Spring 2014

Effect of temporal parameters on the perception of foreign accent in synthesized speech

Ashley Assgari
James Madison University

Recommended Citation

Assgari, Ashley, "Effect of temporal parameters on the perception of foreign accent in synthesized speech" (2014). *Masters Theses*. 136.
https://commons.lib.jmu.edu/master201019/136

Effect of temporal parameters on the perception of foreign accent in synthesized speech

Ashley Assgari

A thesis submitted to the Graduate Faculty of

James Madison University

In

Partial Fulfillment of the Requirements

for the degree of

Masters of Arts

Psychological Sciences

May 2014

Table of Contents

List of Tables

List of Figures

Abstract

Previous research investigating the parameters that affect accent have concentrated mainly on talker characteristics (e.g Flege, 1988). Those studies that do attempt to investigate acoustic parameters rely on post-hoc analysis of signals already judged to be accented. Any acoustic differences between these signals are said to be the basis of accent judgments. The current investigation attempts to rectify this methodological flaw by manipulating acoustic parameter previously implicated in perceived foreign accent within synthesized speech. In a two experiment study we investigate the effect of consonant duration and consonant initial frequency (Experiment 1) along with voice onset time, vowel duration and stop closure duration (Experiment 2). All stimuli were presented to participants over headphones using E-prime 2.0 experimental software. Participants were asked to perform two tasks: an initial phoneme identification task and an accent rating task. Identification accuracy was not expected to change based on our manipulation. Accent ratings were expected to increase as parameters approached non-native values. Consonant duration and initial frequency failed to have any effect on accent due to flaws in our synthesis. This could be a by-product of synthesis issues apparent by unexpected detrimental effects of our manipulation on identification accuracy. VOT and vowel duration significantly impacted accent ratings while stop closure duration did not. These finding suggest that listeners do rely on temporal parameters of speech in their judgment of accent. Suggestions for methodological standards are given.

Effect of temporal parameters on the perception of foreign accent in synthesized speech

Understanding what aspects of the speech signal contribute to the perception of foreign accented speech is vital to understanding and maximizing our ability to communicate with other people. (In fact, the same issue can be argued to be of importance within a given native language as well in terms of understanding acoustic cues to regional dialect). While accent does not eliminate the ability to communicate, it can certainly hinder a listener's comprehension of the speaker's intentions due to lack of accurate perception. As a result, understanding what parameters within speech lead to accent may aid in our ability to correct accented speech through training accented speakers to correctly position their articulators or, at the very least, allow us to understand its cause.

Native speakers can often have difficulty adapting to a second language (henceforth referred to as L2) when there is a discrepancy between the phoneme categories present in the native and non-native language that makes it difficult to accurately produce L2 phonemes (e.g., English /r/ and Japanese /r-l/). The result is often a "thick" and pervasive accent. If the native language (L1) lacks a similar phoneme category, non-native production of phonemes are more accurate. Participants speaking an L2 with no corresponding L1 category are likely forming "new" phoneme categories since there is no existing phoneme to relate to in their native language repertoire (Flege, MacKay & Meador, 1999). However, if L1 has a similar phoneme category as L2 production of both will change to accommodate both categories (Flege et al., 1999). This tendency to accommodate both categories is referred to as a second language carry over effect. A popular example would be Japanese speakers' inability to distinguish the /r/ and

/l/ phonetic categories. Since there is no difference between these phoneme categories in the native tongue, Japanese speakers have difficulty producing a clear distinction in everyday speech (e.g., Idemaru & Holt, 2013).[1]

Prior research has primarily concentrated on speaker-specific characteristics (e.g. length of time in L2 speaking country) in an attempt to explain accent (e.g. Flege, Munro, & MacKay, 1995; Flege, Bohn, & Jang, 1997). This approach often neglects acoustic cues that could contribute to a listener's judgments of speech as accented in favor of higher-order explanations. The existing studies that do attempt to address acoustic parameters contributing to accented speech generally rely on post-hoc explanations of accent based upon acoustic analyses of natural speech signals that have been rated as accented by participants (e.g. Arslan & Hansen, 1997; Mack, 1982). Researchers attribute accent to any reliable differences appearing from the acoustic analysis that tend to correlate with judgments of foreign accent. In contrast, the current investigation represents an initial attempt to pinpoint relevant acoustic cues by directly manipulating parameters that have been previously argued to contribute to accent ratings. Samples containing these manipulations were rated to indicate their perceived level of accent.

---

[1] Researchers investigating accented speech should be cognizant of the relationship between the production and perception of a phoneme when interpreting their results. Past research has shown that there may be a relationship between the ability to perceive a phoneme and the ability to produce it (Flege et al., 1999; Flege & Hillenbrand, 1984). However, it is unclear which of these two aspects of speech, perception or production, needs to come first to produce less accented speech. Depending on the phonemic differences between the languages included in the study, participants may be rating speech as accented simply because the speaker is producing a phoneme category that is not in their native language. In addition, when asked to produce samples, the speaker may be producing, or failing to produce, distinctions between phoneme categories on the basis of their native phoneme inventory. (Note that this prediction might not apply if the producer has been explicitly trained how to generate the non-native category; e.g., see Herd, Jongman, & Sereno, 2013). Since phoneme distinctions (or lack thereof) could affect participant's judgment of accent and the speaker's production of signals, researchers should be cautious when interpreting accent ratings of phoneme categories that are perceived or produced differently between the languages included in their investigation.

The current paper will discuss which parameters are currently said to contribute to accent. It will then address the methodological issues found in the current body of research and how the proposed experiment attempts to correct for these issues. Finally, the goal of the current investigation will be introduced.

Speaker characteristics contributing to accent

There have been many studies that have attempted to look at speaker characteristics that contribute to accent (Flege, 1988; Flege, et al., 1995; Flege, Frieda, & Nozawa, 1997; Flege et al., 1997; Flege, Schirru, & MacKay, 2003). Perhaps the most prolific finding in this body of research is that age of arrival (AOA) in the country where L2 is primarily spoken strongly affects objective ratings of accent. The earlier a person immigrates to the L2 speaking country, the less accented their speech will be (Flege, et al., 1999). Furthermore, no difference in judged pronunciation has been found between participants living in the L2-speaking country for 1 year versus 5 years (Flege, 1988), suggesting that there is a dramatic increase in the ability to produce L2 that asymptotes over time. Thus, while age of arrival in the L2 speaking country has a strong effect on ratings of accent, any benefit to speech from living in that country (also called Length of Residence, LOR) may only be apparent for a short amount of time.

A closely related parameter, age of learning (AOL), has also been shown to impact ratings of accented speech, such that the younger a person learns a language, the less accented their speech will be (Flege et al., 1995; Flege, 1988). However, age of learning and age of arrival are often correlated since earlier arrival in a L2 speaking country is likely to indicate a younger age of learning. As a result, it is possible that age of arrival and age of learning are measuring the same underlying variable, overall L2 experience.

Another parameter that has been shown to affect foreign accent ratings is the amount of L2 usage. As L2 usage increases, accent ratings of L2 speech samples decrease (Flege et al., 1995). In addition, participants who report a relatively higher use of their native language are judged to speak L2 with a stronger accent (Flege et al., 1997). In other words, participants who spend more time communicating in L2 show marked decreases in judged foreign accent relative to participants who spend more time communicating in their native language.

It is important to note that with all of speaker characteristics listed above, even though ratings of accent decrease, those ratings were still higher than for native English speakers. This is an important distinction because throughout the literature there is a running debate about whether or not language acquisition reflects a sensitive or critical period (e.g., see Flege et al., 1995). If a skill is said to have a critical period, this implies that if relevant learning does not take place during this time, then there is no opportunity to properly learn the skill. A sensitive period differs in that there is a period of time when optimal learning will occur, but if learning occurs afterwards, the skill can still be reliably demonstrated at a lower level of mastery compared to someone who learned during that optimal period. For language, this means that one can still learn a language after this period has passed, but it will never be to the degree of fluency (or native like) as someone who experienced the same information during the sensitive period.

There is some evidence to suggest that a linear relationship exists when accent ratings are considered as a function of age, supporting the idea that second language learning is subject to sensitive period learning rather than a critical period (Piske, MacKay, & Flege, 2001). A linear relationship between age of learning and accent also could indicate an effect of experience on accent, such that the more experience a person has with a second language, as apparent by a younger age of learning or more time exposure, the less likely they are to speak with an accent (Evans & Iverson, 2007). This contradicts the argument that second language learning asymptotes over time as suggested by Flege (1988).

Acoustic parameters contributing to accent

In addition to higher-order parameters, researchers have investigated how acoustic parameters may contribute to accent (Kang, Rubin, & Pickering, 2010; Magen, 1998; Sidaras, Alexander, & Nygaard, 2009; Tajima, Port, & Dalby, 1997; Trofimovich & Isaacs, 2012; Vieru, De Mareüil, & Adda-Decker, 2011). These efforts have shown that phoneme-specific parameters, such as phonemic substitution (Trofimovich & Isaacs, 2012), vowel formant center frequencies (Sidaras et al., 2009; Vieru et al., 2011), vowel duration (Sidaras et al., 2009), vowel stress (Braun, Lemhöfer, & Mani, 2011), and consonant duration (Vieru et al., 2011) affect accent ratings. In addition to suprasegmental parameters (or acoustic parameters occurring in a speech signal that are not related to consonant and vowel production) such as, voicing rate (Vieru et al., 2011), and prosody (Pinet & Iverson, 2010) contribute to ratings of accent. A recent study investigating a plethora of acoustic characteristics found that suprasegmental parameters accounted for about 50% of that variance in comprehensibility ratings (Kang et al., 2010). While comprehensibility ratings are different than accent ratings, factors contributing to comprehensibility also contribute to accent ratings (Trofimovich & Isaacs, 2012).

Particularly, a factor called “suprasegmental fluency” has been found to be the most accurate predictor of comprehensibility ratings (Kang et al., 2010). This factor is comprised largely of rate measures such as “mean length of run” (average number of syllables over periods of speech separated by pauses of greater than 100 ms), “phonation time ratio” (percent of the time that speaking occurred in the sample), “syllables per second,” and “articulation rate” (number of syllables produced in 1 s, excluding pauses), etc.

Some production studies have additionally found that parameters critical to vowel identification, specifically, vowel formant center frequencies, vary between accented and non-accented speakers (Sidaras et al., 2009; Vieru et al., 2011). However, there is no predictable pattern to describe the way that the formant center frequencies differed between native and non-native productions (Chan, Hall, & Assgari, 2013). In fact, recent research has contradicted past claims that vowel formant center frequency is an important parameter in perception of accent. For example, Flege & Hillenbrand (1984) found that participants were able to match vowel formant center frequencies when producing L2 consonant-vowel (CV) clusters when the vowel category was not present in the speaker’s native language. Despite a speaker’s ability to produce vowel formants similar to those of native speakers, listeners still would rate these samples as differentially accented. This finding suggests that vowel formant center frequency may not be the most important parameter that listeners rely upon when judging speech as accented, leading to the need to analyze other acoustic characteristics that potentially contribute to accent.

In addition to the suprasegmental and vowel acoustic parameters already discussed, research has investigated the role temporal parameters play in perceived accent. Specifically, unique temporal characteristics of non-native speech contribute to higher ratings of foreign accent. For example, Arslan & Hansen (1997) found that stop closure duration (the time where no voicing occurs between the vowel and the consonant release burst in a final consonant), voice onset time (VOT, a period of aspiration between the release burst and the onset of voicing for an initial consonant), average voicing duration (the time from where voicing begins to where voicing ends across the entire signal), and average word duration all contributed to ratings of accent. In general, they found that accented speakers produced longer durations in L2 than did native speakers. This could be explained by relative unfamiliarity with the proper vocal tract positioning of L2 phonemes compared to L1, resulting in elongated transitions to reach the desired vocal tract configurations. It is very unlikely that a participant will be as familiar with their L2 as their L1.

The fact that AOA has been shown to have such a strong influence on ratings of accent could be seen as supportive of the view that experience with articulation produces more native-like speech. Those who arrive earlier in an L2 speaking country have more experience with L2 production (although still not equivalent to L1) than those who arrive later in life. They could be said to be more familiar with the positioning of the vocal tract in proper production of L2 through experience and practice, resulting in more native-like durations.

Methodological Issues

A fundamental limitation in the current body of research dealing with accented speech is the reliance on sample-based analysis. The current parameters implicated in accented speech have for the most part been derived by having native and non-native speakers of a language produce corresponding words or phrases and then having a different group of participants rate the level of accent of each of the tokens. Once this is completed, the tokens are subjected to a variety of acoustic analyses. Differences found in the tokens are then implicated as causing the potential differences in the ratings of accent. Due to the fact that sample-based measurements rely on post-hoc findings to make their claims about what differences in the signals contributed to accent ratings, the conclusions made based upon this method should be accepted cautiously. This makes it difficult to separate parameters arising from individual differences from those that truly impact perceived accent.

Some researchers attempted to rectify this issue by comparing bilinguals to themselves in production studies (e.g., Mack, 1982). However, due to well-known second language carry-over effects, where production of L2 phoneme categories changes the range of L1 categories (Flege et al., 1997), these studies are still confounded. Since these studies rely on bilinguals, it is possible that both their native and non-native productions have been altered simply by having experience with an L2. Instead, findings based upon experimental manipulations, where the researcher controls the levels of each parameter, would be more robust in their ability implicate which parameters contribute most heavily to judgments of perceived foreign accent.

Another by-product of sample-based research is that accent ratings may be attributed to inherent differences in the individual’s speaking styles or physiology when there may be no direct relationship. For example, fundamental frequency (F_0) is dictated by the length of the vocal folds, a physiological constraint, whereas formant center frequencies are dictated by the shape and size of the vocal tract, which is more fluid and can be trained through experience. One would expect individuals to have differing values on parameters that are indicative of speaking styles or are dictated by individual physiology. In addition, research investigating the effect that speaking presentation context (i.e., with target word displayed or absent) and lexical frequency (the frequency with which a word is encountered in normal conversation) found that words with high levels of lexical frequency are rated as less accented then words with low lexical frequencies (Levi, Winters, & Pisoni, 2007). The authors attribute this to a type of adaptation to variations of words encountered often in all different contexts (Levi et al., 2007). Listening context alone was not found to significantly impact accent ratings, but was expressed as an interaction where samples that were presented with a visual display of the target word produced a smaller range of accent (Levi et al., 2007).

While the studies that attempt to use manipulation to investigate accent are extremely limited (see Tajima et al., 1997; also see Chan et al., 2013), conclusions based upon these studies are clear. One study (Tajima et al., 1997) adjusted temporal parameters of accented speech to those of non-accented speech (and vice versa) to see if there was an effect on intelligibility. As the parameters approached native values, intelligibility increased, supporting the notion that temporal parameters are vital in the perception of foreign-accented speech. However, when manipulating accented tokens in the direction of the native samples, the authors added in phonemes that the speakers failed to produce during recordings, thereby increasing participants' ratings of intelligibility by making accented signals more complete. As discussed earlier, even though accent ratings and intelligibility ratings are not necessarily the same measure, they share contributing parameters. Therefore, the authors may have unintentionally confounded their accent ratings by increasing the token's intelligibility.

In an attempt to confirm the role of vowel formant center frequency, a parameter repeatedly implicated in accent ratings, our laboratory recently had participants rate the accentedness of resynthesized accented and non-accented tokens. In these tokens vowel formant frequencies varied, but the original sample's vowel duration was preserved within both native and non-native utterances (Chan et al., 2013). This is the only study, to our knowledge, that has attempted to directly manipulate vowel center frequencies in order to evaluate their potential impact on perceived foreign accent. Data showed that changing the vowel formant frequency had little effect on our perceived accent. In fact, vowel category could completely shift without any hint of accent. However, native samples were consistently judged as non-accented while non-native samples were consistently judged as accented, indicating that the samples retained their original accent (or lack thereof). The fact that listeners could still perceive the samples as accented (or not) pointed to an uncontrolled parameter as being the cause of the perceived accent. Since our samples were consonant-vowel-consonant clusters (CVCs), and we had controlled for both formant center frequencies and duration of the vowel, the only parameters left to affect accent would be temporally-dynamic parameters, such as consonant formant frequencies. By using manipulation, we were able to show that a parameter that had previously been argued to affect accent was not a key contributor.

Goals of current investigation

The proposed experiments seek to investigate how temporal parameters in speech affect ratings of foreign accent within manipulated, single-syllable synthetic speech. Our approach is unique in that it allows us to directly manipulate temporal parameters that have been argued to contribute to judged accent. This synthesis approach also controls for any individual differences in speakers by using a fixed, simulated glottal pulse as a vibrational source (combined with measured formant frequencies/durations based on samples produced by a native and non-native speaker of American English).

Based upon prior research (Arslan & Hansen, 1997), we chose to manipulate VOT, stop-closure duration, vowel duration, consonant initial frequency, and consonant duration, the temporal parameters said to contribute to accent. When possible, these variables were orthogonally manipulated. However, due to practical limitations of stimulus set size, the variables were evaluated across two experiments. Experiment 1 investigated potential effects of parameters associated with syllable-initial consonants (initial formant center frequency and consonant duration) on accent ratings within CV (consonant-vowel) tokens. Experiment 2 evaluated potential effects of the remaining temporal acoustic parameters that have been associated with accent ratings (VOT, stop-closure duration and word duration) within CVCs. In addition to the parameters indicated above, we looked at two vowels. By including two vowels in our manipulation, we could investigate to what degree the extent of transition between the consonant initial frequency and the vowel initial frequency affects the rating of accent.

Generally, it was expected that durations of formant transitions and steady states would be longer for non-native than native speakers due to relative unfamiliarity of non-native speakers with the vocal tract positioning necessary to produce second language phonemes (Arslan & Hansen, 1997). In other words, it takes longer for non-native speakers to correctly position their articulators when attempting to produce a phoneme in their second language. Any difference in formant center frequencies could be seen as stemming from a similar cause; unfamiliarity with the proper position of articulators to produce second language phonemes. Therefore, small variations in formant center frequencies should be expected although they should not necessarily detract from phoneme recognition.

Based on prior research using manipulation to judge accent, there should have been no effect of any of our manipulations on identification accuracy (Chan et al., 2013). Even though accented speech can be difficult to understand, our tokens should be clearly identifiable as their target phoneme. As consonant initial center frequencies approached non-native values, accent ratings should have increased. In addition, as consonant durations approached non-native values, accent ratings should have increased. No effect of our manipulation of VOT and stop-closure duration on identification accuracy was expected. As VOT and stop-closure duration durations approached non-native values, accent ratings should have increased.

Experiment 1

Experiment 1 investigated the role of the voiced portions of the initial consonant in foreign accented speech. Specifically, initial-consonant formant center frequencies and transition durations were manipulated in synthesized speech. Since these parameters are a by-product of the articulation of the phoneme, differences in these parameters were expected based on differences between non-native and native speaker’s experiences with articulation of target phonemes.

Methods

Participants

Fourteen students from the JMU Participant pool, which consists of undergraduates who are required to participate in experiments to satisfy a requirement of their entry-level psychology course, participated in our experiment. All participants were between the ages of 18 and 40, lowering the risk of presbycusis (loss of high frequency information that naturally occurs with age). All participants self-reported having normal hearing and were native speakers of American English to ensure that their judgments of accent were with reference to American English.

Previous research has established that the match between the L1 of the listener and that of the producer can have an impact on judgments of accented speech. Specifically, non-native listeners often judge tokens from non-native speakers to be more intelligible when they share the same L1 (Bent & Bradlow, 2003). We therefore thought it possible that a listener's experience with Spanish might likewise impact their tendency to rate signals as accented (or not). To permit such an assessment, the extent of each participant's second language experience was assessed using select questions from the Language Experience and Proficiency Questionnaire (LEAPQ, Marian, Blumenfeld, & Kaushanskaya, 2007, see Appendix 1) at the time of informed consent. These questions provided a determination of the AOL of Spanish (if applicable), which could then be submitted to correlational analyses with participants' accent ratings.

The LEAPQ consists of nine general questions to assess the number of languages the participant is familiar with [e.g., "Please list all the languages you know in order of dominance" or "Please list what percentage of the time you are *currently* and *on average* exposed to each language (Your percentages should add up to 100%)”]. Additionally, seven questions are completed for each familiar language in order to evaluate the participant’s experience with each language (e.g., “Please list the number of years and months you spent in each language environment”). We used a subset of 5 of these questions. Participants listed each language they had experience with in order of dominance, acquisition and the percentage of usage in everyday life. Then, for each of the languages indicated, participants provided ages where milestones were reached (e.g., fluency, reading, etc.) and length of times spent in different environments where the language may be encounters (e.g., living in L2 speaking country, going to school at L2 school, etc.) All participants had experience with a language other than English (11 had experience with Spanish; 5 indicated that they were fluent).

Stimuli

The stimuli for Experiment 1 consisted of synthesized CV clusters. Two initial stop consonants and two syllable-final vowels were orthogonally manipulated. To accomplish this, natural CV (specifically /ba/, /da/, /bi/, and /di/) and CVC (Specifically, /bot/ and /dot/for use in Experiment 2) samples were recorded from a female native speaker (the first author), as well as from a female non-native speaker of English whose first language is Peruvian Spanish. Samples were recorded in a single-walled sound-attenuated chamber using a Shure dynamic microphone (Model PG58) at a 44.1 kHz sampling rate recorded with a 24-bit resolution and up-sampled to 32 bit.

Consonant and vowel durations were measured directly from the recorded samples from each speaker. Durations were measured in *Adobe Audition CS6* (Adobe Systems Incorporated, 2012) by determining the time value/sample point where the signal commences, where the consonant starts, where no remaining consonant articulation is present, where the vowel starts, and where the signal terminates. Consonant duration was defined as the time lapse between the first instance of voicing and the last recognizable instance of the consonant. Vowel duration was defined as the time lapse between the last instance of consonant articulation and the commencement of voicing. Consonant initial center frequencies for formant 2 (F_2) and formant 3 (F_3) with the bandwidths were measured using formant analysis in *Praat* (Boersma & Weenink, 2011).

A synthesized glottal pulse in *Praat* was used as our source material. The glottal pulse had a F_0 of 190 Hz that linearly declined over the signal to an endpoint of 170 Hz. Consonant duration and initial center frequencies and bandwidths were manipulated in a Formant Grid within *Praat*. In order to assess the impact of consonant duration on perceived accent, three levels of consonant duration were synthesized for each CV combination (see Table 1.): the native speaker's duration, the non-native speaker's duration, and the intermediate/average duration between the two. Vowel duration will be held constant across stimuli at the average vowel duration produced by the two speakers (see Table 1).

Each stimulus will be synthesized with an intended consonant category. Prior research has indicated that participants can identify consonants based upon distinguishing frequency information coming from formants 2 and 3(e.g., Liberman, Harris, Hoffman, & Griffith, 1957). As a result, only frequency values for formants 2 and 3 were manipulated, whereas corresponding values for formants 1 and 4 will be fixed across stimuli. To see if consonant initial center frequencies and bandwidths can affect accent ratings, we created a continuum of six frequency values in equal Mel steps (e.g., see

Fant,1973) within each consonant category, starting with measured values from the native speaker and ending with measured values from the non-native speaker (see Table 2-5). The choice to manipulate frequency values in Mel steps was an attempt to ensure that manipulation was perceptually equivalent. VOTs were modeled by combining natural samples from each consonant, fixed to an average duration, in *Adobe Audition CS6.* This combination was concatenated to the beginning of each stimulus containing the appropriate consonant.

There were 72 stimuli (2 consonants x 2 vowels x 6 consonant initial center frequencies and bandwidths x 3 consonant durations). Duration of the stimuli varied due to the reliance on parameter durations from the recorded signals. Longer consonant transitions resulted in longer stimuli. The average stimulus duration was 250 ms. The RMS amplitudes of the stimuli were equated after synthesis and prior to presentation to participants to ensure that participants are able to hear each stimulus equally. Stimuli were delivered to participants over Sennheiser headphones (HD25-SPII) at a peak intensity of 80 dB[A] within the aforementioned sound-attenuated chamber.

Procedure

After completing questions from the LEAPQ survey, participants completed two listening tasks: phoneme identification and accent ratings tasks. Both tasks were administered using E-Prime 2.0 (Schneider, Eschman, & Zuccolotto, 2002). In order to ensure that each participant is familiar with the stimuli prior to the accent rating task, all participants completed the phoneme identification prior to the accent ratings. The participants completed 4 blocks of trials (2 phoneme identification blocks and 2 accent rating blocks, with the opportunity for a rest break in between blocks). The experiment was completed in approximately 1 hour or less.

Phoneme Identification. Participants completed a forced-choice identification task to assess the extent to which they perceive each presented stimulus as the intended/target phoneme. They responded by pressing a corresponding button on a serial response box with labels provided on a computer monitor display for each phoneme category (/b/ or /d/). As with the ratings task described below, identification was self-paced insofar as trials did not advance until the participant had responded, although participants were instructed to respond as quickly and accurately as possible. There was an inter-trial interval (ITI) of 250 ms. Each stimulus was block randomized, occurring 10 times across the two blocks of trials (5 repetitions/block). Each block of trials was completed in approximately 11 minutes or less.

Accent ratings. To evaluate if our manipulations impact the relative accent of the stimuli, participants were asked to rate for each stimulus the level of perceived foreign accent on a scale from 1 (most strongly native) to 6 (most strongly non-native). The scale was described as 1, 2, and 3, indicating no accented, with 1 being the most native like. 4, 5, and 6, indicated an accent with 6 being most certainly non-native. They responded on a keyboard using the number keys or the keypad. All other procedural details including task durations and stimulus presentation were as described previously in the identification task.

An effect of consonant duration was hypothesized to occur, such that accent ratings should increase as the consonant duration approaches that of the accented speaker. In addition, we expect ratings of accent to increase when consonant initial center frequencies and bandwidths approach non-native values. The effect of our manipulations was not expected to depend on vowel. Based upon the previously mentioned findings of Chan et al. (2013), we did not expect there to be an effect of consonant duration or consonant initial frequency on identification accuracy. However, if there was any effect, we predicted it would be such that as the parameters approach non-native values, identification accuracy would suffer on the basis that, anecdotally, accented speech is harder to understand and may be confused with other phonemes.

Results and Discussion

For each participant, identification accuracy was calculated as proportion correct scores for each stimulus (i.e., by dividing the number of times the participant responded that the token was from the intended phoneme by the total number of presentations). Each participant's mean rating of accent also was calculated for each token. The resulting scores and averages were submitted to two 2 x 2 x 3 x 6 repeated measure ANOVAs with consonant (/b/, /d/), vowel (/a/, /i/), consonant duration (native, average, non-native), and initial formant center frequency (native, 2, 3, 4, 5, non-native) as factors. All post-hoc, pair-wise comparisons of means were conducted using Bonferroni adjustments.

Identification Accuracy

Identification accuracy varied with consonant. While examples of both consonants were generally identified as the intended phoneme, /d/ was identified accurately significantly more often than /b/ (M = .915, SE = .023 vs, M = .849, SE = .031, respectively). This trend was confirmed by a main effect of consonant, $F(1,13)$ = 4.987, p = .044, η_p^2 = .277.

Identification performance also was influenced by interactions between several manipulated variables. One such interaction can be seen in Figure 1, which summarizes mean accuracy (and corresponding standard errors of measurement) from phoneme identification as of function of vowel duration and consonant. As the figure shows, within the /a/ vowel, the native duration (M = .886, SE = .022) was identified more accurately than the non-native duration (M = .861, SE = .024, p = .001). This contributed to a significant vowel by duration interaction, $F(2,26) = 4.558$, $p = .038$, $\eta_p^2 = .260$. This trend also contributed to a three way interaction between consonant, vowel, and consonant duration, such that within /a/ the extent to which /d/ was more accurately identified the /b/ depended on consonant duration (all *p's*<.02).

Another interaction is apparent in Figure 2, which depicts mean identification accuracy (and corresponding standard errors) as a function of vowel and initial formant center frequency values for the consonants. The figure shows that, for formant step 5, participants tended to identify /a/ (M = .855, SE = .029) less accurately than /i/ (M = .908, SE = .029, p = .041). This tendency led to a significant vowel x initial formant frequency interaction, $F(5,65) = 2.984$, $p = .017$, $\eta_p^2 = .187$. This tendency likewise contributed to a significant three-way interaction between vowel, initial formant frequency, and consonant, $F(5,65) = 2.704$, $p = .028$, $\eta_p^2 = .127$, such that within /a/ the extent to which /d/ tokens were identified more accurately than /b/ tokens depended on formant frequency (all *p's* < .04)

A final interaction is depicted in Figure 3, which summarizes mean identification accuracy (and corresponding standard errors) as a function of initial formant frequency and consonant duration. The graph shows that, within formant step 4, tokens with non-native durations (M = .843, SE = .027) were identified less accurately than were tokens with average durations (M = .877, SE = .027, p = .018) or native durations (M = .900, SE = .024, p = .002). . This contributed to a significant interaction between initial formant frequency and consonant duration, $F(10,130) = 1.996$, $p = .039$, $\eta_p^2 = .133$.

Accent Ratings

In marked contrast, there were no significant effects of variable manipulations on accent ratings (all p's > .1). Anecdotal reports from participants immediately following the experiment indicated occasional difficulty in perceiving the tokens as words or even as speech signals. This could be tied to the participants' ability to identify one phoneme better than the other. While identification was not necessarily poor overall, it was not reliable. If participants are not able to accurately identify the phonemes, we cannot expect them to make systematic ratings of accent involving those phonemes. When asking participants to judge accent, it is vital that they perceive the tokens as speech in order to be able to judge whether or not that speech is accented. This is perhaps a methodological issue when using synthesized speech in an accent task, where deliberate control of acoustic parameters can sometimes produce an impoverished result.

Due to the lack of effect of our manipulations on accent it was not likely that we would see a relationship between age of learning of our participants and their accent ratings. We correlated the participant's age of learning Spanish (M = 12.18, SE = .92) with the average overall accent rating and the average accent rating given to the stimuli we designed to be most accented. As expected, no relationship was found between participants' age of learning Spanish with their overall accent ratings (M = 2.82, SD =

.53*)* and the accent ratings for the stimuli we designed to be the most accented (*M* = 2.89, *SE* = .58*)*, *p* = .190 and *p* = .122 respectively.

Experiment 2

Experiment 2 investigated remaining temporal parameters that are a by-product of articulation and have previously been implicated as important in accented speech. Again, the focus was on information required for consonant perception with an emphasis on closure information. Since we wanted to investigate consonant information at the end of a word, we needed to use CVCs in Experiment 2. The parameters were VOT, stop closure duration, and vowel duration as a way of manipulating over all word duration. Experiment 2 approached synthesis in the same manner as Experiment 1. We predicted that longer durations of these parameters would result in higher accent ratings.

Methods

Participants

Thirteen students from the JMU Participant pool participated. All participants were between the ages 18 and 40 lowering the risk of presbycusis (loss of high frequency information that naturally occurs with age). All participants self-reported having normal hearing and were native speakers of American English to ensure that their judgments of accent were based off of American English. As in Experiment 1 all participants answered select questions from the LEAPQ (see Appendix 1). Twelve participants in Experiment 2 had experience with a language other than English (9 had experience with Spanish: 2 indicated that they were fluent).

Stimuli

Stimuli for Experiment 2 consisted of CVCs. The consonants surrounding the vowels were both stop consonants. We used two initial consonants while holding the vowel and ending consonant constant across the stimulus set. All durations were measured using Adobe Audition as in Experiment 1. VOT (in ms), was measured for both the native and non-native utterances. Stop closure duration, or the period where no voicing occurs in a stop consonant prior to the release burst, was measured for both native and non-native samples.

Consonant formant frequency, vowel formant frequency and vowel duration were modeled in *Praat* using Formant grids and a synthesized glottal pulse for all stimuli. Three vowel durations were synthesized; native, non-native, and an intermediate value (see Table 6). Consonant duration was held constant in Experiment 2.

A continuum of 4 VOT values was synthesized with measured native and non-native values for the intended phoneme as the end points (see Table 6). VOTs were synthesized by mixing natural samples in *Adobe Audition CS6* at the appropriate durations. They were concatenated to the beginning of the sound files.

Three stop closure durations were synthesized; native, non-native and an intermediate value (see Table 6). Silence was generated at the appropriate duration for each stop closure and was appended to the sound files.

For all stimuli containing a stop consonant, consonantal release bursts were synthesized by combining natural bursts in *Adobe Audition CS6*. The combination was appended at the time slice after the silence representing the stop closure. Burst duration was determined by averaging values for naturally produced release bursts by native and non-native speakers.

As in Experiment 1 the duration of the stimuli depended on measurements taken from samples with the average duration being 500 ms. There were 72 stimuli in total.

Procedure

The procedure in Experiment 2 including consent process, tasks, and experimental timing was the same as described in Experiment 1. The principle difference was the substitution of CVCs as a new stimulus set.

If vowel durations do contribute to accent, accent ratings would be expected increase as durations approach non-native values. Identification accuracy was not predicted to change across conditions but if it does, it would suffer closer to non-native values.

Results and Discussion

As in Experiment 1 identification accuracy and accent ratings were the aggregate of the participants' responses to all presentations of that token. These data were respectively submitted to corresponding 2 x4 x 3 x 3 repeated measures ANOVAs with consonant (/b/, /d/), VOT (native, 2, 3, non-native), stop closure duration (native, average, non-native), and vowel duration (native, average, non-native). Post-hoc pair-wise comparisons also were conducted on VOT, stop closure duration, vowel duration and any significant interactions to see where the differences, if any, fall as a function of parameter values. All pair-wise comparisons were made using a Bonferroni adjustment except when the test is too stringent to show inter-group differences. In this case, in order to further understand the nature of the effect, a LSD will be used but cautiously interpreted.

Identification Accuracy

There was a tendency to identify /d/ (*M* = .982, *SE* = .022) more accurately than /b/ (*M* = .932, *SE* = .005). The main effect of consonant on identification accuracy, while statistical significant, should not be regarded as a concern given that all but two participants were able to accurately identify both phonemes, and performance was close to ceiling. This tendency was confirmed by a main effect of consonant, $F(1,12) = 6.772$, $p = .023$, $\eta_p^2 = .361$.

Figure 4 presents average identification accuracy as a function of VOT. As the figure shows, there was a general tendency for identification accuracy to increase with VOT. This contributed to a main effect of VOT, $F(3,36) = 4.731$, $p = .013$, $\eta_p^2 = .283$. When using an LSD, the native VOT (*M* = .937, *SE* = .020) is identified less accurately then v2 ($p = .020$) and the non-native VOT ($p = .013$). The differences between the identification accuracy for each VOT condition are small considering the near-ceiling values. However, the effect of VOT on identification accuracy could be due to the abnormally short duration used for the “native” duration of VOT. Participants may not have been able to perceptually rely on the abnormally short VOT to differentiate the initial consonant as well as the longer VOT.

Accent Ratings

Mean accented ratings as a function of VOT, along with corresponding standard errors, are summarized in Figure 5. The figure shows that participants tended to rate the native VOT duration as more accented. This tendency contributed to a significant main effect of VOT on accented rating, *F*(3,36) = 2.944, $p = .046$, $\eta_p^2 = .197$. When using an LSD, Native VOT (*M* = 3.354, *SE* = .182) is rated as more accented than v2 (*M* = 3.256, *SE* = .171, $p = .019$) and is marginally more accented than the non-native VOT (*M* = 3.250, *SE* = .188, $p = .055$).

The main effect of VOT on accent rating was not initially as hypothesized but is still has interesting implications for the role of durational parameters in perceived accent. When looking at the duration values for VOT, it is apparent that the “native” value is abnormally short for what we would expect in natural speech, while the “non-native” value is just about average. Here, participants rated the abnormally short value as more accented than a value that we would expect to encounter. It is rare that in natural speech, participants would encounter a VOT duration as short as the one in our study, so it is unlikely to have found this result in a non-synthesized speech procedure. This suggests that the effect of VOT on accent may not be as simple as hypothesized, i.e., that longer duration producing higher accent ratings. Instead, any deviation from what participants expect to hear based on their familiarity with their native language could be attributed to accent.

Vowel duration also impacted accent ratings. This can be seen in Figure 7, which displays mean accent ratings (and corresponding standard error bars) as a function of vowel duration. There was a trend for accent ratings to increase with vowel duration. This trend contributed to a significant main effect on accent ratings, *F*(2,24) = 14.310, $p = .001$, $\eta_p^2 = .544$. Participants tended to rate tokens with the longer, native vowel duration (*M* = 3.577, *SE* = .185) as more accented than those with either an average vowel duration (*M* = 3.281, *SE* = .176, $p = .003$) or the shorted non-native duration value (*M* =

2.994, *SE* = .209, *p* = .006). In addition, CVC's with the average vowel duration were rated as more accented than those with the native duration, *p* = .038. Vowel duration had the hypothesized effect on accent rating. Participants rated the longer duration (actually the Native value) as more accented then the shorter duration.

Anecdotal reports from participants indicated that they would rely on the length of the stimulus to make their accent rating. Results supported this tendency with longer durations having higher ratings of accent. Participants were aware that duration was influencing their accent ratings. However, participants claimed that they had difficulty perceiving our stimuli as accented. This is reflected in the relatively low means for the accent ratings, which indicate that overall participants were having difficulty hearing accent in synthesized speech.

An attempt was made to correlate participants' age of learning Spanish (*M* = 10.6, *SE* = 1.07*)* with their overall accent ratings (*M* = 3.28, *SD* = .65) and their ratings of the stimuli designed to produce the highest accent ratings (*M* = 3.62, *SD* = .71*)*. However, this attempt did not show any significant relationship (*p* = .224 and *p* = .313, respectively). This suggests that experience with a language did not change the way in which participants perceived foreign accent within the stimuli of Experiment 2.

General Discussion

The obtained patterns of results from our experiments were not exactly as hypothesized. Experiment 1 failed to show any effect of manipulation on the primary dependent variable, accent ratings. As a result, one might argue that the parameters manipulated in this experiment, consonant initial formant center frequency and consonant duration (Vieru et al., 2011), must not play the role in perceived accent that has been suggested by prior research. However, such an assertion would be hasty. The main effect of consonant on identification accuracy suggests that participants were sometimes having difficulty accurately identifying the intended consonant. Specifically, participants were more variable in their perception of our /b/ tokens (i.e., made errors on a greater percentage of trials than for /d/). Considering that our manipulations were primarily concerned with the initial consonant, the participants' inconsistent performance suggests that they may not have always been receiving enough information from the consonant to identify it, much less rely on that information to make a higher order judgment such as accent rating. If the participants could not accurately identify the token on a given trial, then they could not be expected to be able to make systematic accent ratings of the same tokens.

While the overall identification accuracy in Experiment 1 was not necessarily poor, anecdotal reports from several participants suggest a default to responding /b/ when no initial consonant was perceived. This tendency could have inflated our identification accuracy of /b/ even though participants were not actually gaining any information from the initial consonant. If anecdotal reports are assumed to be valid indicators of perceptual tendency, then this increases the likelihood that the difficulty in assigning accent ratings may have been largely due to problems in perceiving the intended initial consonant.

There are several possible reasons why listeners may have experienced difficulty perceiving the consonants in Experiment 1. The first concerns a limitation in obtaining reasonable formant center frequency and bandwidth measurements at the beginning of the syllables. During measurement, formant values and bandwidths are obtained by averaging parameter values over a set window of time. As a result, each individual data point is influenced by the information before and after the critical sample point. The specific amount of surrounding information that influences an individual data point is based on the length of the time window specified by the researcher (25ms in our experiment). This is especially problematic when the parameter of interest occurs in a dynamic fashion such that the surrounding information can be drastically different than the critical measurement. In our case, a release burst prior to the start of voicing of the initial consonant may have influenced our bandwidth measurements. Specifically, release bursts produce very large bandwidth measurements due to the fact that they are transient noise with energy spanning large frequency ranges. As a result our measured bandwidths at the beginning of voicing were likely inflated, given that by definition, formants have narrow bandwidths. Having unusually large band widths can affect the perception of a critical formant frequencies by making them essentially no longer function as formants.

In addition, our initial formant center frequency could have also been influenced by reliance on measurement. Immediately following the initial consonant, the formant center frequency begins to transition in the direction of the vowel formant frequency. By averaging formant frequency values during this transition with our initial formant frequency measurement, our formant center frequencies could have appeared much closer to the vowel than in the actual utterance. The perceptual result could have been a phoneme that is not as differentially perceived.

One potential way to quickly address this problem would be to rely on fixed bandwidths instead of those obtained through measurement. While such a decision would necessarily reflect a partial sacrifice of natural signal parameters, there is a plethora of research that has been done regarding typical bandwidths at each formant center frequency used in synthesis. For example, Klatt & Klatt (1990) averaged phoneme productions over a variety of speakers to provide recommended formant values for use during synthesis. Reliance on parameters values averaged over a much larger sample than ours, found to be appropriate based on prior research, would allow us to be sure that our values are not being influenced by any non-critical areas of the spectrum and are empirically validated.

In addition, the participants anecdotally reported difficulty in perceiving the tokens as speech. In Experiment 1, we used CVs for the sake of simplicity of our manipulation. This made it impossible for participants to perceive the stimuli in Experiment 1 as words. While there is a long, successful history of using CVs in speech perception research (e.g., see Liberman, Harris, Kinney & Lane, 1961; also see Mattingly, Liberman, Syrdal, and Halwes, 1971), their use in accented ratings may be unintentionally making the task more difficult for participants than is necessary. Anecdotal reports suggested that participants had to first perceive the token as a target word and then decide if that instance of the word was accented. As Trofimovich & Isaacs (2012) reported there is a close link between intelligibility and accent. In their study, they found that less intelligible stimuli were rating as more accented. However in our case, lack of intelligibility may be linked to participants' inability to perceive accent in the signals to begin with, regardless of the manipulation of our parameters. Instead of relying on CVs for accent ratings, where word recognition may be difficult or impossible, CVCs may be necessary to ensure participants are able to perceive the signal as a specific word. Once participants are reliably able to perceive the signals as words, accented ratings should be easier.

Considering that measurement and methodological issues decreased identification accuracy in Experiment 1, the parameters of interest should not be dismissed as having no effect on perceived accent. Instead, follow-up studies are needed that correct for these issues in order to make any conclusive assertion about the effect of consonant initial frequency and consonant duration on perceived accent.

Experiment 2 found that VOT and vowel duration affected accent ratings while stop-closure duration did not. These results provide some support for our general hypothesis that accent ratings should increase as duration of our parameters increased. Additional steps were taken to ensure a higher level of control between speakers when synthesizing the stimulus set for Experiment 2. For example, since manipulation of formant center frequency was not an objective of Experiment 2, formant center frequencies and bandwidths were averaged over speakers. In addition, release bursts were synthesized for final consonants by combining natural signals from both speakers to ensure that the accent rating would not be influenced by the release bursts. By using this method for the synthesis of the release burst, a better approximation of natural speech was obtained. This should aid in participants ability to perceive the stimulus as a word. Furthermore, as already discussed, the use of CVCs in Experiment 2 also made it possible for participants to perceive the tokens as specific words. Both of these attempts at making the stimulus easier to identify should have aided in the participants’ ability to reliably rate the accent of each signal.

The higher accent ratings resulting from longer vowel durations in Experiment 2 confirm the role of vowel duration (Sidaras et al. 2009, Vieru et al., 2011) in accent ratings. In our study, as vowel duration increased, so did accent ratings. Anecdotal reports from participants indicated that they felt that they often relied upon the length of the signal to judge the accent. Participants claimed longer durations were judged to come from accented speakers because the non-native speaker would “take longer” to produce the signal. Signal length should have been most apparent from our manipulation of vowel duration since the vowel constituted the longest segment of the CVC (35 ms consonant vs. 92 ms vowel for /b/, and 52 ms consonant vs. 53 ms vowel for /d/). This is exactly in agreement with our hypothesis that accented speakers are slower in their movements of their articulators suggesting unfamiliarity with non-native phonemes.

The role of VOT (Arslan & Hansen, 1997) in foreign accent also was confirmed in Experiment 2, such that as VOT increased, accent ratings decreased. This was not in concordance with our original hypothesis that longer duration would produce higher accent ratings. Rather, in our experiment syllables with “native” VOT values were rated more accented than those with non-native VOT values. A reasonable explanation for this effect derives from the fact that the “native” VOT duration was abnormally short. This was a by-product of an attempt to adjust the duration of natural samples to ensure that the average length of stimuli was 500 ms. However, it is very unlikely that a person would ever encounter such an abnormally short VOT in typical speech contexts. As a result, it is possible that the effect of an abnormally short VOT may have never been seen outside the laboratory. The “non-native” VOT value was a value more typical of American English, and as a result, should be expected to be unlikely to lead to perception of a heavy foreign accent.

Considering the inability to observe unnatural durations outside of the laboratory, the effect of durational parameters on accent may not be as simple as the longer the duration that more accented the signal will be judged. Instead, it may be that any deviation from what is normally expected by the listener may be judged as accented. As a further test of this suggestion, future research could extend the range of durational parameters both above and below expected/average parameter values derived from large databases of productions. This would permit a more complete evaluation of whether any deviation from (i.e., shorter or longer than) typical values affects accent ratings for other duration-based speech parameters.

We failed to show an effect of stop-closure duration (Arslan & Hansen, 1997) on accented ratings. Previous work with silence durations in speech has shown that duration of silence between syllables can change perceived phonemes (Liberman, Harris, Eimas, Lisker &Bastian, 1961). The fact that phoneme differences can be derived by silence durations suggests that it is an important cue in our perception of speech. While our attempt to manipulate accent through stop-closure duration was restricted to within category ranges of values, we can be sure that participants were able to perceive a noticeable difference in the duration of our stop-closure durations. Abel (1972) found that a difference of 10 ms at 85 dB can be reliably distinguished and our duration differences were above 30 ms. The lack of effect of stop-closure duration on accent suggests that while stop-closure may be an important cue to the discrimination of phonemes, it may not be particularly salient in the perception of accent. It is possible that previous studies showed an effect of stop-closure duration on accent because participants were using stop-closure as a cue to word length (Arslan & Hansen, 1997). If participants were using stop-closure duration as a way to judge the length of the word, then longer stop-closures would have been expected to lead to higher accent ratings. However, in our study, when stop-closure was pitted against a more salient cue to word length, vowel duration, the length of stop-closure failed to have any effect on accent.

In our experiment, stop-closure was synthesized by appending various lengths of silence to the signal. The lack of an effect of this manipulation suggests that participants did not gain any relevant information from this silence and therefore may not have relied on it at all when making accent judgments. Future investigations should further investigate the potential contribution of stop-closure to accent in the absence of other manipulations to provide a fair evaluation of whether, in isolation, stop closure has any effect on perceived accent. If no effect is observed under such conditions, then it can be concluded that stop-closure does not contribute to perceived accent, consistent with our initial findings. If instead an effect of stop-closure duration is found when manipulated in isolation, then this would suggest that it might represent a much less salient cue to foreign accent when presented in the context of other relevant variables.

Extending the suggestion that age of learning (AOL) of the talker may affect the accentedness of the speech, such that younger AOL produces less accented speech (Flege et al., 1995; Flege, 1988) we decided to investigate how listener language experience may affect their judgments of accent. We obtained the age in which our participant started learning Spanish through select questions on the LEAPQ (Marian, et al., 2007) and correlated these ages with the participant’s overall accent ratings and their ratings of the stimuli designed to produce the highest accent ratings in both experiments. We failed to find a relationship between listener AOL and their judgment of accent. However, it is important to note that most of these participants learned Spanish as part of their high school course requirement and have relatively little experience with the language outside of school. This was indicated by most participants failing to reach fluency or living in a region where the language is spoken. So our measure of AOL may not have been reflective of actual years of experience with the language. Since the majority of Flege's research suggests that the amount of experience a person has with L2 is the underlying variable that correlates with their rating of accent, a more concrete measure of language experience may show a stronger relationship.

Conclusion

Our experiments confirmed that VOT (e.g., see Arslan & Hansen, 1997) and vowel duration (Sidaras et al. 2009, Vieru et al., 2011) both contribute to foreign accent ratings. It is apparent by these results that listeners do take durational information into account when making judgments of whether or not speech is accented. Based on our findings for VOT, our original hypothesis that longer durations lead to higher accent ratings had to be revised to include deviation from expected rather than longer duration.

The lack of results for additional parameters cannot necessarily discount their role in accent until subsequent studies with higher levels of control continue to show that they have no effect on participants' accent ratings. It could be that when not in competition with vowel duration, a more salient cue to word duration, stop-closure duration could contribute to accent ratings. This would suggest that some parameters have stronger relative contributions to accent ratings than others. A hierarchy of parameters contributing to accent could only be explored through additional studies aimed at orthogonally manipulating the parameters of interest, while controlling for word perception and extraneous measurements, to see which participants choose to rely upon to make their ratings.

Table 1

Consonant durations and vowel durations (ms) for CVs in Experiment 1

CV	/ba/	/bi/	/da/	/di/
Native	7.9	9.9	22.3	13.6
Average	21.2	23.4	36.8	20.4
Non-Native	34.5	36.9	51.4	27.2
Vowel Duration	222.4	219.7	200.1	210.8

Table 2

Initial consonant and vowel formant center frequencies and bandwidths for /ba/ in Hz

Consonant				
	F1(BW1)	F2(BW2)	F3(BW3)	F4(BW4)
Native	680(129)	1235(225)	2839(179)	3922(518)
2	680(129)	1251(213)	2813(203)	3922(518)
3	680(129)	1266(201)	2787(228)	3922(518)
4	680(129)	1281(189)	2762(254)	3922(518)
5	680(129)	1297(177)	2737(279)	3922(518)
Non-Native	680(129)	1313(166)	2712(306)	3922(518)
Vowel				
	F1(BW1)	F2(BW2)	F3(BW3)	F4(BW4)
Native	848(129)	1275(90)	2677(138)	4039(376)
2	848(129)	1288(105)	2712(160)	4039(376)
3	848(129)	1301(120)	2748(183)	4039(376)
4	848(129)	1314(136)	2783(206)	4039(376)
5	848(129)	1328(151)	2819(229)	4039(376)
Non-Native	848(129)	1341(167)	2856(254)	4039(376)

Table 3

Initial consonant and vowel formant center frequencies and bandwidths for /bi/ in Hz

Consonant	F1(BW1)	F2(BW2)	F3(BW3)	F4(BW4)
Native	363(68)	2259(258)	2825(276)	3902(372)
2	363(68)	2276(292)	2827(272)	3902(372)
3	363(68)	2292(327)	2829(268)	3902(372)
4	363(68)	2308(363)	2831(264)	3902(372)
5	363(68)	2325(400)	2832(260)	3902(372)
Non-Native	363(68)	2341(438)	2834(256)	3902(372)

Vowel	F1(BW1	F2(BW2)	F3(BW3)	F4(BW4)
Native	368(85)	2654(203)	3060(402)	4200(386)
2	368(85)	2699(217)	3183(403)	4200(386)
3	368(85)	2744(229)	3309(405)	4200(386)
4	368(85)	2790(243)	3439(407)	4200(386)
5	368(85)	2837(257)	3573(409)	4200(386)
Non-Native	368(85)	2884(271)	3711(410)	4200(386)

Table 4

Initial consonant and vowel formant center frequencies and bandwidths for /da/ in Hz

Consonant				
	F1(BW1	F2(BW2)	F3(BW3)	F4(BW4)
Native	523(82)	1757(166)	3085(250)	4089(454)
2	523(82)	1744(214)	3054(251)	4089(454)
3	523(82)	1730(264)	3023(251)	4089(454)
4	523(82)	1717(317)	2993(251)	4089(454)
5	523(82)	1704(371)	2962(251)	4089(454)
Non-Native	523(82)	1691(428)	2932(251)	4089(454)
Vowel				
	F1(BW1)	F2(BW2)	F3(BW3)	F4(BW4)
Native	855(186)	1319(116)	2641(238)	3987(266)
2	855(186)	1314(122)	2651(316)	3987(266)
3	855(186)	1308(128)	2660(398)	3987(266)
4	855(186)	1303(134)	2669(486)	3987(266)
5	855(186)	1297(140)	2679(579)	3987(266)
Non-Native	855(186)	1291(147)	2689(679)	3987(266)

Table 5

Initial consonant and vowel formant center frequencies and bandwidths for /di/ in Hz

Consonant	F1(BW1)	F2(BW2)	F3(BW3)	F4(BW4)
Native	316(185)	2388(314)	3183(315)	4451(579)
2	316(185)	2429(323)	3227(388)	4451(579)
3	316(185)	2470(333)	3272(466)	4451(579)
4	316(185)	2512(342)	3317(547)	4451(579)
5	316(185)	2555(352)	3363(633)	4451(579)
Non-Native	316(185)	2597(361)	3409(724)	4451(579)

Vowel	F1(BW1)	F2(BW2)	F3(BW3)	F4(BW4)
Native	335(53)	2739(261)	3084(364)	4299(659)
2	335(53)	2770(239)	3218(377)	4299(659)
3	335(53)	2801(217)	3355(389)	4299(659)
4	335(53)	2832(195)	3497(402)	4299(659)
5	335(53)	2863(174)	3644(416)	4299(659)
Non-Native	335(53)	2895(153)	3795(429)	4299(659)

Table 6.

VOT, consonant, vowel and stop closure durations for CVCs in ms

	VOTs		Consonant duration	Vowel duration		Stop Closure Duration	
bot	Native	7.8	34.9	Native	282.1	Native	95.6
	2	9.2		Average	220.9	Average	137.1
	3	10.7		Non-native	159.6	Non-native	178.6
	Non-Native	12.1					
dot	Native	11.6	52.9	Native	290.1	Native	96.9
	2	16.5		Average	239.5	Average	135.4
	3	21.3		Non-native	188.9	Non-native	173.9
	Non-Native	26.2					

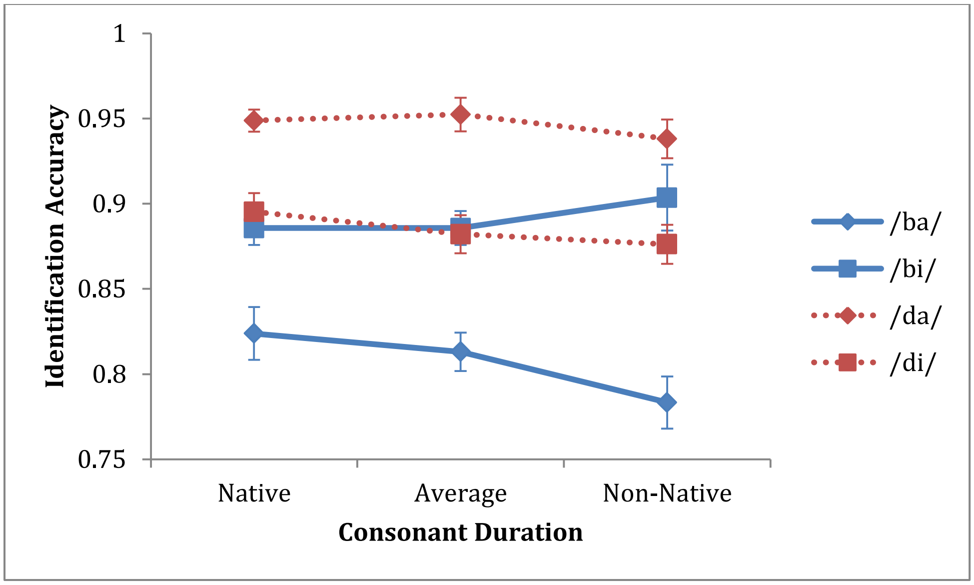

Figure 1. Average identification accuracy by duration, consonant and vowel with standard errors

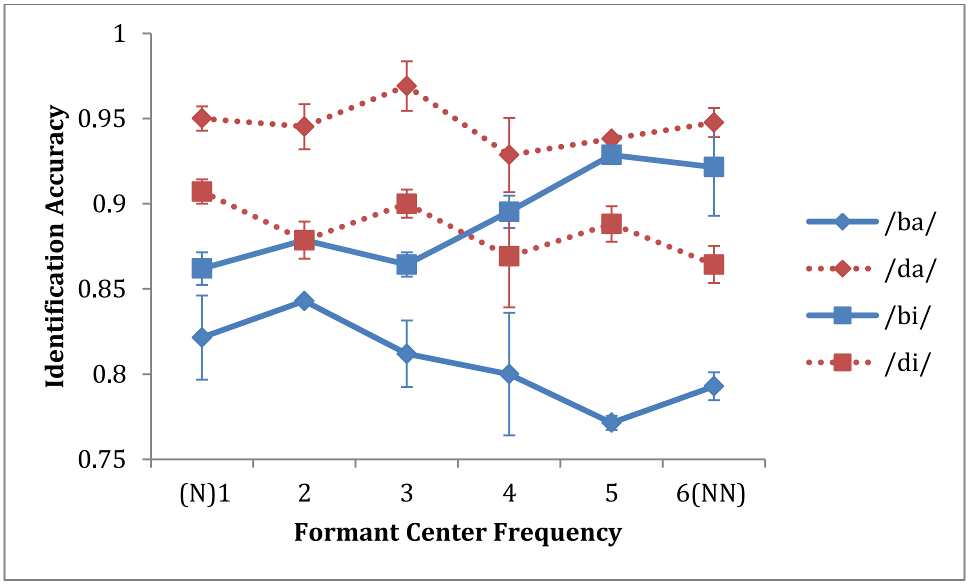

Figure 2. Average identification accuracy by formant center frequency, consonant and vowel with standard errors

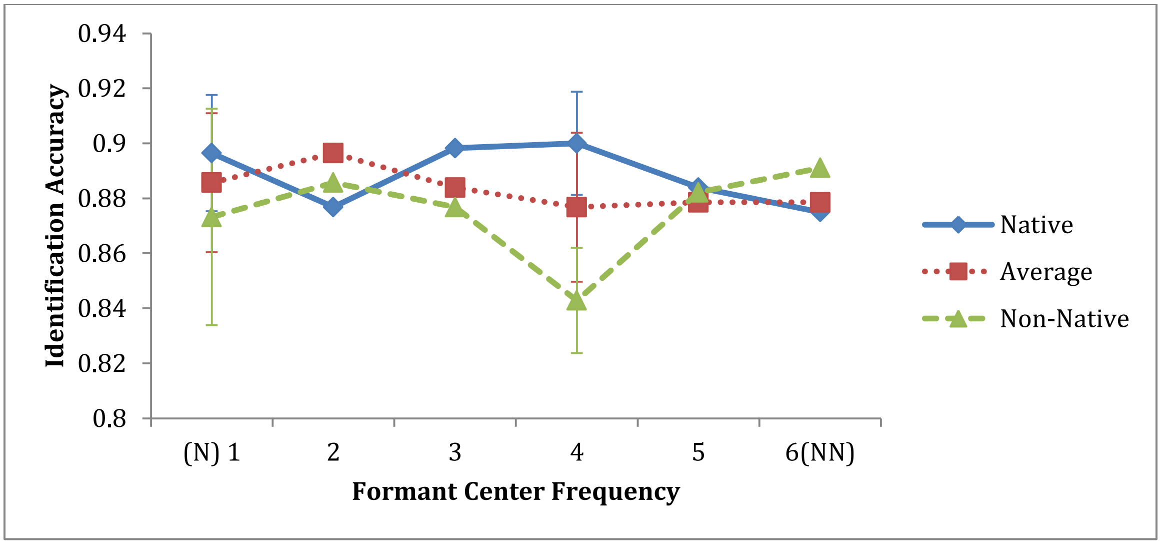

Figure 3. Average identification Accuracy as a function of consonant duration by formant center frequency with standard errors. (Please note: identification accuracy scale has been expanded for clarity).

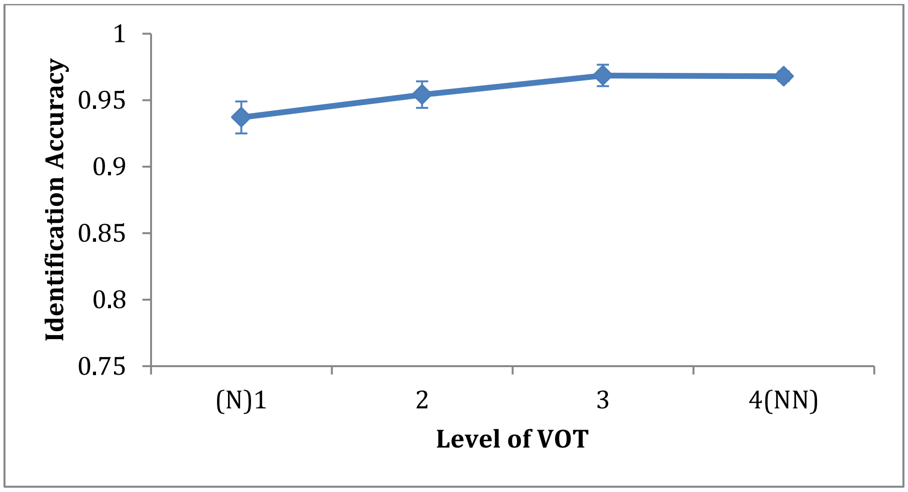

Figure 4. Average identification accuracy by VOT with standard errors

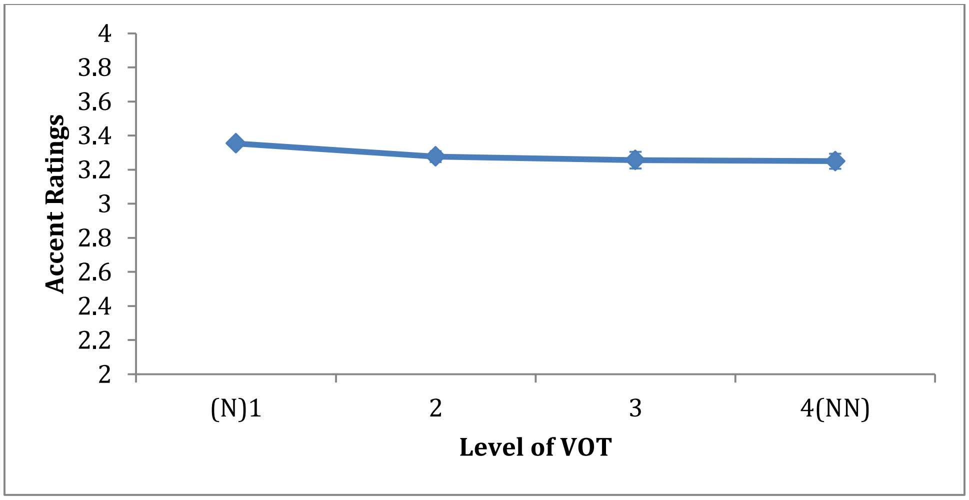

Figure 5. Average accented rating by VOT with standard errors

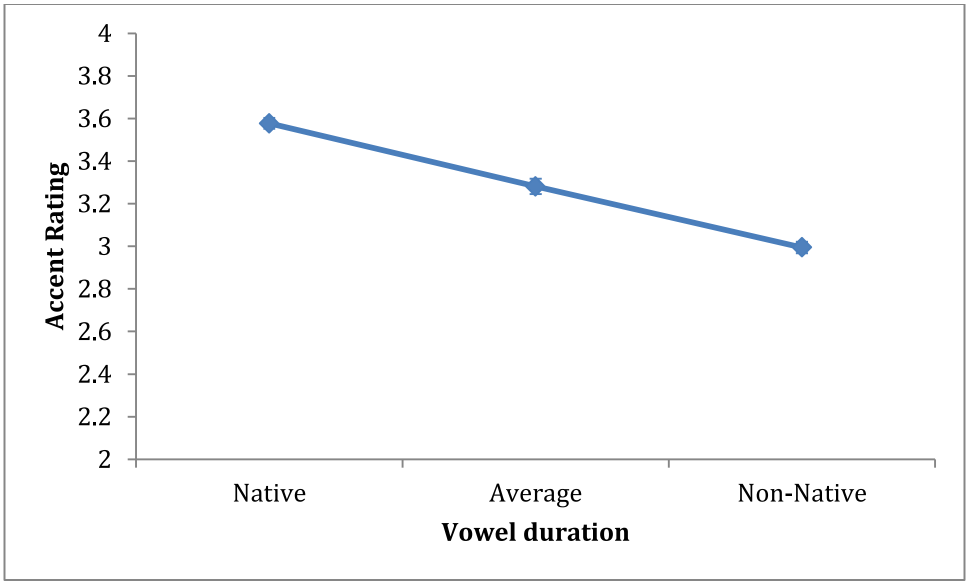

Figure 6. Average accented rating by vowel duration with standard errors

Appendix 1

Select Questions from the LEAPQ

(1) Please list all the languages you know **in order of dominance**:

1	2	3	4	5

(2) Please list all the languages you know **in order of acquisition** (your native language first):

1	2	3	4	5

(3) Please list what percentage of the time you are *currently* and *on average* exposed to each language.

(*Your percentages should add up to 100%*): **List language here:**
List percentage here:

(TO BE COMPLETED FOR EACH LANGUAGE)

Language:
This is my (**native second third fourth fifth**) language.

(1) Age when you…

began acquiring this language:	*became fluent in this language*:	*began reading in this language*:	*became fluent reading in this language*:

(2) Please list the number of years and months you spent in each language environment:

	Years	Months
A country where this language is spoken		
A family where this language is spoken		
A school and/or working environment where this language is spoken		

References

Abel, S. M. (1972). Discrimination of temporal gaps. *Journal of the Acoustical Society of America*, 52, 519-524.

Adobe Systems Incorporated. (2012). Adobe Audition (Version CS6) [computer software]. San Jose, CA: Adobe Systems Incorporated.

Arslan, L., & Hansen, J. (1997a). A study of temporal features and frequency characteristics in American English foreign accent. *Journal of the Acoustical Society of America, 102*(1), 28–40. Retrieved from http://link.aip.org/link/?JASMAN/102/28/1

Bent, T. & Bradlow, A. R. (2003). The interlanguage speech intelligibility benefit. *Journal of the Acoustical Society of America, 144*(3), 1600-1610.

Boersma, P. & Weenink, D. (2011). Praat [computer software]. Available from http://www.fon.hum.uva.nl/praat/

Braun, B., Lemhöfer, K., & Mani, N. (2011). Perceiving unstressed vowels in foreign-accented English. *The Journal of the Acoustical Society of America, 129*(1), 376–87. doi:10.1121/1.3500688

Chan, K. Y., Hall, M. D., & Assgari, A. (2013). An evaluation of the role of vowel formant frequencies on the perception of foreign accent. *Poster presentation at APCAM.*

Evans, B. G., & Iverson, P. (2007). Plasticity in vowel perception and production: a study of accent change in young adults. *The Journal of the Acoustical Society of America, 121*(6), 3814–26. doi:10.1121/1.2722209

Fant, G. (1973) Speech sounds and features. Cambridge: MIT Press.

Flege, J. E., (1988). Factors affecting degree of perceived foreign accent in English sentences. *The Journal of the Acoustical Society of America, 84*(1), 70–9. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/3411056

Flege, J. E., & Hillenbrand, J. (1984). Limits on phonetic accuracy in foreign language speech production. *The Journal of the Acoustical Society of America, 76*(3), 708–721. Retrieved from http://link.aip.org/link/?JASMAN/76/708/1

Flege, J. E., Munro, M. J., & MacKay, I. R. (1995). Factors affecting strength of perceived foreign accent in a second language. *The Journal of the Acoustical Society of America, 97*(5), 3125–34. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/7759652

Flege, J. E., Bohn, O.-S., & Jang, S. (1997). Effects of experience on non-native speakers' production and perception of English vowels. *Journal of Phonetics, 25*(4), 437–470. doi:10.1006/jpho.1997.0052

Flege, J. E., Frieda, E., & Nozawa, T. (1997). Amount of native-language (L1) use affects the pronunciation of an L2. *Journal of Phonetics*, 169–186. Retrieved from http://www.sciencedirect.com/science/article/pii/S0095447096900406

Flege, J. E., MacKay, I. R., & Meador, D. (1999). Native Italian speakers' perception and production of English vowels. *The Journal of the Acoustical Society of America, 106*(5), 2973–87. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10573909

Flege, J. E., Schirru, C., & MacKay, I. R. (2003). Interaction between the native and second language phonetic subsystems. *Speech Communication, 40*(4), 467–491. doi:10.1016/S0167-6393(02)00128-0

Herd, W., Jongman., A., & Sereno, J. (2013) Perceptual and production training of intervocalic /d, ɾ, r/ in American English learners of Spanish. *Journal of the Acoustical Society of America, 133*(6), 4247-4255.

Idemaru, K., & Holt, L. L. (2013). The development trajectory of children's perception and production of English /r/-/l/. *Journal of the Acoustical Society of America, 133*(6), 4232-4246.

Kang, O., Rubin, D., & Pickering, L. (2010). Suprasegmental measures of accentedness and judgments of language learner proficiency in oral English. *The Modern Language Journal*. doi:10.1111/j.1540-4781.2010.01091

Klatt, D. H., & Klatt, L. C. (1990). Analysis, synthesis and perception of voice quality variations among female and male talkers. *Journal of the Acoustical Society of America, 87*(2), 820-857.

Levi, S. V., Winters, S. J., & Pisoni, D. B. (2007). Speaker-independent factors affecting the perception of foreign accent in a second language. *The Journal of the Acoustical Society of America, 121*(4), 2327. doi:10.1121/1.2537345

Liberman, A., Harris, K. S., Hoffman, H. S., & Griffith, B. C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology, 54*(5), 358–368.

Liberman, A., Harris, K. S., Eimas, P., Lisker, L., & Bastian, J. (1961). An effect of learning on speech perception: The discrimination of durations of silence with and without phonemic significance. *Language and Speech, 4*(4), 175-195.

Liberman, A., Harris, K. S., Kinney, J., & Lane, H. (1961). The discrimination of relative onset-time of the components of certain speech and non-speech patterns. *Journal of Experimental Psychology, 61*(5), 379-388.

Marian, V., Blumenfeld, H. K., Margarita, K. (2007). The language experience and proficiency questionnaire (LEAP-Q): Assessing language profiles in bilinguals and multilinguals. *Journal of Speech, Language, and Hearing Research. 50*. 940-967.

Mack, M. (1982). Voicing-dependent vowel duration in English and French: Monolingual and bilingual production. *The Journal of the Acoustical Society of America, 71*(1), 173. doi:10.1121/1.387344

Magen, H. (1998). The perception of foreign-accented speech. *Journal of Phonetics.* Retrieved from http://www.sciencedirect.com/science/article/pii/S009544709890081X

Pinet, M., & Iverson, P. (2010). Talker-listener accent interactions in speech-in-noise recognition: effects of prosodic manipulation as a function of language experience. *The Journal of the Acoustical Society of America, 128*(3), 1357–65. doi:10.1121/1.3466857

Piske, T., MacKay, I., & Flege, J. (2001). Factors affecting degree of foreign accent in an L2: A review. *Journal of phonetics*, 191–215. doi:10.006/jpho.2001.0134

Schneider, W., Eschman, A., & Zuccolotto, A. (2002). E-prime reference guide Psychology Software Tools, Incorporated.

Sidaras, S., Alexander, J., & Nygaard, L. (2009). Perceptual learning of systematic variation in Spanish-accented speech. *The Journal of the Acoustical Society of America. 125*(5), 3306-3316. doi:10.1121/1.3101452

Tajima, K., Port, R., & Dalby, J. (1997). Effects of Temporal Correction on Intelligibility of Foreign-Accented English. *Journal of Phonetics*, 25. Retrieved from http://www.sciencedirect.com/science/article/pii/S0095447096900315

Trofimovich, P., & Isaacs, T. (2012). Disentangling accent from comprehensibility. *Bilingualism: Language and Cognition*. doi:10.1017/S1366728912000168

Vieru, B., De Mareüil, P., & Adda-Decker, M. (2011). Characterisation and identification of non-native French accents. *Speech communication*, *53*, 292–310. doi:10.1016/j.specom.2010.10.002