January 2015

Hesitation Rate as a Speaker-Specific Cue in Bilingual Individuals

Jamie Lynn Armbrecht
University of South Florida, jamiearmbrecht@hotmail.com

Hesitation Rate as a Speaker-Specific Cue in Bilingual Individuals

by

Jamie Armbrecht

A thesis submitted in partial fulfillment
of the requirements for the degree of
Master of Science
Department of Communication Sciences and Disorders
College of Behavioral and Community Sciences
University of South Florida

Major Professor: Ruth Huntley-Bahr, Ph.D.
Nathan Maxfield, Ph.D.
Kyna Betancourt, Ph.D.

Date of Approval:
June 5, 2015

Keywords: hesitation, bilingual, speaker identification, spontaneous speech, reading

Acknowledgments

To my major professor, Dr. Ruth Huntley Bahr, I am eternally grateful for all of your support and guidance in writing this thesis. I would like to thank you for reading my countless number of drafts and answering my endless number of questions. I believe that you truly challenged me to make this thesis my own and to learn how to complete a research project.

I would like to thank Dr. Kyna Betancourt, who was immensely helpful in my data analysis. I am so grateful to you for your assistance in creating the Praat script for my procedures, and all of your support. You truly saved me an enormous amount of time, while educating me on how to utilize speech analysis programs.

I would also like to thank Dr. Nathan Maxfield, for serving on my thesis committee. Your presence in the research field is very inspiring and helped to encourage my interest in completing a thesis.

To the academic and clinical faculty of the Department of Communication Sciences and Disorders, thank you for encouraging the pursuit of research, while being incredibly accommodating to the research process. Without the support of the instructors I have encountered, I do not believe I would have been able to finish this thesis in such a timely manner.

Finally, I would like to thank my family and friends for being so supportive and sympathetic to me during this thesis process. During the completion of this thesis you have listened to me incessantly vent, while continuing to encourage me.

Table of Contents

List of Tables

List of Figures

Abstract

Hesitation use is common among all speakers, regardless of whether they are engaged in their dominant or non-dominant language (Fehringer & Fry, 2007; Reed, 2000). The question is whether a bilingual speaker will engage in the same types of hesitations in both languages. If hesitation patterns can be identified consistently across speakers regardless of language, their use as an acoustic cue for speaker identification may be possible. This study examines differences in hesitation use across languages and speaking contexts (reading vs. conversation) in bilingual speakers.

Twenty Spanish-English bilinguals (ages 19 -31 years) were tested as part of a larger speaker identification project focusing on bilingual speech patterns. These individuals were recorded in a sound-treated booth while speaking extemporaneously and reading a standardized passage in both Spanish and English. Unfilled pause length and speech segment durations were obtained from one minute speech samples using Praat scripts (Boersma & Weenink, 2014). Pause to speaking ratios were computed in Excel. The number of filled pauses were determined from the same one minute speech samples in English and Spanish. Differences in planning style were demonstrated with step graphs which compared both the frequency and length of alternations between speech and pauses in two participants with different planning styles.

Wilcoxon signed ranks tests revealed significant differences in the use of unfilled pauses across speaking contexts in both languages. Both pause to speaking ratios and pause durations were larger in spontaneous speech when compared to read speech. Speech segment durations

were shorter in extemporaneous speech and filled pauses were more common in spontaneous speech.

Cross-language comparisons were considered within each speaking condition. Results indicated few instances where there were significant differences. There were longer speech segment durations in read speech and more filled pause use in spontaneous speech in English. Further demonstration of these patterns was illustrated through step graphs.

The similarities in the hesitation phenomenon between languages suggests that bilingual speakers often use the same planning aspects between languages and carryover aspects of speech production from their first language to their second (Fehringer & Fry, 2007). Therefore, comparisons within and across languages within a specific speaking condition may be useful in speaker identification. However, these findings also indicate the need for caution when comparing speech samples across speaking conditions using unfilled and filled pauses. One should consider hesitation as one of several acoustic cues for use in speaker identification in a cross-language situation.

Chapter One

Literature Review

Hesitation is evident in speakers of every language, and is present whether the speaker is engaged in their dominant or non-dominant language (Fehringer & Fry, 2007; Reed 2000). However, the type and amount of hesitation use may not follow the same pattern across languages, as speakers may increase the number of hesitations used while speaking or alternate their use of filled and unfilled pauses depending on the language being spoken and their familiarity with that language. The presence of a different hesitation pattern across languages would be related to difference in cognitive planning, with speech production in the dominant language requiring less planning time than production in the non-dominant language (Butterworth, 1975; Wiese, Dechert, Möhle, & Raupach, 1984). However, if a speaker becomes more competent in their non-native language, their use of hesitation may also decrease (Fehringer & Fry 2007), causing both languages to become more similar in their fluency and hesitation patterns.

If differences in the use of hesitation can be identified consistently across languages, then this acoustic cue can be used to identify a particular speaker. However, research in the area of bilingual speaker-specific phenomena has been sparse (Fathman, 1980). Therefore, more research is needed in this area to determine the utility of hesitation analysis in the identification of individual speakers and possibly assisting in the forensic identification of speakers.

The aim of this project was to explore the role of hesitation in speaker identification involving bilingual speakers. Differences in the use of hesitation across English and Spanish in

two different speaking contexts (spontaneous speech and reading) were examined. In the following literature review, the presence of hesitation in speech, in both reading and spontaneous productions is first defined. Next, hesitation as a planning device and its variation based on speaking context is discussed. Later language storage in relation to bilingual speakers will be addressed, followed by the difference in bilingual speaker's hesitation patterns. Finally, the notion of speaker identification will be reviewed.

Use of Hesitation Phenomena

Whether a speaker is engaged in oral reading or spontaneous speech, s/he is likely to exhibit hesitation. These interruptions in the flow of speech are meant to help a speaker achieve accurate expression of his/her thoughts, rather than hinder expression (Temple, 1992). Hesitation is used as a quality control device to anticipate errors occurring in speech, as well as to repair them once they have occurred (Reed, 2000).

Because pauses naturally occur between words, it is important to identify the significance of these pauses by their length. While several different pause lengths have been proposed ranging from 200ms (Beattie, 1979; Butterworth, 1975**)** to about 4000ms (Cenoz, 2000)**,** Goldman-Eisler (1972) has suggested that meaningful, unfilled hesitations consist of periods of silence lasting longer than 250ms. This length is significant because it is longer than the time needed for planning, physiological, or articulatory movements and it is suggestive of cognitive processing time. However, pauses lasting longer than 2000ms are not considered meaningful, as they are not indicative of continuous on-line planning, and only occur in select speakers, due to individual speaker variability (Cenoz, 2000).

Pauses less than 25ms are reflective of the more fluent transitions that occur naturally in speech production. Short pauses are often physiologically necessary (Cenoz, 2000), as a speaker

is physically required to interrupt their speech to accommodate breathing (O'Connell & Kowal, 2005), and they account for the time needed for articulatory movements. If a speaker avoids short pauses, speech units become too dense, impacting both the production of intelligible speech and the listener's capacity to understand the spoken message (O'Connell & Kowal, 2005).When planning periods are extended, both the production of speech and transitions become less fluent.

Regardless of their length, silent, unfilled pauses occur frequently during speech production. These pauses can occur in many different places, such as between sentences, clauses, and within clauses. They often act as a precursor for repair devices, which are actions designed to clarify speech attempts, including self-corrections, repetitions, and reformulations (Cenoz, 2000).

Not all pauses are silent, however. Filled pauses occur when the speaker uses words, such as *um* and *uh*, to facilitate speech planning (Clark & Fox Tree, 2002; Wiese et al., 1984). These pauses also can consist of specific sounds to be used as fillers, such as /æ, ɑ, r, ə, m/ (Goldman-Eisler, 1961). Although speakers often alternate between unfilled and filled pauses regularly, each pause type can serve separate and specific functions.

Filled pauses can act as both a precursor to a repair device, or as their own repair device (Cenoz, 2000). When functioning as a repair device, the filled pause will frequently occur in the clause initial position (Cenoz, 2000; Engelhardt, Nigg, & Ferreira, 2013) and will be utilized to correct an already communicated idea. Filled pauses can also be used for planning, such as when a speaker encounters a change in idea or focus while speaking. Hence, the need for greater cognitive activity during difficult speech planning tasks drives speakers to engage in the use of filled pauses (Cenoz, 2000).

When not acting as repair devices or planning strategies, filled pauses can also be used by the speaker pragmatically to introduce new information and maintain proper turn-taking (Englehardt et al., 2013). This use of pause appears to be more of a voluntary action by the speaker. Pause also has a syntactic function, known as juncture. These types of pauses are defined as fluent pauses that are used to mark the boundaries between syntactic units (Cenoz, 2000). It helps to establish shifts in ideas or focus. Some examples of juncture include *I scream* versus *ice cream* or *let's eat, grandma* versus *let's eat grandma*. Occasionally, these pauses will occur in places that are not considered syntactic boundaries. When pauses do not occur at syntactic boundaries, they are judged to be unnecessary and signal the speaker's uncertainty in their message (Cenoz, 2000).

Speakers demonstrate a preference for using either filled or unfilled pauses while planning utterances (Cenoz, 2000; Goldman-Eisler, 1961 Maclay & Osgood, 1959). However, there are speech factors that would cause an individual to be more likely to use one type of hesitation over the other. Unfilled pauses are more likely to occur before lexical as opposed to function words (Maclay & Osgood, 1959).They also are closely associated with a speaker's uncertainty in their word choice (Goldman-Eisler, 1961). Filled pauses are more likely to occur at phrase boundaries, rather than within phrases and are associated with uncertainty of what to say next (Maclay & Osgood, 1959). Although these factors may influence which type of hesitation is used, individual speaker preference for pause type should not be overlooked (Goldman-Eisler, 1961).

Hesitation as a Planning Device

Regardless of whether a speaker is using filled or unfilled pauses, the presence of hesitation is continuously occurring in speech. Since speech is considered to be a higher-level

cognitive activity (Butterworth, 1975), the placement of hesitations compared to periods of fluent speech is not haphazard (Beattie, 1979). Instead, the speaker displays certain patterns of hesitant versus fluent speech, which can be viewed across different speaking contexts, such as spontaneous speech versus oral reading. Additionally, a speaker's intent also will impact their use of hesitation as a planning strategy.

Hesitations tend to occur in a cyclic manner, consisting of alternating periods of fluent speech and pauses (Beattie, 1979; Butterworth, 1975). These alternating cycles are not random, but rather are linked to the planning aspects of all speech production (Butterworth, 1975). First, the speaker is engaged in the process of word choice and semantic planning, which is then followed by syntactic planning. The speaker not only looks ahead in terms of clauses and sentences, but also considers how to combine these clauses and sentences most effectively to convey his/her desired meaning (Butterworth, 1975). This planning process is considered a higher-level creative and cognitive activity and the patterns of hesitation exhibited by the speaker are directly reflective of these higher levels of cognition (Butterworth, 1975). Without engaging in this complex cognitive planning, fluent speech would not be possible (Butterworth, 1975).

The presence of higher-level cognitive planning during spontaneous speech seems to affect the use of hesitation more in spontaneous than in prepared speech. For instance, spontaneous speech is considered to be more disfluent (Clark & Fox Tree, 2002) and lacks fluent transitions between sentences (Goldman-Eisler, 1972). These pauses are likely due to the greater degree of planning necessary for the production of spontaneous speech. On the other hand, prepared speech, such as reading, requires less cognitive planning. In this case, the speaker's only task is generate the speech necessary to convey the written text.

Similar to spontaneous speech, the production of oral narratives also requires a high degree of planning (Chafe, 1980). When a speaker produces a narrative, s/he forms groups of ideas called focus clusters, which are large units comprised of foci of consciousness, or smaller ideas. These clusters are grouped together to produce sentences. When producing speech spontaneously, an individual is likely to move through different focus clusters, as his/her ideas shift. Therefore, the speaker is more likely to hesitate between focus clusters than within a focus cluster (Chafe, 1980). This increase in hesitation length is associated with the mental processing needed to change ideas. However, when the focus change occurs in conjunction with other aspects of spontaneous speech, such as choices in vocabulary or grammatical structure, increases in both hesitation presence and length become even more prominent (Chafe, 1980).

However, a different picture of hesitation is noted when someone is asked to recall something previously seen, read, or heard. In this situation, the speaker is required to reflect on previous information and plan the narrative accordingly. This is where the pre-planning aspects of idea generation, vocabulary selection, and grammatical structure come into play. The pre-planning of recall makes it more like spontaneous speech in that the speaker has to produce an original narrative. The result can be an increased amount of hesitation while speaking.

Another aspect included in speech planning is the structuring of the speaker's message to promote complete understanding by the listener. For example, if a speaker begins giving directions without adequately planning his/her explanation, the listener may not understand the directions and will possibly complete the steps out of order. Therefore, speakers may need extra time when structuring their speech in this situation, resulting in an increased hesitation rate (Chafe, 1980). Further, if the message being communicated involves complex vocabulary or syntax, the degree of hesitation may increase even more (Chafe, 1980).

Hesitation and Planning in Bilingual Speech

Bilingual speakers differ from monolingual speakers in that they store and use two different languages for speech production. According to Levelt's Model of Speech Production, monolingual individuals begin their word selection during a conceptual preparation stage (Levelt, Roelofs, & Meyer, 1999). This stage involves activating specific lexical concepts as part of a larger message and is designed to incorporate the speaker's overall intent. Once the lexical concept has been activated, speakers then need to retrieve the specific word forms from their mental lexicon. This stage is known as the lexical selection stage. In the past, the research literature has not presented a consistent picture with respect to lexical access and language storage in bilingual speakers.

Past opinions of language storage in bilinguals viewed these speakers as being a combination of two complete monolingual speakers within a single person (Grosjean, 1989). Present research, however, views the bilingual speaker as having his/her own distinctive configuration of known languages (Grosjean 1989). When discussing the storage of language within a speaker, the bilingual view proposes that a bilingual speaker is considered a whole, rather than a combination of two separate language parts (Grosjean, 1989). The bilingual speaker is considered to have the ability to use either language or both together depending on the communicative need.

According to the Bilingual Model of Lexical Access (BIMOLA), three levels of nodes for communication exist within bilingual speakers: features, phonemes, and words (Grosjean, 2008). In the BIMOLA model, the bilingual speaker shares the features nodes between both languages spoken (Grosjean, 2008). However, the phoneme nodes and word nodes are not stored identically. Instead, both the phoneme and word nodes are organized individually for each

language, but they are also stored together in one large system (Grosjean, 2008). The storage of the phoneme and word nodes in a large system provides further support for the notion that a bilingual speaker's languages are not stored separately. If the speaker's languages are stored together, there should be a way for him/her to actively choose between both languages when engaged in speech.

The bilingual speaker's ability to choose between speaking in one language, or both interchangeably, may be related to the range of language activation in bilingual speakers, which goes from no activation to total activation (Grosjean, 2001). As a bilingual individual is engaged in speech, s/he will be engaged in the activated language, while the other language is either not activated, or activated but to a lesser extent than the main language being spoken (Grosjean, 2001). Invariably, the bilingual speaker selects and produces words from a specific language without error (Costa & Santesteban, 2004).

When a speaker has both languages activated and intentionally uses two languages interchangeably in speech, it is referred to as code-switching (Hlavac, 2011). Speakers may code-switch when the language in use cannot effectively communicate their intent, or for socially motivated reasons, such as emphasizing emotions or ideas (Poulisse & Bongaerts, 1994). When an individual is engaged in code-switching, the use of the different languages is not considered to be random, but rather an intended choice by the speaker. (Hlavac, 2011).

It is also interesting to note that code-switching does not increase hesitation rate in bilinguals (Hlavac, 2011). In fact, hesitation is likely to occur in the same places as when only one language is being used. Although most forms of code-switching are viewed as intentional, there are often instances where speakers insert words from one language into another unintentionally (Costa & Santesteban, 2004; Poulisse & Bongaerts, 1994). These insertions, also

called performance switches, seem to occur due to language interference, and are often referred to as "slips of the tongue" (Poulisse & Bongaerts, 1994).

Language interference may occur when a bilingual speaker learns his/her second language (L2; Hernandez, Li & MacWhinney, 2005). When a speaker begins to learn the L2 at a young age, s/he is already experiencing entrenchment associated with the first language (L1; Hernandez et al., 2005). This entrenchment causes the L2 words to be learned as parasitic connections to the L1 forms (Hernandez et al., 2005). When the speaker is older, the entrenchment of his/her L1 is even stronger (Hernandez et al., 2005). With the effects of entrenchment, speakers will likely not have their L2 links as strong as their L1 links, affecting the overall level of their language competency (Hernandez et al., 2005).

According to the Unified Competition Model, there even are important differences between L1 and L2 learners. These differences include L1 speakers learning about the world while simultaneously learning language, having a brain which is not dedicated to other tasks, and a large amount of support from caregivers (MacWhinney, 2012). L2 speakers, however, begin to learn their language while already having an understanding of the world, a brain with dedication to other tasks, and decreased supportive L2 serving interactions (MacWhinney, 2012). With these differences present, it is apparent that L2 learners will have a more difficult time acquiring a language as they become older. Although the younger L2 learners may use similar methods to learn L1 and L2, later learners will require progressively more explicit teaching and training (MacWhinney, 2012). These later speakers may not become as fluent as the younger bilinguals, and this may be apparent in their use of hesitation. On the other hand, if a speaker learns their L2 early in life, they are perceived as a more native-like speaker, and should speak the L2 more like a monolingual speaker (Flege, MacKay, & Meador, 1999; Yeni-Komshian, Flege, & Liu, 2000).

The Importance of Speech Fluency. A bilingual speaker's mastery of a language can be judged based on how fluent their speech production sounds in their non-native language (Tavakoli, 2011). Speech fluency, however, can be defined differently, depending on which specific language constraints are encountered. For the purposes of this study, fluency is defined as a stream of smooth, uninterrupted and hesitation-free speech (Tavakoli, 2011). Speech remaining uninterrupted is important in this definition, as native speakers are considered to produce faster speech with fewer hesitations than non-native speakers (Tavakoli, 2011).

In the beginning of a speaker's transition into his/her L2, the odds that s/he engages in more frequent hesitations are larger. However, once the speaker becomes more familiar with the language, his/her number of hesitations should decrease. Hence, hesitation rate might be a speaker-specific characteristic across languages and might assist the listener in determining the degree of fluency that a speakers displays in a particular language. The utility of this type of speech cue also might be helpful in the identification of a speaker by voice alone.

One explanation for the differences in hesitation use across languages is based upon the degree of automatization (Wiese et al., 1984). When bilingual speakers are speaking in their L2, their production of speech is less automatic, which often results in increased planning time and number of corrections (Wiese et al., 1984). Essentially, a speaker with less metalinguistic knowledge and skill in their L2 has constraints placed on them, which are limitations that impacts the speaker's ability to access all aspects of a language (Wiese et al., 1984). Due to these constraints, many speakers utilize hesitation phenomena as a strategy to increase their planning time, thus facilitating successful communication (Wiese et al., 1984).

In addition, extra planning time may be needed for non-native speakers, who are required to adapt their speaking style to that of their L2. Some adaptations include vocabulary (Costa &

Santesteban, 2004; Flege, MacKay, & Meador, 1999; Yeni-Komshian et al., 2000), grammatical structure (Costa & Santesteban, 2004; Flege et al., 1999; Yeni-Komshian et al., 2000), phonotactic differences (Costa & Santesteban, 2004; Flege et al., 1999; Yeni-Komshian et al., 2000), language timing (Dellwo, 2010), and temporal aspects of speech production (Reed, 2000). These temporal aspects include slower speaking rate, phonation-time ratio, and articulation rate, which is total speaking time not including pause time (Reed, 2000). When speakers have to transition between languages and adapt their speaking styles, there may be barriers slowing down their speed of language processing (Runnqvist, Strijkers, Alario, & Costa, 2012). This slowing of cognitive processing will likely increase the speaker's use of compensatory strategies in his/her L2 (Fehringer & Fry, 2007). These strategies can include a speaker's use of additions, deletions, and word exchanges in order to improve speech fluency when encountering areas of difficulty (Poulisse, 2000). Although the use of such strategies will decrease as a speaker becomes more fluent, the likelihood of that speaker reaching the same level of proficiency as that of a native speaker is rare (Fehringer & Fry, 2007).

Bilingual speakers have also been thought to employ long hesitations to prepare for any upcoming planning or speech production difficulties (Reed, 2000). In short, non-native speakers may engage in long, silent pauses to slow their rate of speech thereby facilitating speech planning and production. These types of pauses are considered uncharacteristic and unnatural for a native speaker (Reed, 2000). However, this idea was not supported (Reed, 2000), as non-native speakers of a language were not observed to engage in longer silent pauses. Rather, speakers displayed an increase in the frequency of their use of short silent pauses, and an even larger increase in their number of filled pauses when producing spontaneous speech (Reed, 2000). It is interesting to note that speakers have a tendency to use the same planning features in their L2 as

they do in their L1, as characterized by their carryover of specific types of hesitations (i.e. filled pauses and unfilled pauses) in their speech (Fehringer & Fry, 2007). Hence, a speaker who engages in frequent filled pausing in his/her L1 is likely to continue this practice in their L2. Although the speaker is likely to engage in the same type of hesitations, the number of hesitations used by each speaker is not likely to be consistent. The speaker engaged in his/her L2 may have a significantly greater number of hesitations than a speaker engaged in his/her L1 (Fehringer & Fry, 2007). This may be related to a stronger need for the speaker to access working memory during speech production, which results in an increase in the use of time-buying devices to compensate for an incomplete knowledge of the language being spoken (Fehringer & Fry, 2007).

Although the type of pause used by a speaker may be carried over into the L2, the placement of pauses could change. A non-native speaker of a language is more likely to pause in the mid-clause portion of a sentence, whereas a native speaker is more likely to pause toward the end of a clause (Tavakoli, 2011). This positioning is significant as different placements of pauses are indicative of different functions. Pauses that are located mid-clause indicate that the speaker is either having difficulty with word choice, structure, correct pronunciation, or is pausing to allow increased planning time (Tavakoli, 2011). These pauses are related to the speaker's difficulty with the information processing load associated with L2 acquisition (Tavakoli, 2011). However, pauses that are located toward the end of a clause indicate that the speaker is either allowing time for breathing, organizing their speech, or is about to shift the focus of speech (Tavakoli, 2011).

Pause placement may also be influenced by the characteristics of each language spoken by a bilingual. For instance, Spanish is a syllable-timed language, where syllables are recurring

at regular intervals, with stressed and unstressed syllables having similar durations (Dellwo, 2010). English, on the other hand, is a stress-timed language, where stress patterns on syllables occur regularly (Dellwo, 2010). There also tends to be a larger number of complex syllables and consonant clusters in stress-timed versus syllable-timed languages (Dellwo, 2010). These differences in complexity might cause difficulty for a speaker when transitioning between two languages. The impact of these structural differences may be most obvious when a speaker is engaged in their L2, as the speaker has to maintain the proper structure of the language s/he is currently speaking. If a speaker is struggling with any of the described linguistic differences between languages, the effect may be noted in his/her use of hesitations.

There are many aspects of speech production for bilingual speaker's that may impact the amount of hesitation when engaged in his/her L2. Based on the structural differences and adaptations the speaker must make (Costa & Santesteban, 2004; Flege et al., 1999; Yeni-Komshian et al., 2000), as well as temporal differences between languages (Reed, 2000), a L2 speaker has many barriers to overcome. Although initially early L2 learners will be impacted significantly by these barriers, increased exposure to their L2 should help the speaker become more competent, near the level of a native speaker (Fehringer & Fry, 2007).

Statement of Purpose

If bilingual speakers are able to achieve near-native speaking characteristics in their L2 (Fehringer & Fry, 2007), how does this impact one's ability to identify a speaker by voice alone? Forensic speaker identification is the process where two or more recordings of speech are compared in order to determine if they are from the same speaker (Rose, 2002). This practice can aid in forensic investigations where speech samples are considered as evidence. Research on speaker-specific factors has focused on features such as fundamental frequency, intensity, vowel

formants, and other spectral characteristics (Kunzel, Masthoff, & Köster, 1995; Leeman, Kolly, & Dellwo, 2014; Rose, 2002). Although these features are considered important in speaker identification, other speech production factors, such as hesitation, should be considered, as they are likely to have speaker-specific characteristics as well. Though the use of hesitation is not yet regarded as a common practice in speaker identification, the use of temporal factors is increasing (Leeman et al., 2014).

Research has indicated that speakers are likely to have a preference for the type of hesitation used in their speech (Goldman-Eisler, 1972). The question for a forensic examiner then becomes whether this hesitation preference remains constant across different languages and speaking conditions. If so, then hesitation patterns may be useful in comparing speech samples to determine whether they were produced by the same speaker or different speakers across languages. Comparison of speech samples is also complicated by the planning differences evident between speaking conditions (spontaneous vs. read speech). This comparison of speakers becomes even more complicated when dealing with bilingual speakers.

The present study seeks to determine the utility of hesitation analysis in forensic speaker identification among bilingual speakers. Since current research on bilingual speakers reveals similar use of hesitation patterns across languages, quantitative and qualitative analyses were completed with bilingual speakers in two different languages (English and Spanish), and in two different speaking conditions (spontaneous speech and read speech). If these speakers display similar patterns across languages within a speaking condition, there is a possibility of using hesitation for forensic speaker identification. This finding could improve the process of speaker identification, as it could increase the reliability of comparisons of speech samples when paired with other speech factors. However, a factor that appears to impact speaker identification is the

differences between speaking conditions. Since planning differs across speaking conditions, as in reading vs. spontaneous speech, it may not be possible to compare across conditions for speaker identification. Both language and speaking conditions were addressed in the research questions.

Three research questions were asked:

1: Do bilingual speakers exhibit similar patterns of unfilled and filled pauses across their languages?

2: Do bilingual speakers exhibit similar patterns of unfilled and filled pauses across speaking conditions?

3: Can hesitation patterns be used as a speaker-specific cue in speaker identification tasks?

Chapter Two

Methods

Participants

Twenty Spanish-English bilingual males were selected from a larger database of bilingual speakers. The speaker's ages ranged from 19 to 31 years of age, with an average age of 21.15 years. The speakers were recruited from a bilingual fraternity on a university campus in southwest Florida. Participants were compensated with a small donation to their fraternity. No hearing or speech deficits were reported by the speakers. All participants volunteered to have their voice recorded for cross-language comparisons across languages and dialects of Spanish.

Each participant completed a questionnaire regarding their language history and background (see Appendix A). They were asked to answer questions about their country of origin to determine nationality and the particular dialect of Spanish spoken. The speakers came from Puerto Rico, the Dominican Republic, Venezuela, Cuba, Guatemala, Columbia, Mexico, Peru, and the United States. Demographic characteristics of the participants' parents, which may account for culture and dialect, are listed in Table 1.

Participants were also asked to answer how long they have lived in the United States and the age when they learned English to determine length exposure to English. The mean amount of time for participants living in the United States was 18.25 years, with a standard deviation of 5.31 years. The mean age at which the participants began learning English was 3.95 years of age, with a standard deviation of 3.97.

Table 1. *Summary Statistics of Participants' Familial Demographics*

Country of Origin	Mother	Father
Columbia	4	4
United States	4	3
Dominican Republic	3	3
Puerto Rico	3	3
Mexico	2	2
Peru	1	2
Cuba	1	0
Venezuela	1	2
Guatemala	1	1

Finally, participants were asked how often they spoke each language and which language they preferred when speaking. The majority of participants indicated that they primarily spoke English, and they preferred speaking English (see Table 2).

Table 2. *Participants' Indication of Language Spoken and Preference*

Language	Language Spoken Primarily	Language Preference
English	14 (70%)	15 (75%)
Spanish	1 (5%)	2 (10%)
Equal Amounts of English and Spanish	5 (25%)	3 (15%)

Materials

Recording Conditions. Three different recording conditions were used to produce these recording as part of a larger project: lab quality, landline telephone in a private office, and cellular telephone within a sound-proof booth. The laboratory portion of the recordings occurred in an Industrial Acoustics Company (IAC) sound-proof booth with an AKG C240 microphone. For the telephone recording tasks, an AT&T landline phone and a Sony Ericson cellular phone were used. Only the lab quality booth recordings were utilized for this study.

Speech Samples. Recordings consisted of the speakers reading the *Rainbow Passage* (Fairbanks, 1960). They also were recorded speaking extemporaneously about their favorite dish, as well as how to prepare and cook it. Each speaker was recorded in both English and Spanish for the reading and spontaneous speech conditions.

Procedures

Recording. Upon arrival at the speech laboratory, participants were asked to complete the language/dialect questionnaire. Once the questionnaire was completed, participants were recorded in a sound-treated booth using a Sony PCM-MI Digital Audio Recorder. Two different speech samples (conversation and reading) in English and Spanish were recorded. The language and the sample type of the recording conditions was balanced for each participant. This procedure minimized any speaking effects that may occur due to order of recording.

Each participant was recorded spontaneously speaking in both English and Spanish about their favorite food. Participants were also recorded reading the *Rainbow Passage* (Fairbanks, 1960), which was translated into Spanish for cross-language voice comparisons.

Participants were given the spontaneous speech topic a few minutes prior to recording, giving them opportunity to prepare a narrative. Each participant was required to speak for 2 ½ minutes on the topic. Participants were also given *The Rainbow Passage* in advance in order to familiarize themselves with it. The passage was provided in English as well as Spanish.

Digital Analysis. Each participant had a total of four speech samples, two spontaneous speech samples and two reading samples, one in each both English and Spanish. Each audio file was transferred from a Sony PCM-MI Digital Audio Recorder to a Dell Desktop computer and digitized at 22050 Hz using Pratt (Boersma & Weenink, 2014). The original samples were approximately 2 ½ minutes. However, only the speech recorded during the middle of the

passage, i.e., 1:05-1:15 (minutes: seconds), was utilized. Once the recordings were digitized onto the computer, analyses were completed.

Hesitation Analysis. Two types of analyses were conducted. The first analysis considered silent pauses. This involved the measurement of pause to speaking ratios within each speech sample. The second analysis considered the use of filled pauses. These analyses will be described below.

A script for the identification of speaking time and pauses was created using Praat (Boersma & Weenink, 2014). This script was used to identify segments of silent pauses lasting longer than 250 ms. Hesitations shorter than 250 ms were not relevant for this analysis, as they are not indicative of meaningful, planning pauses (Goldman-Eisler, 1972). The script produced markings on each participant's spectrogram indicating moments of speech and hesitation, which were then automatically measured for length. These intervals were measured on the spectrogram for each speech segment. A second Pratt script was developed to export speech and pause lengths into participant-specific Excel files. These time lengths were utilized to create individual pause-to-speech ratios for each recording, which were then averaged to develop a mean pause to speaking ratio for each speech sample for each participant.

Filled pauses (i.e., "um," "uh,") were also counted to note differences in the number of filled pauses between the English and Spanish samples. Each speech segment was played and the researcher tallied all moments of filled hesitations present in each sample. The number of hesitations were stored in an Excel document in order to make comparisons across both languages and speaking contexts. Because the primary researcher was not a fluent speaker in Spanish, a second graduate student was recruited in order to aid in the identification of filled pauses in Spanish. The primary researcher initially identified the filled pauses in Spanish, and the

second clinician either confirmed or refuted the number of pauses identified. If the clinician refuted the pauses, the samples were listened to for a third time and the bilingual researcher determined the exact number of pauses.

Statistical Analyses. Four speech samples were analyzed for each speaker (2 languages x 2 speaking conditions) using non-parametric statistics. Separate Wilcoxon tests were run, with SPSS Statistics 21.0 software (IBM Corp., 2013) to note differences in pause to speaking ratios, length of speech segments, length of pauses, and the number of filled pauses across languages and speaking conditions. A supplementary qualitative analysis for two speakers was also completed by creating step graphs demonstrating differences in planning style. These speakers were chosen based on the large differences in their pause to speaker ratios between either languages or speaking conditions. The step graphs were created by displaying the length of speech segments compared to the length of pause segments.

Chapter Three

Results

The present study examined patterns of hesitation within bilingual speakers, across Spanish and English, in both spontaneous and read speech. The length and number of unfilled and filled pauses were compared across languages and contexts. The overall goal of these comparisons was to identify situations in which hesitation might be useful in the speaker identification process. The present study incorporates both quantitative and qualitative methodologies. Wilcoxon signed-ranks tests and the appropriate effect sizes (Corder & Foreman, 2009) were applied during the quantitative analysis. Step graphs were utilized to compare both the frequency and length of alternations between speech and pauses in two participants with different planning styles.

Quantitative Results

Wilcoxon signed-ranks tests were applied for the analysis of hesitation in bilingual speakers across speech contexts. SPSS Statistics 21.0 software (SPSS, Inc., 2012) was used for the statistical analyses. The independent variables included language spoken (English/Spanish) and speaking condition (conversational vs. read speech). The dependent variables were pause to speaking ratio, duration of pause units, duration of speaking segments and number of filled pauses. Effect sizes for the Wilcoxon tests were calculated using the following formula (Corder & Foreman, 2009, p. 39): $ES = |z|/\sqrt{n}$. The quantitative analyses were conducted to answer the following three research questions:

1. Do bilingual speakers exhibit similar patterns of unfilled and filled pauses across their languages?
2. Do bilingual speakers exhibit similar patterns of unfilled and filled pauses across speaking conditions?
3. Can hesitation patterns be used as a speaker-specific cue in speaker identification tasks?

Speech Cycles and Speaking Condition

Pause to Speaking Ratio. This measurement is a comparison of the pause time from the end of one speech segment to the beginning of the following speech segment. These ratios were computed over the entire duration of the speech sample and then were averaged to generate a mean pause to speaking ratio. These ratios quantify the cycles of speech involved in speech planning (Beattie, 1979; Butterworth, 1975).

Scores for the pause to speaking ratio ranged from 0.261 to 1.447 (minimum possible score = 0). Ratios greater than 1 suggest that the pause duration was greater than the speech segment duration across all speaking conditions (N = 80). Table 3 provides the medians and interquartile ranges of the pause to speaking ratios for each language and speaking condition.

Wilcoxon signed ranks tests revealed no significant differences in the pause to speaking ratios across languages for spontaneous speech, $Z(1) = -0.187$, $p = .852$, $ES = .042$ or reading, $Z(1) = -1.083$, $p = .279$, $ES = .242$. The tests did reveal significant differences in duration of pause to speaking ratios across speaking conditions in English, $Z(1) = -3.435$, $p = .001$, $ES = .768$, and in Spanish, $Z(1) = -2.389$, $p = .017$, $ES = .534$. As illustrated in Figure 1, these results suggest that there were no differences across language for either speech sample, but there were significant differences attributable to speaking condition. The reading condition had smaller

pause to speaking ratios than spontaneous speech in both languages, with moderate to strong effect sizes.

Table 3. *Medians and Interquartile Ranges for Pause to Speaking Ratios*

Sample Type	Language	Percentiles		
		25th	50th (Median)	75th
Spontaneous Speech	English	0.438	0.526	0.696
	Spanish	0.437	0.532	0.510
Reading	English	0.301	0.399	0.797
	Spanish	0.398	0.429	0.511

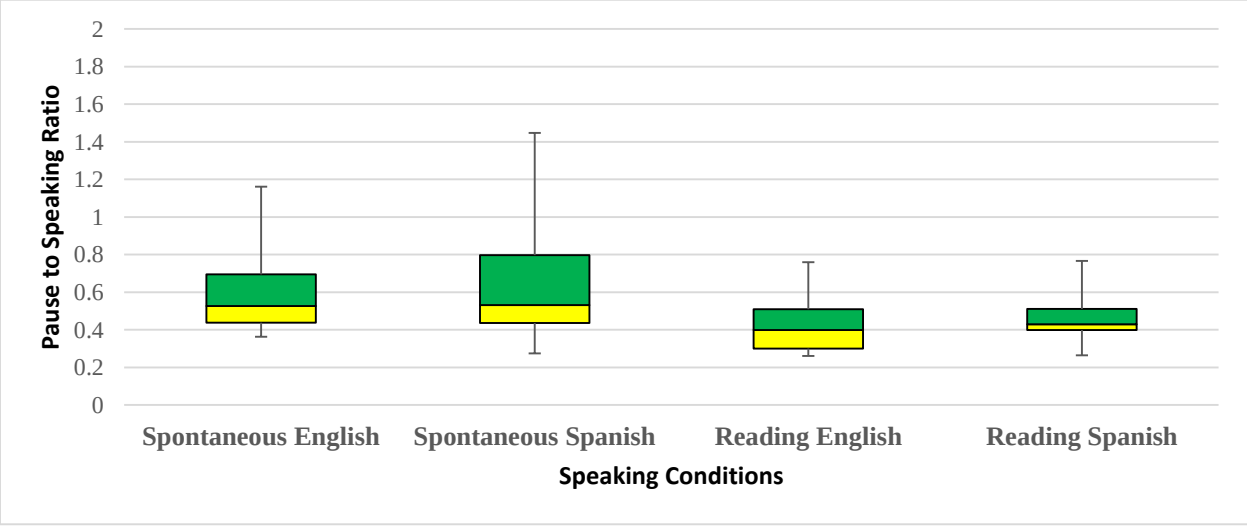

Figure 1. Median Pause to Speaking Ratio Values across Languages and Speaking Conditions.

Duration of Pause Units. Scores for the pause durations ranged from 0.373 to 1.140 ms, across all speaking conditions (N = 80). A minimum value for a planning pause was 0.250 ms. Values less than 250 ms are presumed to represent articulatory timing and not cognitive planning (Goldman-Eisler, 1972). Table 4 provides the medians and interquartile ranges of the duration of pause units for each language and speaking condition.

Table 4. *Medians and Interquartile Ranges for Duration of Pause Units*

Sample Type	**Language**	**Percentiles**		
		25^{th}	50^{th} (Median)	75^{th}
Spontaneous Speech	English	0.632	0.662	0.764
	Spanish	0.602	0.688	0.692
Reading	English	0.484	0.560	0.835
	Spanish	0.522	0.564	0.656

Wilcoxon signed ranks tests revealed no significant differences in the duration of pause units across languages for spontaneous speech, $Z(1) = -1.008$, $p = .313$, $ES = .226$ or reading, $Z(1) = -.037$, $p = .970$, $ES = .008$. The tests revealed significant differences in duration of pause units across speaking conditions in English, $Z(1) = -3.061$, $p = .002$, $ES = .685$, and in Spanish, $Z(1) = -2.987$, $p = .003$, $ES = .668$. As illustrated in Figure 2, these results suggest that there were no differences across language for either speech sample, but there were significant differences attributable to speaking condition within a language with moderate to strong effect sizes. Specifically, speakers used shorter pauses while they were reading.

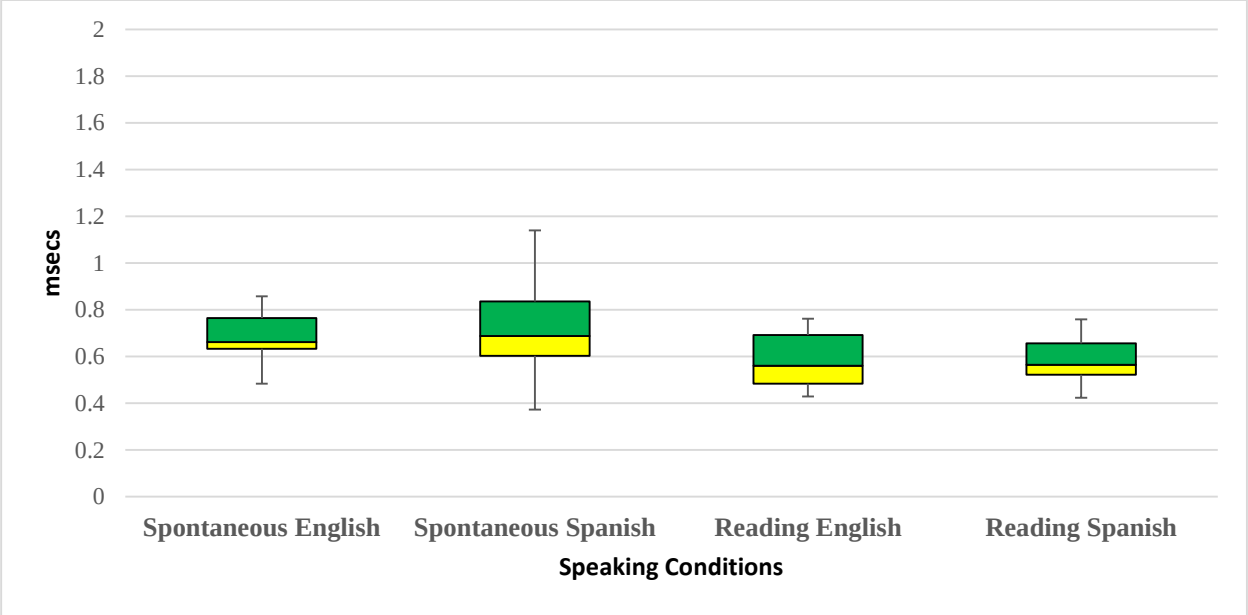

Figure 2. Median Pause Duration Values across Languages and Speaking Conditions.

Duration of Speech Segments. Durations of the speech segments ranged from 1.285 to 3.549 seconds across all speaking conditions (N = 80). Table 5 provides the medians and interquartile ranges for the durations of speech segments for each language and speaking condition.

Table 5. *Medians and Interquartile Ranges for Duration of Speech Segments*

Sample Type	**Language**	**Percentiles**		
		25th	50th (Median)	75th
Spontaneous Speech	English	1.630	1.893	2.278
	Spanish	1.597	1.759	2.397
Reading	English	.1.786	2.240	2.744
	Spanish	1.577	2.021	2.210

Wilcoxon signed ranks tests revealed no significant differences in the duration of speech segments across languages for spontaneous speech, Z(1) = -.373, p =.709, ES = .083 or across speaking conditions in Spanish, Z(1) = -.187, p =.852, ES = .042. The tests revealed significant differences in duration of speech segments across languages for reading, Z(1) = -.2.389, p =.017, ES = .534 and across speaking conditions in English, Z(1) = -2.763, p =.006, ES = .618.

As illustrated, Figure 3, these results suggest there were no differences between English and Spanish for the speech segment durations in spontaneous speech and there were no differences in speech segment durations in either reading or spontaneous speech conditions in Spanish. On the other hand, speech segment durations for English reading were longer than in the production of spontaneous speech in English and the read segments of Spanish were shorter than the read segments in English.

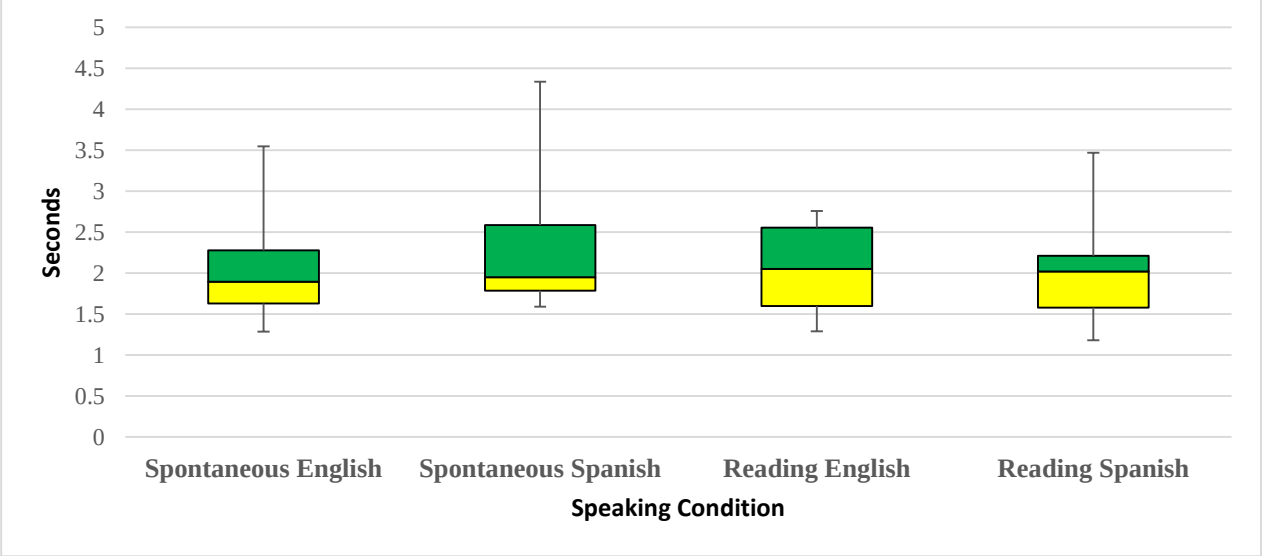

Figure 3. Median Values for the Duration of Speech Segments across Languages and Speaking Conditions.

Number of Filled Pauses. The frequency of filled pauses ranged from 0 to 19 across all speaking conditions (N = 80). Table 6 provides the medians and interquartile ranges of the number of filled pauses for each language and speaking condition.

Table 6. *Medians and Interquartile Ranges for Number of Filled Pauses*

Sample Type	**Language**	**Percentiles**		
		25th	50th (Median)	75th
Spontaneous Speech	English	5.000	8.000	10.750
	Spanish	3.000	5.500	8.000
Reading	English	.000	.000	.000
	Spanish	.000	.000	.000

Wilcoxon signed ranks tests revealed significant differences in the number of filled pauses across languages for spontaneous speech, $Z(1) = -2.786$, $p = .005$, $ES = .623$ and across speaking conditions in English, $Z(1) = -3.926$, $p < .001$, $ES = .878$ and Spanish, $Z(1) = -3.926$, $p < .001$, $ES = .878$. There were no significant differences in the number of filled pauses across languages for reading, $Z(1) = .000$, $p = 1.000$, $ES = 0$. As illustrated, in Figure 4, these speakers used more filled pauses in English than in Spanish during spontaneous speech and more pauses in spontaneous speech than reading in both English and Spanish. Moderate to strong effect sizes were noted for all significant comparisons.

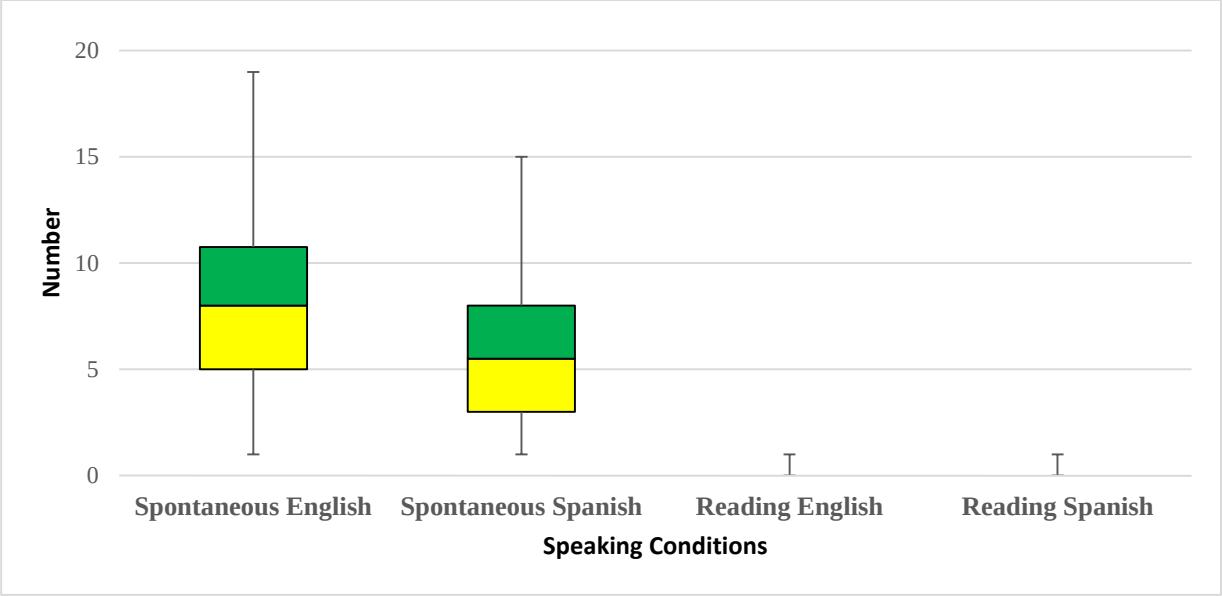

Figure 4. Median Values for Number of Filled Pauses across Languages and Speaking Conditions.

Qualitative Analysis

Step graphs were created for two selected speakers to demonstrate their planning style. A step graph illustrates how speech production and planning pauses are expressed over time. Slope measurements were computed to note individual differences in speech production.

Speaker 47. Speaker 47's questionnaire results indicated that he learned English as an infant, speaks English primarily throughout the day, and prefers to speak English. This speaker had the largest difference in pause to speaking ratios across languages in spontaneous speech. Specifically, his mean pause to speaking ratio was .437 for spontaneous English and 1.447 for spontaneous Spanish. As illustrated in Figure 5, this speaker demonstrated shorter speech segments and more pauses when speaking in English. This observation is supported by the low slope value (0.245), indicating frequent bursts of speech separated by short pauses. Although the

slope of the graph remains relatively low, the speaker does have occasional increased needs for planning, which appear as small surges increasing the slope, as described by Reed (2000).

A different speech pattern is illustrated in Figure 6. While speaking Spanish, this speaker utilized longer speech segments with bursts of planning, characterized by brief alternations between pauses and speech. The slope for spontaneous Spanish was much steeper (0.730). This pattern suggests a need for greater planning time to produce a narrative in Spanish. The speech slope appears inconsistent, with bursts of rapid speech and hesitation occurring throughout the sample. In both figures 5 and 6, moments where the speaker alternates quickly between speech and hesitation are identified. These alternations are likely related to the speaker's greater need for planning following longer speech segments.

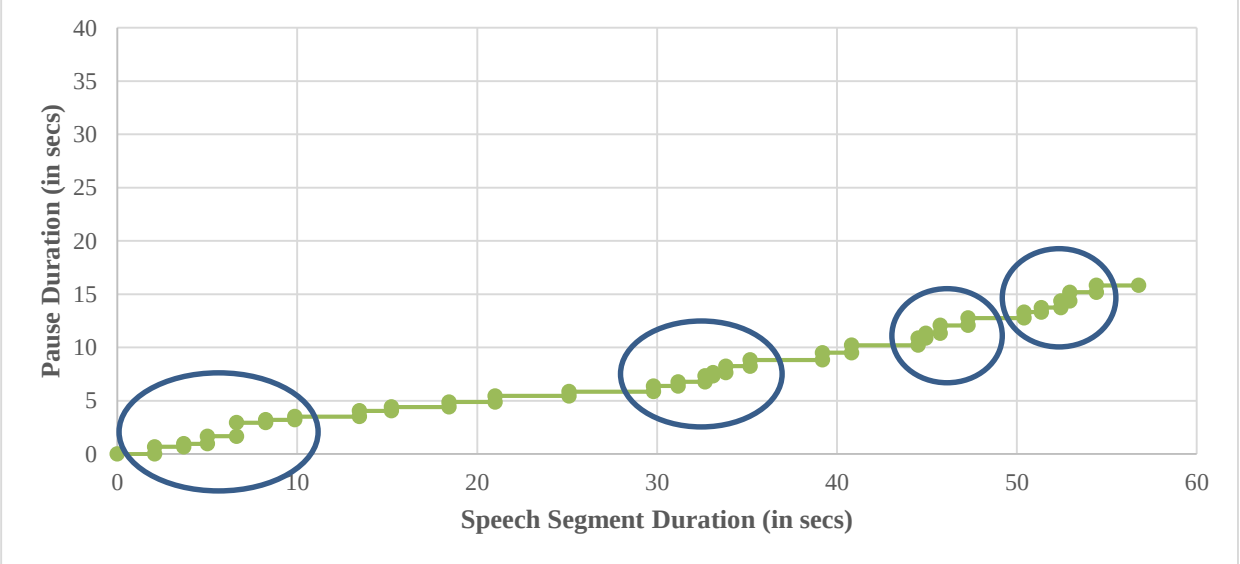

Figure 5. Duration of Pause and Speech Segments for Speaker 47 during Spontaneous Speech in English with Bursts of Speech Identified.

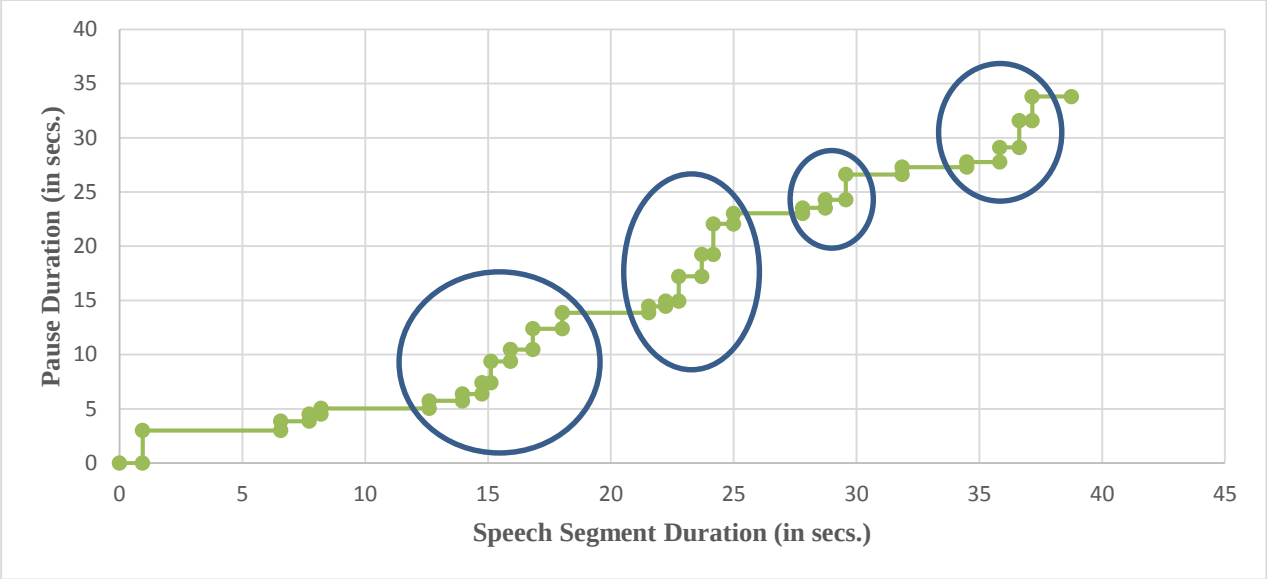

Figure 6. Duration of Pause and Speech Segments for Speaker 47 during Spontaneous Speech in Spanish with Bursts of Speech Identified.

Speaker 33. Speaker 33's questionnaire results indicated that he learned English as an infant, speaks English primarily throughout the day, and prefers to speak English. This speaker was selected based on the magnitude of difference in pause-to-speaking ratios between languages during the reading task. This speaker's mean pause-to-speaking ratio while reading was 0.373 in English and 0.631 in Spanish. As illustrated in Figure 7, this speaker read in English with short bursts of speech and frequent use of small pauses, which resulted in a slope of 0.215. The slope of this speech segment displays consistent alternations between speech and hesitation.

However, when reading in Spanish, this speaker utilized more frequent pausing to plan his speech. The slope for Spanish reading was higher (0.322) than in English reading. As observed in Figure 8, there are identified moments of dramatic alternations between speech and hesitations. This finding suggests that the speaker may have been less familiar with written Spanish and needed more time to plan his utterances.

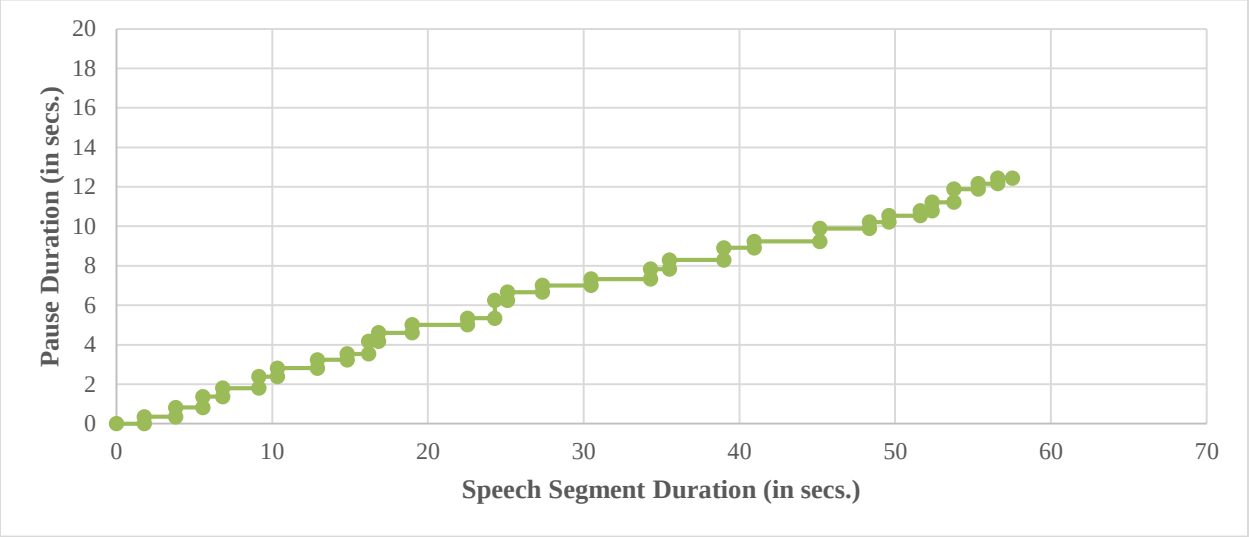

Figure 7. Duration of Pause and Speech Segments for Speaker 33 Reading in English.

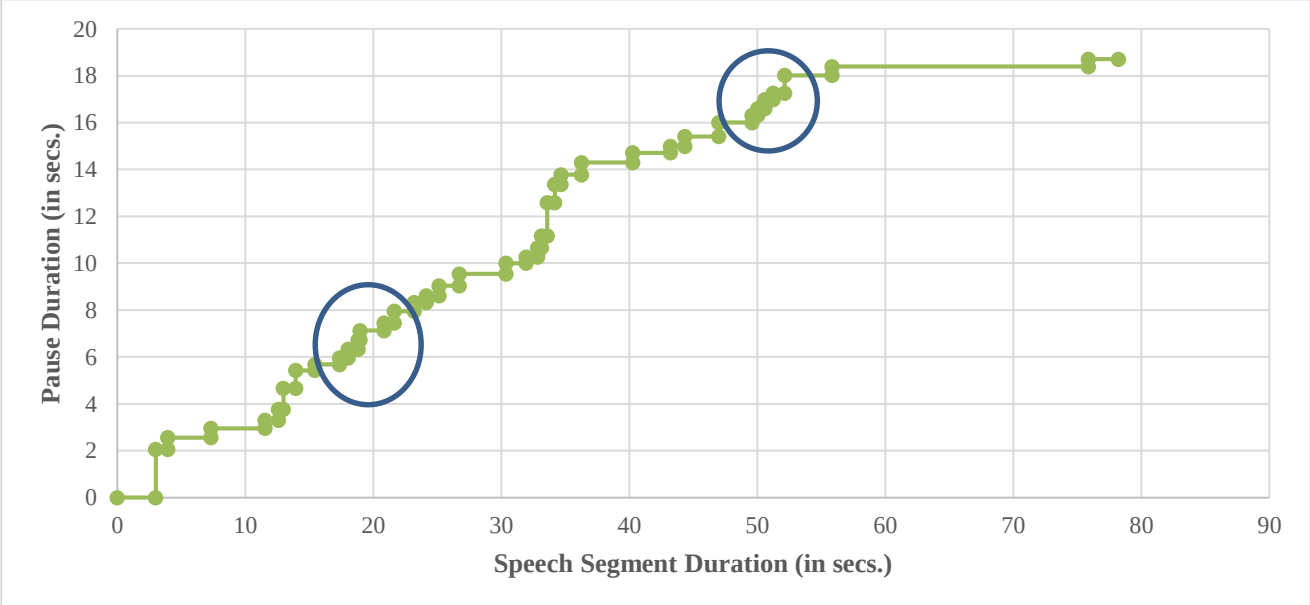

Figure 8. Duration of Pause and Speech Segments for Speaker 33 Reading in Spanish with Bursts of Speech Identified.

Speaker 33's spontaneous speech in English and Spanish was also used for qualitative comparisons. This speaker's mean pause-to-speaking ratio while spontaneously speaking was 1.162 in English and 0.487 in Spanish. As illustrated in Figure 9, this speaker spontaneously spoke in English with longer segments of speech, followed by short bursts of speech and use of short pauses, which resulted in a slope of 0.362. The speaker appears to follow regular cycles of alternating patterns of speech and hesitations (Beattie, 1979; Butterworth, 1975).

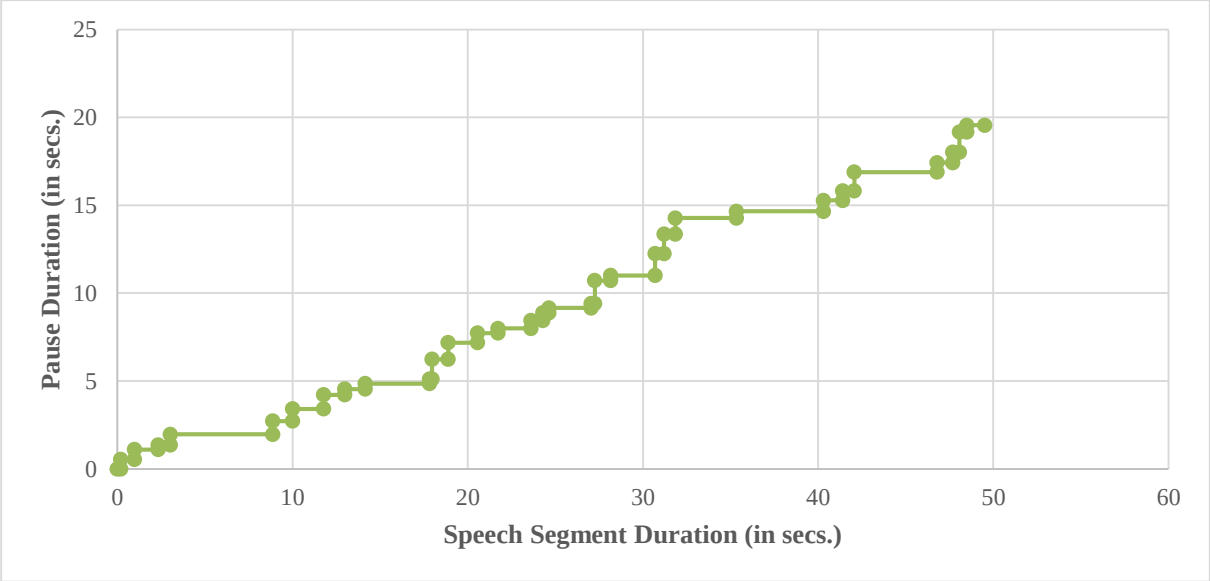

Figure 9. Duration of Pause and Speech Segments for Speaker 33 Spontaneous Speech in English

However, as illustrated in Figure 10, when spontaneously speaking in Spanish, this speaker utilized more frequent pausing and shorter speech segments to plan his speech. The slope for spontaneous speech in Spanish was lower (0.318) than in English (0.362). In Figure 10, there is a consistent alternation of short segments of speech and short segments of hesitations. Both the speech and hesitations appear to occur in a cyclic manner (Beattie, 1979; Butterworth,

1975). This finding suggests that the speaker may have been less comfortable in his spontaneous Spanish speech, and required more frequent pausing.

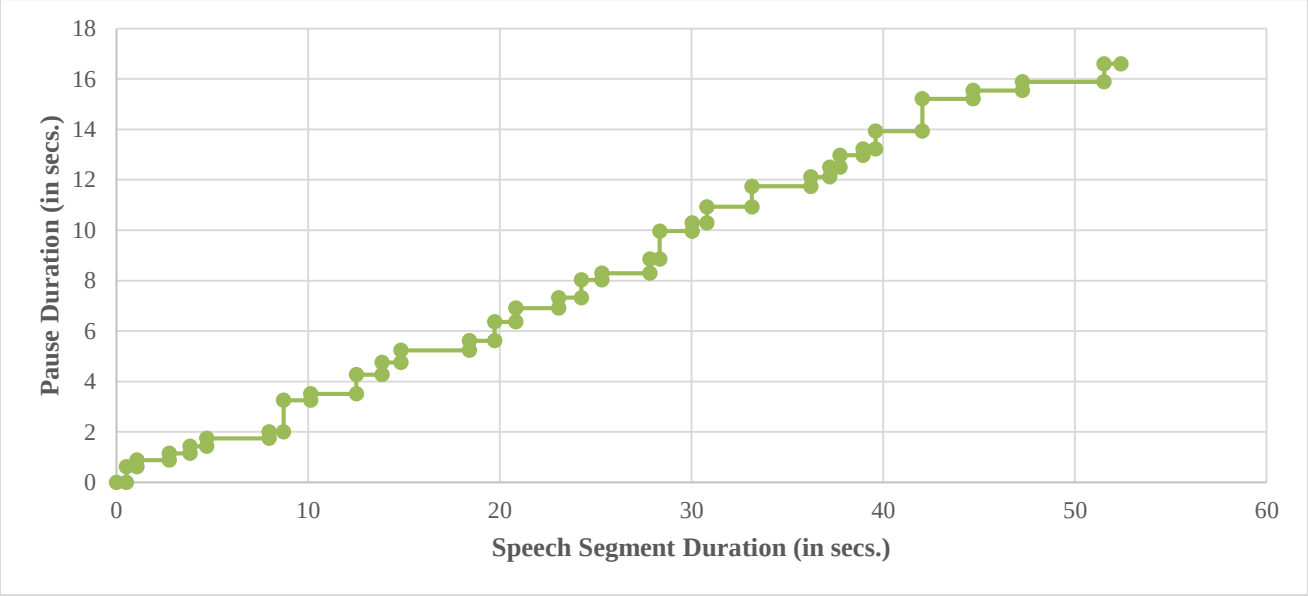

Figure 10. Duration of Pause and Speech Segments for Speaker 33 Spontaneous Speech in Spanish.

Summary

The quantitative results indicated significant differences across speaking conditions for pause to speaking ratio, pause duration, and filled pauses in both English and Spanish. A significant difference was also indicated across speaking conditions for speech segment duration in English. Table 7 illustrates the significant differences within each language based on speaking condition. These findings would suggest that comparisons across speaking conditions would not be appropriate when only using parameters associated with hesitation. This is true for both English and Spanish.

The difference in conditions within a language were also apparent in the qualitative analysis. For instance, in Speaker 33, the comparison of pause-to-speech slopes in English

reading versus English spontaneous speech was very different. The English reading slope was not as steep as the English spontaneous speech slope, which seemed to be affected by the increased number of pauses in spontaneous speech in English. This increased number of pauses is likely related to the differences in planning among these two speaking conditions. Both the quantitative and qualitative results suggest that due to the significant planning differences associated with spontaneous and read speech, comparing segments recorded in different conditions would not be beneficial for speaker identification.

Table 7. *Significant Differences for English and Spanish for each Dependent Variable.*

	English	Spanish
Pause to speaking ratio	Reading< Spontaneous Speech	Reading< Spontaneous Speech
Pause Duration	Reading< Spontaneous Speech	Reading< Spontaneous Speech
Speech Segment Duration	Spontaneous Speech< Reading	
Filled Pauses	Reading< Spontaneous Speech	Reading< Spontaneous Speech

Cross-language comparisons were considered within each speaking condition (see Table 8). The results indicate very few instances where there were significant differences between Spanish and English. In read speech, there were longer speech segment durations in English than in Spanish and speakers used more filled pauses in English than in Spanish during spontaneous speech.

Qualitative cross-language comparisons revealed differences in slopes in two speakers. For instance, in speaker 33, the English reading slope was not as steep as the reading slope in Spanish. The Spanish segment was affected by an increased number of pauses, and shorter

speech segments. This difference in slopes may be related, however, to the speaker's decreased experience with Spanish reading. Since these speakers were not as familiar reading Spanish as they were in English, there is a likelihood of increased pauses related to the speaker's inexperience with both the words and content of the reading passage. This speaker did not display as much of a difference in their slopes for spontaneous English speech and spontaneous Spanish speech.

Table 8. *Significant Differences for Spontaneous and Read Speech for each Dependent Variable.*

	Spontaneous Speech	Read Speech
Pause to speaking ratio		
Pause Duration		
Speech Segment Duration		English > Spanish
Filled Pauses	English > Spanish	

These findings suggest the need for caution when comparing speech samples across speaking conditions using hesitation. One should consider hesitation as one of several acoustic cues for use in speaker identification in a cross-language situation.

Chapter Four

Discussion

Hesitation use is common among all speakers, regardless of whether they are engaged in their dominant or non-dominant language (Fehringer & Fry, 2007; Reed, 2000). The question is whether a bilingual speaker will engage in the same types of hesitations in both languages. If hesitation patterns can be identified consistently across speakers regardless of language, their use as an acoustic cue for speaker identification may be possible.

Current research, however, has not sufficiently addressed the use of hesitation as a speaker-specific phenomenon. This study incorporated a mixed methods design to address the consistency of hesitations across languages and speaking conditions in a bilingual speaker. This discussion first describes the study results in relation to the three research questions. Next, the study strengths and limitations will be addressed. Finally, directions for future research will be outlined.

Unfilled and Filled Pauses across Languages

Measures of Unfilled Pauses. Three different measures of unfilled pauses were computed. The first was an overall measure of speech planning, pause to speaking ratio, which was a measure that reflected the individual contributions of pause and speech production to planning. The other two measures of unfilled pause considered the roles of pause and speech segment durations separately.

Comparisons of pause to speaking ratio and pause duration across languages were not significant across English and Spanish. This similarity between these factors may be related to

the level of English competency of the speakers in this study. When a speaker has a near-native level of speech in their L2, they are likely to display very similar hesitation patterns across both languages (Flege et al., 1999; Yeni-Komshian et al. 2000). Since the majority of these speakers indicated that they speak English primarily throughout the day, it is probable that their patterns are very similar in both English and Spanish.

Speech segment durations exhibited a different pattern. This speech feature was measured to represent the length and fluency of speech following a planning pause. These competent speakers of English and Spanish evidenced a significant difference during the reading condition, but not during spontaneous speech. They produced longer speech segment durations during reading in English. This finding is likely related to their greater familiarity with reading in English since they were college students at an American university.

Along with increasing their knowledge of their L2, speakers also have a tendency to carryover planning aspects from their L1 to their L2 (Fehringer & Fry, 2007). This carryover consists of not only type of hesitation (unfilled versus filled), but also duration and frequency of pause. This consistency of carryover between languages is evidenced in the similarities between the hesitations in the English and Spanish speech samples.

Filled Pauses. Filled pause use was noted to differ across languages. Speakers utilized more filled pauses when speaking English than when speaking Spanish. This increase in filled pause use in English may be related to automatization. The majority of speakers in this study acquired Spanish early and learned English as a second language, anywhere from birth to the age of 12. Since the majority of these speakers learned English in early childhood and are now college students, they are comfortable speaking English on a regular basis, and many reported a preference for speaking English. This, however, was not the case for all speakers. Those who are

less fluent in English as their L2, likely have decreased linguistic knowledge (Wiese et al., 1984), which impacts their ability to access all aspects of a language. If a speaker is unable to access all aspects of language, they will require more planning time to communicate their message (Wiese et al., 1984). Hence, speakers may use hesitation phenomenon in order to combat the limitations they face when communicating in their L2.

Unfilled and Filled Pauses across Speaking Conditions

Measures of Unfilled Pauses. The same measures of unfilled and filled pauses were used to note differences in hesitation across spontaneous speech and reading. In this case, comparisons of pause to speaking ratio and pause duration across speaking conditions were significant in both English and Spanish. Consistent with previous research (Chafe, 1980), speakers hesitated longer in spontaneous speech than in read speech in both languages. These findings are likely related to the difference in planning time between spontaneous speech and read speech. Spontaneous speech is considered more disfluent than read speech (Clark & Fox Tree, 2002) and requires more planning and organization of ideas (Chafe, 1980). Spontaneous speech also requires higher-level creative and cognitive planning, therefore speakers tend to use a larger number of hesitations (Butterworth, 1975; Chafe, 1980). On the other hand, read speech requires decreased planning, as the speaker is more focused on just generating speech to produce the written text orally.

Speech segment durations exhibited a different pattern. This speech feature was measured to represent the length and fluency of speech following a planning pause. Findings suggested no significant difference in the mean speech segment durations between read and spontaneous speech when the speakers were engaged in Spanish. This was not true for English, as speakers had significantly longer speech segments in reading that in spontaneous speech. This difference

between English and Spanish may be based on the speaker's competency with English as their L2. Although these speakers are considered near-native speakers of English, the likelihood of them ever reaching native speaker status is very unlikely (Fehringer & Fry, 2007). The increased competency these speakers have in Spanish is likely related to the longer speech segments with less need to hesitate for planning. When engaged in English, the speakers have a higher planning need which accounts for the shorter speech segments.

Filled Pauses. The speakers demonstrated more filled pauses in their spontaneous speech than in their read speech in both English and Spanish. This finding, as noted with unfilled pauses, may be related to the planning differences between spontaneous speech and reading. Since spontaneous speech requires more planning (Chafe, 1980), it is no surprise that speakers used more hesitation devices.

Hesitation Use as a Speaker-Specific Cue

The comparison of pause to speaking ratio, pause duration, and speech segment duration provides insight into the possibility of using hesitation as a cue in forensic speaker identification. Since the speakers are likely to present with similar alternating cycles of speech and hesitations in the same speaking condition (Beattie, 1979; Butterworth, 1975), it was predicted that this finding might hold true when using a second language. This prediction is based on the speaker's carryover of planning aspects between languages (Fehringer & Fry, 2007), with each speaking condition having its own aspects of pause use. The lack of significant differences in pause to speaking ratio, pause duration, and speech segment between English and Spanish, when considering the speaking context, supports the idea that speaker identification using unfilled pauses is possible if only one speaking context is used. For instance, a speaker should be only compared within a language or between two different languages as long as the speaking

condition (spontaneous speech or read speech) is maintained for both speech samples. The same cannot be said for analysis comparing two different speaking conditions.

The significant differences in pause to speaking ratio, pause duration, and speech segment duration in English when comparing speaking contexts also has implications for forensic speaker identification. Due to the increased hesitation length and the need for planning in spontaneous speech, speaker comparisons using unfilled and filled pauses are not possible across speaking conditions. These differences between speaking conditions also were apparent in the qualitative analysis presented in the step graphs. Analysis of speaker 33 revealed an increased number of pauses in English spontaneous speech when compared to read speech in English. The difference in the number of pauses displayed and the slopes demonstrated the overall planning differences between the two language samples. The significant differences in the need for planning between reading and spontaneous speech provide evidence that comparing speech samples across different speaking contexts is not plausible.

Along with comparing across contexts, using filled pauses as a cue for forensic speaker identification is not possible. Analysis of filled pauses produced significant differences across both languages and speaking conditions for separate reasons. Although a speaker may be competent speaking a language, such as Spanish, the speaker may not be familiar reading in the language. This unfamiliarity with reading in a specific language is going to impact hesitations. Speakers who also have difficulties accessing all aspects of their L2, such as vocabulary and grammatical structure, are also going to exhibit an increased number of filled pauses. Due to these differences, there is no instance where an individual speaker could be identified using filled pauses. Even if the speakers being analyzed were considered to be fluent bilinguals, they are not

likely to achieve the same level of competency as native speakers (Fehringer and Fry, 2007), the differences in filled pauses between languages may still be present.

Study Strengths and Limitations

This study provides much needed information on the use of hesitation in the identification of a speaker. Study strengths and limitations will be presented below.

Study Strengths. The first strength was the use of the Pratt script (Boersma & Weenink, 2014) for speech analysis. This script was successful in identifying the location of unfilled pauses, as well as their length. The program also identified the length of speech segments surrounding the unfilled pauses. This program made it possible to have exact frequency and duration measures to create pause to speaking ratios. The identified segments could then be confirmed by the researcher as appropriate moments of hesitation.

The second strength of this study involved the diversity in cultural background of the population sample. The familial heritage of the speakers involved many different Spanish-speaking nations in the western hemisphere. Some differences in culture and dialect were evident in a speaker's vocabulary choices, rate of speech, grammatical structures, accent, and timing, as well as some suprasegmental aspects of a speaker's L1 may be carried over to their L2 (Flege et al., 1999). The use of speakers from diverse background is important for the representation of differences in Spanish dialects used in Central America, South America, and the Caribbean.

Study Limitations. Two study limitations may have affected the outcome of this project.

The first limitation involves language preference and the frequency of each language spoken throughout the day. The majority of speakers chosen selected English as both their language preference and the language most frequently spoken. This limited the speakers to predominantly English speaking bilinguals, who only utilized Spanish at home, or in the

presence of other Spanish speakers. A larger sample size with speakers that varied in both language preference and frequency with which they spoke English and Spanish would account for differences in speech production aspects, such as fluency, automatization, language constraints, and language adaptations. Integrating speakers who were less competent in English may result in different hesitation patterns between English and Spanish and across speaking conditions. These types of speakers may present with hesitation patterns that are different than a bilingual speaker who utilizes both languages equally throughout the day.

The second limitation involved the recording process. Since the samples were recorded for a larger study on dialect, the current researcher did not have input into the recordings that were created. Researcher input may have included the topics for spontaneous speech, the specific reading passage chosen, and the cueing provided to the participants. This limitation primarily affected the spontaneous speech recordings. Often during the recording of oral narratives, the speaker required cues from the clinician to continue speaking, as he would abruptly stop before speaking for three minutes. The clinician would often be heard on the recordings encouraging the speaker to continue their narrative. The cues from the clinician could have affected the flow and fluency of speech, rendering some segments unusable. Only segments of speech absent of redirection could be used for this study.

Utility of Findings

The similarities in the hesitation phenomenon between languages is further evidence that bilingual speakers often use the same planning aspects between languages, and carryover aspects from their L1 to their L2 (Fehringer & Fry, 2007). Since the speakers studied were very near-native speakers of English and native speakers of Spanish, they were not under the same constraints as a speaker who is not as fluent in English. Therefore, the speakers displayed very

similar pause to speaking ratios, pause durations, and speech segment durations across languages when speaking condition was held constant. This finding provides evidence that forensic speaker identification can be conducted between two different languages as long as they occur within the same speaking condition.

An extremely important aspect in forensic speaker identification with bilingual speakers involves the speaking condition. Although the results suggest that comparisons can occur between two different languages, it cannot occur between two different contexts when considering hesitation patterns. The significant differences in the planning and hesitations between spontaneous speech and read speech indicate that the alternating cycles of fluent and hesitant speech do not remain consistent in all speaking conditions. Comparisons between two different speaking conditions would be considered irresponsible and unreliable.

The use of hesitation as a speaker-specific cue, however, must be utilized with speakers who are fluent in both languages. Less fluent speakers are more likely to have a higher number of hesitations in their non-dominant language (Fehringer & Fry, 2007), as they may be under a larger number of language constraints reflecting their decreased linguistic knowledge in L2 (Wiese et al., 1984). For instance, non-native speakers are likely to produce slower speech with an increased number of hesitations compared to a fluent speaker (Tavakoli, 2011). This increase in hesitations may be related to less automatic speech, and decreased metalinguistic knowledge and skill in their L2 (Wiese et al., 1984). On the other hand, if the speaker is more fluent in their L2, their patterns of hesitations should be more similar to a native speaker (Fehringer & Fry, 2007).

Forensic scientists may be able to compare speech samples from different languages in a single speaking condition to determine the similarities of speech and hesitation cycles for

individual speakers. Although hesitation is beginning to appear as a plausible factor for speaking identification, it cannot be the only factor considered for forensic speaker identification. Hesitation must be used in conjunction with other factors such as fundamental frequency, intensity and voice quality to create a more specific picture of individual speakers. Nevertheless, hesitations should not be discounted for forensic speaker identification, but rather considered as an additional factor when combined with other speech characteristics.

References

Beattie, G. W. (1979). Planning units in spontaneous speech: Some evidence from hesitation in speech and speaker gaze direction in conversation. *Linguistics, 17*, 61-78.

Boersma, P. & Weenink, D. (2014). Praat: doing phonetics by computer [Computer program]. Version 5.4.08, retrieved 30 December 2014 from http://www.praat.org/

Butterworth, B. (1975). Hesitation and semantic planning in speech. *Journal of Psycholinguistic Research, 4*, 75-87.

Cenoz, J. (2000). Pauses and hesitation phenomena in second language production. *ITL. Institut voor Togepaste Linguistik, 127-128*, 53-69.

Chafe, W. (1980). Some reasons for hesitating. In H. W. Dechert & M. Raupach (Eds.), *Temporal Variables in Speech. Studies in Honour of Frieda Goldman-Eisler* (pp. 169-180). Berlin: Mouton Publishers.

Clark, H. H., & Fox Tree, J. E. (2002). Using uh and um in spontaneous speaking. *Cognition, 84*, 73-111.

Corder, G. W., & Foreman, D. I. (2009). *Nonparametric Statistics for Non-Statisticians*. Hoboken, NJ: John Wiley & Sons, Inc.

Costa, A., & Santesteban, M. (2004). Lexical access in bilingual speech production: Evidence from language switching in highly proficient bilinguals and L2 learners. *Journal of Memory and Language, 50*, 491-511.

Dellwo, V. (2010). Influences of speech rate on the acoustic correlates of speech rhythm: An experimental phonetic study based on acoustic and perceptual evidence (Doctoral dissertation). *Universität Bonn*, Bonn, Germany.

Engelhardt, P., Nigg, J., & Ferreira, F. (2013). Is the fluency of language outputs related to individual differences in intelligence and executive function? *Acta Psychologica, 144*, 424-434.

Fairbanks, G. (1960). *Voice and Articulation Drillbook*. New York: Harper and Row.

Fathman, A. K. (1980). Repetition and correction as an indication of speech planning and execution processes among second language learners. In H. W. Dechert & M. Raupach (Eds.), *Towards a Cross-Linguistic Assessment of Speech Production* (pp. 77-85). Frankfurt am.Main: Peter Lang GmbH.

Fehringer, C., & Fry, C. (2007). Hesitation phenomena in the language production of bilingual speakers: The role of working memory. *Folia Linguistica, 41*, 37-72.

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

Goldman-Eisler, F. (1961). A comparative study of two hesitation phenomena. *Language & Speech, 4*, 18-26.

Goldman-Eisler, F. (1972). Pauses, clauses, sentences. *Language and Speech, 15*, 103-113.

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

Grosjean, F. (2001). The bilingual's language modes. In J. Nicole (ed.). *One Mind, Two Languages: Bilingual Language Processing* (pp. 1-22). Oxford: Blackwell.

Grosjean, F. (2008). The Léwy and Grosjean BIMOLA model. In *Studying Bilinguals* (pp. 201-212). Oxford: Oxford University Press.

Hernandez, A., Li, P., & MacWhinney, B. (2005). The emergence of competing modules in bilingualism. *TRENDS in Cognitive Sciences, 9(5),* 220-225.

Hlavac, J. (2011). Hesitation and monitoring phenomena in bilingual speech: A consequence of code-switching or a strategy to facilitate its incorporation? *Journal of Pragmatics, 43,* 3793-3806.

IBM Corp. Released 2013. IBM SPSS Statistics for Windows, Version 22.0. Armonk, NY: IBM Corp.

Künzel, H. J., Masthoff, H. R., & Köster, J. P. (1995). The relation between speech tempo, loudness, and fundamental frequency: An important issue in forensic speaker recognition. *Science & Justice: Journal of the Forensic Science Society, 35,* 291-295.

Leemann, A., Kolly, M. J., & Dellwo, V. (2014). Speaker-individuality in suprasegmental temporal features: Implications for forensic voice comparison. *Forensic Science International, 238,* 59-67.

Levelt, W. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences, 22,* 1-38.

Maclay, H. & Osgood, C. E. (1959). Hesitation phenomena in spontaneous English speech. *Word, 15,* 19-44.

MacWhinney, B. (2012). The logic of the unified model. In S. Gass & A. Mackey (Eds.), *The Routledge Handbook of Second Language Acquisition* (pp. 211-227). New York: Routledge.

O'Connell, D., & Kowal, S. (2005). Uh and um revisited: Are they interjections for signaling delay? *Journal of Psycholinguistic Research, 34*, 555-576.

Poulisse (2000). Slips of the tongue in first and second language production. *Studia Linguistica, 54*, 136-149.

Poulisse, N., & Bongaerts, T. (1994). First language use in second language production (English). *Applied Linguistics, 15*, 36-57.

Reed, M. (2000). He who hesitates: Hesitation phenomena as quality control in speech production, obstacles in non-native speech perception. *Journal of Education, 182*, 67-91.

Rose, P. (2002). *Forensic Speaker Identification*. London: CRC Press.

Runnqvist, E., Strijkers, K., Alario, F., & Costa, A. (2012). Cumulative semantic interference is blind to language: Implications for models of bilingual speech production. *Journal of Memory and Language, 66*, 850-869.

Tavakoli, P. (2011). Pausing patterns: Differences between L2 learners and native speakers. *ELT Journal, 65*, 71-79.

Temple, L. (1992). Disfluencies in learner speech. *Australian Review of Applied Linguistics, 15*, 29-44.

Wiese, R., Dechert, H., Möhle, D., & Raupach, M. (1984). Language production in foreign and native languages: Same or different. In Dechert, H. W., Moehle, D., & Raupach, M. (Eds.) *Second Language Productions* (pp. 11-25), Amsterdam: John Benjamins Publishing Co.

Yeni-Komshian, G. H., Flege, J. E., & Liu, S. (2000). Pronunciation proficiency in the first and second languages of Korean-English bilinguals. *Bilingualism: Language and Cognition, 3*, 131-49.

Appendix A

Bilingual Questionnaire

Subject Number:_____ Gender: Male Female

1. How old are you? _____
2. Where were you born? _____
3. Where were your parents born? Mom:_____

   Dad: _____
4. When did you learn English? _____
5. How long have you lived in the US? _____
6. When did you first come to the US? _____
7. During a typical day, how often do you speak Spanish? English?

   _____
8. Which language do you prefer to communicate in? _____