STARS

Comprehending Synthetic Speech Personal And Production Influences

Jingjing Wang-Costello

University of Central Florida

Part of the Psychology Commons

Find similar works at: https://stars.library.ucf.edu/etd

University of Central Florida Libraries http://library.ucf.edu

Showcase of Text, Archives, Research & Scholarship

COMPREHENDING SYNTHETIC SPEECH: PERSONAL AND PRODUCTION
INFLUENCES

by

JINGJING WANG-COSTELLO
B.S. Syracuse University, 2002
M.S. Syracuse University, 2004

A dissertation submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy
in the Department of Psychology
in the College of Sciences
at the University of Central Florida
Orlando, Florida

Summer Term
2011

Major Professor: Richard D. Gilson

ABSTRACT

With the increasing prevalence of voice-production technology across societies, clear comprehension while listening to synthetic speech is an obvious goal. Common human factors influences include the listener's language familiarity and age. Production factors include the speaking rate and clarity. This study investigated the speaking comprehension performance of younger and older adults who learned English as their first or second language. Presentations varied by the rate of delivery in words per minute (wpm) and in two forms, synthetic or natural speech. The results showed that younger adults had significantly higher comprehension performance than older adults. English as First Language (EFL) participants performed better than English as Second Language (ESL) participants for both younger and older adults, although the performance gap for the older adults was significantly larger than for younger adults. Younger adults performed significantly better than older adults at the slow speech rate (127 wpm), but surprisingly at the medium speech rate (188 wpm), both age groups performed similarly. Both young and older participants had better comprehension when listening to synthetic speech than natural speech. Both theoretical and design implications are provided from these findings. A cognitive diagnostic tool is proposed as a recommendation for future research.

This effort is dedicated to my husband, Tony. Thank you for all your love and support. Without

you, this accomplishment would not be possible.

ACKNOWLEDGMENTS

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ACRONYMS AND ABBREVIATIONS

EFL	English as First Language
ESL	English as Second Language
WPM	Words Per Minute
IV	Independent Variable
DV	Dependent Variable
Std Dev	Standard Deviation

CHAPTER ONE: INTRODUCTION

Voice production technologies with synthetic speech applications are now everywhere, built into our computers, phones, cars, home appliances, answering services, automated teller machines, building elevators, transportation centers, and much more. People seem to either love these technologies or hate them. The people who like them typically have many different reasons for it. However, the people who hate them generally have one reason: "They're too hard to understand and use!" This research effort focuses on the latter group, in particular, older adults, who are the most self-proclaimed technology-averse segment of the population, who 'hate' synthetic speech. Whether it is because of the cognitive slowing or the decline in working memory, synthetic speech has been reported to be much more difficult to understand for older adults than natural speech (Duffy & Pisoni, 1992). If older native English speakers have trouble adapting to synthetic applications, presumably it is even harder for non-native English speakers to understand and deal with computerized English-speaking tools. Therefore, the general goal of this study is to investigate the differences between older and younger adults when comprehending synthetic speech, including non-native English speakers who may have more barriers in understanding English synthetic speech. This study will explore theoretical views to provide insights as to the bases of comprehending synthetic speech as well as design recommendations to expand the acceptability of such applications across larger segments of the population.

Listening to synthetic speech may be more difficult when compared to natural human speech, and require longer processing time. As of the year 2000, the amount of information

conveyed by natural and synthetic speech at the phonemic level is different (Paris, Thomas, Gilson, & Kincaid, 2000). Specifically, synthetically generated phonemes may have been 'lower quality' in comparison with natural phonemes. In other words, many acoustic cues are either poorly represented or not represented at all (Pisoni, 1981). Also, older synthetic speech systems may have limited capability to utilize prosodic cues, such as stress, rhythm, intonation, voicing, etc. According to Paris et al. (2000, p. 422), "prosodic cues provide perceptual segmentation and redundancy, speeding the real-time processing of continuous speech. They also guide expectancies, causing search processes to end when contact is made between an acoustic representation and a cognitive representation." Without adequate prosodic cues, synthetic speech may be much more difficult for listeners to understand than natural speech.

However, despite the limitations of synthetic speech, computer-generated speech has been extensively integrated into today's society, reaching people with varying levels of language proficiency level. For instance, English as Second Language (ESL) learners, as compared to English as First Language (EFL) learners, may find it more difficult to comprehend English-based synthetic speech. There is evidence that ESL speakers might experience more difficulty deciphering synthesized speech than EFL speakers (Mack, Tierney, & Boyle, 1990, as cited in Axmear et al., 2005). Mack et al. (1990) reported that ESL adult speakers made significantly more errors than EFL speakers in identifying consonant-vowel-consonant synthesized words. However, there were no significant differences between the two groups with stimuli presented in natural speech.

Beyond previous findings, research investigating language background and proficiency as contributors to synthetic speech comprehension is limited. More research is required to gain a

better understanding of ESL speakers and their comprehension of synthetic speech. According to studies focused on human speech comprehension and second language learning, all language learners face some level of difficulty when listening to the target language (Goh, 2000). "Other issues have also been linked to non-native speaker listening difficulties of a targeted language. These range from text structure and syntax to personal factors such as insufficient exposure to the targeted language" (Goh, 2000, p. 56). Some general factors that may cause these listening difficulties include speech rate (Griffiths, 1992; Zhao, 1997), lexis (Kelly, 1991), phonological features (Henrichsen, 1984), and background knowledge (Chiang & Dunkel, 1992). Brown (1995) acknowledged the relevance of all these issues, and further argued that listener difficulties may also related to the levels of cognitive demand made by the content of the texts.

Aging also should be considered when developing or adopting synthetic speech systems, given their processing demands and the known cognitive declines with aging. Older adults regardless of language background tend to have more difficulties performing listening comprehension tasks as compared to younger adults. This performance difference may be caused by the decrease of cognitive processing resources associated with aging (e.g., Murphy, Craik, & Li, 2000; Salthouse, 1988; Stine & Wingfield, 1987). Therefore, for older ESL adults (i.e., age 65 to 80) comprehending synthesized English language speech may be even more challenging.

Figure 1 shows the fast growth of the U.S. immigrant population in recent decades. Although, most of the growth is driven by working age immigrants, the number of older immigrants who are defined as people age 65 years or older who were not born in the U.S. and who reside in the United States at a given time has almost doubled since 1990 (Leach, 2008). Presently, older immigrants account for approximately 11 percent of the overall immigrant

population. Furthermore, 8 percent of all older Americans are older immigrants (Treas &

Batalova, 2007). Thus, as the number of older immigrants in the United States increases every

year, it is important to know the specific characteristics of older immigrants when interacting

with synthetic speech applications. Note in the following, these older immigrants are referred to

as older ESL adults.
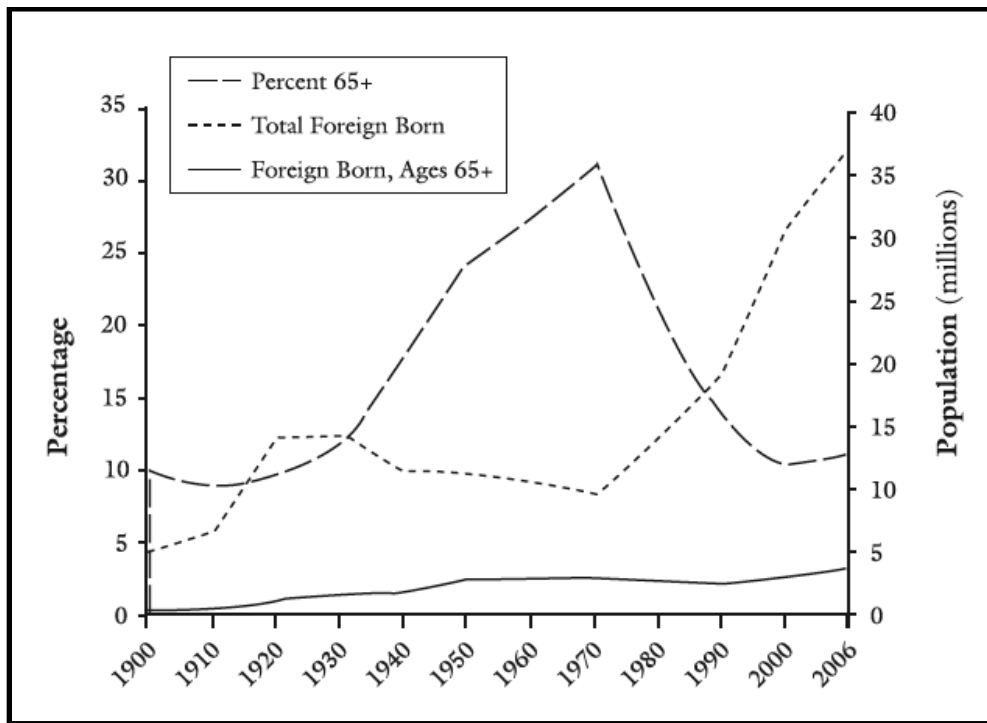
Figure 1. Size of foreign-born population and percentage age 65 or older, United States (Leach, 2008)

Speech rate also influences speech comprehension. After passing a certain speed in words

per minute (wpm), speech comprehension declines as the speech rate increases (Griffiths, 1992).

Consequently, speech rate may play an important role regarding the acceptance of synthetic

speech application. For example, presenting synthetic speech at a rate that is too fast for older

users may cause them to tend to avoid the use of synthetic applications. Similarly, ESL speakers may have a different speech rate preference from EFL speakers. For instance, Griffiths (1990) reported that speech faster than 200 wpm was difficult for lower-intermediate ESL speakers to comprehend, with ESL speakers performing the best at a rate of 127 wpm. Therefore, this aspect of synthetic speech is important to evaluate.

Although synthetic speech systems may provide many advantages to enhance communication by converting often ambiguous information from a variety of sources to more easily comprehensible verbal communication, existing research does not address whether synthetic speech can truly benefit older adults (ages 65-80), as well as ESL adults. This is the first study that focuses on how one's age and language background may affect his or her performance on synthetic and natural speech comprehension. Accordingly, the goals of this study are to investigate: (1) if ESL speakers and EFL speakers differ in their ability to comprehend English-based synthetic speech; and (2) if these differences are more pronounced for older adults. In particular, older ESL speakers and older EFL speakers will be compared in terms of their ability to comprehend synthetic speech at three different speech rates (slow, medium, and high) to vary the difficulty.

CHAPTER TWO: LITERATURE REVIEW

Synthetic Speech vs. Natural Speech Comprehension

When discussing synthetic speech comprehension, it is essential to review how people comprehend natural speech. Anderson (1995) proposed a comprehension model that consists of *perception*, *parsing*, and *utilization*. *Perception* is a process that encodes incoming audio or a written message. During listening, the perceptual process separates phonemes from the continuous speech stream. At the *parsing* stage, the meaning of words are combined and transformed into mental images. The parsing process starts with an utterance being separated into smart segments based on syntactic structures. These segments are then being recombined to create a meaningful representation of the original message. This mental image created during the parsing stage is related to an individual's existing knowledge that is stored in his long-term memory. During the *utilization* stage, a listener may use his existing knowledge to draw different types of inferences to complete the interpretation. The three stages of speech comprehension, perception, parsing, and utilization, represent different levels of processing, with perception being the lowest. They are interrelated and recursive. They could also occur concurrently during a speech comprehension task (Goh, 2000).

Past research has concluded that synthetic speech in general is more difficult to understand than natural human speech (Mirenda & Beukelman, 1987). The main reason for this difficulty may be explained in terms of an increase in attentional processing demands. Since synthetic speech does not contain as much redundant acoustic and prosodic cues as natural speech, it is presumed to overburden the attention allocation system (Duffy & Pisoni, 1992).

According to Slowiaczek and Nusbaum (1985), when prosodic cues were removed from a natural speech sentence, people's ability to recognize words in the sentence diminished. When the same procedure of removing prosodic cues was performed in synthetic speech, the effect was even more pronounced. Despite the considerable amount of effort that has been invested into making synthetic speech more intelligible, most of the synthetic applications still have difficulty replicating natural sentence prosody (Paris et al., 2000; Roring et al., 2007).

Synthetic speech may also take longer for people to process than natural speech (Ralston, Pisoni, Lively, & Greene, 1991). However, researchers have not reached an agreement regarding this statement. Reynold and Given (2001) measured participants' speed and accuracy performance of synthetic and natural speech using a sentence verification task. They concluded that synthetic speech does not require more processing time or resource than natural speech (Roring et al, 2007). In contrast, Pisoni, Manous, and Dedina (1987) found that at comfortable speech rates, synthetic speech requires more processing time than natural speech. "The mixed findings have been explained in terms of differing methodologies (Reynolds & Givens, 2001), though a deeper investigation into the problem is needed to disentangle whether and how speech rate affects the comprehension of synthetic speech, especially when considering users of different ages" (Roring et al., 2007, pp. 26).

Factors that Affect Synthetic Speech Comprehension

Although a number of factors may affect synthetic speech comprehension, this study will focus on the following major factors: age, language background, speech rate, and intelligibility. Each of these will be discussed in turn.

Age

> *Hearing Declines Associated with Aging*

Hearing loss is one of the most common chronic disabilities among older adults. Approximately half of the older adults aged between 75 and 79 have some degree of measurable threshold hearing loss (Schneider & Pichora-Fuller, 1999). Specifically, hearing loss in the higher frequency ranges (e.g., above 1000 Hz) that are essential for the accurate perception of speech may create more difficulties for older adults to comprehend audio messages than for younger adults (Cruickshanks et al., 1998; Macoy et al., 2005). In addition, noise is commonly present in daily conversations and speeches. Since older listeners may have difficulty filtering out the background noise, their perception of speech could also be affected (Tun, O'Kane, & Wingfield, 2002). Thus, when studying older adult participants, it is important to account for hearing declines.

> *Cognitive Declines Associated with Aging*

The cognitive decline associated with normal aging has been well documented in the cognitive literature. Some of the major cognitive changes include reductions in working memory capacity, attentional difficulties in inhibiting irrelevant information, and slowing in perceptual and cognitive processes (McCoy et al., 2005). In addition, slowed information processing has been considered as a main factor that causes age-related differences in reading comprehension (Pichora-Fuller, 2003). The effect is assumed to be even more damaging for listening comprehension, because a reader usually has full control of his/her reading rate, whereas a listener normally does not have control of a speaker's speaking rate (Winfield & Tun, 2001).

Thus, age-related processing losses of working memory and attention may affect older adults' ability to comprehend speech, specifically synthetic speech. It is important to note that though many cognitive theories have been proposed to explain older adult listening comprehension performance, one single theory on its own does not sufficiently account for this performance loss.

Working memory may be considered as a modern view of traditional short-term memory. Unlike short-term memory, which is viewed as a passive storage space, working memory is an active component that not only stores, but also processes information. One of the most influential models of working memory was proposed by Baddeley and Hitch in 1974, which consists of four major components: central executive, phonological loop, visuo-spatial sketchpad, and episodic buffer. The *central executive* may be considered as the 'boss' in the working memory model. It sends and retrieves information from long-term memory as well as controls the information that flows inside of working memory. The *phonological loop* is mainly responsible for processing auditory information. The *visuo-spatial sketchpad* processes and stores both visual and spatial information. Visual and spatial information are then sent to the central executive. The *episodic buffer* was later added to the working memory model by Baddeley in 2001 to accommodate temporary storage of auditory and visual information and more importantly, to integrate information from the visual and auditory channels. Therefore, according to Baddeley's model, it may be concluded that working memory is a key component of receiving, maintaining, and processing of information.

Past research has shown that there is a close relationship between working memory and aging. Craik (1994) stated that working memory declines are related to the amount of

information that is being remembered or processed by an individual. In other words, as the amount of incoming information increases, it becomes more difficult for older adults to remember and manipulate this information than for younger adults (Hardee & Mayhorn, 2007; Kempter, Herman, & Lian, 2003).

Another sign of aging in relation to working memory is the slower processing speed that is often observed in older adults (Smither, 1993). Salthouse and Babcock (1991) proposed two possible mechanisms that might account for the impact of speed on working memory. One mechanism was that increased age might be associated with a more rapid loss of information. The second mechanism suggested that aging might lead to a slower encoding or activation of information. "In either case, the amount of simultaneously active information, which can be considered equivalent to working memory capacity, would be smaller among older adults than among younger adults" (Salthouse, 1994, pp. 540-541).

Capacity shrinkage is another common explanation for the decline of working memory. Craik and Byrant (1982) suggested that because older adults have less 'mental energy,' or have decreased 'processing resources' available to them, this explains, or at least partially explains, why older adults perform more poorly than younger adults during memory tasks (Park & Payer, 2006). Zacks, Hasher, and Li (1999) explained capacity shrinkage in a more specific fashion. They indicated that the age-related decline in processing resources suggests that older adults experience more difficulties carrying out resource-demanding encoding and retrieval operation than younger adults.

Since working memory has limited capacity, the ability to stop irrelevant information from entering working memory is essential for working memory to function efficiently. In

addition, the ability to delete no-longer-useful items from working memory to free up space for new and relevant information has also been proven to ensure the productivity of working memory (Park & Payer, 2006). Older adults have difficulties not only in filtering out irrelevant information from entering their working memory, but also in deleting information that is no longer useful (Grady & Craik, 2000; Hasher & Zack, 1988). When the ability of filtering/deleting irrelevant information declines, it creates a kind of 'mental clutter.' This mental clutter refers to how extraneous thoughts and plans can interfere or even crowd out task-relevant information in working memory (Zacks et al., 1999). These attributes have been commonly used to explain the decline of working memory due to age.

A number of cognitive aging theories have suggested that attention deficits may be a contributing factor to all or most age-related changes in cognition (McDowd & Shaw, 1999). Among these cognitive theories, the inhibition-deficit view has been particularly prevalent. Hasher and Zacks (1988) suggested that age-related processing decline in a variety of cognitive functions might be caused by a decrease in the ability to inhibit task irrelevant information. In other words, inefficient inhibition could lead to inefficient selective attention, which could result in the intrusion of irrelevant information into working memory (Kane & Engle, 2003). The consequence of having 'noise' information in working memory may include increased processing time and poor performance on recognition and recall of relevant information (Kramer, Humphrey, Larish, Logan, & Strayer, 1994). Since speech comprehension often takes place in the context of several resources competing for attention, the ability to focus on relevant information while inhibiting irrelevant information is vital to speech comprehension tasks.

Synthetic Speech and Older Adults

Past research on synthetic speech comprehension has shown that people are able to perceive and respond to synthetic speech under ideal environmental conditions (e.g., in a quiet room with no distractions). However, in conditions that are closer to real world environments (e.g., with environmental noise and other sources that compete for the listeners' attention), listeners must devote much more effort to perceive the synthetic speech, which causes comprehension to suffer (Drager & Reichle, 2001; Duffy & Pisoni, 1992). Since age-related cognitive and physical declines are often observed in older adults, it is likely that older listeners would have more difficulty understanding synthetic speech.

Language Background

Language Processing in Relation to One's Language Background

A traditional view of second language learning assumes that second language learners have more difficulties with the grammar rather than with the lexicon (Clahsen & Felser, 2006). Lexicon may be defined as the vocabulary of a language and grammar may be seen as the characteristic system of inflections and syntax of a language (Merriam Webster Dictionary, 2010). However, this opinion has been proven by recent research as an inaccurate or incomplete statement (Clahsen & Felser, 2006). It has been shown that even 'late' learners who acquired a second language after puberty were able to achieve native-like processing in some domains of grammar. Therefore, grammar may not be the sole contributor to the differences between a native speaker and a non-native speaker regarding language processing. The authors proposed the following three major factors to explain how non-native and native speakers differ when

processing a language.

Factor 1: Limitations of the second language grammar. In human languages, grammar may be considered as a system of rules that govern the formation of words, phrases, and sentences. Linguistic research has found that acquisition of grammar by late second language learners tends to be less successful than first language acquisition (Bley-Vroman, 1990). It is also evident that grammar systems developed by late second language learners may be somewhat different from the native grammar system (Clahsen & Muysken, 1996). Thus, assuming that successful language processing requires native-like grammar skills, non-native speakers may perform less well than native speakers due to their grammatical insufficiency.

Factor 2: The role of first language transfer. The non-native speakers' native language may affect their processing of a second language. Past research has shown that phonological, orthographic, morpholexical, and lexical-semantic properties of one's first language could affect his or her second language processing (Frenck-Mestre & Pynte, 1997). For example, when a native German speaker who speaks English as a second language processes an agentive noun ending with –er in English, he or she may automatically activate the masculine gender features associated with the German agentive suffix –er (Scheutz & Eberhard, 2004).

Factor 3: Cognitive resource limitations. Past research has used event-related potentials to study whether non-native speakers are less automatic than native speakers during grammatical processing. These studies have found that automaticity is indeed lower in second language sentence processing. One possible explanation for automaticity differences between native speakers and non-native speakers is that having to identify words and phrases in a second language drains more cognitive resources from working memory (Ardila, 2003). In behavioral

experiments, the response times of second language speakers tended to be slower than native speakers (Clahsen & Felser, 2006). In addition, the results from neuro-imaging studies have shown higher cortical activation for structurally difficult sentences in second language speakers. This result suggested that comprehending a second language requires greater computational effort than first language comprehension (Hasegawa, Carpenter, & Just 2002).

Language Comprehension in Relation to One's Language Background

Even though Anderson's (1995) three-phase speech comprehension model is based on first language comprehension, the same theory may also be applied to second language comprehension. According to Færch and Kasper (1986), the fundamental cognitive processes in first and second language comprehension are similar despite that non-native speakers would face more linguistic and sociolinguistics challenges.

Goh (2000) conducted a study to investigate the difficulties that intermediate ESL speakers encounter when comprehending speech. One of the problems is that ESL speakers tend to quickly forget what is heard. According to Anderson's three-phase speech comprehension model, this observation represents a parsing problem. One possible explanation for this issue is the limited capacity of the listeners' short-term memory. Goh (2000) observed that the times when ESL speakers experience such memory problems the most is when the parsed information was followed by unfamiliar input, such as new vocabulary. Since processing unfamiliar information may consume more cognitive capacity, the parsed information received prior the unfamiliar content would not be rehearsed enough to be stored into long-term memory.

The second problem indicated by Goh's study (2000) was that non-speakers do not recognize words they know. This difficulty is related to a fundamental aspect of comprehension

14

– perceptual processing. ESL speakers were found to be unable to recall the meaning of familiar words immediately, which causes them to be unable to process messages using those words. A possible cause for this slow recognition may be that ESL speakers could not match the sounds they heard with the script they stored in their long-term memory. In other words, the sound-to-script link might not be fully automatized.

The third difficulty found by Goh (2000) was that non-native listeners could understand words but not the intended message. In other words, although listeners were able to understand the meaning of each word, they could not make useful inferences from the message. Inference is not only a process that is essential to speech comprehension (Eysenck & Keane, 1995), but also a crucial mental activity during utilization (Aderson, 1995).

Synthetic Speech and Language Background

Since many studies have shown that ESL speakers have more difficulties comprehending a natural speech passage than native speakers, it is likely that the same pattern may be observed in synthetic speech comprehension. Greene (1986) found that non-native listeners of English have significantly more difficulty perceiving synthetic speech (in English) than native speakers. Greene tested non-native and EFL speakers using the Modified Rhyme Test and a sentence transcription task, using speeches that were generated by a natural voice and MITalk. MITalk is a tool that can be used to generate text-to-speech synthetic audio stimuli. The ESL speakers only performed slightly worse than the EFL speakers in the natural speech condition. However, in the MITalk synthetic speech condition, ESL speakers performed significantly worse in terms of accuracy than EFL speakers. The performance gap between the ESL and EFL speakers was much greater for the sentence transcription task than for the Modified Rhyme Test. Greene

(1986) also found that ESL speakers' performance was highly correlated to their English proficiency level. He concluded that the ability of ESL speakers to perceive synthetic speech depended greatly on their proficiency in the targeted language (Winters & Pisoni, 2004).

Reynolds, Bond, and Fucci (1996) also found that ESL listeners transcribed synthetic speech sentences less accurately than EFL listeners. The authors asked listeners (both non-native and native) to transcribe synthesized English sentences in quiet and noise conditions. They found that ESL listeners' performance, which was measured by the percent correct score, dropped 8.7% between the quiet and noisy conditions, while the EFL listeners only dropped 2.8%. Reynolds et al. (1996) concluded that the drop in performance might be due to the possibility that ESL listeners were less familiar with the English language and they were less able to interpret unusually pronunciations that were associated with the language. Similar to Greene (1986), Reynolds et al. (1996) also suggested that ESL listeners' ability to perceive synthetic speech heavily depended on their proficiency in the English language (Winters & Pisoni, 2004).

The second language learning literature has summarized a variety of reasons of why non-native speakers may experience more difficulties when comprehending a targeted natural language (e.g., English) than native speakers. Rubin (1994) stated five factors that effect second language listening comprehension: 1) text characteristics; (2) interlocutor characteristics; (3) task characteristics; (4) listener characteristics; and (5) process characteristics. *Text characteristics* involve the variation in a listening passage/text, such as speech rate. *Interlocutor characteristics* are the variation in the speaker's personal characteristics (e.g., a speaker's gender). *Task characteristics* involve the different types of listening tasks. For example, Sohamy and Inbar (1991) studied how different types of questions might influence non-native speakers' listening

comprehension. They found that participants were better at questions referring to local cues in the text than at those that refer to global cues. The authors concluded that generalizing, inferring, and synthesizing information is more difficult for non-native speakers than looking for data specific information. *Listener characteristics* appear to have considerable impact on one's listening comprehension. These characteristics include language proficiency, memory, age, and gender, etc. Finally, *process characteristics* involve how non-native listeners' process auditory messages of a targeted language. More specifically, process characteristics are the strategies that people use to process information, such as using top-down processes, bottom-up processes, or both. A top-down process suggests that the listener uses his/her knowledge of the world or situations to focus on meaning. A bottom-up process involves the listener using his/her knowledge in words, syntax, and grammar to work on sentence structure.

Previous research has shown that when listening to synthetic speech, non-native listeners tend to experience even more difficulties than listening to natural speech. However, factors that could cause this effect have rarely been investigated. The present study theorizes that because synthetic speech tends to miss natural speech cues or sounds machine-like, non-native listeners may have more difficulties matching the words they heard to the words that are stored in their long-term memory. Long (1989, p. 32) described a typical listening comprehension process: "…listeners construct meaning during the comprehension process by segmenting and chunking input into meaningful units, actively matching the results, known as intake, with their existing linguistic and world knowledge, and filling in the gap with logical guesses." Since non-native speakers are generally less proficient with a targeted language than native speakers, they may be likely to have more matching problems after parsing and chunking incoming words. In other

words, non-native listeners may experience more matching losses than native speakers during synthetic speech listening comprehension.

Speech Rate

Speech rate is another important factor that could affect synthetic speech comprehension. Griffiths (1992) suggested that after passing a certain rate level, speech comprehension declines as the speech rate increases. Researchers have studied speech rate preferences using natural human speech. Lass and Prater (1973) studied speech rate preferences using younger adults and found that participants most preferred 175 wpm and least preferred 100 wpm for oral reading (Sutton, King, Hux, & Beukelamn, 1995). Similar research on speech rate preference has also been conducted using synthetic speech. Sutton et al. (1995) reported that older adults (M = 69 years of age) preferred speech rates between 130 and 210 wpm. Among younger adults (M = 23.5 years of age), speech rates below 150 wpm were judged as too slow and rates above 220 were too fast (Axmear et al, 2005).

According to Sutton et al. (1995), speech rate is an important social variable that may differentiate communication partners in their willingness to accept synthetic speech. For instance, delivering synthetic speech to older adults at a rate that is too fast for them to process may cause them to reject the use of synthetic speech applications in the future.

In addition, ESL speakers may have a different preference for speech rate as compared to EFL speakers. As cited in Rubin (1994), Griffiths (1990) found potential evidence that speech faster than 200 wpm was harder for lower-intermediate ESL speakers to understand. This level of ESL learners performed best at 127 wpm speech rate level. Similarly, Kelch (1985) also found that listening at a lower speech rate (e.g., 124 wpm) significantly improved ESL speakers'

comprehension performance. It is plausible that advanced ESL learners may prefer similar

speech rate as EFL speakers. However, more research is required to investigate the speech rate

differences, specifically synthetic speech rate preferences between ESL and EFL speakers.

Following the same notion, due to the fact that the ESL speakers may be less proficient in

the English language than the EFL speakers, they may have different preferences for speech rate.

Griffiths (1990) found potential evidence that speech faster than 200 wpm is hard for lower-

intermediate ESL speakers to understand. He suggested that this level of ESL speakers

performed best at 127 wpm.

In addition, older ESL adults' synthetic speech rate preferences are likely to differ from

their younger counterparts. When listening to synthetic speech at a rate that is not appropriate for

a particular ESL age group, they may have difficulty comprehending the message. Therefore, it

is essential to study the speech rate preferences of both younger and older ESL adults so that

future synthetic speech applications can be designed to accommodate their specific needs.

Intelligibility of Synthetic Speech

Since the literature has not formally defined the attribute of *intelligibility* as it relates to

synthetic speech, a formal definition was developed by the author. Intelligibility, as it relates to

synthetic speech, may be defined as the clarity of the speech itself as well as the clarity of the

message to its listeners. Besides the actual synthetic voice used, there are several factors that

may influence the intelligibility of synthetic speech: speech rate, competing background noise,

and the age of listeners (Axmear et al., 2005). Venkatagiri (1991) studied the correlation between

speech rate and synthetic speech intelligibility. He reported that slowing syllable production

from 201 syllables per minutes (spm) to 139 spm significantly improved intelligibility. Reynolds

et al. (1996) studied background noise in relation to intelligibility using EFL and ESL speakers. They concluded that synthetic stimuli were less intelligible when background noise was present for both speaker groups. However, ESL speakers experienced significantly more difficulty understanding the stimuli. Whether or not the age of listeners has an effect on synthetic speech intelligibility has been inconclusive. Humes, Nelson, and Pisoni (1991) suggested that age did not significantly impact one's intelligibility performance during a synthetic speech task. In contrast, Kangas and Allen (1990) stated that age was a significant factor for synthetic speech intelligibility. Although intelligibility is not the point of interest in the present study, it will be measured as a potential confounding variable due to its direct relations to listening comprehension.

CHAPTER THREE: METHODOLOGY

Experiment Design

The present study was intended to study whether older and younger adults, as either EFL speakers or EFL speakers, differ in their ability to comprehend English-based synthetic speech. Difficulty was varied by in three speech rate (slow, medium, and high). Natural speech was used as a control condition. To investigate these research questions, this study employed a mixed between-within group experimental design (see Table 1). Independent variables were indicated as IVs and Dependent Variable was shown as DV. The variables are listed follows:

IV1 (between-groups) Age (Younger vs. Older adults): younger adults defined as age 18-25, older adults defined as age 65 to 80.

IV2 (between-groups) Language background (ESL vs. EFL)

IV3 (within-groups) Speech type (Synthetic Speech vs. Natural Speech)

IV4 (within-groups) Speech rate (Low, Medium, High): Low speech rate defined as 127 wpm, Medium speech rate defined as 188 wpm, and High speech rate defined as 225 wpm. The selection of speech rates for this study was based on two rationales. First, the low and medium speech rates were adopted from a study by Griffiths (1992), which examined younger ESL adults' listening comprehension when listening to passages presented at 127 wpm, 188 wpm, and 250 wpm. Second, since older adults tend to prefer a slower speech rate than younger adults (Wingfield & Ducharme, 1999), the present study used 225 wpm for its high speech rate condition.

DV: Comprehension accuracy: a participant's comprehension accuracy will be measured by the number of correct responses on the Experimental Comprehension Task.

Table 1 *Experimental Design*

Language	Age Group	Speech Type		Speech Rate		
		Synthetic	Natural	Low	Medium	High
EFL	Younger					
	Older					
ESL	Younger					
	Older					

Hypotheses

The following four hypotheses were investigated in this study.

<u>Hypothesis 1</u>: A main effect for age was hypothesized such that, across all other conditions, younger adults would exhibit higher comprehension accuracy than older adults.

Rationale for Hypothesis 1: In comparison to younger adults, age-related processing losses of working memory and attention, and other factors such as hearing loss of high frequency sounds (above 1000 Hz) may negatively affect older adults' ability to comprehend speech.

<u>Hypothesis 2</u>: A main effect for language was also hypothesized, such that across all conditions, EFL participants would exhibit higher comprehension accuracy than ESL participants.

Rationale for Hypothesis 2: Since non-native speakers are generally less proficient with a targeted language than native speakers, they may experience more language-related matching losses when listening to synthetic and natural speech.

Hypothesis 3: A two-way interaction between age and language background was hypothesized. Older ESL adults would have significantly lower comprehension accuracy than older EFL adults. No significant differences were hypothesized on comprehension accuracy between younger ESL and younger EFL adults (Figure 2). Specifically, age was hypothesized to have a stronger effect on speech comprehension than language background.
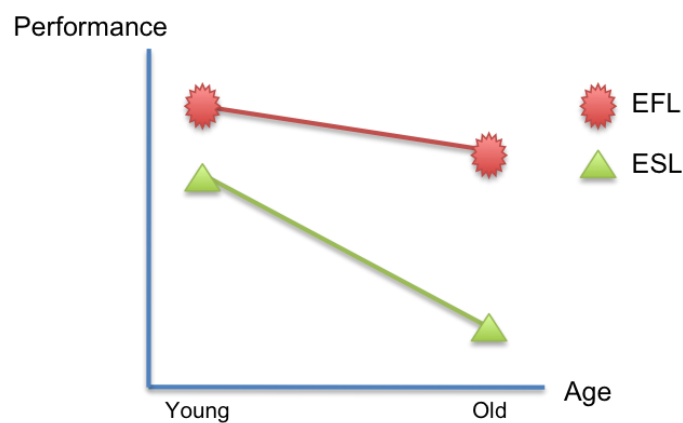
Figure 2. Hypothesized interaction effect between age group and language background.

Rationale for Hypothesis 3: Older EFL listeners' performance may suffer from age-related processing losses. However, older ESL listeners' performance may be hindered by both

age-related processing losses and language-related matching losses. Therefore, older ESL listeners may perform significantly worse than other groups.

Hypothesis 4: A three-way interaction between age, language, and speech rate was hypothesized. Across language groups, at a low speech rate, older adults would exhibit only slightly worse comprehension accuracy than younger adults. As speech rate increases, however, older adults' comprehension accuracy would exhibit significantly lower comprehension accuracy than younger adults. The effect of speech rate on age would be greater for ESL participants than for EFL participants (Figure 3 & 4).
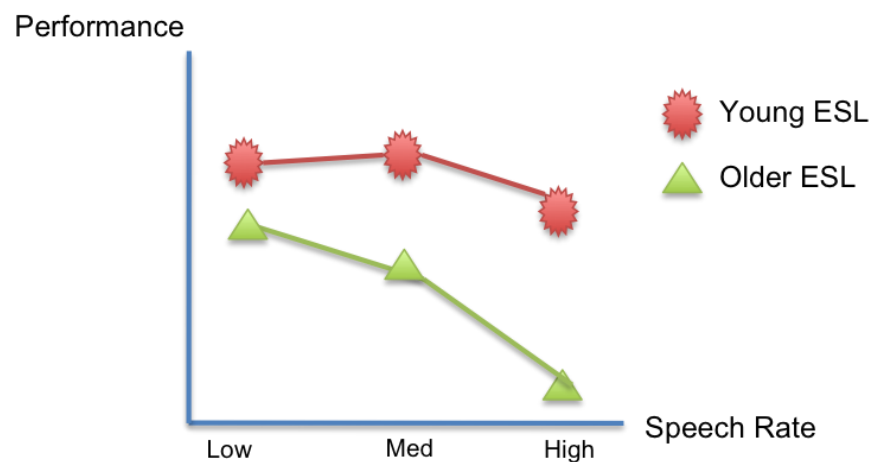
Figure 3. Hypothesized interaction effect between age, language, and speech rate: ESL speakers
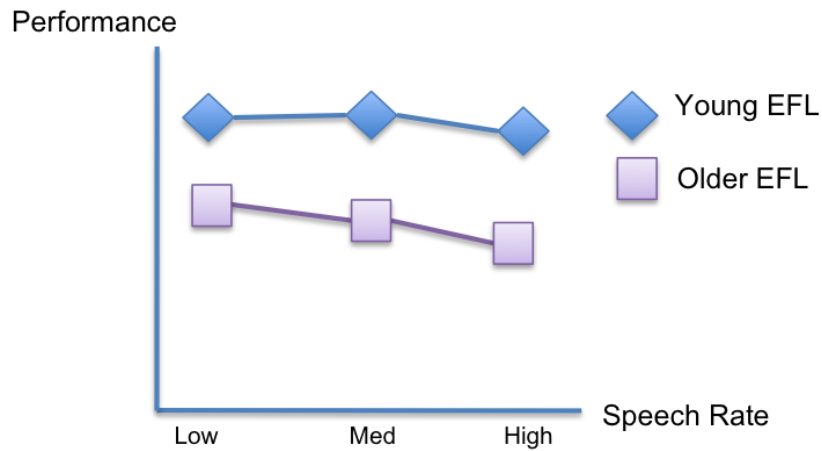
Figure 4. Hypothesized interaction effect between age, language, and speech rate: EFL speakers

Rationale for Hypothesis 4: According to past literature, younger adult listeners tend to prefer faster speech rate than older adult listeners. For example, Sutton et al. (1995) reported that older adults preferred speech rates between 130 and 210 wpm. Among younger adults (M = 23.5 years of age), speech rates below 150 wpm were judged as too slow and rates above 220 were too fast (Axmear et al., 2005). In addition, Griffiths (1990) suggested that speech faster than 200 wpm was harder for lower-intermediate ESL speakers to understand. This level of ESL learners performed best at 127 wpm speech rate level. Similarly, Kelch (1985) also found that listening at a lower speech rate (e.g., 124 wpm) significantly improved ESL speakers' comprehension performance. Since age-related processing losses and language-related matching losses may both hinder ESL older adults' performance, as speech rate increase, this participant group may suffer significantly greater performance decline as compared to the other groups.

Participants

An a priori power analysis was conducted using the G*Power 3 program (Faul, Erdfelder, Lang, & Buchner, 2007). With the effect size set to f = .25, α = .05, and desired power level at .80, 112 total participants (28 per condition) were needed to achieve a power level of .81. See Appendix B for a detailed illustration. A total of 117 people participated in the study (48 males and 69 females).

Participants for the older adult group were recruited from the Learning Institute for Elders (LIFE) at the University of Central Florida (LIFE@UCF), a nonprofit organization which provides continuous education programs designed for older adults, as well as from local active retirement communities. Monetary compensation was offered to these participants as recruitment incentives. The majority of older adult participants (n = 59) fell within the age range of 65 to 80. Participants for the younger adult group were recruited from the Psychology Department at the University of Central Florida. Extra course credit was offered to these participants as recruitment incentives. The majority of younger adult participants (n = 58) fell within the age range of 18 to 27. Participants for the ESL group were non-native English speakers who reported speaking Spanish as their first language and English as their second language. All participants had normal hearing. Additional demographic information can be found in Appendix I.

Materials

Demographics Questionnaire

A paper-based demographics questionnaire contained questions regarding age, gender, education, hearing, and language background information. In particular, information regarding when a participant learned English, and the years of English practice that a participant had were documented. This information was used to establish each participant's self-reported English proficiency level. See Appendix B for the complete version of this questionnaire. See Appendix I table 8 for participant demographic information on age, gender, education and race.

Speech-in-Noise Hearing Test

A 5-minute computer-based speech-in-noise hearing test, adopted from the Hearing-It Organization (Hearing Test, 2010), was used to measure how well all participants could hear and understand speech in the presence of distracting ambient noise. The ambient noise levels varied from low to high. During the test, participants were instructed to wear a noise-canceling headset. Participants first went through a speech sample, which provided them the opportunity to adjust the headset volume and became familiar with the test. During this hearing test, numeric numbers were read aloud at random order. The voice was overlaid with various degrees of background noise. When the numeric number was presented, participants were asked to select the matching number by clicking on the corresponding number button, or guess the best answer if they were not sure.

English Proficiency Test on Listening Comprehension

After consulting with Dr. Mihai (F. Mihai, personal communication, May 20, 2010) who is an expert in language assessment for English language learners, and Ms. Farina (M. Farina, personal communication, May 26, 2010) who is a senior ESL instructor from the University of Central Florida, a condensed version of the listening section of a TOEFL Sample Test (4th edition) (Educational Testing Service, 1990) was adopted to measure all participants' English proficiency on listening comprehension. According to the Education Testing Service (2009), TOEFL is a valid and reliable test for measuring how well a person uses English. In this 10-minute English listening assessment test, every participant was evaluated on his or her English listening levels. After listening to each short sentence, conversation, or speech, participants were instructed to answer one or more multiple-choice questions. The test consisted of five short sentences, four short conversations, and one longer speech. See Appendix D for transcripts of this test.

Shadowing Task

A shadowing task was generated for this study to assess participants' intelligibility of the synthetic speech, which could be a potential confounding variable. The participants were instructed to listen to ten synthesized sentences. After each sentence, they were asked to repeat the sentence back to the experimenter word by word. The purpose of this shadowing task was to ensure that participants were able to hear the synthesized sentences well. In other words, this task was used to measure intelligibility of the synthetic speech. An accuracy score, which was measured by the percentage of correct words that were shadowed, were generated upon the completion of this task for each participant.

Post-Trial Questionnaire

The post-trial questionnaire was intended to measure participants' perceived mental workload and their preference regarding speech rate. The questionnaire consisted of two questions. The first question asked participants to rate the difficulty of the passage that they just heard on a 7-point scale from "very easy" to "very difficult." The second question asked participants to evaluate the speed at which the passage or trial was presented (also on a 7-point scale) from "too slow" to "too fast." See Appendix G for more details.

Apparatus

An Apple Macbook Pro computer with Mac OS X 10.5.8 operating system was used for this study. Synthetic speech materials were generated using the "Text–to-Speech" function provided by the Mac operating system. In the Text-to-Speech function, a list of male or female voices was available. Since a large portion of male older adults suffer from some degrees of hearing loss, high frequency acoustic signals which are often found in the female voice may cause more difficulty for male older adults than for female older adults (Lines & Hone, 2002). For this reason, a male voice was more suitable than a female voice for the present study. Thus, a synthesized male voice named "Alex" was selected for reading the synthesized passages. To generate natural speech passages, a 34 year old male native English speaker's voice was used. According to Cruickshanks et al. (1998), audio stimuli with frequencies higher than 1,000 Hz were much more difficult for older adults to hear. Therefore, it is essential that the hearing stimuli used in this study did not exceed 1,000 Hz. These passages were presented using a Sony

noise-canceling headset with an adjustable volume control. The noise-canceling feature on the headset was used to block out any background noise.

Experimental Tasks

The experimental tasks included a training section and an experiment section. The training section was designed to not only help participants to become familiar with the experiment section, but also to provide an opportunity for participants to ask questions. This procedure ensured that participants had a clear understanding of the experiment section. The training section consisted of one synthesized passage and one natural human speech passage. Both passages were recorded at the medium speech rate. Once a passage was played for the participants, they were asked to answer a multiple-choice question.

The experiment section contained a total of 12 short computer-generated passages, of which six passages were synthetic and the other six passages were natural speech. The readability of these experimental passages was measured by the Flesch-Kincaid Grade Level Readability Formula, which was provided by Microsoft Word 2007. The experimental passages scored an 8.8 on the Flesch-Kincaid Grade Level test, which was appropriate for American 8th grade level readers (Si & Callan, 2001). Each passage was followed by several questions that were associated with the information of the passage. Two passages were recorded for each speech rate (i.e., Low, Medium, or High) for both synthetic and natural speech types (Table 2).

Table 2 *Audio Passage Assignments*

Speech Rate	Natural Speech Type	Synthetic Speech Type
Low	2 passages	2 passages
Medium	2 passages	2 passages
High	2 passages	2 passages

The order of the 12 passages were counterbalanced, but the content of each passage was held constant, fixed to a given speech rate and speech type. For example, Passage One on the "TV program" was always presented using synthetic speech and at a low speech rate, Passage Two on "yawning" was always presented using synthetic speech and at a medium speech rate, and so on. Keeping the passage content fixed to the presentation format controlled for a possible interaction effect between the content and the other independent variables (speech rate or speech type). After listening to a passage, participants were instructed to answer 2-3 multiple-choice questions regarding the information presented in the passage. Participants were instructed to select the best answer.

Every passage covered a specific topic, such as zoology, literature, and astronomy, etc. Due to the concern that participants' background knowledge might play an important role in their listening comprehension, a list of topics that were used in the study was included in the demographics questionnaire. Participants were asked to rank these topics based on their familiarity with a particular topic. This way, a topic familiarity level for each participant was established.

Procedure

Upon arrival, participants were briefed as to the purpose of the study and were then asked to read and complete the Informed Consent, followed by the demographic questionnaire. Participants were then asked to take the 5-minute computer-based speech-in-noise hearing test. At the end of test, the percentage of correct answers was presented to the participants. The experimenter then recorded this number to establish each participant's hearing level. An English proficiency test on listening comprehension was administered after the hearing test. Each participant's performance was graded on the percentage of correct answers. If a participant was unable to score 60% or higher on this test, he/she was eliminated from this study, due to the concern that the participant's listening comprehension of English may not have been at the proficiency level needed to participate in this study.

Once participants completed the English proficiency test on listening comprehension, they were asked to complete the shadowing task. Participants were scored based on the percentage of correct words that were repeated. All participants were asked to take a 5-minute break after the shadowing task. Participants then proceeded to the training section of the experimental task. Once participants understood the task objective, they were then asked to complete the 12 experiment section passages. Once participants completed the multiple-choice questions after each experiment section passage, the post-trial questionnaire was administered. Finally, participants were debriefed.

CHAPTER FOUR: RESULTS

Data Analysis Plan and Data Screening

Analyses were conducted using SPSS 11.0 for Mac OS X. An alpha level of .05 was used for all analyses, unless otherwise noted. Before the analyses were performed, the data were screened for any potential issues that could affect the results of the statistical analyses.

Normality was checked for the dependent variable – the Average Comprehension Score, using the Kolmogorov-Smimov (K-S) test for normality. It is found that the DV was not normally distributed. According to Tabachnick and Fidell (2000), in large samples (over 100), a not normally distributed DV is fairly common. Therefore, analysis was continued.

The scores for the speech-in-noise hearing test, the English proficiency test on listening comprehension, and the shadowing task were generated by calculating the percentage of correct answers. The means and standard deviations for each test performance are presented in Appendix I, Table 7. The purpose of the shadowing task was to control for the intelligibility of the synthetic generator. If participants performed poorly on the shadowing task, the intelligibility of the synthetic speech might be a confounding variable. Since all participants scored highly in the shadowing task (everyone scored higher than 94% out of 100%), intelligibility was not considered to be a confounding variable.

Participants' performance on the Experimental Comprehension Task was coded using ones and zeros. Participants received 1 point for a correct answer and 0 points for an incorrect answer. Correct answers were averaged to arrive at an overall performance score (the total performance score of all 12 passages), and again for each passage, each speech rate, and each

speech type. The Post-Trial Questionnaire asked participants to rate their perceived difficulty and perceived speed after listening to each experimental passage.

Listening comprehension literature has shown that a participant's background knowledge may play an important role in their listening comprehension (e.g., Schmidt-Rinehart, 1994). A series of correlations between the familiarity of a particular topic and the Experimental Comprehension Task performance on the corresponding passage was conducted to explore this relationship. Only 4 of the 12 topics were significantly correlated to performance (see Appendix J).

Participants' education level was significantly correlated with Experimental Comprehension Task performance ($r = .207$, $p = .025$). This medium correlation was expected. Since the passages were about Arts and Science, participants' with higher education levels would have more exposure to these topics. Participants' scores on the English Proficiency test were significantly correlated with the Experimental Comprehension Task scores ($r = .455$, $p < .001$). Specifically, participants who scored high on the English Proficiency test also tended to score high on the passages for the Experimental Comprehension Task. Also, participants' perceived difficulty and their perceived speed for each passage were also significantly correlated ($r = .408$, $p < .001$). This medium correlation suggests that when participants reported a passage as being difficult, they also perceived that the speed of the passage was fast. In addition, participants' medical condition was negatively correlated with Experimental Comprehension Task performance ($r = -.185$, $p = .046$). Fifty participants reported that they have one or more medical conditions. Their medical conditions were coded as ones (at least one medical condition) and zeros (no medical conditions). This negative correlation suggests that participants who reported a

medical condition(s) tended to do less well on the comprehension test. It is important to note that this correlation was very small. There was a significant correlation between age and medical condition ($r = .730$, $p < .001$). This result suggests that older adults tend to report that they have medical condition(s).

The results show that participants' hearing scores and experimental listening comprehension task scores had a medium correlation ($r = .372$, $p < .001$). Since younger adults as a whole had better hearing scores and higher comprehension performance than older adults, the correlation result was not unusual. To investigate whether the hearing score of older EFL speakers differed from older ESL speakers, an independent samples t-test was conducted. There was no significant difference in hearing scores for older EFL speakers ($M = .530$, $SD = .152$) and older ESL speakers ($M = .485$, $SD = .111$), $t (54) = 1.569$, $p = .216$ (two-tailed).

Hypotheses

To test the proposed hypotheses, a 2x2x2x3 mixed-model ANOVA was conducted using listening comprehension accuracy on Experimental Comprehension Task as the DV, with age (young vs. old) and language background (EFL vs. ESL) as the between-groups variables, and speech type (natural vs. synthetic) and speech rate (low vs. medium vs. high) as the within-groups variables. Table 3 presents the means and standard deviations for comprehension accuracy by age group and language background.

Hypothesis 1 predicted a main effect for age. Across all other conditions, it was hypothesized that younger adults would have higher comprehension scores than older adults. As hypothesized, results revealed a significant difference between younger and older adults on the

listening comprehension test, F (1,113) = 20.635, p < .001, partial η^2 = .154, power = .995. Younger adults scored significantly higher on the comprehension test than older adults.

Hypothesis 2 predicted a main effect for language. Across all conditions, EFL participants would have higher comprehension accuracy than ESL participants. Results supported Hypothesis 2. Listening comprehension accuracy was significantly affected by the language background, F (1,113) = 38.975, p < .001, partial η^2 = .256, power = 1.000. EFL speakers scored significantly higher on the comprehension test than ESL speakers.

Hypothesis 3 predicted a two-way interaction between age and language. It suggested that older ESL adults would have significantly lower comprehension accuracy of synthetic speech than older EFL adults. No significant differences were hypothesized on comprehension accuracy between younger ESL and younger EFL adults. More specifically, age was hypothesized to have a stronger effect on synthetic speech comprehension than language background. An interaction effect was found between age and language background, F (1,113) = 31.954, p < .001, partial η^2 = .220, power = 1.000. Young ESL adults' comprehension accuracy of synthetic speech was comparable to young EFL speakers. However, older ESL speakers performed much worse than older EFL speakers on the synthetic speech test portion (see Figure 6).
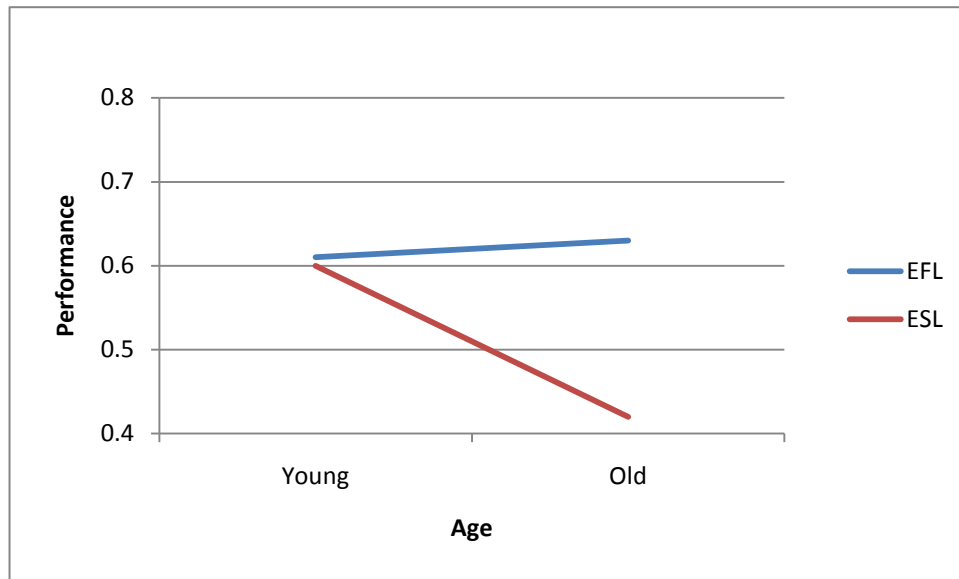
Figure 5. The interaction between age group and language background on comprehension accuracy

Table 3 *Descriptive Statistics – Means and Standard Deviations for Comprehension Accuracy by Age Group and Language Background*

Age Group	Language	Mean	Std Dev	N
Younger	EFL	.605	.093	30
	ESL	.595	.095	28
	Total	.600	.091	58
Older	EFL	.625	.076	30
	ESL	.416	.097	29
	Total	.521	.092	59
Total	EFL	.615	.116	60
	ESL	.506	.113	57

Hypothesis 4: A three-way interaction among age, language, and speech rate was hypothesized. Across language groups, at a low speech rate, older adults would have only slightly worse comprehension accuracy than younger adults. As speech rate increases, however, older adults' comprehension accuracy would exhibit significantly lower comprehension accuracy

than younger adults. The effect of speech rate on age would be greater for ESL participants than for EFL participants. A significant main effect was found for speech rate, Wilks Lambda = .192, F (2, 112) = 234.947, $p < .001$, partial $\eta^2 = .808$, power = 1. Across age and language background, participants scored the highest in the medium speech rate condition ($M = .721$, $SD = .141$), scored the second in the low speech rate condition ($M = .616$, $SD = .151$), and scored the least in the high speech rate condition ($M = .344$, $SD = .141$) (see Table 4).

The data from the mixed-model ANOVA suggested that the hypothesized three-way interaction among age, language, and speech rate was not significant, Wilks' Lambda = .953, F (2,112) = 2.741, $p = .069$, partial $\eta^2 = .047$, power = .532. The data also had a different distribution than predicted in Hypothesis 4. For the EFL speakers, older adults performed only slightly worse than younger adults at a low speech rate. For the medium speech rate, however, older EFL speakers outperformed younger EFL speakers. In the high speech rate condition, both age groups had a dramatic drop in performance. Unlike the performance pattern predicted in Hypothesis 4, older EFL speakers actually performed slightly better than younger EFL speakers in the high speech rate condition (see Figure 6).

For the ESL speakers, the younger adults performed much better across all speech rates than older adults. However, the results did not match the hypothesized pattern. The results suggested that the performance gap between young and older ESL speakers for each speech rate was consistent. Both younger and older ESL speakers performed the best in the medium speech rate condition, the second in the slow speech rate, and the worst in the high speech rate. It is interesting to see that older adults, even the ESL older adults, did not perform well in the slow

speech rate condition (see Figure 7). Descriptive statistics for Hypothesis 4 are presented in

Table 5.
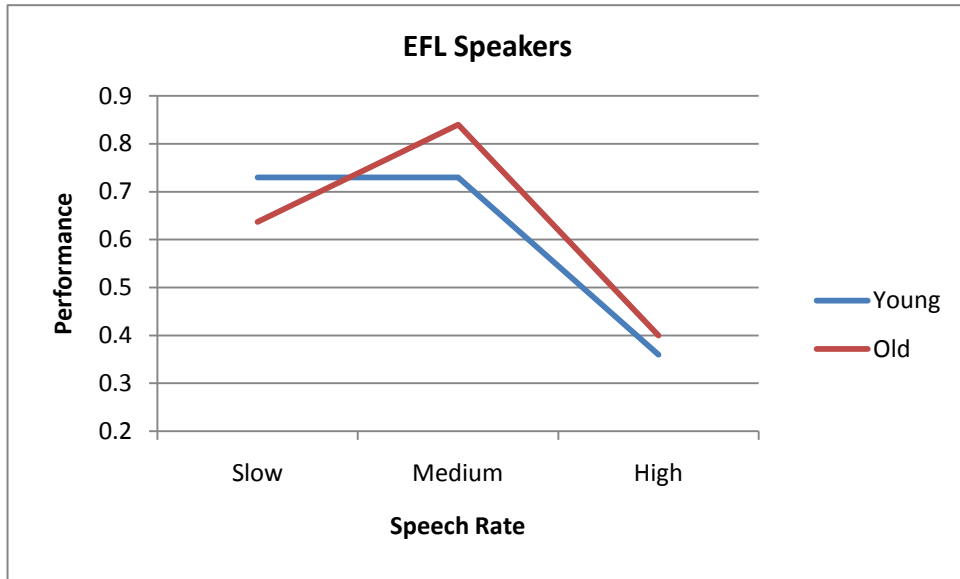
Figure 6. Performance of younger and older EFL speakers across speech rates.
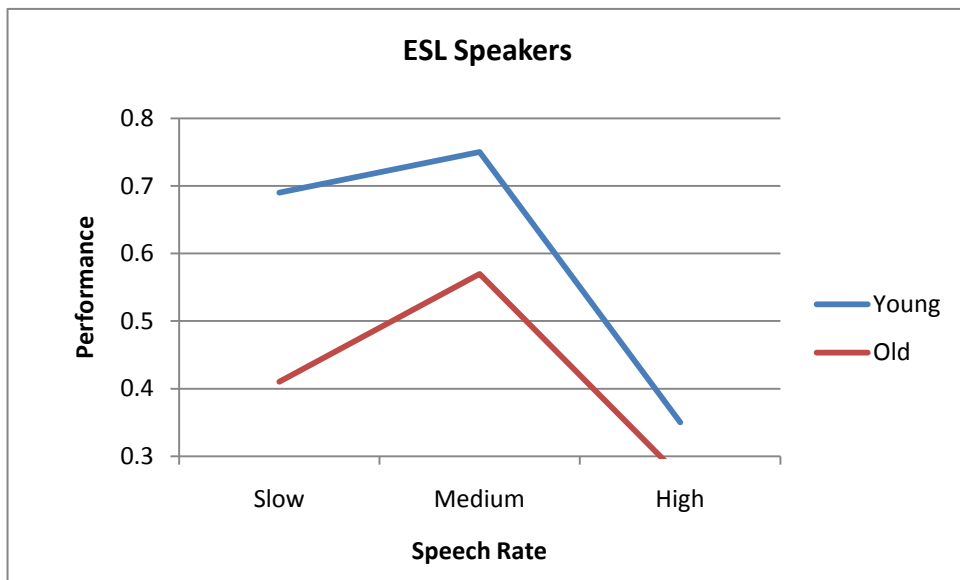
Figure 7. Performance of younger and older ESL speakers across speech rates.

A significant interaction between age and speech rate was found from the mixed-model ANOVA, Wilks' Lambda = .792, F (2,112) = 14.711, $p < .001$, partial $\eta^2 = .208$, power = .999. Across the language condition, at the slow speech rate condition, younger adults scored much higher on comprehension accuracy than older adults. However, the performance gap was much smaller for the medium and high speech rate conditions (Descriptive statistics are presented in Table 5). Post-hoc comparison using the Bonferroni Adjustment was conducted on the interaction of age and speech rate. The results are listed in Table 6. A follow-up independent t-test was conducted on comprehension scores in the slow speech rate condition for older EFL adults and older ESL adults, t (57) = 5.228, p < .001 (two tailed). There was a significant difference in scores for older EFL speakers ($M = .652$, $SD = .142$) and ESL speakers ($M = .652$, $SD = .142$. The magnitude of the differences in means (mean difference = .230, 95% CI: .142 to .319) was very large ($\eta = .324$).
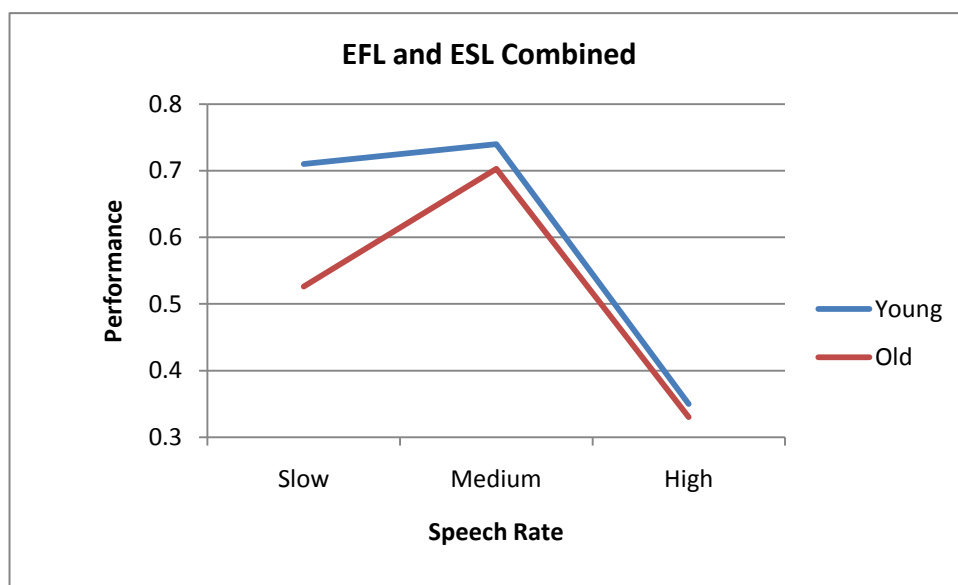
Figure 8. Performance of younger and older adults across speech rates

Table 4 *Means and Standard Deviations for Comprehension Accuracy by Age Group, Language Background, and Speech Rate*

Speech Rate	Age Group	Language	Mean	Std Dev	N
Slow	Younger	EFL	.725	.148	30
		ESL	.687	.148	28
		Total	.706	.145	58
	Older	EFL	.637	.148	30
		ESL	.414	.151	29
		Total	.526	.146	59
Medium	Young	EFL	.733	.142	30
		ESL	.746	.138	28
		Total	.740	.137	58
	Old	EFL	.840	.142	30
		ESL	.566	.140	29
		Total	.703	.138	59
High	Young	EFL	.357	.137	30
		ESL	.351	.138	28
		Total	.354	.137	58
	Old	EFL	.397	.137	30
		ESL	.270	.140	29
		Total	.333	.138	59

Table 5 *Means and Standard Deviations for Comprehension Accuracy by Age Group and Speech Rate*

Age Group	Speech	Mean	Std Dev	N
Younger	Slow	.706	.145	58
	Medium	.740	.137	58
	High	.354	.137	58
Older	Slow	.526	.146	59
	Medium	.703	.138	59
	High	.333	.138	59

Table 6 *Post Hoc Analyses of Age and Speech Rate Interaction*

Age Group	(I) Speech Rate	(J) Speed Rate	Mean Difference (I-J)
Younger	Slow	Medium	-3.36
		High	.352*
	Medium	High	.386*
Older	Slow	Medium	-.177*

		High	.192*
	Medium	High	.370*

Note: *. The mean difference is significant at the .05 level.

Additional Findings

Analysis on Speech Type

A significant main effect was found for speech type, Wilks Lambda = .433, F (1, 113) = 86.237, $p < .001$, partial η^2 = .433, power = 1.000. Across age and language background, participants scored significantly better in the synthetic speech condition than in the natural speech condition (see Table 7).

A significant interaction effect between age and speech type was found, Wilks' Lambda = .901, F (1, 113) = 12.423, $p = .001$, partial η^2 = .099, power = .938 (see Figure 8). Younger adults performed significantly better in the synthetic speech condition than they did in the natural speech condition. Older adults also scored higher in the synthetic speech condition than in the natural speech condition. However, their performance gap was much smaller than younger adults (see Table 7).
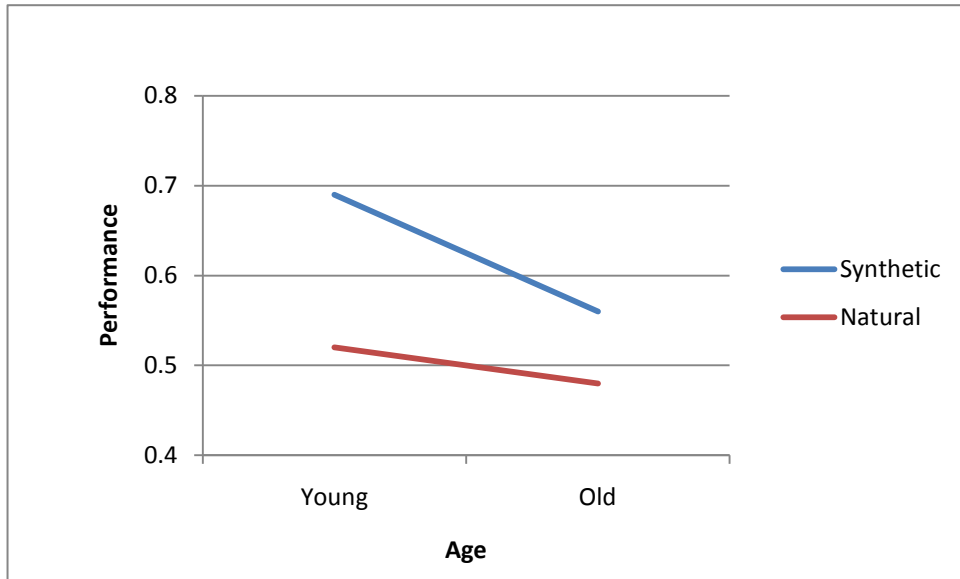
Figure 9. Performance of younger and older adults across speech types.

Table 7 *Means and Standard Deviations for Comprehension Accuracy by Age Group and Speech Type*

Speech Type	Age Group	Mean	Std Dev	N
Synthetic	Younger	.685	.114	30
	Older	.559	.115	28
	Total	.622	.141	58
Natural	Younger	.515	.122	30
	Older	.482	.123	29
	Total	.499	.130	59

CHAPTER FIVE: GENERAL DISCUSSION AND CONCLUSION

General Discussion

In this section, the goal, approach, hypotheses and related findings are examined as to their theoretical and practical implications. Then design implications that can be derived from these findings are discussed. Finally, a cognitive diagnostic tool is proposed along with recommendations for future research.

The goal of this study was to determine whether personal factors of age and language background have an effect on listening comprehension and if so, which would have the larger effect. The approach was to challenge an individual's performance by altering task difficulty with different independent variables including presentations at different speech rates and passages spoken in one of two types of speech, synthetic or natural voice. Both the goal and the approach were successful in establishing significant effects. Figure 8 shows that both chronological age and history of language acquisition have an interactive effect on listening comprehension, with age having the larger direct effect. Figures 8 and 9 show, respectively, that speech rate and speech type have a large effect on listening comprehension, with the speech rates chosen having the largest effect.

Hypotheses Related Findings

Hypothesis 1 Findings

The results of this study revealed that older adults, across all other conditions, scored lower on the listening comprehension test than younger adults. This finding supported Hypothesis 1. Literature has shown that in comparison to younger adults, older adults tend to experience reductions in working memory capacity, attentional difficulties in inhibiting irrelevant information, and slowing in perceptual and cognitive processes (e.g., McCoy et al., 2005). These cognitive changes that are associated with normal aging could affect their ability to comprehend speech.

Hypothesis 2 Findings

The results of this study also supported Hypothesis 2. Non-native English speakers, across all other conditions, had lower performance on the listening comprehension test than native English speakers. The explanation for this performance pattern may be that non-native speakers are generally less proficient with a targeted language than native speakers. ESL speakers may have to translate the English sentence to their native language prior to comprehension. This translation process may cause them to hold information in their working memory longer than EFL speakers, so they can find the counterparts of the English words in their native language. Ardila (2003) used a similar reasoning to explain the different listening comprehension performance between ESL and EFL speakers. He stated that ESL speakers have to identify words and phrases in a second language. This process drains more cognitive resources from working memory.

Hypothesis 3 Findings

The research findings also supported Hypothesis 3. A significant two-way interaction between age and language background was found. Younger and older EFL speakers and younger ESL speakers had similar levels of performance on comprehension accuracy. However, older ESL speakers performed significantly worse than all other groups. This finding suggests that both age and language background have an interactive effect on listening comprehension. Specifically, age seems to have a stronger effect than language background.

Hypothesis 4 Findings

The hypothesized three-way interaction among age, language, and speech rate was not supported by the research findings. First of all, this three-way interaction was not statistically significant, but the result was trending toward significance. By increasing the sample size (e.g., N = 200), statistical significance may be achieved. Second, the interactions in the findings processed different patterns from the hypothesis. For the EFL speakers, older adults showed higher comprehension performance in medium and high speech rates than younger adults. Younger EFL speakers performed better only for the slow speech rate. This performance pattern may be explained by participants' motivation for participating in this study. The majority of the older EFL speakers were recruited from a continuous education program at UCF. This group was very eager to learn and to participate in research. In contrast, younger EFL speakers were required to participate in department research to gain extra course credits. Thus, they might not have been as motivated as older ESL adults to perform well in this experiment. In addition, the experiment passages might not be challenging enough for older EFL speakers. If the experiment was designed to challenge each participant by increasing the difficulty level of stimuli until they

could not get any correct answers, a significant performance difference might be seen between the younger and older EFL adults.

For the ESL speakers, younger adults performed better in all three speech-rate conditions than older adults. It seems that older ESL speakers performed much better in the medium speech rate (188 wpm) than in slow (127 wpm) or high speech (225 wpm) rate. Additionally, by way of comparison, Griffiths (1990) suggested that lower-intermediate ESL speakers performed best at 127 wpm speech rate level. However, the results of this study suggest that older ESL speakers may perform better at 188 wpm rather than 127 wpm.

The hypotheses discussed individually above were predominately supported. However, unlike previous studies (e.g., Duffy & Pisoni, 1992; Mirenda & Beukelman, 1987; Paris, Thomas, Gilson, & Kincaid, 2000), Figure 8 shows that all participant groups performed better in the synthetic speech condition as compared to the natural speech condition. This outcome may be caused by recent technology improvements of synthetic speech generators or synthesizers or possibly by the chance selection of the average difficulty of the passages in this within-subject design. That is, passages that were used in the natural speech condition might be more difficult than in the synthetic speech condition.

With regard to speech rate, in contrast to what is commonly found in the literature, Figure 7 clearly shows that older adults did not benefit from slow speech rate (127 wpm). To the contrary, slow speech rate actually hindered their listening comprehension with strong and significant improvements arriving at the medium speech rate of 188 wpm, regardless of language background (EFL or ESL). It seems reasonable to assume that this was because of age-related working memory deficits rather than losses in processing speed. The reasoning is as follows,

unlike younger listeners, performance for older listeners improved at the medium speech rate as opposed to the slow speech rate. Therefore, this result may implicate working memory because of longer storage times required at the lower speech rates. Further analysis looked at whether older EFL and ESL performed differently in the slow speech rate. The result suggested that older EFL speakers performed significantly better than older ESL speakers. Since older ESL speakers have to use more cognitive resource than older EFL speakers to comprehend audio passages (due to disadvantages associated with language background and age), this result further suggests that slow speech rate may hinder speech comprehension for older adults.

Design Implications

Based on the findings of present study, design implications are listed below. The medium (188 wpm) speech rate may be the most optimal speech rate for all age and language groups. Younger adults (both ESL and EFL) may perform well on both 127 wpm and 188 wpm. In addition, synthetic speech seems to improve listening comprehension performance of all groups. When designing an application or tool that requires listening comprehension for older adults or non-native English speakers, designers could use the recommendations below as guidelines to maximize user performance.

Design for Younger ESL and EFL Speakers

- When designing for younger ESL and EFL speakers, both 127 wpm and 188 wpm (slow and medium speed for this study, respectively) speech rates may be used to ensure comprehension of synthetic voice.

- When designing for younger ESL and EFL speakers, the use of synthetic voice at a 225 wpm (fast speed for this study) speech rate should be avoided as it may result in poor comprehension.

- When designing for younger ESL and EFL speakers, the use of synthetic voice as opposed to natural voice may result in significantly higher comprehension.

Design for Older EFL speakers

- When designing for older EFL speakers, a speech rate of 188 wpm (medium speed for this study) may be used to ensure comprehension of synthetic voice.

- When designing for older EFL speakers, the use of synthetic voice at a speech rate of 127 wpm or a rate of 225 wpm (slow and high speed for this study, respectively) should be avoided as it may result in poor comprehension.

Design for Older ESL speakers

- When designing for older adult ESL speakers, a speech rate of 188 wpm (medium speed for this study) using synthetic voice should be used to optimize listening comprehension performance.

Study Limitations and Future Research Recommendations

The present study collected data on when ESL adults learned English, but did not control for this language variable. In other words, the study did not exclude participants who learned English when they were children. Therefore, some ESL speakers who participated in this study learned English when they were very young (e.g., 3-5 years old) and were highly proficient in the English language. This language variable might have caused some insignificant findings when comparing the comprehension performance of younger ESL and EFL speakers. It would be beneficial for future research to control for when participants learned English or to set a cutoff point to exclude ESL speakers who learned English before a certain age.

The older ESL and EFL speakers who participated in this study might have possessed comparatively different cognitive capabilities. The majority of the older EFL adults were recruited from a continuous education program (LIFE@UCF) offered by the University of Central Florida. Older adults who were members of this program enjoyed learning and participating in research. The older ESL adults, on the other hand, were recruited from the Orange County Senior Center and seemed to be more timid towards participating in research as compared to the LIFE@UCF participants. Some of them expressed that they had not read a science related book since high school or college. By examining the demographic data on education, it is clear that older EFL speakers in general have higher education than the ESL speaker group (see Appendix I, Table 8). For example, 4% of the older EFL speakers were high school graduates. In comparison, 11% of the older ESL speakers reported that the highest education they had completed was high school. This non-language background related difference between the two groups might have caused the comprehension performance gap to be even

larger. Future research should normalize this population by selecting older ESL and EFL adults who are from similar community backgrounds.

The present study used passages to test participants' comprehension accuracy. In the medium speech rate condition, older EFL speakers performed as well as younger EFL speakers (see the discussion section on Hypothesis 4). It is possible that the passages in the medium speech rate may not have been challenging enough for older EFL speakers. Future research could manipulate the passage difficulty levels for the medium speech rate. Specifically, by giving them harder and harder questions, participants should be challenged until they could not get any more corrected answers. This design might magnify the performance difference between younger and older EFL adults.

Proposing a Cognitive Diagnostic Tool

Continuing the earlier discussion associated with listening comprehension and slow speech rates, Figure 7 also suggests an idea for creating a simple diagnostic tool. It may be feasible to use synthetic speech at slow and moderate listening rates as a comparison test to differentially assess cognitive functioning. If there are large differences in comprehension performance with the poorest performance at the slow rate, then this suggests a working memory decline. If there are large differences in comprehension performance with the poorest performance at the moderate rate, then this suggests a processing speed decline. Conversely, if comprehension performance is good and steady at both slow and moderate speech rates, then this suggests generally normal functioning in both working memory and processing speed.

Small et al.'s (1997) findings seemed to support this idea. They studied working memory capacity of three Alzheimer's patients by measuring their listening comprehension using two speech rates: slow (155 wpm) and normal (188 wpm). They found that the effect of speech rate for comprehension performance was determined by the extent of working memory capacity. Patients with the most severe working memory impairment performed the worst on the slow speech rate. Speech rate did not affect comprehension for patients with moderate impaired working memory. Only patients with the most preserved working memory benefited from the slow speech rate. Thus, based on Small's work and the findings of the present study with its much larger sample size to draw conclusions, it appears that people with severe working memory decline would perform poorly on a comprehension test at slow speech delivery rates. People who have no or minimum working memory decline would perform relatively the same in the slow or

medium speech rates. Of course, this idea is only speculation at this point. Further research involving support and replication of these findings, as well as investigating other cognitive declines similarly associated with speech delivery rate is required to take such a diagnostic concept from a validated concept to an accepted instrument.

Conclusion

This study explored the ability of older and younger adults to comprehend English synthetic speech, including those who had learned English as their first language (EFL) and those who had learned English as their second language (ESL). Presentations of passages were made at three different speech rates (slow, medium, and high) with natural speech as a control condition. The findings indicated that overall, younger adults scored higher than older adults in listening comprehension; with EFL listeners performing significantly better than ESL listeners for both age groups. As predicted, older ESL adults had the most difficulty comprehending English synthetic speech. However, the differences between older and younger listeners diminished as speech rate increased, with older EFL listeners actually outperforming younger EFL listeners at the medium speech rates of 188 wpm. This suggests that performance deficits are more related to age-related working memory losses than to losses in processing speed. The finding that participants generally performed better in the synthetic speech condition than the natural speech condition may be a reflection of recent technological improvements to synthetic speech generators; however, the chance selection of more difficult passages with natural speech cannot be ruled out. Both theoretical implications and design recommendations are included,

along with a proposed cognitive diagnostic tool to measure working memory capacity as a suggestion for future research.

APPENDIX A: POWER ANALYSIS

Power Analysis

F tests - ANOVA: Repeated measures, between factors

Analysis: A priori: Compute required sample size

Input: Effect size f   = 0.25
            α err prob   = 0.05
            Power (1-β err prob)   = 0.80
            Number of groups   = 4
            Repetitions   = 5
            Corr among rep measures   = 0.5

Output: Noncentrality parameter λ   = 11.6666667
            Critical F   = 2.6886915
            Numerator df   = 3.0000000
            Denominator df   = 108
            Total sample size   = **112**
            Actual power   = 0.8136020

Central and noncentral distributions | Protocol of power analyses

critical F = 2.6887
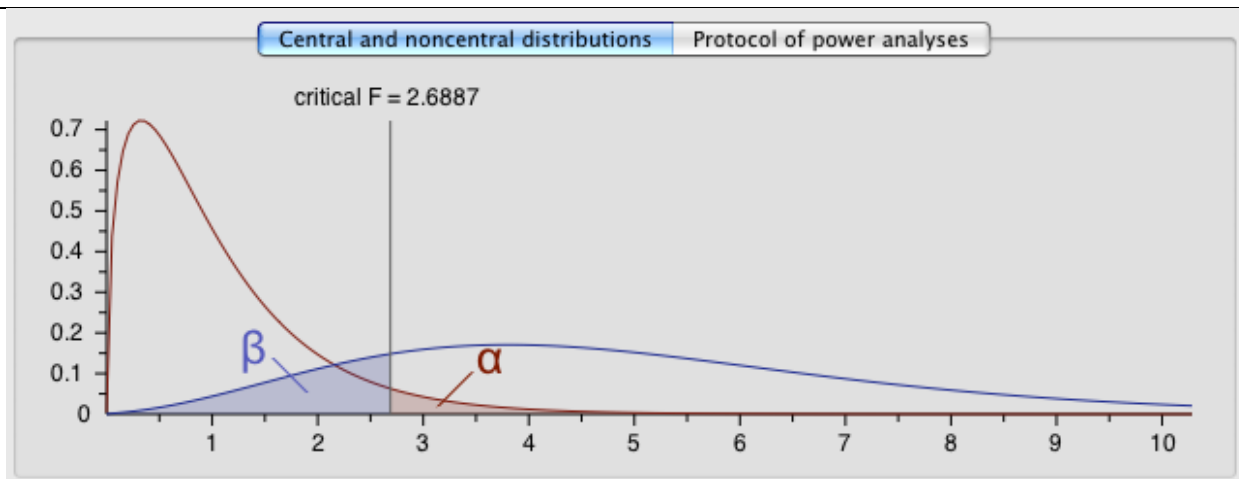
APPENDIX B: INFORMED CONSENT

Effects of Age and Language Proficiency on the Comprehension of Synthetic Speech

Informed Consent

Principal Investigator(s):   *Jingjing Wang Costello, M.S.*

Faculty Supervisor: *Richard Gilson, Ph.D.*

Investigational Site(s):   *University of Central Florida, Department of Psychology*

Introduction: Researchers at the University of Central Florida (UCF) study many topics. To do this we need the help of people who agree to take part in a research study. You are being invited to take part in a research study which will include about 120 people at the University of Central Florida. You have been asked to take part in this research study because you are either a native English speaker or a non-native English speaker who is either between 18 to 25 years old or 65 to 80 years old. You must be 18 years of age or older to be included in the research study.

The person doing this research is Mrs. Jingjing Wang Costello from the Univeristy of Central Florida, Department of Psychology. Because the researcher is a graduate student, she is being guided by Dr. Richard Gilson who is a UCF faculty supervisor in the Psychology Department.

What you should know about a research study:
- Someone will explain this research study to you.
- A research study is something you volunteer for.
- Whether or not you take part is up to you.
- You should take part in this study only because you want to.
- You can choose not to take part in the research study.
- You can agree to take part now and later change your mind.
- Whatever you decide it will not be held against you.
- Feel free to ask all the questions you want before you decide.

Purpose of the research study: The purpose of this study is to understand how people with English as a *second* language are able to understand computer-generated English language speech compared to people with English as their native language. Since computer-generated speech, also known as synthetic speech has been widely used in various applications such as

Global Positioning Systems (GPS), warning systems in aircraft cockpits, etc., this study is designed to explore the listening comprehension performance of old and younger native English and non-native English listeners when listening to synthetic audio passages.

What you will be asked to do in the study: Following an informal briefing about the experiment, you will be asked if you are comfortable to proceed. If you are comfortable, you will be asked to perform several tasks. First, you will be instructed to fill out a demographics questionnaire. You will be randomly assigned to a short or a long version of this experiment. The short version will take approximately 60 minutes and the long version will take approximately 90 minutes. Next, you will conduct a computer-based listening task using a headset. For the third task, you will listen to several English conversations and speech using a headset. You will be asked to answer several multiple-choice questions after each conversation or speech. The following task is called a shadowing task. Several computer-generated sentences will be play to you. You will be instructed to repeat the sentences back to the experimenter word by word. Lastly, you will listen to some computer-generated passages and some human-generated passages through a headset. You will be asked to answer two to three multiple-choice questions regarding the content of the passage that you have just heard. The passages will be presented at various speech rates. Once you complete the multiple-choice questions of each passage, you will be asked to rate the difficulty level and the speed of the passages. Please answer each question to your best ability, but you do not have to answer every question or complete every task. You will not lose any benefits if you skip questions or tasks.

Location: Volunteer participation in this research project will take place in the UCF Psychology Building, located in Room 303J.

Time required: We expect that you will be in this research study for only one session, which last approximately 90 minutes (1.5 hours).

Risks: There are no reasonably foreseeable risks or discomforts involved in taking part in this study.

Benefits: There are no expected benefits to you for taking part in this study.

Compensation or payment: Participants may expect to spend approximately 60-90 minutes performing experimental tasks, for which they may elect to receive either course credit for the amount of time they participate, or, if not participating for course credit, cash payment at a rate of $10.00 per hour. Maximum course credit will be 90 minutes, while maximum cash credit will be $15.00.

For course credits, if you choose not to participate, you may notify your instructor and ask for an alternative assignment of equal effort for equal credit. There will be no penalty.

Confidentiality: Your identity will be kept confidential. Your name will not be used in any report. The recorded data will be assigned a code number. All documents, including the informed consent documents, the paper-based pre/post-test questionnaires and a list correlating participant names and code numbers will be stored in a locked cabinet separate from all other study documents for a minimum of three years, after which the information will be destroyed.

Study contact for questions about the study or to report a problem: If you have questions, concerns, or complaints, or think the research has hurt you, talk to Mrs. Jingjing Wang Costello, Graduate Student, Experimental Psychology and Human Factors, Psychology Department, (407)-968-3330 or Dr. Richard Gilson, Faculty Supervisor, Department of Psychology at (407)-823-2755 or by email at wangjjj@gmail.com.

IRB contact about your rights in the study or to report a complaint: Research at the University of Central Florida involving human participants is carried out under the oversight of the Institutional Review Board (UCF IRB). This research has been reviewed and approved by the IRB. For information about the rights of people who take part in research, please contact: Institutional Review Board, University of Central Florida, Office of Research & Commercialization, 12201 Research Parkway, Suite 501, Orlando, FL 32826-3246 or by telephone at (407) 823-2901. You may also talk to them for any of the following:
- Your questions, concerns, or complaints are not being answered by the research team.
- You cannot reach the research team.
- You want to talk to someone besides the research team.
- You want to get information or provide input about this research.

APPENDIX C: DEMOGRAPHIC QUESTIONNAIRE

Demographic Questionnaire

Instructions: Please complete the following questions. Any information you provide is voluntary and will be kept strictly confidential. A participant number will be assigned to your responses and in no way will your name be associated with the data. The information you provide will be used only for the purpose of this study.

1. What is your age? _____

2. What is your gender? ☐ Female ☐ Male

3. Please indicate the highest education level you have completed. Check only one box.
    ☐ High school or equivalent   ☐ Graduate degree
    ☐ Some college   ☐ Other (please specify)
                                         _____

    ☐ Bachelor degree

4. How would you classify yourself? Check only one box.
    ☐ Caucasian/White   ☐ Hispanic (white)   ☐ Multiracial
    ☐ Black   ☐ Hispanic (non-white)   ☐ Other
                                                       _____
    ☐ Asian/Pacific Islander ☐ Arab   ☐ Would rather not say

5. Do you use a hearing aid?   ☐ YES   ☐ NO
If you answered YES, please explain what kind of hearing issue(s) you have:

_____
_____

6. How often do you experience difficulties hearing people on the phone? Please circle the
    number that represents you the most.

1	2	3	4	5
Never	Sometimes	Frequently	Almost Always	Always

If you circled 3 or higher, please indicate why you think you experience difficulties. Check all that apply.
☐ Background noise in the phone
☐ Voice on the other end is too low
☐ Other, please explain:

_____

7. Is English your native language?   ☐ YES   ☐ NO
If you answered YES, please skip to Question 8.

If you answered NO, please answer Questions 7a to 7d.

   7a. What is your native language? _____

   7b. How many years have you used or practiced the English language? _____

   7c. Have you ever taken a class to learn English? ☐ YES ☐ NO

   7d. Have you ever taken a Test of English as a Foreign Language (TOEFL)? ☐ YES ☐ NO

  8. How often have you interacted with computer- or machine-generated speech applications? Examples include: Global Positioning Systems (GPS); text-to-speech function on a computer; E-reader; automatic customer service systems. Please circle the number that represents you the most.

1	2	3	4	5
Never	Rarely	Sometimes	Often	Very Often

9. Please indicate if you have been diagnosed or are currently taking medications for any of the following medical conditions. Check all that apply.

☐ Diabetes	☐ Stroke	☐ Cancer
☐ Heart disease	☐ Hearing impairment	☐ Other
		_____
☐ High blood pressure	☐ Insomnia	☐ Would rather not say

10. How many hours of sleep did you get last night? _____

11. Using the 7-point scale below, please rate your familiarity on the following topic areas by writing the number that represents you the most in the space next to each topic.

1	2	3	4	5	6	7
Not At All Familiar			Somewhat Familiar			Very Familiar

A) Biology _____

B) Anatomy _____

C) English Literature _____

D) Astronomy _____

E) Aviation _____

F) Zoology _____

G) Human Anatomy _____

H) Current Events_____

I) Medical devices_____

J) Medical science_____

K) Psychology_____

L) Geography_____

APPENDIX D: ENGLISH PROFICIENCY TEST HANDOUT

English Proficiency of Listening Comprehension Test – Handout Portion

The following content was adopted from Educational Testing Service (1990) TOEFL sample test, 4th edition.

PART A

<u>Directions</u>: For each question in Part A, you will hear a short sentence. Each sentence will be spoken just one time. The sentences you hear will not be written out for you. Therefore, you must listen carefully to understand what the speaker says.

After you hear a sentence, read the four choices in your test book, marked (A), (B), (C), and (D), and decide which one is closest in meaning to the sentence you heard. Then circle the letter of the answer you have chosen.

Example:

      You will hear: (play audio)
      You will read: (A) Mary outswam the others.
                     (B) Mary ought to swim with them.
                     (C) Mary and her friends swam to the island.
                     (D) Mary's friends owned the island.

      The speaker said "Mary swam out to the island with her friends." Sentence (C), "Mary and her friends swam to the island", is the closest in meaning to the sentence you heard. Therefore, you should choose answer (C).

PART A Test questions start here:

      1. (A) Didn't you have two appointments?
         (B) I assumed you called to make an appointment.
         (C) Did you call them all today?
         (D) I didn't call you up today.

      2. (A) Ken's doctor is old, but he couldn't be more active.
         (B) Ken should learn to play tennis better.
         (C) Ken's doctor wants him to get more exercise
         (D) Ken has to stop playing tennis.

      3. (A) Susan learned German without taking lessons.
         (B) Susan did her German homework in school.
         (C) Susan's German is poor because she never practices it.
         (D) Susan went to Germany to study the language.

      4. (A) How old do you think he is?
         (B) He thought you were very young.
         (C) He looks younger than he really is.

(D) Is he older than you are?
5. (A) May I please borrow your car?
    (B) I'm really grateful you lent me your car.
    (C) I realize you won't let me drive your car.
    (D) Could you tell me how much the car has depreciated?

PART B

Directions: In Part B you will hear short conversations between two speakers. At the end of each conversation, a third person will ask a question about what was said. You will hear each conversation and question about it just one time. Therefore, you must listen carefully to understand what each speaker says. After you hear a conversation and the question about it, read the four possible answers and decide which one is the best answer to the question you heard. Then, circle the letter of the answer you have chosen.

Example:

You will hear: (play audio)
You will read: (A) Present Professor Smith with a picture.
                (B) Photograph Professor Smith.
                (C) Put glass over the photograph.
                (D) Replace the broken headlight.
From the conversation you learn that the women thinks Professor Smith would like a photograph of the class. The best answer to the question "What does the woman think the class should do?" is (A), "Present Professor Smith with a picture." Therefore, you should choose answer (A).

PART B Test questions start here:

6. (A) The desk doesn't need cleaning.
    (B) There is a hole in the desk.
    (C) There isn't another room.
    (D) The room is very messy.

7. (A) He' also interested in seeing it.
    (B) It's his own work.
    (C) He wanted to get it for himself.
    (D) It looks like him.

8. (A) The sale was not advertised.
    (B) She wants to buy some read jeans.
    (C) She doesn't know which sales tag to read
    (D) The jeans with the red tag are the ones on sale

9. (A) Erase his entire document.
   (B) Put his work on another computer.
   (C) Go back to the very beginning of his work.
   (D) Find someway to restore power to the computer.

PART C

Example:

> You will hear: (play audio)
> You will read: (A) They are impossible to guide.
>   (B) They may go up in flames.
>   (C) They tend to leak gas.
>   (D) They are cheaply made.

The best answer to the question "Why are gas balloons considered dangerous?" is (B), "They may go up in flames." Therefore, you should choose answer (B).

PART C Test questions start here:

10. (A) Nineteenth-century political activists.
    (B) The work of Clara Barton.
    (C) A comparison of Clara Barton and Florence Nightingale.
    (D) The ratification of the Geneva Treaty.

11. (A) After hearing political speeches.
    (B) While studying in medical school.
    (C) During the Civil War.
    (D) After the Geneva Treaty.

12. (A) Zoologist.
    (B) Hospital nurse.
    (C) Military adviser.

(D) Schoolteacher.

13. (A) She became famous in her own lifetime.
    (B) She lived according to her beliefs.
    (C) She was a talented storyteller.
     (D) She was a fictional character.

APPENDIX E: ENGLISH PROFICIENCY TEST HANDOUT

English Proficiency of Listening Comprehension Test – Transcribed Audio Script

The following content was adopted from Educational Testing Service (1990) TOEFL Sample Test, 4th Edition.

Part A

Question 1: You called for an appointment, didn't you?
Question 2: The doctor told Ken he shouldn't play tennis anymore.
Question 3: Susan has never studied German but speaks fluently.
Question 4: He is older than you think.
Question 5: If you could lend me your car, I would really appreciate.

Part B

(Men) We really need to clean up this desk today.
(Women) Why not the whole room?
Question 6: What does the women imply?

(Women) What an interesting portrait!
(Men) Thank you! I painted it myself.
Question 7: What does the man say about the portrait?

(Men) Excuse me, are these the jeans on sale for 20 dollars?
(Women) Do they have a red sales tag?
Question 8: What does the women imply?

(Women) Did the power failure affect the work you were doing on the computer?
(Men) It sure did. It erased my whole document and now I have to start it again from scratch.
Question 9: What does the man have to do?

Part C

As a follow up to our discussion on the political activists of this country, today I want to talk about a very interesting woman, Clara Barton, one of the most determined 19th century political activists in the United States. She found the American Red Cross, but strangely enough, she was originally a schoolteacher and a writer not a hospital nurse like her famous contemporary Florence Nightingale.

During the Civil War, Clara Barton turned to nursing in order to help the thousands of wounded solders receiving little or no medical care in the understaffed military hospitals. She openly defied military authorities who did not allow female nurses on the battlefield maintaining the

wounded soldiers who often died because of lack of food, clothing, or simple medication and bandages, not just medical problems that required medical doctor's expertise.

Even though it was against orders, she followed the troops and helped medical personnel with vital supply she gathered herself. She became a living legend and the soldiers called her the "angel of battlefield".

After the war, she worked for the Ratification by United States of Geneva Treaty, which officially established the International Red Cross. She also served as the first president of the American Red Cross for over 20 years and it was she who convinced that organization to extend the relieve efforts to victims of natural disaster as well as victims at war.

Question 10: What is the main subject of the talk?

Question 11: When did Clara Barton first involved in medical relieve?

Question 12: What was Clara Barton's original profession?

Question 13: What does the speaker mean by referring to Clara Barton as a living legend?

Answer Key

Q1: B
Q2: D
Q3: A
Q4: C
Q5: A
Q6: D
Q7: B
Q8: D
Q9: C
Q10: B
Q11: C
Q12: D
Q13: A

APPENDIX F: EXPERIMENTAL TASK QUESTIONS HANDOUT

Experimental Task Questions – Handout Portion

The following content was adopted from Rymniak, Shanks, et al., 2002, TOEFL Workbook, 2nd Edition.

Section 1: Listening Training

Directions: In this section, you will listen to two training audio passage. One passage was recorded using human voice. One was generated by a computer. After each audio passage is over, you will answer one question regarding the passage.

Passage 1: Computer generated passage

Where does the Alaska pipeline start?
    A. The frozen edge of the Arctic Ocean
    B. The ice-free seaport village
    C. Glacier bay
    D. None of the above

Passage 2: Human generated passage

The passage implies that New York
    A. Has lost its pre-eminent place in American society
    B. Has permanently escaped the threat of bankruptcy
    C. Has managed to avoid asking the national government for assistance
    D. Enjoyed a period of unusual economic strength prior to 1970

Section 2: Experimental Speech Listening Task

Directions: In this section, you will hear 12 short passages. Each passage is followed by several questions about it. . Half of these passages are recorded using human voice and half are generated by a computer voice. These passages will be presented at various speeds. Some are at the same speed as the training passages. Some will be slower than or faster than the training passage. Each question in this part has four answer choices. You should circle the best answer to each question. Answer the questions on the basis of what is stated or implied by the speakers.

Computer-generated Speech

Passage 1:

1. What is the main purpose of the program?
    A. To demonstrate the latest use of computer graphics.

B. To discuss the possibility of an economic depression.
C. To explain the working of the brain.
D. To dramatize a famous mystery story.

2. Why does the speaker recommend watching the program?
   A. It is required of all science majors.
   B. It will never be shown again.
   C. It can help views improve their memory skills.
   D. It will help with course work.

Passage 2:

1. What is the speaker's main point?
   A. Animal yawn for a number of reasons.
   B. Yawning results only from fatigue or boredom.
   C. Human yawns are the same as those of other animals.
   D. Only social animals yawn.

2. According to the speaker, when are hippos likely to yawn?
   A. When they are swimming.
   B. When they are quarreling.
   C. When they are socializing.
   D. When they are eating.

3. What physiological reason for yawning is mentioned?
   A. To exercise the jaw muscles.
   B. To eliminate fatigue.
   C. To get greater strength for attaching.
   D. To gain more oxygen.

Passage 3:

1. What theme did Hemingway use for many of his books?
   A. War
   B. Romance
   C. Travel
   D. Sports

2. What was the Hemingway style?
   A. Long descriptions
   B. Imaginative details
   C. Short sentences
   D. Difficult symbolism

3. What advice would Hemingway probably give to other writers?
    A. Write for a newspaper before you begin writing novels
    B. Create your own style of literature
    C. Write from experience about things you have seen and people you have known
    D. Travel in order to meet interesting people

Passage 4:

1. In what course is this lecture probably given?
    A. Philosophy
    B. Meteorology
    C. Astronomy
    D. Photography

2. According to the speaker, which of the following occurs during a lunar eclipse?
    A. The Earth's shadow moves across the Moon.
    B. Clouds block the view of the Moon.
    C. The Moon moves between the Earth and the Sun.
    D. The Sun is too bright to be observed without special equipment.

Passage 5:

1. What is the main topic of the talk?
    A. Energy conservation.
    B. New kind of transportation.
    C. Strip cities.
    D. Advantages of air transportation over railroads.

2. When are airplanes not fuel-efficient?
    A. On short trips.
    B. On long trips.
    C. When flying over cities.
    D. When flying at high altitudes.

Passage 6:

1. What is the main topic of this passage?
    A. Birds that live in colonies
    B. How birds defend their territory
    C. The behavior of birds
    D. Territoriality in birds

2. According to the passage, male birds defend their territories primarily against
    A. Female birds

B. Birds of other species
C. Male of their own species
D. Carnivorous mammals

3. It can be inferred from the passage that gulls and penguins
   A. Do not claim a feeding area as part of their territories
   B. Share their territories with many other birds
   C. Leave their colonies during their nesting season
   D. Do not build nest

Human generated Speech

Passage 7:
1. According to the passage, the human liver is composed of how many lobes?
   A. 2
   B. 4
   C. 5
   D. 50,000 to 100,000

2. According to the passage, the most important work within the lobule is
   A. Performed by a central vein
   B. Inside the sinusoids
   C. The work done by the lobes
   D. Done by the sheets

Passage 8:

   1. What is the author's main purpose?
      A. To describe a new cure for ear infections
      B. To inform the reader of a new device
      C. To urge doctors to use a new device
      D. To explain the use of a magnet

   2. The word "relief" in the last sentence means:
      A. Less distress
      B. Assistance
      C. Distraction
      D. Relaxation

Passage 9:
1. The nuclear accident described in the movie
   A. Was successfully concealed by power industry leaders and officials
   B. Was caused by a series of coincidences

C. Was a surprisingly accurate foreshadowing of actual events

D. Took place at the Three Mile Island

2. Officials of the nuclear power industry in real life
   A. Have committed murders to make possible a cover-up of the incident at Harrisburg
   B. Had predicted that nuclear accidents were likely to occur
   C. Have been reluctant to reveal the full story about the Three Mile Island incident
   D. Have tried to make all the facts freely accessible to those concerned

3. According to the passage, public concern over the accident near Harrisburg
   A. Had no effect on the subsequent investigation
   B. Was lessened by the quick response of industry leaders and officials
   C. Prompted widespread panic throughout Pennsylvania
   D. Persisted at many questions were left unanswered

Passage 10:

1. What is the main topic of this reading?
   A. Women and drugs
   B. The dangers of pregnancy
   C. The fetus and alcohol
   D. Drinking and the human body

2. How much time can it be inferred that it takes alcohol to enter a woman's bloodstream after she takes a drink?
   A. About one hour
   B. A few seconds
   C. Several minutes
   D. At least 24 hours

Passage 11:

1. It can be inferred from the passage that the matching process in visual recognition is
   A. Not a natural activity
   B. Not possible when an object is viewed for the first time
   C. Not possible if a feature of a familiar object is changed in some way
   D. Only possible when a retinal image is received in the brain as a unitary whole
   E. Now fully understood as a combination of the serial and parallel process

2. In terms of its tone and form, the passage can best be characterized as
   A. A biased exposition
   B. A speculative study
   C. A dispassionate presentation
   D. An indignant denial
   E. A dogmatic explanation

Passage 12:

1. The peripheral furrows or deeps are found
    - A. Only in the Pacific and Indian oceans
    - B. Near earthquakes
    - C. Near the shore
    - D. In the center of the ocean
    - E. To be 14,000 feet in depth in the pacific

2. The largest ocean is the
    - A. Atlantic
    - B. Pacific
    - C. Aleutian deep
    - D. Arctic
    - E. Indian

APPENDIX G: EXPERIMENTAL TASK PASSAGES

Experimental Task Passages

The following content was adopted from Rymniak, Shanks, et al., 2002, TOEFL Workbook, 2nd Edition.

Training Passages

Training passage 1:
The Alaska pipeline starts at the frozen edge of the Arctic Ocean. It stretches southward across the largest and northernmost state in the United States, ending at a remote ice-free seaport village nearly 80 miles from where it begins. It is massive in size and extremely complicated to operate.

Training passage 2:
New York successfully averted bankruptcy in the mid-1970s through the creation of the Municipal Assistance Corporation, which issued nearly $2 billion in special bonds to help finance city spending, and through a loan program offered by the federal government later in the 1975. But the essential problems of the city are still present and continue to play a role in how the city is managed, even today. In the 1970s, income from taxation was limited by the shrinking of New York's middle-class population, which resulted in revenues inadequate for supporting the vast expensive city government. A number of major businesses left the city, taking with them jobs, income, and more tax revenues. The wealth of New York during this financial crisis slowly dwindled.

Realizing that the remarkable prosperity of the 1950s and 1960s had faded, New York's leaders had to closely reexamine the city's financial position. As a first step, the city's wasteful practices had to be halted, for only a new era of thrifty and efficient government could enable New York to retain its position as the economic, social, and cultural capital of the United States.

Experimental Task Passages

Computer-generated passages

Passage 1:
I would like to tell you about an interesting TV program that will be shown this coming Thursday. It'll be on from 9 to 10 p.m. on Channel 4. It's part of a series called "Mysteries of Human Biology." The subject of the program is the human brain — how it functions and how it can malfunction. Topics that will be covered are dreams, memory, and depression. These topics are illustrated with outstanding computer animation that makes the explanations easy to follow. Make an effort to see this show. Since we've been studying the nervous system in class, I know you'll find it very helpful.

Passage 2:
This discussion is about a common animal reaction — the yawn. The dictionary defines a yawn as "an involuntary reaction to fatigue or boredom." That's certainly true for human yawns, but not necessarily for animal yawns. The same action can have quite different meanings in different species.

For example, some animals yawn to intimidate intruders on their territory. Fish and lizards are examples of this. Hippos use yawns when they want to settle a quarrel. Observers have seen two hippos yawn at each other for as long as two hours before they stop quarreling.

As for social animals like baboons or lions — they yawn to establish the pecking order within social groups, and lions often yawn to calm social tensions. Sometimes these animals yawn for a strictly physiological reason — that is, to increase oxygen levels. And curiously enough, when they yawn for a physical reason like that, they do what humans do — they try to stifle the yawn by looking away or by covering their mouths.

Passage 3:
Ernest Hemingway began his writing career as an ambitious young American newspaperman in Paris after the First World War. His early books, including The Sun Also Rises, were published in Europe before they were released in the United States.

Hemingway always wrote from experience rather than from imagination. In Farewell to Arms, published in 1929, he recounted his adventures as an ambulance driver in Italy during the war. In For whom the Bell Tolls, published in 1940, he retold his memories of the Spanish Civil War.

Perhaps more than any other twentieth-century American writer, he was responsible for creating a style of literature. The Hemingway style was hard, economical, and powerful. It lured the reader into using imagination in order to fill in the details.

Passage 4:
Students, this evening we'll have a chance to observe a phenomenon that we've discussed several times in class. Tonight there will be a lunar eclipse. As we've said, when an eclipse of the Moon occurs, the Earth passes between the Sun and the Moon. Therefore, the shadow of the Earth moves across the surface of the Moon and obscures it. Because you won't be looking at the Sun, it is not necessary to use the special lenses and filters that you need when observing a solar eclipse. You can observe a lunar eclipse with your unaided eye or with a telescope and photograph it with an ordinary camera. So if the weather's not cloudy tonight, go out and take a look at this eclipse of the Moon. I'm sure you will find it interesting.

Passage 5:
Although I think the United States generally has an excellent system of transportation, I do

not think that it does a good job of transporting people between cities that are only a few hundred miles apart. A person commuting between Detroit and Chicago, or between San Francisco and Los Angeles, so-called strip cities, may spend only a relatively short time in the air while spending several hours getting to and from the airport. This situation makes flying almost as time-consuming as driving. Moreover, airplanes use a lot of their fuel just getting into the air. They simply are not fuel-efficient on short trips. High-speed trains may be an answer. One fairly new proposal for such a train is for something called a "maglev," meaning a magnetically levitated train. Maglevs will not actually <u>ride</u> on the tracks, but will <u>fly</u> above tracks that are magnetically activated. This will save wear and tear on the tracks. These trains will go faster than one hundred fifty miles per hour — at that speed, conventional trains have trouble staying on the tracks. As you can see, maglevs offer exciting possibilities for the future.

Passage 6:
A bird's territory may be small or large. Some male claim only their nest and the area right around it, while others claim far larger territories that include their feeding area. Gulls, penguins, and other waterfowl nest in huge colonies, but even in the biggest colonies, each male and his mate have small territories of their own immediately around their nest. Males defend their territory chiefly against other males of the same species. In some cases, a warning call or threatening pose may be all the defense needed, but in other cases, intruders may refuse to leave peacefully.

Human-generated passages

Passage 7:

The human liver is composed of three parts: the right lobe (the largest), the left lobe, and two small lobes that are located behind the right lobe. Each lobe is composed of small, multisided units called lobules. Most livers have between 50,000 and 100,000 lobules. The essential work in each lobule is performed by a bundle of liver cells that surround the central vein found in each lobule. These bundles, often called sheets, are separated by cavities known as sinusoids. It is the presence of these sinusoids that accounts for the liver's somewhat spongy texture and its ability to absorb large amounts of blood.

Passage 8:
A new hearing device is now available for some hearing-impaired people. This device uses a magnet to hold the detachable sound-processing portion in place. Like other aids, it converts sound into vibrations. But it is unique in that it can transmit the vibrations directly to the magnet and then to the inner ear. This produces a clearer sound. The new device will not help all hearing-impaired people only those with a hearing loss caused by infection or some other problem in the middle ear. It will probably help no more than 20 percent of all people with hearing problems. Those people who have persistent ear infections, however, should find relief and restored hearing with the new device.

Passage 9:

A highly acclaimed motion picture of 1979 concerned a nearly disastrous accident at a nuclear power plant. Within a few weeks of the film's release, in a chilling coincidence, a real-life accident startlingly similar to the fictitious one occurred at the Three Mile Island plant near Harrisburg, Pennsylvania. The two incidents even corresponded in certain details; for instance, both in the film and in real life, one cause of the mishap was a false meter reading caused by a jammed needle.

Such similarities led many to wonder whether the fictional movie plot had been prophetic in other ways. The movie depicted officials of the power industry as seriously corrupt, willing to lie, bribe, and even kill to conceal their culpability in the accident. Did a similar cover up occur in the Three Mile Island accidents? Perhaps we will never know. We do know that, despite the endeavors of reporters and citizen groups to uncover the causes of the accident, many of the facts remain unknown. Although they declare that the public is entitled to the truth, many of the power industry leaders responsible have been reluctant to cooperate with independent, impartial investigators.

Passage 10:

One of the most dangerous drugs for pregnant women to consume is   alcohol. Because alcohol is delivered quickly into the blood and passes   quickly into the tissues and membranes, the human fetus is particularly   vulnerable to its effects. In fact, the negative effects on a fetus are so   pronounced that babies born after exposure to alcohol are said to be   suffering from fetal alcohol syndrome.   As a pregnant woman drinks alcohol, the alcohol is passed into her   her bloodstream almost simultaneously. Moreover, because the   bloodstream of the fetus is inextricably tied to that of the mother,   the alcohol passes directly into the bloodstream of the fetus as well.   And, what is more, the concentration of alcohol in the fetus is exactly   the same as in the mother.   For the mother, this concentration is not a problem because her liver   can remove one ounce of alcohol from her system per hour. However,   the fetus's liver is not completely developed (how developed it is   depends on its stage of development). The rate at which it is able   to eliminate the alcohol from the blood of the fetus is much slower.

   Eventually, the alcohol will be returned to the mother's system by   passing across the placenta, but this process is slow. By the time   this takes place, major neurological damage may have already   occurred. Research has shown that as little as one drink of   alcohol can produce significant, irreversible damage to the fetus.   Babies born after exposure to alcohol generally exhibit facial   distortion, inability to concentrate, and difficulty in remembering.   Simply speaking, it is imperative that pregnant women avoid alcohol.

Passage 11:

Visual recognition involves storing and retrieving memories. Neural activity, triggered by the eye, forms an image in the brains memory system that constitutes an internal representation of the viewed object. When an object is encountered again, it is matched with its internal representation and thereby recognized. Controversy surrounds the question of whether

recognition is a parallel, one-step process or a serial, step-by-step one. Psychologists of the Gestalt school maintain that object are recognized as wholes in a parallel procedure: the internal representation is matched with the retinal image in a single operation. Other psychologists have proposed that internal representation features are matched serially with an object's features. Although some experiments show that, as an object become familiar, its internal representation becomes more familiar, its internal representation becomes more holistic and the recognition process correspondingly more parallel, the weight of evidence seems to support the serial hypothesis, at least for objects that are not notably simple and familiar.

Passage 12:
From the 197 million square miles, which make up the surface of the globe, 71 percent is covered by the interconnecting bodies of marine water. The Pacific Ocean alone covers half the Earth and averages near 14,000 feet in depth. The portions which rise above sea level are the continents-Eurasia, Africa; North America, South America, Australia, and Antarctica. The submerged borders of the continental masses are the continental shelves, beyond which lie the deep-sea basins.

The ocean are deepest not in the center but in some elongated furrows, or long narrow troughs, called deeps. These profound troughs have a peripheral arrangement, notably around the borders of the pacific and Indian oceans. The position of the deeps, like the highest mountains, are of recent origin, since otherwise they would have been filled with waste from the lands. This is further strengthened by the observation that the deeps are quite often, where world-shaking earthquakes occur. To cite an example, the "tidal wave" that in April, 1946, caused widespread destruction along Pacific coasts resulted from a strong earthquake on the floor of the Aleutian Deep.

APPENDIX H: POST-TRIAL QUESTIONNAIRE

Post-Trial Questionnaire

Participant ID _____

Passage No. _____

Instruction: Please circle the number that best corresponds to the way you feel concerning each of the following questions.

1. Overall, how easy or difficult did you find it to comprehend the information presented in this audio passage?

1	2	3	4	5	6	7
VERY EASY			FAIRLY EASY			VERY DIFFICULT

2. Please evaluate the speed at which this audio passage was presented.

1	2	3	4	5	6	7
TOO SLOW			JUST RIGHT			TOO FAST

APPENDIX I: DEMOGRAPHIC TABLES

Table 8 Biographic Data for Age Group and Language Background Conditions

Biographic Data	EFL				ESL			
	Younger		Older		Younger		Older	
	Mean (SD)	Percentage %	Mean (SD)	Percentage %	Mean (SD)	Percentage %	Mean (SD)	Percentage %
Age	18.20(.48)	n/a	70.80 (4.54)	n/a	21.04 (2.78)	n/a	68 (3.42)	n/a
Gender								
M	n/a	.05	n/a	.10	n/a	.12	n/a	.14
F	n/a	.21	n/a	.15	n/a	.12	n/a	.11
Education								
High school or equivalent [a*]	n/a	.14	n/a	.04	n/a	.07	n/a	.11
Some college [a*]	n/a	.12	n/a	.08	n/a	.12	n/a	.09
Bachelor degree [a*]	n/a	n/a	n/a	.07	n/a	.03	n/a	.02
Graduate degree [a*]	n/a	n/a	n/a	.07	n/a	.02	n/a	.03
Race/Ethnicity								
Caucasian/White [a*]	n/a	.15	n/a	.26	n/a	.01	n/a	n/a
Black [a*]	n/a	.03	n/a	n/a	n/a	n/a	n/a	n/a
Asian/pacific islander [a*]	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Hispanic (white) [a*]	n/a	.03	n/a	n/a	n/a	.14	n/a	.21
Hispanic (non-white) [a*]	n/a	.01	n/a	n/a	n/a	.09	n/a	.04
Arab [a*]	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Multiracial [a*]	n/a	.02	n/a	n/a	n/a	.01	n/a	n/a
Others [a*]	n/a	.01	n/a	n/a	n/a	n/a	n/a	n/a
Hearing Aid Usage [a*]	n/a	0	n/a	.21	n/a	0	n/a	.24
Difficulty Hearing b*	1.7 (.37)	n/a	1.75 (.52)	n/a	1.87 (.68)	n/a	1.69 (.54)	n/a

a* Indicate % of participants that responded "yes" to these questions (Value range from .00 to 1.0).

b* Responses range from 1(Never) to 5 (Always)

Table 9 Individual Differences Variables Related to Hearing Comprehension Performance

Individual Differences Variables	EFL				ESL			
	Younger		Older		Younger		Older	
	Mean (SD)	Percentage %	Mean (SD)	Percentage %	Mean (SD)	Percentage %	Mean (SD)	Percentage %
English as First Language?								
Yes (Native Speakers)	n/a	.26	n/a	.26	n/a	n/a	n/a	n/a
No (Non-Native)	n/a	n/a	n/a	n/a	n/a	.24	n/a	.25
For non-native speakers: Spanish speaker [a*]	n/a	n/a	n/a	n/a	n/a	1	n/a	1
For non-native speakers: Years of practice of English	n/a	n/a	n/a	n/a	14.43 (4.95)	n/a	47.72 (14.67)	n/a
For non-native speakers: have taken English classes [a*]	n/a	n/a	n/a	n/a	n/a	.08	n/a	.11
For non-native speakers: have taken TOEFL test [a*]	n/a	n/a	n/a	n/a	n/a	.19	n/a	.19
Medical Condition [a*]		.01		.21		.02		.19
Hours Slept	7.32 (1.82)	n/a	7(.83)	n/a	7.46 (1.30)	n/a	7 (1.35)	n/a

Previous Experience with Synthetic Applications b*	2.24(.87)	n/a	2.17(.95)	n/a	2.82 (.94)	n/a	1.79 (1.07)	n/a
Hearing Score c*	.85 (.069)	n/a	.53 (.15)	n/a	.81(.11)	n/a	.50(.11)	n/a
English Proficiency Test Score c*	.9 (.09)	n/a	.84 (.11)	n/a	.92 (.10)	n/a	.77 (.09)	n/a
Shadowing Task Score c*	.99 (.013)	n/a	.99(.022)	n/a	1 (.008)	n/a	.95 (.10)	n/a

a* Indicate % of participants that responded "yes" to these questions (Value range from .00 to 1.0).
b* Responses range from 1(Never) to 5 (Always)
c* Score are calculated as percentage of correct answers

Table 10 Topic Familiarity Ratings by Age Group and Language Background

Topic Area	EFL				ESL			
	Younger		Older		Younger		Older	
	Mean	Std Dev	Mean	Std Dev	Mean	Std Dev	Mean	Std Dev
Biology	4.07	1.17	3.93	1.55	4.04	1.17	2.21	1.52
General Anatomy	3.50	1.36	4.10	1.47	3.29	1.44	2.80	1.95
English Literature	5.43	0.86	4.23	1.30	4.29	1.27	3.55	1.74
Astronomy	2.10	1.21	2.87	1.46	3.18	1.56	2.03	1.35
Aviation	1.53	1.41	3.43	1.98	2.07	1.63	1.72	1.44
Zoology	1.63	1.03	2.50	1.43	1.93	1.39	2.03	1.38
Human Anatomy	3.77	1.45	4.27	1.34	3.46	1.57	3.07	1.79
Current Events	4.10	1.37	5.73	1.14	4.54	1.26	4.21	2.30
Medical Devices	2.77	1.45	4.23	1.41	2.61	1.29	3.10	1.97
Medical Science	2.60	1.45	4.13	1.48	2.54	1.37	2.72	1.79
Psychology	4.57	1.50	3.83	1.29	4.54	1.55	3.17	2.04
Geography	3.83	1.60	5.03	1.19	4.04	1.69	3.59	1.93

Note: Response recorded from 1(Not at all familiar) to 7 (very familiar).

APPENDIX J: CORRELATION TABLE ON TOPIC FAMILIARITY RATING AND

PASSAGE PERFORMANCE

Table 11 Correlations Between Reported Topic Familiarities with Performance

Passage #/ Topic (Conditions)	EFL		ESL	
	Younger	**Older**	**Younger**	**Older**
Passage 1/ Biology (ST & SL)	-.311	.019	.368	.142
Passage 2/ Anatomy (ST & SL)	-.261	.180	.263	.185
Passage 3/ English Literature (ST & MD)	.143	.142	.151	.221
Passage 4/ Astronomy (ST & MD)	.443*	.019	.302	.157
Passage 5/ Aviation (ST & HI)	.120	.133	-.345	.115
Passage 6/ Zoology (ST & HI)	-.067	.120	-.034	.107
Passage 7/ Human Anatomy (NA & SL)	-.053	-.032	.062	-.070
Passage 8/ Current Events (NA & SL)	-.030	-.039	-.375*	.372*
Passage 9/ Medical Devices (NA & MD)	-.122	-.006	-.265	.280
Passage 10/ Medical Science (NA & MD)	-.038	.410*	-.161	-.301
Passage 11/ Psychology (NA & HI)	-.212	.197	-.278	.169
Passage 12/ Geography (NA & HI)	.352	.313	.053	.116

Note: *p value < .05

ST: Synthetic Speech Condition

NA: Natural Speech Condition

SL: Slow Speed Condition
MD: Medium Speed Condition
HI: High Speed Condition

APPENDIX K: IRB OUTCOME LETTER

University of Central Florida Institutional Review Board
Office of Research & Commercialization
12201 Research Parkway, Suite 501
Orlando, Florida 32826-3246
Telephone: 407-823-2901 or 407-882-2276
www.research.ucf.edu/compliance/irb.html

Approval of Human Research

From: **UCF Institutional Review Board #1**
FWA00000351, IRB00001138

To: **Jingjing Wang Costello**

Date: **July 06, 2010**

Dear Researcher:

On July 6, 2010, the IRB approved the following human participant research until 07/05/2011 inclusive:

Type of Review:	UCF Initial Review Submission Form
Project Title:	Effects of Age and Language Proficiency on The Comprehension of Synthetic Speech
Investigator:	Jingjing Wang Costello
IRB Number:	SBE-10-07007
Funding Agency:	N/A

The Continuing Review Application must be submitted 30days prior to the expiration date for studies that were previously expedited, and 60 days prior to the expiration date for research that was previously reviewed at a convened meeting. Do not make changes to the study (i.e., protocol, methodology, consent form, personnel, site, etc.) before obtaining IRB approval. A Modification Form **cannot** be used to extend the approval period of a study. All forms may be completed and submitted online at https://iris.research.ucf.edu .

If continuing review approval is not granted before the expiration date of 07/05/2011, approval of this research expires on that date. When you have completed your research, please submit a Study Closure request in iRIS so that IRB records will be accurate.

Use of the approved, stamped consent document(s) is required. The new form supersedes all previous versions, which are now invalid for further use. Only approved investigators (or other approved key study personnel) may solicit consent for research participation. Participants or their representatives must receive a copy of the consent form(s).

In the conduct of this research, you are responsible to follow the requirements of the Investigator Manual.

On behalf of Joseph Bielitzki, DVM, UCF IRB Chair, this letter is signed by:

Signature applied by Janice Turchin on 07/06/2010 10:41:26 AM EDT

IRB Coordinator

Page 1 of 1

REFERENCES

Ardila, A. (2003). Language representation and working memory with bilinguals. *Journal of Communication Disorders, 36*, 233-240.

Axmear, E., Reichle, J., Alamsaputra, M., Kohnert, K., Drager, K., & Sellnow, K. (2005). Synthesized speech intelligibility in sentences. *Language, Speech, and Hearing Services in School, 36*, 244-250.

Baddeley, A. D. (1990). *Human memory: Theory and practice*. Needham Heights, MA: Allyn and Bacon.

Bley-Vroman, R. (1990) The logical problem of second language learning. *Linguistic Analysis, 20*, 3–49.

Bickmore, T. W., Caruso, L., & Clough-Gorr, K. (2005). Acceptance and usability of a relational agent interface by urban older adults. *Proceedings of the 2005 Conference on Human Factors in Computing Systems*, 1212-1215. New York: ACM.

Brown, G. (1995). Dimensions of difficulty in listening comprehension. In: D. Mendelsohn, and J. Rubin (Eds.), *A Guide for the Teaching of Second Language Listening* (pp. 59-73). San Diego, CA: Dominie Press.

Chiang, C. S., & Dunkel, P. (1992). The effect of speech modification, prior knowledge, and listening proficiency on EFL lecture learning. *TESOL Quarterly, 26*, 345-374.

Clahsen, H. and Muysken, P. (1996) How adult second language learning differs from child first language development. *Behavior and Brain Sciences, 19*, 721–723.

Clahsen, H., & Felser, C. (2006). How native-like is non-native language processing? *TRENDS in Cognitive Sciences, 10*(2), 564-570.

Craik, F. I. M. (1994). Memory changes in normal aging. *Current Directions in Psychological Science, 3*, 155-158.

Craik, F. I. M., & Salthouse, T. A. (Eds.). (1999). *The handbook of aging and cognition* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.

Cruickshanks, K. J., Wiley, T. L., Tweed, T. S., Klein, B. E. K., Klein, R., Mares-Perlman, J. A., & Nondahl, D. M. (1998). Prevalence of hearing loss in older adults in Beaver Dam, Winconsin. American Journal of Epidemiology, 148(9), 879-886.

Drager, K. D. R. & Reichle, J. E. (2001). Effects of discourse context on the intelligibility of synthesized speech for young adult and older adult listeners. *Journal of Speech, Language, and Hearing Research, 44*, 1052-1057.

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

Educational Testing Service (1990). *TOEFL sample test* (4th ed.). Princeton, NJ: Educational Testing Service.

Educational Testing Service (2009). *The official guide to the TOEFL test* (3rd ed.). New York, NY: McGraw-Hill.

Eysenck, M.W., Keane, M.T., 1995. *Cognitive Psychology*, 3rd Edition. Psychology Press, Sussex.

Faul, F., Erdfelder, E., Lang, A. G. & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*, 175-191.

Frenck-Mestre, C. and Pynte, J. (1997) Syntactic ambiguity resolution while reading in second and native languages. *The Quarterly Journal of Experimental Psychology, 50A*(1), 119–148.

Færch, C., Kasper, G., 1986. The role of comprehension in second language learning. *Applied Linguistics 7*, 257-274.

Goh, C. C. M. (2000). A cognitive perspective on language learners' listening comprehension problems. *System, 28*, 55-75.

Grady, C. L., & Craik, F. I. M. (2000). Changes in memory processing with age. *Current Opinion in Neurobiology, 10*(2), 224-231.

Greene, B.G. (1986). Perception of synthetic speech by nonnative speakers of English. In *Proceedings of the Human Factors Society,* 1340-1343. Santa Monica, CA.

Griffiths, R. (1990). Speech rate and nonnative speaker comprehension: A preliminary study in the time-benefit analysis. *Language Learning, 40*, 311-316.

Griffiths, R. (1992). Speech rate and listening comprehension: Further evidence of the relationship. *Teachers of English to Speakers of Other Languages (TESOL) Quarterly, 26*(2), 385-390.

Hardee, J. B. & Mayhorn, C. B. (2007). Reexamining synthetic speech: intelligibility and the effects of age, task, and speech type on recall. *Proceedings of the Human Factors and Ergonomics Society 51st Annual Meeting*, Baltimore, MD, October 1-5, 1143-1147.

Hasegawa, M., Carpenter, P., & Just, M. A. (2002). An fMRI study of bilingual sentence comprehension and workload. *Neuroimage, 15*, 647-660.

Hasher, L., & Zacks, R. T. (1988). Working memory, comprehension, and aging: A review and a new view. In G. Bower (Ed.), *The psychology of learning and motivation* (Vol. 22, pp. 193-225). New York: Academic Press.

Hearing Test. (2010). Speech-in-Noise Test. [Software]. Available from http://hear-it.org/page.dsp?page=5224

Henrichsen, L. E. (1984). Sandhi-variation: a filter of input for learners of ESL. *Language Learning, 34*, 103-126.

Humes, L. E., Nelson, K. J., Pisoni, D. B., & Lively, S. E. (1991). Effects of age on serial recall of natural and synthetic speech. *Journal of Speech and Hearing Research, 34*, 634–636.

Kane, M. J., & Engle, R. W. (2003). Working-memory capacity and the control of attention: the contributions of goal neglect, response competition, and task set to stroop interference. *Journal of Experimental Psychology: General, 132*(1), 47-70.

Kangas, K. A., & Allen, G. D. (1990). Intelligibility of synthetic speech for normal-hearing and hearing-impaired listeners. *Journal of Speech and Hearing Disorders, 55*(4), 751–755.

Kelly, P. (1991). Lexical ignorance: The main obstacle to listening comprehension with advanced FL learners. *IRAL, 29*(2), 135-150.

Kelch, K. (1985). Modified input as an aid to comprehension. *Studies in Second Language Acquisition, 7*, 81-89.

Kemper, S. Herman, R. E., & Lian, C. H. T. (2003). The costs of doing two things at once for younger and older adults: talking while walking, finger tapping, and ignoring speech or noise. *Psychology and Aging, 18*(2), 181-192.

Kline, D., Scialfa, C. (1997). Sensory and Perceptual Functioning: Basic Research and Human Factors Implications. In: A. D. Fisk and W. A. Rogers WA (Eds.), *Handbook of Human Factors and the Older Adult,* (pp.27-54)*.* San Diego, CA: Academic Press.

Konstantinos, P., Koutsoklenis, A., Katemidou, E., & Okalidou, A. (2009). Perception of synthetic and natural speech by adults with visual impairments. *Journal of Visual Impairment and Blindness*, 103(7), 403-414.

Kramer, A. F., Humphrey, D. G., Larish, J. F., Logan, G. D., & Strayer, D. L. (1994). Aging and inhibition: beyond a unitary view of inhibitory processing in attention. *Psychology and Aging, 9*(4), 491-512.

Lass, N. J., & Prater, C. E. (1973). A comparative study of listening rate preferences for oral reading and impromptu speaking tasks. *The Journal of Communication, 23*, 95–102.

Leach, M. A. (2008). America's older immigrants: A profile. *Generations, 32*(4), 34-39.

Macoy, S. L., Tun, P. A., Cox, L. C., Colangelo, M., Stewart, R. A., & Wingfield, A. (2005). Hearing loss and perceptual effort: downstream effects on older adult's memory for speech. *The Quarterly Journal of Experimental Psychology, 58A*(1), 22-33.

Mack, M., Tierney, J., & Boyle, M. E. T. (1990). The intelligibility of natural and LPC-coded words and sentences of native and non-native speakers of English. (Tech. Rep. No. 869) Lexington, MA: Massachusetts Institute of Technology Lincoln Laboratory.

McDowd, J. M., & Shaw, R. J. (1999). Attention and Aging: a Functional Perspective. In F. I. M. Craik and T. A. Salthouse (Eds.), *The handbook of aging and cognition* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.

Mirenda, P., & Beukelman, D. R. (1987). Acomparison of speech synthesis intelligibility with listeners from three age groups. *ACC: Augmentative and Alternative Communication, 3,* 120–128.

Murphy, D. R., Craik, F. I. M., & Li, K, Z. H. (2000). Comparing the effects of aging and background noise on short-term memory performance. *Psychology and Aging, 15*(2), 323-334.

Paris, C. R., Thomas, M. H., Gilson, R. D., & Kincaid, J. P. (2000). Linguistic cues and memory for synthetic and natural speech. *Human Factors, 42,* 421–431.

Park, D. C., & Payer, D. (2006). Working Memory Across the Adult Lifespan. In E. Bialystok and F. I. M. Craik (Eds.), *Lifespan cognition: mechanisms of change*, (pp. 128-142). Oxford, England: Oxford University Press.

Pichora-Fuller, M. K. (2003). Processing speed and timing in aging adults: psychoacoustics, speech perception, and comprehension. *International Journal of Audiology, 42*, S59-S67.

Pisoni, D. B. (1981). Speeded classification of natural and synthetic speech in a lexical decision task. *Journal of the Acoustical Society of America, 70*, S98.

Pisoni, D. B., Manous, L. M., & Dedina, M. J. (1987). Comprehension of natural and synthetic

speech: Effects of predictability on the verification of sentences controlled for intelligibility. *Computer Speech and Language, 2,* 303–320.

Pitts, D. G. (1982). The effects of aging upon selected visual functions: Dark adaptation, visual acuity, stereopsis, and brightness contrast. In R. Sekuler, D. Kline, & K. Dismukes (Eds.), *Aging and human visual function*. (pp. 131-160). New York: Liss.

Pollack, M. E., Engberg, S., Matthews, J. T., Thrun, S., Brown, L., Colbry, D., et al. (2002). *Pearl: A mobile robotic assistant* (American Association for Artificial Intelligence Tech. Rep. WS-02-02, pp. 85–92). Menlo Park, CA: AAAI Press.

Reynolds, M. ED., Bond, Z. S., & Fucci, D. (1996). Synthetic speech intelligibility: Comparison of native and non-native speakers of English. *Augmentative and Alternative Communication, 12*(1), 32.

Reynolds, M. E., & Givens, J. (2001). Presentation rate in comprehension of natural and synthetic speech. *Perceptual and Motor Skills, 92,* 958–968.

Roring, R. W., Hines, F. G., & Charness, N. (2007). Age differences in identifying words in synthetic speech. *Human Factors, 49*(1), 25-31.

Rubin, J. (1994). A review of second language listening comprehension research. *The Modern Language Journal, 78*(2), 199-221.

Rymniak, M. J., Shanks, J. et al. (2002). *TOEFL workbook* (2nd ed.). New York, NY: Simon and Schuster

Salthouse, T. A. (1988). The role of processing resources in cognitive aging. In M. L. Howe & C. J. Brainerd (Eds.), *Cognitive development in adulthood: Progress in cognitive development research* (pp. 185-239). New York, NY: Springer-Verlag.

Salthouse, T. A. (1994). The aging of working memory. *Neuropsychology, 8*(4), 535-543.

Salthouse, T. A., & Babcock, R. L. (1991). Decomposing adult age differences in working memory. *Developmental Psychology, 27,* 763-776.

Scheutz, M. and Eberhard, K. (2004) Effects of morphosyntactic gender features in bilingual language processing. *Cognitive Science: A Multidisciplinary Journal, 28*, 559–588.

Schieber, F. (2006). Vision and Aging. In J. E. Birren and K. W. Schaie (Eds.), *Handbook of the psychology of aging* (6th ed.) (pp. 129-161). Burlington, MA: Elsevier Academic Press.

Schmidt-Rinehart, B. C. (1994). The effect of topic familiarity on second language listening comprehension. *Modern Language Journal*, 78(2), 179-189.

Schneider, B. A. & Pichora-Fuller, M. K. (1999). Implications of Perceptual Deterioration for Cognitive Aging Research. In F. I. M. Craik and T. A. Salthouse (Eds.), *The handbook of aging and cognition* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.

Shohamy, E., & Inbar, O. (1991). Validation of listening comprehension tests: the effect of text and question type. *Language Testing, 8*, 23-40.

Si, L., & Callan, J. (2001). A statistical model for scientific readability. *Proceedings of the 10th International Conference on Information and Knowledge Management,* 574-576. New York, NY: AMC.

Slowiaczek, L. M., & Nusbaum, H. C. (1985). Effects of speech rate and pitch contour on on perception of synthetic speech. *Human Factors,27,* 701–712.

Small, J. A., Andersen, E. S., & Kempler, D. (1997). Effects of working memory capacity on understanding rate-altered speech. *Aging, Neuropsychology, and Cognition, 4*(2), 126-139.

Smither, J. A. (1993). Short term memory demands in processing synthetic speech by old and young adults. *Behaviour and Information Technology, 12*(6), 330-335.

Stine, E. L., & Wingfield, A. (1987). Process and strategy in memory for speech among younger and older adults. *Psychology and Aging, 2*(3), 272-279.

Sutton, B., King, J., Hux, K., & Beukelman, D. R. (1995). Younger and older adults' rate performance when listening to synthetic speech. *Augmentative and Alternative Communication, 11*(3), 1477-3848.

Syntax. (2010). In *Merriam-Webster Online Dictionary*. Retrieved March 23, 2010, from http://www.merriam-webster.com/dictionary/syntax

Tabachnick, B. G., & Fidell, L. S. (2001). *Using multivariate statistics* (4th ed.). Massachusetts: Allyn and Bacon.

Treas, J., & Batalova, J. (2007). Older Immigrants. In K. W. Schaie and P. Uhlenberg (Eds.), *Social structures: the impacts of demographic changes on the well-being of older persons*, (pp. 309-330). New York, NY: Spring Publishing Company.

Tun, P. A., O'Kane, G. O., & Wingfield, A. (2002). Distraction by competing speech in young and older adultlisteners. *Psychology and Aging*, *17*, 453–467.

Venkatagiri, H. S. (1991). Effects of rate and pitch variations on the intelligibility of synthesized speech. *Augmentative and Alternative Communication, 7*(4), 284.

WebMD. (2009). Cataracts – topic overview. Retrieved on May 10, 2010, from:
http://www.webmd.com/eye-health/cataracts/cataracts-topic-overview

Wingfield, A., & Ducharme, J. L. (1999). Effects of age and passage difficulty on listening-rate preferences for time-altered speech. *Journal of Gerontology: Psychological Sciences, 54B*(3), 199-202.

Wingfield, A., & Tun, P. A. (2001). Spoken language comprehension in older adults: interactions between sensory and cognitive change in normal aging. *Seminars in Hearing, 22*, 287-301.

Winters, S. J., & Pisoni, D. B. (2004). Perception and comprehension of synthetic speech. In *Research on spoken language processing report No. 26.* (pp. 95 – 138). Bloomington, IN: Speech Research Laboratory, Indiana University.

Zacks, R. T., Hasher, L., & Li, K. Z. H. (1999). Human Memory. In F. I. M. Craik and T. A. Salthouse (Eds.), *The handbook of aging and cognition* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.

Zhao, Y. (1997). The effects of listeners' control of speech rate on second language comprehension. *Applied Linguistics, 18*, 49-68.