch post-consonantal [b] was also significantly longer than English post-consonantal [b] ($p < 0.0001$). There were 129 tokens of Dutch post-consonantal [b] and 104 tokens of English post-consonantal [b]. In general, the trend can be seen that English /p/ VOT is longer than Dutch /p/ VOT, and there is longer prevoicing of Dutch /b/ than of English /b/.

These tests for significance prove that based on a sample of ten Dutch-English bilinguals, most, but not all, have four separate categories for bilabial plosives, /b/ and /p/, one voiced and voiceless bilabial plosive, in each language. There is a certain age distinction that is at play, since the four category pattern is contested mainly by Speaker 5, and to a lesser extent, speaker 7, whose categories overlap in areas that other speakers' categories do not, as shown in Charts 1 and 2. Speaker 5 learned English at age 6 and Speaker 7 learned English at age 3 or 4. Speaker 8 shows overlap in /b/ productions and Speaker 9 shows overlap in /p/ productions and those speakers both learned both languages from birth. In the case of Speaker 8, his English /b/ productions are more similar in both positions. Speaker 9, however, appears to have nearly the same average VOTs for Dutch and English intervocalic productions. Both speakers 8 and 9 grew up living half of his or her life in an English-speaking country and the other half in a Dutch-speaking country.

Chart 1: Comparative Spline Chart of Average VOTs English and Dutch [b] per Speaker

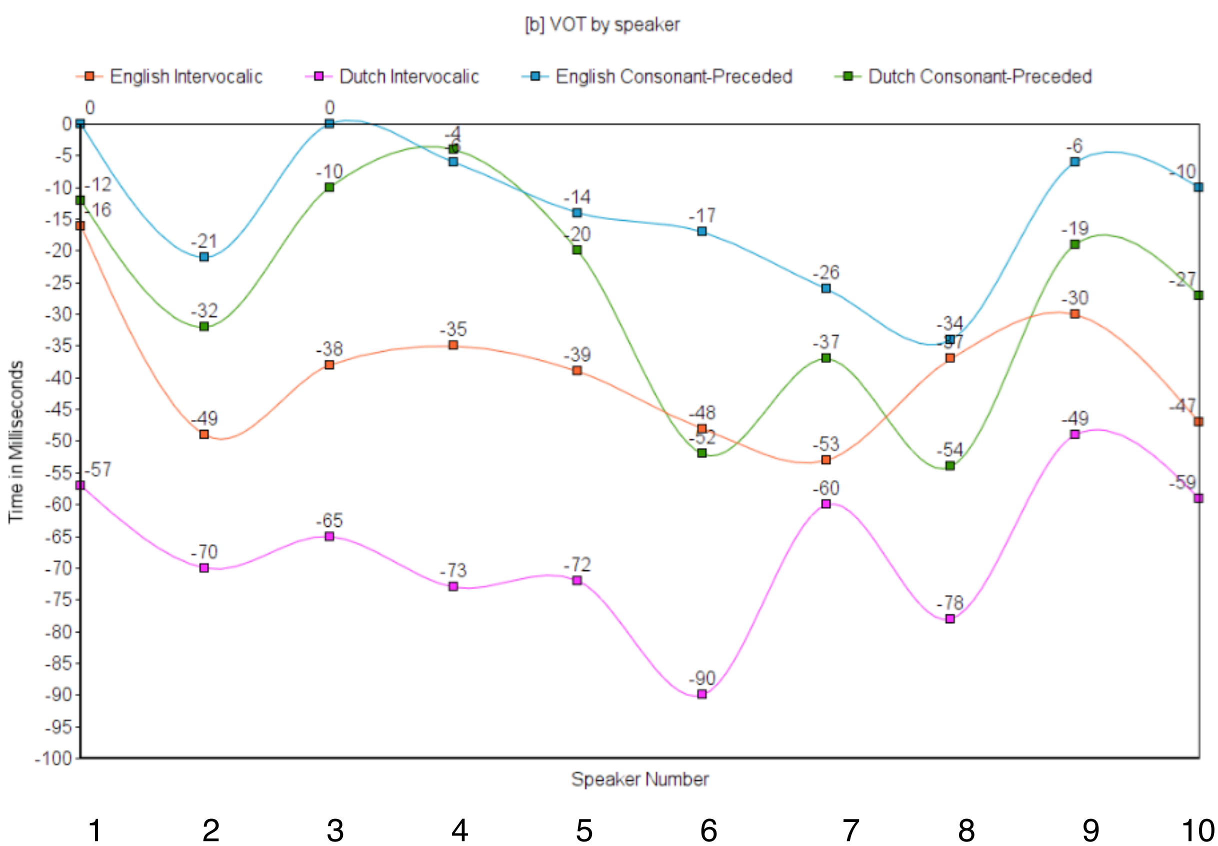
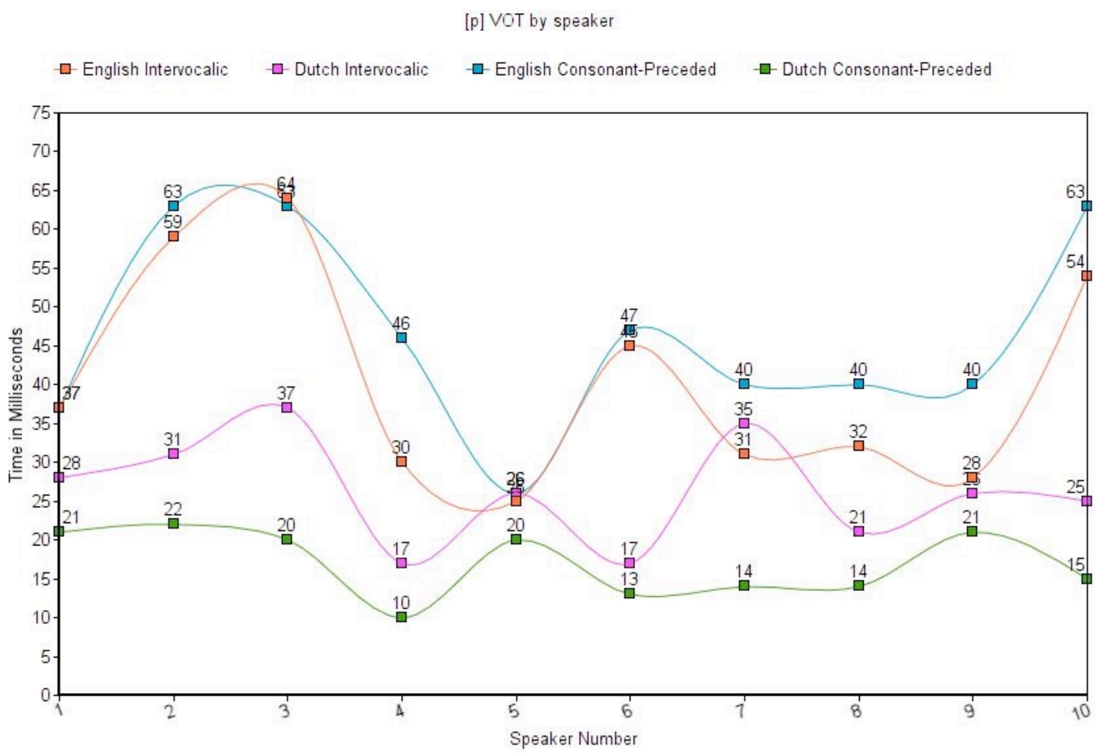


Chart 1 above shows the average VOT distribution for different productions of [b] in each speaker. While clearly some speakers have a more even spacing between productions of sounds in different languages, three speakers show intersections. The Dutch intervocalic appears to be the most stable and consistent, remaining the most prevoiced sound for each speaker. Only one speaker, number 7, produced the English and Dutch intervocalic [b] nearly the same. That speaker's [b] productions were all within a rather tight range, compared to the other speakers. The second from the top line, Dutch post-consonantal, shows the most variation. Speaker 4 has nearly no prevoicing at all in that category. It is reasonable and expected however that the post-consonantal [b] productions would have the most variation, since they follow an interruption in the speech signal. The preceding interruptions, the consonants, elicit more variety in voicing and friction characteristics, which can impact the following plosive production. The more reliable and comparison-worthy categories, English and Dutch intervocalic, are represented by the blue and pink line in the bottom of the chart. As apparent, those categories are always separate in each speaker, even if it be to varying degrees.

Chart 2: Comparative Spline Chart of Average VOTs English and Dutch [p] per Speaker



As shown in Chart 2 above, most speakers showed clear distinction between average VOT in Dutch and English productions of /p/. The average English /p/ VOT was 40.6 and 46.2 milliseconds in intervocalic and post-consonantal position, respectively.

The average Dutch /p/ VOT was 26.3 and 16.8 milliseconds in intervocalic and post-consonantal position, respectively. However, speakers 5, 7, and 9 did not show a significant difference between their production of Dutch and English intervocalic /p/. Speaker 5 has averages in both positions in both languages that are within 5 milliseconds of each other. Speaker 5 is the one speaker that learned English beginning at age 6, but she is one of three speakers who attended only English-speaking schools before university. It appears that the daily exposure at school had less impact than the sequential acquisition of English after Dutch. As for speaker 5's production of /b/, there did appear to be a significant boundary between the productions in each of the two languages. Speaker 5 uses Dutch-like VOT for voiceless plosives, but she uses both Dutch and English like VOT in the appropriate positions when it comes to voiced bilabial plosives.

As shown above, in the production of [p], both Dutch intervocalic and Dutch post-consonantal [p] had consistently shorter VOT lengths than their English counterparts. The exceptions to this consistency were speakers 5 and 7. Speaker 5's English and Dutch [p] productions intersected, so they seem to be indistinguishable into two different categories. Speaker 5 is the speaker who began learning English at age 6, which explains why she uses the Dutch category for her English productions. Speaker 7, who also showed less spacing between language categories in productions of [b], also shows unusual patterns for production averages of intervocalic Dutch and English [p]. Speaker 7 had a more limited source of English than other participants, but did begin speaking English around age 3 or 4. This speaker also had an accent, which may contribute to his abnormal [p] and [b] productions.

In the Chart 2, the orange and green lines, which represent the post-consonantal productions, tend to stay close to the individual language line it is associated with. The Dutch post-consonantal productions hover below and close to the Dutch intervocalic productions. A similar pattern is visible in the relationship of the English intervocalic and post-consonantal productions. This could be due to the nature of intervocalic voiceless consonants. If a voiceless sound is wedged between two voiced sounds, it may be necessary to hold the voiceless sound a bit longer for clarity, whereas the post-consonantal sounds were followed by vowels. When the stop consonant is preceded by another closure or semi-closure, halting or lessening the flow of speech, then the following burst, and maybe accompanying aspiration, are more immediately evident. Table 4 lists the ages and sequence of acquisition of each speaker, so it may be cross-referenced with the charts above for clarity.

Table 4: Speakers Language Acquisition Order by Age (Additional Information)

Age of Language Acquisition

Speaker #	English	Dutch
1	0	2
2	0	0
3	0	2 - 3
4	0	0
5	6	0
6	0	0
7	3 - 4	0
8	0	0
9	0	0
10	0	4

- Bilingual First Language
- English First
- Dutch First

4.4 Comparisons to VOT Averages in Literature

The average VOTs of English /b/ are much longer than the majority of results yielded in the earlier study by Lisker and Abramson. In that study, out of 51 samples, found an average of 1 millisecond average VOT and out of the minority sample set of 17 tokens, found a -101 milliseconds average VOT (see Tables 5 and 6 below). In English [b] produced by the participants in this study, the average for intervocalic [b] was -39 for intervocalic position and -13 for post-consonantal position. As for /p/, Lisker's average VOT was 10 milliseconds, while the data here resulted in an average of 16 milliseconds in post-consonantal position and 26 milliseconds in intervocalic position. The average VOT of /p/ is within 5-15 milliseconds of Lisker's measurements. The English /p/ is in Lisker's study averaged at 58 millisecond VOT and the current data has the VOT at around 40 milliseconds, 20 milliseconds shy of their data.

Table 5: Dutch VOT in Milliseconds (Lisker and Abramson 1964: 392)

TABLE 1. Voice Onset Time in Msec: Dutch
(1 speaker)

	/b/	/p/	/d/	/t/	/k/
Av.	−85	10	−80	15	25
R.	−145: −50	0:30	−115: −45	5:35	10:35
N.	22	46	32	56	60

Shown in Table 5 are the average VOTs of one native speaker of Dutch, in pre-vowel, word-initial positions only (Lisker and Abramson 1964: 388). The average was taken from 22 tokens and the range was from -145 to -50 milliseconds for Dutch [b]. In the current study, the average prevoicing period for intervocalic Dutch [b] was approximately 10 milliseconds less than the speaker shown here at -67 milliseconds. Dutch [p] here had a 10 millisecond positive VOT. This figure closely resembles the post-consonantal Dutch [p] produced by the speakers in this study, which had a 16 millisecond positive VOT, but the average VOT of Dutch intervocalic [p] was 26 milliseconds.

Table 6: English VOT in Milliseconds (Lisker and Abramson 1964: 394)

TABLE 6. Voice Onset Time in Msec: English
(4 speakers)

	/b/	/p/	/d/	/t/	/g/	/k/
Av.	1/−101	58	5/−102	70	21/−88	80
R.	0:5/−130: −20	20:120	0:25/−155: −40	30:105	0:35/−150: −60	50:135
N.	51/17	102	63/13	116	53/13	84

As for English [b], the situation is a bit more complicated. Lisker and Abramson provide two values, shown in Table 6, for all of the voiced stop consonants in their research because of the occurrence of several positive VOT instances as well as positive VOT instances. They chose to separate the VOT values into separate categories. Lisker and Abramson also mention it was evident that the individual speakers “do not randomly produce such stops with positive and negative values of relative onset time; rather, each speaker, in isolated words at least, always or nearly always produces a single kind of /b d g/” (Lisker and Abramson 1964: 395). Out of four speakers, only one prevoiced /b/. This speaker also prevoiced /d/ and /g/ and consistently prevoiced the sounds, producing a positive VOT only once out of 42 productions of voiced stop consonants. The post-consonantal productions tended to be composed of more 0 values, those with voicing beginning at the time of the burst, than the intervocalic category. While /b/ then cannot be accurately compared to the Lisker and Ambramson figures, the average VOT for English /p/ was 40.6 milliseconds in intervocalic positions and 46.2 milliseconds in post-consonantal positions. This is less than half of the VOT given in the aforementioned article.

Table 7: VOT of /p/ in French-English Bilinguals versus French and English Monolinguals (Caramazza 1973)

TABLE II. Means of VOT values (in milliseconds) for the voiceless consonants in the production test.

	/p/	/t/	/k/
UF	18	23	32
UE	62	70	90
BF	20	28	35
BE	39	48	67

Table 7 is taken from Caramazza et al. (1973), from their study on French-English bilingual production of stop consonants. Unilingual English (UE) speakers in this experiment had a longer average VOT in the production of [p] than the Bilingual English (BE) speakers, while the French average was nearly the same. This is true for the other two voiceless stop consonants as well. The Bilingual English production is about 20 milliseconds less than the Unilingual English, while the Bilingual French production is within 5 milliseconds of the Unilingual French. This research was conducted in Montreal, where the language situation is comparable to Amsterdam, since English has a place in the city as the second most common language, although it is much less common than French. The authors of the study say that in the case of their bilingual participants, 20 total, all had begun learning English before 7 years of age, but French was their first language. It appears from the table that the average VOT productions for BE speakers was approximately half way between the average BF and UE VOTs. Since the trend extends across all of the voiceless stops, it seems that the BE English average VOT length is generally closer to the original French consonant VOT length. Lengthened VOT versions of French consonants suffice.

This interpretation of the trend assumes that the English consonant VOT is constructed in relationship to the French consonant, since the participants' first language was French. However, the authors also note that in the perception part of their research, UE subjects did not show overlap when distinguishing between phonemic categories, while UF subjects did. They then assert that "VOT appears to be an important variable for voicing distinctions in Canadian English but not in Canadian French" (425). In the current research, in Chart 2, speakers 2, 3, and 10 show averages closest to the UE averages from the Caramazza research, which is interesting because two of the speakers were of those who learned English before Dutch (speakers 3 and 10) and the third speaker grew up mostly in England. The sequential acquisition of English after French was the author's explanation for abnormal VOT productions of stop consonants. Here the effect only appears to impact two of three speakers that it should, as well as one speaker that it should not. The majority of speakers (1, 4, and 6-9) show VOT productions that actually match closely those of the BE speakers in the Caramazza article.

4.5 Explanations through Interference

Antoniou et al. lay out the possibilities for inter-language interference in bilinguals. The authors explain in which directions interference usually occurs, and the reasons one language will interfere with another. The first possible case, which is observed very often, is that the L1 interferes with L2, which of course results in accented speech. Antoniou et al expand on this point to say that age of L2 acquisition typically dictates how strong the accent is, since when the L2 learner is older, “the L2 is learned through the ‘filter’ of the L1” (Antoniou et al. 2011). This first case arises in the speech of sequential bilinguals, those who learn one language before the other. This type of influence is termed “unidirectional,” (Antoniou et al. 2011). Antoniou et. al cite a study by Flege & Eefting 1987 which examined VOT averages of voiceless stop consonants in early Spanish-English bilinguals, who ultimately exhibited signs of this unidirectional influence. The hypothesis of the current paper is that Dutch-English bilinguals will show signs of interference in English because of living in a Dutch environment, interacting with mostly Dutch-accented English speakers. However, the effect of unidirectional influence may need to be considered as the cause of influence in some participants, as two speakers learned Dutch before English. Three participants learned English before Dutch, though, so it is possible to see an influence on Dutch from English.

Another possibility is “bidirectional” influence, which is when both L1 and L2 effect each other. Researchers Flege and Eefting (1987) found that Dutch-English bilinguals with good English accents “had more English-like VOTs . . . and also produced Dutch /t/ with even shorter VOTs than other Dutch speakers” (Antoniou et al. 2011). Some speakers from the same study produced shorter VOT of English /t/ than monolingual English speakers (Antoniou et al. 2011). In this scenario, the bilingual seems to produce sort of intermediate sounds, not having a full accent in either language. They seem to adapt the sounds in one language slightly to match the other language without losing the full integrity of those sounds in the original language. The age of acquisition is a factor here, as well as whether the bilingual is a simultaneous or sequential learner. In one case cited, English-French simultaneous bilinguals produced longer VOT than French-English sequential bilinguals, both of which were longer than French monolingual productions of the French stops (Antoniou et al. 2011). The French-English sequential learners in that study had learned both languages before 5 years of age, so they could still be considered native speakers by some definitions, and depending on the regularity of their use.

The third scenario discussed by Antonio et al. is that the L1 ceases to be used as much as the L2 and the L2 then interferes with the L1. This seems to be displayed in the case of speaker 1, whose was exposed to English from birth and Dutch from 2 years old. Unlike the other two speakers, 3 and 10, who learned English before Dutch, speaker 1 has a shorter VOT of [p] in both intervocalic and post-consonantal position than those with a similar linguistic background to her. Speaker 1 also had the lowest intervocalic [b] VOT of all the speakers. Speakers 1, 3, 4, 9 and 10, a group which includes all three of the English-Dutch sequential learners, had the lowest VOT of English post-consonantal productions of [b]. Speaker 1 in particular has only ever lived in the Netherlands, and her source of English is her family. Since in most aspects of her life are conducted in Dutch and she speaks English only about once per week, it appears that Dutch influenced her English and that is why she produces [p] with a shorter than average VOT in English. Antoniou et al. cite a case in which a Portuguese-English speaker exhibits the same behavior. They describe the situation as the speaker shifting her speech “toward the norms of the language environment, demonstrating that temporary language-context-dependent changes occur at a detailed phonetic level in a bilingual speaker’s L1 and L2 VOT productions” (Antoniou et al. 2011). This means that the language environment has previously been found to have an impact on the bilingual speaker’s ability to perform in their (first) native language, and speaker 1 seems to be an extension of that proof. The fourth possibility that Antoniou et al. discuss is that utter lack of any interference at all, which is rare but does indeed occur. Here it appears that Dutch influenced productions of English /b/, since the averages in this category were less stable than their Dutch counterparts. For some speakers, that means unidirectional influence, for others it is bidirectional, depending on which language they learned first.

5. Discussion & Conclusion

This study of the phonetic aspect VOT in Dutch-English early acquisition bilinguals yielded significant results in regards to whether such speakers typically have separate categories for voiced and voiceless bilabial plosives in both languages. The speakers who learned both languages before the age of four did have all separate categories, while one speaker who learned English at age six, tended to produce /p/ in a Dutch-accented way, with a a shorter VOT than her peers. In general, it was found that these bilinguals have a longer prevoicing period of Dutch /b/ than English /b/ and their VOT for English /p/ is longer than for Dutch /p/. This is due to their early age of acquisition. However, English VOT tended to be much more variable across speakers in /b/, which suggests that these bilinguals experience some interference from Dutch /b/, the more consistently produced sound, based on the lower number of approximations of this sound. The main findings of this study were that there are substantial differences between the sequential and simultaneous learners. The average /p/ and /b/ VOTs in Dutch and English were significantly different in most Dutch-English bilinguals, but the sequential learners showed the most overlap in production boundaries. In following with former assertions, Dutch /b/ had the longest negative VOT (prevoicing), although English /b/ also had negative VOT. The English /p/ had the longest positive VOT. The Dutch /b/ was found to have significantly longer average prevoicing than English /b/ and the English /p/ was found to have significantly longer VOT than Dutch /p/. The hypothesis beginning this study was that the speakers would show interference from Dutch in their productions of English bilabial plosives. The findings show that based on age of acquisition, Dutch could influence English productions. The average English VOTs produced by these speakers do not match monolingual values in the literature (Lisker and Abramson 1964), but do match bilingual production values in the literature (Caramazza 1973).

Future Research

Probably the most valuable extension to the current research would be the addition of perceptual analysis. For instance, playing the collected recordings to monolingual Dutch and monolingual English speakers and find out if the speakers detected any differences. One could isolate both more average realizations as well as more questionable productions and then play them to monolinguals and ask them to categorize the sounds as either /p/ or /b/. That would give a nice baseline to then extend the research to bilinguals. One could play /b/ or /p/ with only the two surrounding sounds, rather than in a whole word, to a bilingual and ask them to choose which language they believe the snippet is from.

Another interesting line of research would be to measure the data produced for this study against a larger set of monolingual speakers in each language. In fact, Caramazza (1973) did measure their data against baseline samples from monolingual speakers. They also had a more involved screening process for choosing participants, which included asking potential participants to perform in a reading speed test, the criteria being set at a 180 words per minute rate, before taking part in any of the experiments. Since the participant's objective in the current study was to read aloud from a page, it would be possible to confirm their fluency after the fact, but it may have been useful to assess each speaker's fluency beforehand to clear doubts.

It should be noted that the use of female and male speakers in the same data set could have affected the results, since it may have increased the range. Voicing may come across as more apparent in male speakers with perhaps louder voices, as their voicing can be clearer in some occasions in the spectrogram and wave signals. However, even in the case of many male speakers, the voicing could become fainter just before the burst, even coming to full silence in the wave signal. In both males and females, some of those with breathier voices also had an impact on the visible signals, where voicing appeared quite faint or so broken that it was difficult to discern, especially in cases with short prevoicing. When the longest VOT averages were sought for /b/, the males had the longest prevoicing periods in post-consonantal positions in both Dutch and English as shown in Table 8. The male speakers numbers are highlighted in yellow. In the intervocalic position, the effect is not strong enough to be significant.

Table 8: Average VOT of /b/ in Order of Length in Milliseconds

S #: Speaker Number
IVC: Intervocalic
CP: Post-Consonantal

S #	English IVC	S #	English CP	S #	Dutch IVC	S #	Dutch CP
1	-16.2	1	-0.0	9	-49.3	4	-3.6
9	-29.8	3	-0.0	1	-57.0	3	-10.0
4	-34.8	9	-5.5	10	-58.7	1	-12.2
8	-37.0	4	-5.8	7	-59.6	9	-18.5
3	-38.2	10	-9.5	3	-65.3	5	-19.9
5	-39.1	5	-14.2	2	-70.2	10	-26.8
10	-46.8	6	-17.3	5	-72.1	2	-31.5
6	-48.4	2	-21.1	4	-73.2	7	-37.3
2	-48.6	7	-25.7	8	-78.2	6	-51.9
7	-53.4	8	-34.2	6	-89.7	8	-53.5

While comparing males and females for length of /b/ VOT showed some possible effects, the same comparison made for /p/ showed no patterns, as pictured in Table 9. So, while being male may predict a longer VOT of post-consonantal /b/, sex does not seem to have any bearing on other VOT productions of bilabial plosives in Dutch-English bilinguals.

Table 9: Average VOT of /p/ in Order of Length in Milliseconds

S #	English IVC	S #	English CP	S #	Dutch IVC	S #	Dutch CP
5	25.2	5	25.8	6	16.5	4	10.4
9	28.3	1	36.5	4	17.2	6	12.6
4	30.4	7	39.5	8	21.2	8	13.5
7	30.8	9	39.5	10	25.4	7	13.6
8	32.3	8	40.3	9	25.9	10	14.7
1	37	4	45.6	5	26.2	5	19.6
6	44.6	6	46.8	1	27.7	3	19.9
10	54.4	10	62.5	2	30.8	1	20.7
2	58.9	2	62.9	7	35.4	9	21.2
3	64	3	63	3	36.7	2	21.7

S #: Speaker Number
IVC: Intervocalic
CP: Post-Consonantal

There are any number of physical characteristics that could have contributed to the results and individual speaker differences. Even in the case of the male speakers with longer prevoicing of post-consonantal /b/ in Dutch and English, could possibly be contributed to another cause, not necessarily physiological or to do with sex. Studying the correlation between sex and VOT was not the aim of this study, but it would be interesting to expand the speaker base to see if there are actually any differences. Lisker and Abramson explain that the base frequency a speaker produces when their vocal folds begin to vibrate not including higher formant frequencies or noise in the signal, or the fundamental frequency, corresponds to the "length, mass and tension of the vocal folds in conjunction with varying amounts of air pressure from the lungs" (415) (Lisker 4and Abramson 1964). They explain that these features of the glottis correspond to the adjacent cartilages and muscles that control the movements of the glottis. These features are individual and the glottis is generally larger and the vocal folds longer in males, producing generally lower fundamental frequencies. It could be interesting to separate the data into male and female participant data sets, but more participants would be required to have a significant set of values to analyze. The involuntary conditions present in each individual speaker can affect group data sets, if the speakers have very different sizes and shapes of vocal folds. This is particularly interesting because the physical mechanisms of speech, may be the involuntary cause of certain patterns of speech, like a low fundamental frequency.

To the native (American) English-speaking researcher, none of the participants seemed to have a Dutch accent. This of course is in accordance with the results which the analysis yielded. A short conversation was necessary to assess the participants individually before the recording process started. The only exception to this filtration could have been speaker 7, since this speaker had what seemed to be a Scottish or Irish accent. The foreign (to the researcher) accent could have masked a Dutch accent, if very slight. One would-be participant was wholly discarded from the study as he spoke very good English, but clearly was not a native speaker, although he was able to imitate an American accent.

Environment

The two of the four speakers with the most approximates of English [b] only speak English once per week, even if they read it daily. It also seems that those who speak English at home instead of work had more approximates of English [b] than those who conduct work in English. This may indicate that when English is used in a more relaxed environment, it is spoken with less precision.

It is possible that the lack of Dutch spoken between the researcher and the participants could have affected the results. During language acquisition, many parents with separate native languages choose to adopt the One-Parent-One-Language (OPOL) method, so the child learns to distinguish between the appropriate environments for using one language or the other (E. Venables et al 2013). In fact, six of the ten participants spoke English exclusively with one parent. The principle behind this method may also bleed into normal adult life, where multilingual people get into the habit of using one language for certain interactions and another for other interactions. For example, a speaker may use English when speaking with their mother if it is her native language, even if she also knows Dutch, but the speaker and her/his mother may both use Dutch to speak to a supermarket clerk. This is logical and an effective approach, but in following that, it may be a bit off-putting to require the participants to interact with the researcher only in Dutch during the Dutch mode when they are aware that the researcher is in fact a native English speaker and a non-Dutch speaker. It is also possible that the researcher's native English speaking background had an affect on the participants' mental orientation and speech style. These are possible affects of the research methods upon the results. However the results did yield quantifiable differences between all eight variables.

Obviously, it was not evident to the researcher whether an English accent was detectable in the participants' Dutch speech, but the participants were accepted in good faith since the research was conducted in a Dutch-speaking country. Seven of the ten speakers stated in the questionnaire that they began learning Dutch from birth. In fact, the data analysis showed that the Dutch productions yielded more consistent average VOTs than in English, which supports the assumptions made in consideration of the research environment, that the participants would speak a non-English-accented variety of Dutch.

Another possible oversight is that the participants were not asked about other languages they spoke. This means that if any of the participants in this research spoke a third language, the effects on their speech in both/either Dutch and/or English could have come from that source, rather than from each other.

Speaker 4 was born in France and lived there until she was 2.5 years old, when she moved to the Netherlands. She lived in the Netherlands for most of her life and went to university in Ireland for 6 years. Her source of English was her mother and her source of Dutch was her Father, and she made no mentions of speaking French or any other language.

Speaker 5 was born in Saudi Arabia, after six months, her family moved to India where they stayed for four years. She may have been influenced by the English spoken there. They then moved to Belgium, where she began primary school in Dutch. After living in Belgium for three years, she lived in Norway for one year, but attended school in English. She continued school in English when her family finally settled in the Netherlands. Her parents are both Dutch and speak Dutch to her, but her father also taught her basic English when she was a child.

Speaker 8 lived in Switzerland for just under one year at age 9 and did not specify which language his school was instructed in. He said he attended school instructed in Dutch from age 10. Before that, he lived in the UK and attended school taught in English. All other speakers had lived only in English- or Dutch-speaking countries. Also, all speakers were asked to specify in which environments they use each language, such as at home, at work, at university, with friends, or with family. Any evidence of speaking a third language was absent in the questionnaires. The three speakers discussed (4, 5, and 8), who resided in countries where the official language was not Dutch or English, had very similar intervocalic /b/ productions in both English and Dutch (See Tables 2 and 3 in the Results section). They showed no unusual similarities in the productions of the other sounds.

It remains unknown whether any of the participants spoke other languages. This in turn may even have some bearing on the effectiveness of the language modes employed in the recording process. However, the Dutch productions of [b] prevoicing from the data set were actually more consistent than the English productions. A factor that may have hampered the efforts to preserve a monolingual environment during the recording sessions was the presence of a Dutch speaker, who was used to assist during the Dutch portion of the recordings. She was present during the recording sessions of the first three speakers (7, 9, and 10) and introduced herself in English, but gave instructions in the Dutch mode section in Dutch. Grosjean indicates that the presence of another bilingual sets the bilingual participant up to access both of her or his languages rather than just one (Grosjean 2011: 14). The Dutch assistant's presence was intended to maintain the Dutch mode better, so the participant felt no need to communicate with the English-speaking researcher in English during the Dutch mode portion and those boundaries were adhered to during the recording sessions. Grosjean goes as far as suggesting researchers include monolingual participants as well if the bilinguals will be in contact with other participants (Grosjean 2011: 14). Then, that hurdle was overcome using different methods as discussed in the Data and Methods section.

Conclusion

Many different factors are at play in how these bilinguals produced bilabial plosives in their two languages, from age of acquisition to living environment to the recording methods used in this experiment. There were significant findings that most speakers had distinct productions of voiced and voiceless plosives in Dutch and English, but that after age 5, a bilingual can no longer be classified as an early acquisition bilingual, as shown in the results here. There were both very individual results and group results that play to the previously mentioned factors. In light of the literature, it seems that bilinguals have their own distinct productions. Both languages showed influence, varying with the speaker's individual backgrounds. The research here is based on a small sample of unique speakers, but lends support to previous findings in the field of phonetic studies of bilinguals.

6. References

Antoniou, M., Best, C. T., Tyler, M. D., & Kroos, C. (2011). Inter-language interference in VOT production by L2-dominant bilinguals: Asymmetries in phonetic code-switching. *Journal of Phonetics*, 39(4), 558-570.

Boersma, P. & Weenink, D. (2013). Praat: doing phonetics by computer [Computer program]. Version 5.3.59, retrieved June 2013 from http://www.praat.org/

Bosch, L., & Ramon-Casas, M. (2011). Variability in vowel production by bilingual speakers: Can input properties hinder the early stabilization of contrastive categories? *Journal of Phonetics*, 39(4), 514-526.

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. *The Journal of the Acoustical Society of America, 54*(2), 421-8.

Du, L. (2010). Initial Bilingual Development: One Language or Two? *Asian Social Science*, 6(5), 132-139. Retrieved March 15, 2015, from Directory of Open Access Journals (DOAJ).

Flege, J. (1995) The Phonetic Study of Bilingualism. In G. Bloothooft, V. Hazan, D. Huber & J. Llisterri (Eds) *European Studies in Phonetics and Speech Communication*. Utrecht: OTS Publications, Pp. 98-103.

Genesee, F., & Nicoladis, E. (2006). Bilingual acquisition. In E. Hoff & M. Shatz (eds.), *Handbook of Language Development*, Oxford, Eng.: Blackwell.

Grosjean, F. (n.d.). An attempt to isolate, and then differentiate, transfer and interference. *International Journal of Bilingualism, 16*(1), 11-21.

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

Lisker, L., and Abramson, A. S. A Cross-Language Study of Voicing in Initial Stops: Acoustical Measurements. *Word* 20.3 (1964): 383-422.

Lively, S., Pisoni, D., & Carr, T. H. (1997). On Prototypes and Phonetic Categories: A Critical Assessment of the Perceptual Magnet Effect in Speech Perception. *Journal of Experimental Psychology: Human Perception and Performance*, 23(6), 1665-1679.

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, 18(1), 1-86.

Miller, R., Sanchez, M., & Rosenblum, K. (2010). Alignment to visual speech information. *Attention, Perception, & Psychophysics*, 72(6), 1614-1625.

Plag, Ingo, Kunter, Gero, & Schramm, Mareile. (2011). Acoustic correlates of primary and secondary stress in North American English. *Journal of Phonetics*, 39(3), 362-374.

Saffran, J., Thiessen, E., & Dannemiller, J. L. (2003). Pattern Induction by Infant Language Learners. *Developmental Psychology*, 39(3), 484-494.

Schulpen, B., Dijkstra, T., Schriefers, H., Hasper, M., & Rosenbaum, David A. (2003). Recognition of Interlingual Homophones in Bilingual Auditory Word Recognition. *Journal of Experimental Psychology: Human Perception and Performance, 29*(6), 1155-1178.

Vihman, M. M. (2002). Getting started without a system: From phonetics to phonology in bilingual development. *International Journal of Bilingualism*, 6(3), 239.

Walsh, M., Möbius, B., Wade, T., & Schütze, H. (2010). Multilevel Exemplar Theory. *Cognitive Science, 34*(4), 537-582.

Yavaş, M. (2006). *Applied English phonology*. Malden, MA: Blackwell.

7. Appendix

7.1 Experiment Texts

A pair of bears gathered apples and pears to bake in a pie. They worried about being able to open the honey jar and operate the oven because they had clumsy bear paws. They were having fun though, so they began to feel optimistic. After they put the pie in the oven, they began to prepare another one with strawberries, raspberries and peaches.

Een paar ijsberen plukten appelen en peren om een taart te bakken. De beren maakten zich zorgen of zij in staat waren de honingpot te openen en de oven te bedienen omdat ze onhandige berenpoten hadden. Uiteindelijk vonden ze bakken zo leuk waardoor ze toch optimistisch werden. Toen de taart in de oven was, begonnen ze met de voorbereidingen voor het volgende gebak met aardbeien, frambozen en perziken.

7.2 Participant Background Information

	Age of acquisition		Source		Frequency		Environment		Countries				Schools	
Speaker	English	Dutch	English	Dutch	English	Dutch	English	Dutch	English S	Dutch Spe	Other	Education	English	Dutch
1	0	2	M	S F	W(s) D(w)	D	H F	H W	0	all		undergrad		all
2	0	0	S	M F	D	W	W C H	P	15	6		undergrad	all	
3	0	2/3yo	M F	S	D	D	C P W	H	0.4	all		undergrad		all
4	0	0	M	F S G	D	D	P W F	F W	6	19.5	France 2.5	2 grad		all
5	6	0	F S	M F S	D	D	H F S	W S H	0	21	SA 0.5, India 4, Norway 1	grad	age 7-18	2 years
6	0	0	M	F S	D	D	H	W	0	27		grad		all
7	3/4yo	0	M's BF	M F S	W(s) D(w)	D	H	W C P F	0	all		high school		all
8	0	0	F	M S	D	D	H W	W H	13	11	Switz'd 1	undergrad	before 10	8 years
9	0	0	M F S	M F S N	D	D	C H	W H F	14	15		undergrad	before 14	some
10	0	4	M F	S	D	D	P C	C F	0	all		undergrad		all
			S - school		D - daily		W - work		in years				Before uni	Before uni
			M - mother		W - weekly		C - college							
			F - father		s - spoken		H - home							
			G - grandparents		w - written		P - parents/family							
			N - nanny				F - friends							

Appendix 7.3 English /b/ Approximates and Zero VOT

			Speaker	1	2	3	4	5	6	7	8	9	10	approx	zero VOT
	[b] initial	pre	post												
	Bake	schwa/u	e	0	71	0	27	69	82	42	64	0	77		3
	Bear	i	ɛ	0	38	74	28	23	62	67	35	0	75		2
consonant	Bears	v	schwa	0	0	0	0	52	0	49	26	0	26		6
consonant	Because	n	i	0	11	A	21	0	0	6	A	20	12	2	3
	Began	eɪ	i	21	28	37	A	21	37	A	47	54	25	2	
	Began	eɪ	i	A	54	25	A	A	28	A	33	49	12	4	
consonant	Being	t	i	0	52	0	0	0	64	60	71	0	0		6
	[b] medial														
	Able	eɪ	l	A	A	30	31	A	A	A	10	34	25	5	
	About	schwa	aʊ	A	40	47	A	18	44	41	46	34	41	2	
consonant	Raspberries	z	ɛ/schwa	A	22	0	0	0	0	0	A	0	0	2	7
	Strawberries	ɔː	schwa	20	45	41	A	A	22	A	A	33	26	4	
approx				4	1	1	4	3	1	4	3				
zero				5	1	4	3	3	3	1		5	2		

Appendix 7.4 Dutch /b/ Approximates and Zero VOT

			speaker	1	2	3	4	5	6	7	8	9	10		
	b-initial	pre	post											Approx	Zero VOT
	Bakken	ə	ɐ	42	62	73	65	77	84	46	76	52	22		
	Beren	ə	ɛ	54	88	74	74	77	108	69	79	48	72		
	Bedienen	ə	ə	57	57	37	77	40	69	53	72	72	58		
	Berenpoten	ə	ɛ	51	82	65	81	88	102	69	87	77	80		
	Bakken	ə	ɐ	86	66	75	83	72	87	53	76	0	78		1
consonant	Begonen	s	ə	37	0	0	12	42	42	35	A	0	45	1	3
	b-medial														
consonant	ijsberen	s	ɛ	A	0	0	0	0	64	16	65	0	0	1	6
consonant	Voorbereidinge	r	ə	A	37	29	A	A	43	A	32	56	24	4	
	Gebak	ɜ	ɐ	52	66	68	77	67	89	47	53	47	67		
consonant	Aardbeien	d	a	0	76	24	0	37	72	74	41	A	65	1	2
consonant	Frambose	m	o̜	0	44	0	0	0	39	24	16	0	0		6
approx				2	0		1	1		1	1	1			
zero				2	2	3	3	2				4	2		