Synthesized Speech Intelligibility and Preschool Age Children: Comparing Accuracy for Single Word Repetition with Repeated Exposure.

A THESIS
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL
OF THE UNIVERSITY OF MINNESOTA
BY

Carrie Lynn Pinkoski

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
MASTER OF ARTS

Joe Reichle Ph.D, advisor

August 2010

Acknowledgements

I would like to thank my committee member Dr. LeAnne Johnson for her thoughtful insights. Thank you to Dr. Sarah Angerman and Dr. Edward Carney for their time and assistance. Thank you to my husband for his kind support and patience throughout. I would especially like to thank my thesis advisor Dr. Joe Reichle and committee member Dr. Benjamin Munson for their invaluable guidance, assistance and dedication to this project. Much obliged.

Abstract

Purpose: This investigation examined the effect of repeated exposure to novel and repeated single spoken words in typical listening environments on the intelligibility of two synthesized speech voices and human recorded speech in preschool age children.

Methods: Eighteen preschool aged participants listened to and repeated single words presented in human recorded speech, DECtalk™ Paul and AT & T Voice ™ Michael during five experimental sessions. Stimuli consisted of repeated and novel vocabulary items presented in each speech output condition for each session. Experimental sessions took place in the presence of background noise in participants' classroom or home settings.

Results: There was a significant main effect for voice as participants accurately identified significantly more words in the human recorded speech and AT & T Voice ™ than in the DECtalk™ speech output conditions. When averaged across speech output conditions, children increased their accuracy as they participated in additional sessions. There was a statistically significant effect for the interaction between session and voice. DECtalk ™ had a slightly larger effect of session than AT&T Voice ™ and human recorded speech. There was a non-significant interaction between session and vocabulary type (repeated/ novel). When averaged across each voice type, repeated vocabulary words resulted in more accurate responses in later sessions than novel vocabulary words.

Table of Contents

List of Tables

Augmentative and alternative communication (AAC) systems that utilize synthetic speech output, provide individuals with little to no functional speech with the potential to communicate effectively in their home, school, work and community environments. They have been demonstrated to increase speech output in individuals with severe communication impairments (DiCarlo & Banajee, 2000; Light, Drager, Hayes, Kristiansen, May, Page, et al. 2004). With communication delays or disorders detected in early childhood, AAC interventions have been successfully implemented to avoid reduced social and learning opportunities that may result without an effective communicative means (Beukelman & Mirenda 2005; Kangas & Lloyd, 1988; Weitz, Dexter, & Moore, 1997).

The communicative partners of speech generating device (SGD) users must be patient in that synthesized speech messages can be much slower to produce and less intelligible than typical conversational speech. The rate of typical conversational speech varies from 150 words to 250 words per minute (Goldman – Eisler, 1986). The rate of speech for an AAC system users who use aided symbols has been reported to be 15 wpm or less (Foulds, 1980; 1987). Often, young children with significant disabilities who use a speech generating device have fewer repair strategies available to them when their original message is not understood by their listener. Typically, SGD users have fewer vocabualry items accessible to them, reduced control over the volume of their message and are more likely to repair communication breakdowns with unconventional gestures, vocalizations, or other behaviors that are less easily understood by their communication partners (Halle, Brady & Drasgow, 2004). Potentially, these factors can decrease the quality and quantity of meaningful communication between an individual who is using an augmentative communication system and their communication partners.

Additionally, the intelligibility of synthetisized speech poses a potential problem as a result of its impoverished signal that contains fewer of the acoustic – phonetic cues that are present in natural speech. Natural speech consists of acoustic redundancies that provide a boost to intelligibility (Reynolds, Isaacs – Duvall, Sheward & Rotter, 2000). Interpreting a synthesized speech signal requires a greater use of attention and concentration than is typically used when interpreting natural speech (Drager & Reichle, 2001b; Duffy & Pisoni, 1992).

Variables Affecting Intelligibility of Synthesized Speech

Although much of the research in synthesized speech intelligibility has been with adult participants (Drager, Reichle, and Pinkoski, in press; Fucci, Reynolds, Bettagere, & Gonzalez, 1995; Higginbotham, Drazek, Kowarsk, Scally, & Segal, 1994; Higginbotham Scally, Lundy, & Kowarsky, 1995), many of the same factors affect the intelligibility of synthesized speech among children as well. A number of investigations have reported that synthesized speech intelligibility is even more difficult for children (e.g. McNaughton, Fallon, Tod, Weiner, and Neisworth 1994; Mirenda & Beukelman, 1987). Listening conditions have been found to result in significantly poorer intelligibility for young adult listeners (Fucci, Reynolds, Bettagere, & Gonzales, 1995). Sentences or words presented with context as well as predictable words and sentences have been found to be more intelligible than single words with limited context (Hoover, Reichle, Van Tasell, & Cole, 1987; Mirenda & Beukelman, 1987, 1990; Oshrin & Siders, 1987; Slowiaczek & Nusbaum, 1985). For example, Drager & Reichle (2001b) found that short stories were more intelligible than isolated sentences in their study of typical elderly participants. Not all synthesized speech has similar characteristics. Until recently, DECtalk™ voices have been indentified as one of the more intelligible of synthetic speech software packages (Koul & Hanners, 1997; Mirenda & Beukelman 1987, Von Berg, Panorska, Uekn, & Qeadan, 2009). A wide variety of variables have been shown to effect synthesized speech intelligibility. Those of interest in the current investigation include, chronological age, voice synthesizer type, listening conditions and repeated exposure to synthesized speech. Each of these will be discussed in greater detail as they relate to the independent variables in this study.

Synthesized speech intelligibility and background noise. In order for speech generating devices to facilitate communicative experiences, device users and their communication partners must be able to understand the synthesized speech output of an AAC system in a variety of settings including school, work and home. "AAC devices are implemented more frequently with preschool age children in an effort to increase opportunities for success in inclusive settings by providing a communication system that permits them to participate in typical curricular activities" (Beukelman & Mirenda 2005, p 392). As the number of preschool age children exposed to synthesized speech increases, there is a need to demonstrate its intelligibility with young preschoolers.

To date, the majority of research examining synthesized speech intelligibility has frequently been conducted in sound controlled environments (Hustad, Kent & Beukelman, 1998; McNaughton et al. 1994; Pisoni, Manous, & Dedina, 1987). Virtually no investigations have examined synthesized speech intelligibility with preschoolers in natural environments (elementary classrooms) where natural noise levels can range from 55 – 65 dBA (Nelson, Soli, & Seitz, 2002).

Synthesized speech intelligibility and chronological age. Typically developing children have been found to have greater difficulty understanding synthesized speech than adults (Axmear, Reichle, Alamsaputra, Kohnert, Drager, & Sellnow, 2005; Beukelman & Mirenda 1987; Greene, Logan, & Pisoni, 1986; McNaughton et al. 1994). Mirenda and Beukelman (1987) compared the intelligibility of natural speech and seven speech synthesizer voices among participants in three age groups consisting of adults ages 22 to 50 years old, children ages 11 to 12 years old and children ages 7 to 8 years old. All of the participants were monolingual speakers of English with no significant previous exposure to synthetic speech. The dependent measure was the number of correct single words and number of correct words in sentences repeated by participants. Their findings indicated that there were significant differences in synthesized speech intelligibility across age groups and synthesized speech voices. Additionally, the authors reported an overall decrease in single word intelligibility with increasingly younger groups of participants.

Von Berg, Panorska, Uekn, & Qeadan (2009) compared synthesized speech voices DECtalk™ Paul (male), Betty (female) and Kit (child) and VeriVox™ Michael (male), Sarah (female) and Jamie (child) among participants in four age groups (6 to 13 years old, 14 to 24 years old, 25 – 65 years old and 66 – 85 years old). The investigators found that DECtalk™ Kit was the least intelligible voice for child and adult participants, however the DECtalk™ male and VeriVox™ child voices revealed significant differences in word repetitions and sentence verification tasks across age groups. The younger adult group performance was significantly more accurate for single words presented in DECtalk ™ male voice than the older adult group and the younger child group. The younger adult group performance was significantly more accurate for single words presented in VeriVox™ child voice than the performance of the older adult group.

Drager, Clark-Serpentine, Johnson, & Roeser (2006) compared word and sentence repetition accuracy of three age groups of young children: 3 -year olds, 4-year olds and 5-year olds in the presence of background noise. Overall intelligibility was low for all groups, however, there was a statistically significant difference in intelligibility scores between children in the 3 year old group and children in the 4 and 5 year old groups for single words and sentences in DECtalk™, MacinTalk™, and digitized speech.

McNaughton, Fallon, Tod, Weiner, and Neisworth (1994) examined the effect of repeated exposure to synthesized speech as a variable in synthesized speech intelligibility. The researchers reported on two separate experiments, one with children and one with adults. The experiment with children included 24 between the ages 6-10 years while the experiment with adults included 24 between the ages of 19 – 44 years. The dependant measure was accurate repetition of single words presented in DECtalk™ and ECHO™ speech synthesizers. There was a significant effect for session for both the child and adult participants, indicating that both groups significantly improved their accuracy over the duration of five experimental sessions. Chronological age was not considered an independent variable in this study, thus the difference in group performances was not statistically analyzed. However, when comparing the intelligibility for the two age groups, the children were less accurate than the adult participants. For example, the group mean accuracy after five experimental sessions for DECtalk™

repeated single words is 81.4% for children and 91.2% for adults. The child mean accuracy were lower than the adult mean accuracy scores for single words in each condition: ECHO™ novel, ECHO™ repeated, DECtalk™ novel and DECtalk™ repeated, although both groups significantly improved their scores for novel and repeated vocabulary in experimental sessions 2 through 5.

Speech synthesizer type. There is a wide range of speech synthesizers commercially marketed. One of the most salient variables affecting performance accuracy in both adults and children is the synthesized software package utilized, thus a variety of synthesized speech voices have been studied and compared in the current literature (Axmear et.al. 2005; Drager et al., 2006; Helsel –Dewert & Van Den Meiracker, 1987; Koul & Hanners, 1997; Koul & Hester, 2006; McNaughton et al., 1994; Mirenda & Beukelman 1987,1990; Reynolds & Fucci, 1998; Reynolds & Jefferson, 1999; Von Berg Panorska, Uken & Qeadan 2009). To date, DECtalk™ synthesized adult voices have been among the most intelligible when compared to other speech synthesizers (Koul & Hanners, 1997; Mirenda & Beukelman 1987; Von Berg et al. 2009). Often, the adult male voice, *Perfect Paul* has been reported to be the most intelligible amongst different synthesizers (DECtalk™, VeriVox™) when compared to the adult female voices and child voices. Additionally, child synthesized voices are often reported as being the least intelligible when compared to adult male synthesized speech voices amongst the DECtalk™ voices. (Drager, et al., 2006; Koul & Hanners, 1997; McNaughton et al., 1994; Mirenda & Beukelman 1987; Von Berg et al., 2009).

Mirenda and Beukelman (1987) compared synthesized speech intelligibility for seven speech synthesizer voices and natural human speech in three age groups (children

6-8 years old, children 10-12 years old and adults). When synthesized speech was presented in sentences, the most intelligible voice for adults was the DECtalk™ male voice (Paul), while the DECtalk™ female voice (Betty) was the most intelligible for both of the child groups. The mean accuracy scores for sentences decreased to an even greater degree with the age of participants. There were different patterns of intelligibility for voice in each of the groups. In the adult group, natural human speech, DECtalk™ male, female, and child voices were statistically equivalent in intelligibility. In the 10 – 12 year old group, natural human speech and all three DECtalk™ voices were statistically equivalent. The results for the 6-8 year old group were similar to the 10 – 12 year old group with all of the DECtalk™ voices statistically equivalent to natural human speech. However the mean accuracy scores of intelligibility for the 6-8 year olds were considerably lower than the scores of the other age groups.

Results of synthesized speech intelligibility studies comparing children and adult participants have suggested that there are differences in how children perceive synthetic speech. Studies have indicated differences and decreases in performance accuracy for younger children when compared to older children and adults. The difference in performance accuracy for young children indicates a need to further investigate the variables affecting synthesized speech intelligibility with this population.

Synthesized speech intelligibility and repeated exposure. McNaughton et al. (1994) (described earlier), the intelligibility of two synthesized speech voices (DECtalk™ child and ECHO™ II+) and natural recorded speech for a single word intelligibility task across five sessions for each participant. One experiment described adult performance while a second described child performance. Novel and repeated speech stimuli were presented for each session to measure whether improved accuracy would be specific to trained items (repeated stimuli) or would generalize to untrained items (novel stimuli). Novel speech stimuli were items that were only presented for one of the five experimental sessions. Repeated speech stimuli were presented for each of the five experimental sessions, resulting in four repeated exposures to the same item (repeated for sessions 2 through 5). Both groups of participants significantly increased accuracy across sessions for novel and repeated stimuli in both synthesized voices. There was a subtle, but not statistically significant decrease in intelligibility in both synthesized voices for novel and repeated stimuli for children when compared to the adult participants. In comparing results for synthetic speech, intelligibility scores for DECtalk™ were significantly higher than intelligibility scores for Echo™ synthesized speech for child and adult participants. The McNaughton et al. (1994) findings indicate that both children and adults are able to improve intelligibility accuracy by recognizing words that they have heard before and also improve intelligibility accuracy via repeated exposure to synthetic speech.

In an earlier study Greenspan, Nusbaum, & Pisoni (1985), hypothesized an explanation for this phenomenon by suggesting that adults may improve intelligibility accuracy of synthetic speech with repeated exposure or training (practice effect) in that through perceptual learning, they develop "an abstract set of acoustic-phonetic rules for encoding synthetic speech" (p. 430). The encoding rules developed by listeners may allow them to create a sound inventory and subsequently organize the distorted synthesized speech sounds into categories according to natural speech phonetics. The authors suggested that studies with younger children or individuals with cognitive impairments might yield different results related to practice effect.

Reynolds & Jefferson (1999) examined differences in synthesized speech intelligibility between 24 children that were divided into two groups of children that included 6 to 7 and 9 to 11 years olds. The researchers compared response latency performance as a measure of intelligibility on sentence verification tasks. Sentence stimuli were presented in recorded natural speech and synthesized speech voice DECtalk™ Betty across two listening sessions spaced three to seven days apart. The dependent measure of response latency was used to examine practice effect between the first and second sessions. The researchers found a significant effect for voice and group in that both groups of children had longer response latencies for synthetic speech. The young child group experienced a decrease in response latency for synthesized speech across sessions. Reynolds and Jefferson suggested that the performance variable related to differences in processing speed and working memory capacity may be due to differences in general development of cognitive skills between older and younger children (Hale, 1990; Hale et al., 1993; Miller & Vernon, 1997; Mirenda & Beukelman, 1987,1990). Response latencies for natural speech did not significantly decrease from session 1 to session 2 for either group. Response latencies for synthetic speech only decreased from session 1 to session 2 for the younger child group. Overall, the older child group was significantly faster than the young child group with shorter response latencies for natural and synthetic speech for sessions 1 and 2.

Koul and Clapsaddle (2006) examined the differences in synthesized speech intelligibility between 10 typically developing children, between 4 and 8 years and 18 adults with mild to moderate intellectual disabilities between 22 and 55 years. The researchers examined whether individuals with intellectual impairment demonstrated improved accuracy with synthesized speech intelligibility across sessions as a result of repeated exposure to speech stimuli. ETI Eloquence™ was used to create the synthesized speech stimuli that included novel and repeated single words and sentences that were presented during each of three experimental sessions. Adults with mild to moderate intellectual disabilities group significantly improved their accuracy across the three experimental sessions. There was no significant interaction between session and stimuli type (novel/repeated) as intelligibility improved for both repeated and novel word and sentences, indicating that accuracy of synthesized speech intelligibility generalized to novel or untrained stimuli.

In summary, research related to synthesized speech intelligibility in young children is sparse. There is a need for additional research to expand on the findings from previously conducted studies in order to support the increase in numbers of children exposed to synthesized speech. Studies with high ecological validity should be conducted in order to best inform clinical and educational practice.

Purpose of Current Investigation

The current investigation was designed to compare the intelligibility of repeated exposure to synthesized speech and practice effect (exposure to novel and repeated vocabulary items) among preschool age children using synthetic speech voices DECtalk ™ Paul, AT & T Voice ™ Michael and a human recorded male voice in the sound level environment of a typical preschool classroom. A review of the literature addressing synthesized speech intelligibility and young children (Drager, Reichle & Pinkoski, in press) reported that few studies that examined variables related to listener, speech output and environment among preschool or school aged children. Based on the findings of this scoping review, there is evidence to support that, (1) children as young as 3 years old (Drager et. al 2006) have demonstrated comprehension of synthesized speech messages, (2) synthesized speech intelligibility is lower than natural speech intelligibility for children, (3) synthesized speech intelligibility is lower for children than for adults, (4) background noise that results in decreased synthesized speech intelligibility for adults also has the same effect on intelligibility for children (Drager et al. 2006), (5) synthesized speech intelligibility may be lower for bilingual adults and children than for monolingual adults and children (Axmear et al. 2005), (6) repeated exposure may contribute to an increase in intelligibility for synthesized speech in adults and children (McNaughton et al. 1994), (7) linguistic redundancy and contextual speech increases intelligibility for synthesized speech in adults and children (Drager et al. 2006).

Given that preschool age children are often the users of a SGD or listeners of synthetic speech educational software applications (Binger & Light, 2006; Mirenda & Iacono, 2009), additional evidence addressing synthesized speech intelligibility with this population represents an important need. With a new generation of synthesized speech packages beginning to be used in speech generating devices, it is important to evaluate the relative intelligibility of newly available packages with those that have been traditionally used.

In the current investigation we compared the speech intelligibility of DECtalk ™ Paul, AT & T Voice ™ Michael and human recorded speech among children aged 2 years 7 months to 3 years 6 months. Additionally, experimental sessions included novel and repeated single word vocabulary items to measure improved accuracy with repeated exposure to both the single word vocabulary items as well as the synthesized speech voices.

Methods

Participants

Eighteen (18) typical monolingual preschoolers participated. Participants' ages ranged from 2 years and 7 months to 3 years and 6 months. In investigators asked childcare providers to identify children who were between the ages of 30 – 42 months old and native speakers of English. Upon parents' consent, children meeting age and language requirements participated in screenings for hearing, language and verbal imitation. Characteristics for all participants are displayed in appendix A.

Screening Procedures

Hearing acuity. Hearing was assessed using a portable audiometer *Maico MA 25*. The children practiced responding to the tones they heard by first listening to an 80 dB pure tone at 1000 and 2000 Hz without wearing the headphones. Children were instructed to drop a toy in a bucket or raise their hand each time they heard a tone. After listening and responding to two different frequencies in practice, the investigators began the screening. Pulse tone signals were presented at 25 dB (HL) for 500 Hz and 20 dB for 1000, 2000, and 4000 Hz. Participants were required to accurately respond to a single presentation of each of the four frequencies presented to each ear. Children were excluded from the study if they failed to respond to one or more frequency level in either ear.

Vocabulary comprehension. The experimenter implemented Peabody Picture Vocabulary Test – Fourth Edition (PPVT- IV) (Dunn & Dunn, 2007) to assess vocabulary comprehension in order to insure that differences in vocabulary comprehension skills did not account for performance variation across participants. A standard score between 85 and 115 (i.e. within +/- 1 standard deviation for the child's age) was required. The mean standard score for the participants was 110 with a range of 101 to 114.

Verbal imitation. All children participated in a verbal imitation screening consisting of 10 words chosen from the MacArthur -Bates Communicative Development Inventories (Fenson, Dale, Reznick, Bates, Pethick, Hartung & Reilly, 1993). Stimuli were selected from a list of words that 70% or more of parents reported that their child understood or produced at 30 months. Children were told, "I'm going to say some words. I want you to say what I say. Say exactly what I say." Children were required to intelligibly imitate at least nine out of the ten words spoken by the investigator. Responses containing phonological errors were accepted as correct provided that the word was a lexical match to the model.

Setting. All screenings and experimental sessions were completed at community daycare centers and homes. The screenings took place in quiet rooms, away from other children. In most cases, children were seated in chairs facing the investigators at a table. When that seating arrangement was not possible, the participants and investigators sat on the floor. Sound levels were measured during experiment sessions using a Radio Shack model no. 33-2055 sound level meter. Measurements were obtained at the beginning of each experimental session and found to between 50 and 67 dB (Mean of 56.4 dB).

With synthesized speech applications increasingly used in classroom settings in the form of educational software and speech generating devices, the investigators felt that conducting the experimental sessions in rooms in which some background noise was present would increase the social validity of the study. The American Speech Language and Hearing Association (ASHA, 1995) has recommended that the noise level in an occupied classroom should not exceed 40-50 dB. However, many American school classrooms have higher noise levels in unoccupied classrooms. Knecht, Nelson, Whitelaw, Feth (2002) reported noise levels ranging from 34.4 dB to 65.9 dB in empty classrooms and suggested that noise levels in occupied classrooms would be significantly higher.

Materials

The 210 words used as experimental stimuli were selected from the MacArthur-Bates CDIs (Fenson et al. 1993) because they were words that had been documented to be comprehended by 88.6% of children ages 30 months and older. The stimuli were quasi randomly arranged into lists in that the randomized lists were balanced for the frequency of occurrence in English language (Kucera and Francis 1967), phonological neighborhood density (Hoosier Mental Lexicon Database. Pisoni, Nusbaum, Luce, & Slowiaczek 1985), phoneme length, syllable length, and frequency rating that refers to the proportion of children at age 30 months who comprehend each of the words on the MacArthur-Bates CDIs (Fenson, et al., 1993).

Phonological neighborhood density refers to the number of words that can be created by adding, deleting or substituting one phoneme in the word. In their study of Gierut, Morrison and Champion (1999) found that training in high frequency words facilitated generalization to untreated words and that training sounds in high-density words inhibited generalization to untreated sounds. A multivariate analysis of variance showed no significant effect of any of the descriptive measures (log-transformed word frequency, number of phonological neighbors, length in phonemes or syllables, or percentage of children on the MacArthur who recognized the word, all F [20,188]'s < 1, all p's > 0.50. Twenty one (21) different lists of 10 words were constructed using these stimuli.

For each experimental session, participants were assigned two different stimuli lists for each experimental condition. One list of stimuli was repeated for each subsequent session and one list of stimuli was presented during only one session. Each participant had a unique set of lists for each condition to minimize the impact of variations of complexity within the word lists. The lists were randomly assigned to one of the three speech output conditions for each session for each participant using a Latin Squares design that was modified for 18 participants using 21 possible lists. No list occurred more than once in any of the conditions for any participant across their five experimental sessions. Additionally, the word order presentation was arranged randomly from list to list and participant to participant. During the five experimental sessions, participants were presented 300 total words randomly assigned in lists from the 210 different words that comprised the stimuli. All of the words presented in the first sessions were novel, but ten words in each speech output condition were repeated for sessions 2 through 5. The repeated vocabulary items totaled 120 words for each participant. There were 180 words that were novel in that they were only heard once during the experimental sessions.

The investigators created stimuli that were likely to be easily comprehended while controlling for lexical characteristics that have been demonstrated to influence speech intelligibility. Previous research on adult word recognition has established that words' frequency of use (Log frequency of usage (using the Kucera and Francis, 1967, count of written text), familiarity to listeners and phonological neighborhood density (based on approximately 20,000 words in Webster's Pocket Dictionary) affect speech intelligibility. Phonological neighborhood density refers to a measurement of the number of real words that can be created by adding, deleting, or substituting phonemes in a target word.

Human Speech

A 29 - year old adult male who spoke a Midwestern dialect produced the natural speech stimuli. A single file containing the words produced using natural speech was recorded to a Marantz CDW300 CD recorder. The recordings were performed in a sound treated room. The speaker wore an AKG-C420 head-mounted condenser microphone with phantom power (Rolls PB23). The recording was digitized at a sampling rate of 44.1 kHz, with 16-bit quantization, and an anti-aliasing filter with a cut-off frequency of 22.05 kHz. Each word was stored in a separate file and each file was set to an equal RMS value.

Synthesized Speech

The synthesized speech stimuli were digitized on a Windows PC using online demonstrations of AT&T Voice Michael ™ adult male voice (AT&T Labs, Inc. – Research) and DECtalk *Perfect Paul* ™ adult male voice version 4.6. (GW Micro.com). The stimuli were digitized at 22.05 kHz using a PC sound card and stored in WAV format. RMS level was set to a common value for all files in order to equate the intensity.

The stimuli were organized according to previously determined word lists. Each session's lists were presented using iTunes digital media player on a Macintosh laptop computer with external, powered speakers. The equality of the intensity of human recorded and synthesized speech stimuli was verified using a sound level meter (Bruel & Kjaer model no. 2203). For convenience of administration, tick marks were indicated on the speakers to corresponded to four intensity levels: 55, 65, 75, and 85 dB SPL (within +/- 2 dB).

Independent Variables

The independent variables were three speech output conditions, vocabulary types (repeated or novel) and repeated exposure with feedback across a total of five experimental sessions. The three speech output conditions were human recorded speech, DECtalk *Perfect Paul* ™ synthesized speech, or AT & T Voice Michael ™ synthesized speech.) The two vocabulary types included repeated words and novel words. Repeated stimuli were presented in each of five experimental sessions; novel stimuli were presented once during the course of five experimental sessions.

Dependent Measure

The dependent variable was speech intelligibility (measured as accurately repeating the single word modeled by the experimenter within 15 seconds of presentation).

Procedures

Practice opportunities. Before beginning the first session, children were presented with three spoken training words, one in each output condition (human recorded, DECtalk *Perfect Paul* ™, and AT & T Voice Michael ™) that would not be included in their experimental session word lists. Up to three additional training words were presented for participants who required additional instruction for the task. Children were instructed, “You will hear a word. After you hear the word, I want you to say the word. Are you ready?” After successfully following the instructions by repeating two words from the training items, the experimental sessions began.

Experimental session on novel stimuli. Participants were informed that they would receive stickers or trinkets during and at the end of the experimental session (and each of the four remaining experimental sessions) based on participation (stickers were delivered with spoken praise such as, “I like how you are working,” or “nice job listening to the words”). No tangible reinforcers were delivered contingent on performance accuracy. At the beginning of each experimental session, the investigator documented ambient sound levels using a sound level meter. When compared to a sound controlled environment, there is evidence that increased background noise levels may result in reduced intelligibility of synthesized speech for adults (Koul & Allen, 1993). To compensate for any ambient noise level, participants chose a comfortable listening level (described below) prior to experimental procedures. The sound level meter recordings were taken to facilitate a replication of noise levels in a sound treated environment at a subsequent date.

The loudness levels of the three speech conditions were verified using a sound level meter. Tic marks were made on the speakers that corresponded to four loudness levels: 55dB, 65 dB, 75 dB and 85 dB (within +/- two dB). After measuring the ambient noise level, the participant chose a comfortable listening level. All stimuli were presented at the decibel sound level chosen by the child. After the children selected their preferred listening level, the investigators instructed the child to repeat the word that they heard. Six practice items were presented, two in each speech output condition.

Participants were then presented with one list of 20 words for each of the speech output conditions. Presentation of three stimuli types was counterbalanced across participants in that participants were randomly assigned to hear stimuli presented in three different orders: (1) DECtalk *Perfect Paul* ™, AT & T Voice Michael ™ and human recorded speech (2) Human recorded speech, DEC talk *Perfect Paul* ™ and AT& T Voice Michael ™ (3) AT&T Voice Michael ™, Human recorded speech and DECtalk *Perfect Paul* ™. Participants completed a list of one speech type before beginning a list of the next speech type. Participants repeated each word they heard within 15 seconds following the word presentation. The investigator provided spoken feedback regarding the accuracy of the response (e.g. That's right the word was "cat" or good try, but the word was "cat"). Investigators immediately transcribed whether the response was accurate or inaccurate. A response was tallied as accurate if it was a lexical match to the word, whether it was produced with a phonological error or not. If the participant did not respond within the 15 -second time limit, the investigator indicated "NR" for no response on the data sheet. Each experimental session was completed in 10 – 20 minutes.

Experimental sessions 2 through 5 with novel and repeated stimuli. Each child participated in five total sessions. For the last four sessions, the stimuli were comprised of ten novel words and ten repeated words for each speech output condition with a total of 60 words presented in each experimental session (20 words in each of the conditions). Experimental sessions 2 through 5 lasted approximately 10 – 20 minutes. The investigators scheduled six sessions (one screening, five experimental). All participants completed screenings and all five experimental sessions within a two-week period.

Procedures for experimental sessions 2 through 5 were identical to the procedures for the first experimental session. The presentation order for the type of voices remained the same throughout the duration of the sessions based on the counterbalanced order assigned to the participant. Children continued to hear 60 words, 20 in each speech output condition (human recorded, DECtalk *Perfect Paul* ™, and AT & T Voice Michael ™). Each list of 20 words contained ten repeated words, randomly distributed throughout the list. Children heard the repeated list of words for each experimental session, however as the word order in each list was randomized, the repeated words were not heard in the same order for each session. The remaining ten words in each list were novel words that were only presented once during the course of the experiment.

Reliability

A speech language pathology graduate student attended 10% of experimental sessions to establish procedural fidelity. The observer monitored a checklist of the steps of the experiment. The independent observer noted whether a step was followed or not. Point by point agreement (agreement/agreement + disagreements x 100) was calculated using the data collected by the independent observer. Procedural reliability was 100%. The same speech language pathology graduate student also scored participants responses in 16.7 % of experimental sessions. Point by point agreement was used to compare the scoring between the primary investigator and the independent observer. Results indicated a response reliability of 99.67%.

Measures and Data Analyses

The dependent measure in this study was speech intelligibility measured by the number of words correctly repeated by the subjects within 15 seconds after the word was presented. In order to achieve the strongest and most valid results, the data were analyzed using a logit mixed-effects models (LME) with crossed random effects subjects and items (Jaeger, 2008; Baayen, Davidson, & Bates, 2008) to test the main effects and interaction effects of voice, session (practice effect) and vocabulary type (repeated or novel). A detailed justification for these models is given by Jaeger and by Baayen et al. Briefly, when categorical outcomes—such as the 'correct' and 'incorrect' recognition scores in this study—are analyzed using subject and item data over proportions or percentages, ANOVAs may create difficulty in the interpretation of the data. ANOVAs over proportions potentially create confidence intervals that extend beyond 0 and 1, resulting in uninterpretable outcomes (Jaeger, 2008 p. 435). For example, in this study a child's response to a vocabulary item was coded as an accurate (1) or in accurate response (0); a confidence interval resulting in a score above one or below zero cannot be clearly defined or interpreted in terms of level of accuracy. Moreover, ANOVA is used under the assumption that the amount of variance between two or more experimental conditions will be equal, which is not the case in the study. The experimental conditions of voice output (DECtalk ™, AT & T voice ™ and human recorded speech), sessions 1 through 5 and vocabulary items (novel, repeated) do not result in an identical amount of variance (as shown in Table 1). Logit mixed-effect models have the ability to model random effects without treating the experimental conditions as having an identical amount of variance (Jaeger, 2008). Statistical simulations have found mixed logit models to be more likely to detect true effects than ANOVA (Dixon, 2008).

The reporting of results differs between the two different types of statistical models. Nominal or categorical data is recoded into numerical values for ANOVAs. In an LME, nominal data is recoded as contrasts between one variable and the others. With the use of ANOVA, data results are typically reported in an F ratio for the sampling distribution of main effects and a p value for the probability. With the use of LME, the data results are reported in terms of significance tests for coefficients in the model. In this case, these are Wald's Z values (and their associated p-values) are used to assess the statistical significance of components of the model. These values are calculated by dividing the estimate of a coefficient by its standard error (Jaeger, 2008: 440).

Recall that a multivariate analysis of variance showed that the word lists did not differ significantly in any of the descriptive measures (log-transformed word frequency, number of phonological neighbors, length in phonemes or syllables, or percentage of children on the MacArthur who recognized the word). However, these variables are known to affect word-recognition accuracy. Consequently, it is possible that the various experimental conditions, such as voice output type, repeated versus novel presentation, and session, interacted with these variables in interesting ways. To examine this, a second set of logit mixed-effects models was constructed to examine these as mediating factors. Here, the use of logit mixed- models is critical, as this is the only class of statistical models that allow an examination of the interaction between voice and lexical characteristics such as neighborhood density and word frequency.

Results

Statistical analyses were conducted to examine the effects and interactions of the independent variables of the three speech output conditions of DECtalk ™, AT&T Voice ™ and human recorded speech, session (practice effect), vocabulary type (repeated/novel), as well as chronological age and the lexical characteristics of word frequency occurrence in the English Language (Kucera and Francis 1967), phonological neighborhood density and recognition accuracy from the MacArthur CDI, (Hoosier Mental Lexicon Database. Pisoni, Nusbaum, Luce, & Slowiaczek, 1985).

Speech Output Type

Table 1 presents the mean and standard deviation for the percentage of words correctly repeated by participants in each speech output condition; synthesized voices DECtalk ™ Paul, AT&T Voice ™ Michael and human recorded speech spoken by an adult male with a Midwestern dialect, for novel and repeated vocabulary words during sessions 1 through 5. Pooled across participants and across five sessions, raw data intelligibility scores for each speech output condition were as follows: DECtalk ™ (M = 84.75%, SD 3.55, range 74.4% - 93.3%), AT & T Voice ™ (M =91.5%, SD 2.75, range 87.7% - 93.8%), and human recorded speech (M = 97.3%, SD 1.55, range 95.5% - 98.8%). Table 2 displays increases and decreases in mean accuracies across 5 sessions for each of the three speech output conditions. Human recorded speech had the highest mean accuracy for session 1 of 96.1% for novel vocabulary and 95.5% for repeated vocabulary. Participants exhibited nonsignificant improved accuracy from session 1 to 5 with a 0.7% increase for novel vocabulary items and 2.3% for repeated vocabulary items. The most improvement for human recorded speech occurred between sessions 3 and 4 for novel vocabulary with an increase of 2.7% for the mean accuracy. Participants' accuracy for repeated vocabulary in human recorded speech remained similar from sessions 2 through 5. The greatest improvement from session to session occurred with DECtalk ™, particularly from session 1 to session 2 for repeated vocabulary words. Participants significantly improved in accuracy for DECtalk ™ with an increase in mean word repetition accuracy between session 1 and session 5 of 8.3% for novel vocabulary and 10.6% for repeated vocabulary.

The first model was a LME with session, speech output condition, and vocabulary type as fixed factors, and subject and word as random factors. The full output of this model is shown in Table 3. The model predicts the log accuracy score for each individual response for each participant. As described earlier, the LME by default uses treatment coding to transform nominal data (which, in this experiment, describes the speech output type and vocabulary type factors) as contrasts between one variable and the remaining variables. For example, in Table 3 *Voice: D* refers to the contrast between the DECtalk ™ condition and the other two speech output conditions. In this model the coefficient indicates the slope of the relationship between the dependent measure and independent measure for each item excluding the intercept. The coefficient that contrasted the DECtalk ™ voice to the other two was significant, indicating that DECtalk ™ was less intelligible than AT & T voice ™ and human recorded speech. The coefficient contrasting human voice with the other two was also significant, indicating that the AT & T Voice ™ was less intelligible (though not statistically significantly so at the $\alpha < 0.05$ level) than human recorded speech (Wald Z = 1.73, p = 0.08).

The effect for session fell just short of statistical significance at the conventional $\alpha < 0.05$ level (Wald Z = 1.9, p = 0.06) averaged across speech output conditions in that subjects increased their accuracy as they participated in additional sessions, resulting in repeated exposure. There was a strong effect for the interaction between session and voice (Wald Z = -3.91, p < 0.001). For this interaction vocabulary items presented in DECtalk ™ were identified significantly less accurately than were those presented in the other speech output types and that this difference decreased across sessions. Items presented in human recorded speech and AT & T Voice ™ were identified more accurately overall, and had a more gradual increase relating session to accuracy than did the items presented in DEC Talk™.

To further understand this interaction, three post-hoc LME models were completed separately for each voice. In the model examining human voice, there was no statistically significant effect related to session, nor of vocabulary novelty, nor was there an interaction. The coefficients suggest that overall performance was very high with little difference related to variables of session and novelty vs. repeated single word items. This is consistent with the raw data in Tables 1 and 2.

The second model examined the effects of session and novelty for AT & T Voice ™. In this model, there were no significant effects of session or of novelty, but these did interact significantly, (Wald Z = 2.13, p = 0.03). An inspection of the coefficients suggest that repeated exposure resulted in increased performance for the AT & T™ voice, while the perception of novel stimuli decreased, albeit minimal, over sessions. Again, this was supplemented with analyses of the raw data. There were smaller increments of improvement, particularly for repeated vocabulary from session 4 to 5 (a decrease in mean word repetition accuracy of 1.7% for novel stimuli) with a decrease in mean word repetition accuracy between session 1 and session 5 of 1.2% for novel vocabulary and an increase of 4.3% for repeated vocabulary. It should be noted that group performance did not improve session to session for all five sessions, particularly for AT & T Voice ™ and human recorded speech, which both had relatively high mean scores for intelligibility for session 1 (89.95% AT & T Voice ™ and 95.8% human recorded speech). Participants occasionally exhibited a decrease in performance accuracy from session to session for repeated or novel stimuli in AT & T Voice ™ and human recorded speech.

The final model examined the influence of session and novelty on the perception of DECtalk™ output stimuli. In this model, there were significant effects of session (Wald Z = 2.09, p = 0.04) and novelty (Wald Z = -2.10, p = 0.04), and these two factors interacted significantly (Wald Z = 2.60, p = 0.01). The coefficients suggest that both the repeated and novel words improved over sessions and that there was a greater improvement in the repeated words than in the novel words.

Vocabulary Type

There was an interaction between session and vocabulary type (repeated/ novel). When averaged across each voice type, repeated vocabulary words resulted in more accurate responses in later sessions than novel vocabulary words. However, this interaction between session and voice failed to reach statistical significance using the conventional significance criterion of $\alpha < 0.05$. DECtalk ™ (Wald Z = 1.90, p = 0.06) had a slightly larger effect of session than the other speech output conditions.

Chronological Age

A second set of analyses examined the possible effects of chronological age on performance. The goal of these analyses was to determine whether the modest differences in participants' ages mediated the effects of repeated exposure to stimuli and the discrepancies among different voice-output conditions. In the first analysis, the linear mixed-effects model was re-run with chronological age as an additional fixed factor. In this model, there was only one higher-order interaction that was significant. Specifically, age interacted with voice and session. A series of post-hoc models showed that for one voice-vocabulary combination, DECtalk™ repeated stimuli, the older children showed significantly greater growth across sessions than did the younger children. In this model participants are binned into four age groups, which were quartile groupings of the observed ages. The four quartiles had average ages of 2.58 years, 3.0 years, 3.25 years, and 3.5 years. As this shows, this interaction arose because the older participants had unexpectedly low accuracy scores for the DEC Talk™ repeated stimuli in session one, however all of the stimuli for session one are novel in that the participants had no previous exposure to them. It should be noted that the three oldest participants were inadvertently assigned to a voice output condition order in which DEC Talk™ was presented last. Speech output condition order had been counterbalanced in order to control for the potential of effects of voice presentation order on word repetition accuracy. Based on the raw data for session one, participants performed slightly less accurately when presented with DEC Talk™ after hearing the other speech output conditions than participants who were assigned to the other speech output condition orders. This may explain the unexpected low accuracy scores for the oldest participants.

The results of this model should be interpreted with caution as the chronological age range of participants is narrow, with an age range of 17 months and there was a heavier distribution of participants at the 3 years old to 3 years 6 month old end of the scale than participants at the 2 year 6 month old to 3 year old end of middle to high range of inclusionary ages and fewer participants of the sample representing the youngest age range of inclusion in the study.   No other effects of age were found in this analysis.

In a second analysis, the linear mixed-effects model was re-run to generate the two sets of slope coefficients for individual subjects. The first set of slopes reflected the growth in accuracy (pooled across the three speech-output types) across the 5 sessions. The second set of slopes indexed the effect of speech-output condition on accuracy (pooled across the 5 sessions) for individual subjects. Neither set of slopes correlated with age. Together, these analyses show that age did not mediate the main effects and interactions noted in the previous analysis. This point is revisited in the discussion.

Lexical Characteristics

The final set of analyses examined the possible effects of lexical characteristics on performance. The goal of these analyses was to determine whether the effects of voice-output type and session were equivalent for words that varied in factors that have been shown previously to affect word-recognition accuracy. A series of linear mixed-effects models were run with each of three lexical factors added as an additional fixed effect: word frequency occurrence in the English Language (Kucera and Francis 1967), phonological neighborhood density, (Hoosier Mental Lexicon Database. Pisoni, Nusbaum, Luce, & Slowiaczek, 1985) and recognition accuracy from the MacArthur CDI (Fenson, et al., 1993).

Phonological neighborhood density refers to the amount of words that can be created by adding, deleting or substituting one phoneme in the word. Words that are high in neighborhood density have many phonological neighbors while words that are low in neighborhood density have few phonological neighbors. Words that are high in frequency and low in phonological neighborhood density have been found to facilitate generalization to untrained words or sounds (Gierut, Morrison and Champion, 1999).

Neither word frequency nor MacArthur accuracy percentage predicted recognition accuracy, nor did they interact with any of the factors that were found to be significant previously. In the model that included phonological neighborhood density, there was a significant interaction among the number of neighbors, session, speech-output type, and vocabulary type. This interaction was explored in six post-hoc LME models that examined the interaction between phonological neighborhood density and session separately by speech-output condition and by vocabulary type. In four of these models, neighborhood density had a significant main effect, a significant interaction with session, or both.

The first two analyses examined DECtalk™ repeated and novel, respectively. In both of these, the number of neighbors affected accuracy rates, albeit only at the less-stringent a < 0.10 level (repeated: Wald Z = 1.664, p =0.096; novel: Wald Z = -1.701, p = 0.089). As in previous research, the words with the smallest number of neighbors were identified significantly more accurately than those with greater numbers of neighbors across the five sessions.

In the model examining the AT & T™ repeated vocabulary, there was a significant interaction between the number of neighbors and session (Wald Z = 2.894, p = 0.004). The accurate repetition of words with the highest number of neighbors actually decreased over sessions, while those with relatively fewer neighbors increased, though overall rates of accuracy were exceedingly high in this condition.

Finally, the results of the model for the human voice repeated vocabulary showed a significant effect of number of neighbors (Wald Z = -1.845, p = 0.065), as well as a significant interaction between number of neighbors and session (Wald Z = 2.135, p - 0.033). This model shows exactly the opposite of what is shown by the model for the AT & T ™ voice. For human-voice words, only those with low neighborhood densities increased in recognition accuracy over sessions, while those with high densities actually decreased, though again, overall accuracy rates were exceptionally high.

Based on statistical analyses using a LME to examine the effects and interactions of the independent variables of the three speech output conditions, session (practice effect), vocabulary type (repeated/novel), as well as chronological age and lexical characteristics, significant effects and/or interactions were indicated, particularly for DECtalk ™ speech output condition as a main effect as well as interaction between the other independent variables and fixed effects. The most important findings of this study are that based on the results of statistical analyses, DECtalk ™ was significantly less intelligible than AT & T voice ™ and human recorded speech. This finding is consistent with an examination of the raw data. There was a strong effect for the interaction between session and voice for DECtalk ™ with participants significantly improving word imitation accuracy across five experimental sessions.

Discussion

In this study, the effects of repeated listening opportunities and repeated listening vocabulary on preschool aged children's speech intelligibility performance between two different speech synthesized voices, DECtalk ™ Paul and AT&T Voice ™ Michael compared to recordings of male human speech were examined. Participants performed as had been hypothesized; with greater word repetition accuracy for human recorded speech than AT&T Voice ™ and DECtalk ™ and greater word repetition accuracy for AT&T Voice ™ than DECtalk ™. Although it didn't reach statistical significance, participants improved in accuracy across sessions for both novel and repeated vocabulary words, suggesting that they may have been able to generalize their learning to become more familiar with the specific synthesized speech voices, rather than particular vocabulary words that were presented several times. These findings are similar to previous synthesized speech intelligibility findings with elementary and preschool aged students (Drager et al. 2006; McNaughton et al. 1994; Mirenda & Beukelman, 1987).

An increase in word repetition accuracy of was achieved with minimal feedback given to the child ("That's right" for accurate repetitions and "Good try, but the word was ____"for inaccurate repetitions). Previous studies suggested that single word synthesized were difficult for child and adult listeners to understand (Mirenda & Beukelman, 1987, 1990) as a result of the lack of contextual cues. The results of this study indicate that based on repeated exposure to novel synthesized speech vocabulary items, typically developing children as young as 2 years 7 months to 3 years 6 months can generalize their improved abilities to understand novel single word vocabulary items presented in synthesized speech with minimal training. However, it should be noted that all of the words included in the experimental sessions were selected based on a high level of recognition accuracy by young children indicating that the words were likely to be recognized by participants in this investigation. The outcome may have been different had the researchers selected words that were less familiar to the participants.

Repeated Exposure

Repeated exposure occurred across five experimental sessions. The effects of repeated exposure varied as a function of the speech output conditions. Although it did not reach statistical significance, across the three speech output conditions, repeated vocabulary items were more accurate than novel vocabulary items in later sessions. The interaction of vocabulary and session was greatest for DECtalk ™ followed by AT&T Voice ™ and human recorded speech, respectively. This variation was most likely related to the amount of initial errors made in each speech output condition, in that the participants made the most word repetition errors in session one for DECtalk ™, followed by AT&T Voice ™ and finally the fewest initial word repetition errors were made in the human recorded speech output condition. Throughout all five sessions, DECtalk ™ was the least intelligible of the speech output conditions, which created more opportunities for improvement. Practice effects were the least noticeable in human recorded speech, which was highly intelligible for all five sessions, resulting in a ceiling effect.

When comparing the results with those obtained by McNaughton et al. (1994) child for 6 – 10 years old, the group mean word repetition accuracies for DECtalk ™ are similar. The greatest increases in word repetition accuracy across sessions occurred from session one to session two for repeated vocabulary. The participants in the McNaughton et al. (1994) study had lower word repetition accuracy for the initial DECtalk ™ session, which provided additional opportunities to improve their accuracy across sessions. Participants in this study decreased their performance accuracy in between sessions three and four for novel DECtalk ™ vocabulary items. This phenomenon was noted in McNaughton et al. (1994) between sessions four and five for repeated DECtalk ™

vocabulary items. In this study this inconsistent pattern of growth of mean accuracies between sessions is evident for AT&T Voice ™ and human recorded speech as well. It is possible that this pattern can be attributed to the young age of the participants, which may have limited the quality and quantity of their engagement in the experimental task over time. It may be that the participants in this study became habituated with the task and reduced their level of engagement earlier in their sessions than the slightly older child participants in the McNaughton et al. (1994) study.

Chronological Age

Chronological age did not affect accuracy of word repetition, but did interact with voice and session in that the older children showed significantly greater growth across sessions for DECtalk ™ than the younger children. These findings should be interpreted with caution, as the age range of participants (2 years 7 months to 3 years 6 months) was very narrow and limited to young children with typically developing vocabulary as measured by the PPVT- IV (Dunn & Dunn, 2007). Considering these factors, chronological age was likely not a significant factor for word repetition accuracy in the current investigation. However, comparing performance longitudinally or crossectionally across groupings of a wider range of ages represents an important area for future inquiry.

The effect of limited attention may have a greater influence on young children's performance with synthesized speech than with less degraded natural speech. Drager & Reichle (2001a) examined this variable in young and older adults. They found that listeners' comprehension of synthetic speech significantly decreased when attention was divided. Young children have a more limited working memory capacity than older children and adults, affecting the attentional resources that they have available to them. It is possible that fewer available attentional resources inherent among young children may have a greater effect on their perceived intelligibility of a degraded speech signal. This cognitive difference may affect age differences in synthesized speech intelligibility tasks that have been noted in earlier studies (Greene, 1983; Mirenda & Beukelman, 1987, 1990). Young children are also less equipped to take advantage of the benefits of experience and background knowledge when trying to decipher a synthetic speech message and must rely more heavily on the information contained within the signal (Case, 1985; Dempster, 1981).

Lexical Characteristics

Vocabulary items were arranged in balanced lists based on the results of a multivariate analysis of variance as a precaution to minimize the effect of the variables of phonological neighborhood density, phoneme length and syllable length, word frequency in the English language, and frequency rating on the MacArthur-Bates CDIs (Fenson, et al., 1993). These precautions appeared to have been successful in that there was a group effect for voice, which suggests that performance accuracies across speech output conditions were not related to the lexical characteristics of the vocabulary words.

In previous studies, phonological neighborhood density has been shown to impact children's word recognition ability (Munson, 2001). Perhaps due to the young age and limited vocabulary size of the participants', it did not appear to affect word recognition in this study. As stated in the methods, a MANOVA was used to create statistically equivalent word lists in terms of the variables related to lexical characteristics (log-transformed word frequency, number of phonological neighbors, length in phonemes or syllables, or percentage of children on the MacArthur who recognized the word, all F

[20,188]'s < 1, all p's > 0.50.) However, within each wordlist, phonological neighborhood density was allowed to vary. When analyzed separately, phonological neighborhood density did interact with a number of fixed factors, particularly in the human recorded speech condition. Words having more phonological neighbors decreased in accuracy over the course of the five sessions, while words with fewer phonological neighbors tended to increase, suggesting that training words from sparse neighborhoods may result in generalizations to untrained words in more dense neighborhoods. This may have been due to a reduction in the quality and quantity of attention to this speech output condition over time. Participants experienced the greatest initial success in this condition. Consequently, it appears that the ease of the task may have resulted in reduced attention when compared to their behavior during the synthesized speech output conditions.

The interaction between session and phonological neighborhood density for human recorded speech is exactly opposite of the same interaction for AT & T ™ voice. For human recorded speech, participants increased word repetition accuracy over sessions for vocabulary items with low neighborhood densities and decreased in accuracy for those with high phonological neighborhood densities. This asymmetry may have been due to the effect of neighborhood density on fine phonetic detail in natural speech; phonetic information that is likely not accounted for in synthetic speech, even when it is highly intelligible such as AT & T ™ voice. It is possible that these differences in phonetics can mediate the effect of neighborhood density on word recognition.

Listening Conditions

Drager et al. (2006) examined intelligibility of single words in context, single words out of context, sentences in context and sentences out of context within the presence of background noise with 3, 4 and 5 year olds using digitized speech, DECtalk ™ synthesized speech, MacinTalk™ synthesized speech and human recorded speech. For single out of context words in DECtalk ™, intelligibility for the 3 - year old group was 45%. During session one in the current investigation, the participants novel vocabulary items in DECtalk ™ were 80.5% intelligible while repeated vocabulary items in DECtalk ™ were 74.4% intelligible. Given the limited number of studies with which to compare these results, participants in this study appeared to perform with greater accuracy than in other studies. However, this may be related to the highly familiar words selected for the current investigation.

This is the first study examining the intelligibility of AT&T Voice ™, which was more intelligible than DECtalk ™ among the preschool age participants in this study. AT&T Voices ™ are now commercially available for the most recent DynaVox Mayer Johnson SGDs V ™ and Vmax™ as well as the Prentke Romich ECO2 ™. Until recently, AT&T Natural Voices ™ hadn't been widely available, which may explain why it has not been included in synthesized speech intelligibility studies. DECtalk ™ continues to be relevant as it still used in commercially available devices.

Participants in this study were slightly more accurate for novel and repeated DECtalk ™ vocabulary items than the older children in the McNaughton et al. (1994) study. This difference occurred under less optimum listening conditions than those in McNaughton et al. (1994) in which children wore headphones, through which the researchers presented the stimuli in a sound controlled environment. In this study there was a background noise level between 50 – 67 dB with a mean of 56.4 dB.

In McNaughton et al. (1994), participants listened to 80 vocabulary items per session compared to 60 items per session in this study. This difference may have resulted in slightly longer experimental sessions, which, in turn, may have increased attention demands on the participants.

Limitations and Future Research

One of the limiting factors in this study was the size of the sample. Additionally, single word stimuli were utilized. Using sentences or short narratives would have provided a greater level of social (external) validity in that most communication is comprised of sentences. On the other hand, although sentences provide the listener with additional linguistic context to assist in intelligibility, synthesized speech users cannot rely solely on preprogrammed sentences. Young children often use single words for communication, making them a relevant population for study.

Future investigations should compare the performance between children of different age groups. The children in this study were typically developing and in the prescreening Peabody Picture Vocabulary Test- IV (Dunn & Dunn, 2007), all of the participants scored within the high average range of 101 to 114. It is possible that language comprehension status may influence synthesized speech intelligibility in that children with more rudimentary comprehension skills may be working within the constraints of a more limited working memory capacity than children with better comprehension skills (Case, 1985; Dempster, 1981). This difference may impact the attentional resources available for deciphering synthesized speech.

The single word stimuli used in this study were selected from the MacArthur-Bates CDIs (Fenson et al. 1993) because they were words that had been documented to be comprehended by 88.6% of children ages 30 months and older. Performance accuracy may be quite different with more challenging vocabulary items. Training may require sessions with more explicit training in order to attain similar results related to practice effect. Since preschool and school age children are frequently introduced to new vocabulary in school that is challenging in that it is novel to them, it would be valuable to examine the influence of challenging vocabulary.

This study examined the effects of repeated exposure over the duration of five experimental sessions. There are currently no studies published that examine these effects for more than five experimental sessions. Additional studies should examine this variable as the significant difference between DECtalk ™ and AT&T Voice ™ may become less significant with additional exposure. In this study DECtalk ™ never attained the same level of intelligibility as AT&T Voice ™ and human recorded speech after five experimental sessions. However, participants exhibited a slower rate of increase of word imitation accuracy for DECtalk ™ and consistent high levels of performance accuracy for AT&T Voice ™ and human recorded speech across the five experimental sessions. It is possible that performance accuracy for DECtalk ™ and AT&T Voice ™ could continue to increase with additional exposure, ultimately resulting in equivalent intelligibility performances between DECtalk ™ and AT&T Voice ™ or AT&T Voice ™ and human recorded speech. If there were sufficient evidence to support equivalent intelligibility over time, it could potentially provide an indication that measures for assessments should be collected over a period of time instead of a single event.

There are very few studies that examine the effect of previous exposure to synthetic speech with an intervening period of non -systematic exposure to it (Helsel – Dewert, & van den Meiracker, 1987). Based on the results of this study, it is impossible to accurately predict whether or not improved synthesized speech intelligibility accuracy can be maintained over time. In this study, participants experienced repeated exposure within one to three days apart for a total of five experimental sessions. Practice effect may not result in a maintained improvement over time when there are periods of non-exposure to synthesized speech during the maintenance period. Future studies should examine the effect of previous exposure with an intervening period of non- systematic exposure on the maintenance of newly acquired skills.

Educational Implications

The minimal amount of exposure required to improve synthesized speech intelligibility for the young participants in this study, may be useful when evaluating SGDs to implement with preschool age children. It may be possible to perform brief assessments or performance probes over the duration of a week or more to attain a baseline level of synthesized speech intelligibility. The minimal amount of performance feedback required by participants in this study suggests that this type of training could easily be incorporated into a learner's typical routine with staff trained to provide feedback.

As the numbers of young children with ASD who use a SGD increase (Binger & Light, 2006; Mirenda, 2009), one should have a measure of intelligibility and factors that affect performance accuracy for implementing an AAC system with synthetic speech output. The social communication network of children with ASD is likely to include other children who have communication or cognitive impairments (i.e. Early Childhood Special Education classroom). The recommendations proposed by the NRC Committee of Educational Interventions for Children with Autism[1] emphasize the importance of developing appropriate social skills with peers, thus peers must be able to understand the SGD user's message. One would expect that young children with cognitive and/or communication impairments may require additional training opportunities, more explicit feedback related to performance and additional context for the communicative message.

In order to inform evidence based practices, additional synthesized speech intelligibility research should be conducted in socially valid environments, such as environments where young children will potentially need to listen to and understand synthesized speech. Performance accuracies that are the result of experimental sessions that occur in sound controlled audio booths do not reflect the conditions of typical listening environments and there is not enough evidence currently available to inform practice of optimum listening levels for typical sound environments, such as elementary and preschool classrooms.

[1] The NRC Committee on Educational Interventions for Children with Autism recommended that emphasis be placed on the use of evidence-based instructional techniques in six main instructional area: 1) functional, spontaneous communication using speech and/or AAC; 2) developmentally appropriate social skills with parents and peers; 3) play skills with peers; 4) various goals for cognitive development, with emphasis on generalization; 5) positive behavior supports; 6) functional academic skills, as appropriate (2001).

However, additional sound controlled studies should be conducted to examine the differences between DECtalk ™ and AT&T Voice ™ by as a function of different levels of background noise. The overall higher performance accuracy for AT&T Voice ™ may allow individuals to more accurately perceive synthesized speech in the presence of higher levels of background noise than for DECtalk ™ before the signal begins to degrade. This may prove to be an important distinction between the two synthesized speech voices as elementary school classroom noise levels can range from 55 – 65 dBA (Nelson, Soli, & Seitz, 2002). Individuals who use a SGD in these environments may potentially be more intelligible to their peers and educators when they use AT&T Voice ™ for their speech output. This may be one of several important considerations to take into account when assessing voice outputs for school age children who use speech-generating devices.

An additional important factor to consider is that the population of public school students represents increasingly diverse races, cultures, ethnic groups and languages. Although the research is sparse, there is some evidence to support that bilingual children experience greater difficulty than monolingual children in for synthetic speech intelligibility. (Axmear et al. 2005). Future research should consider synthesized speech intelligibility from the perspective of listeners who are learning English as a second language.

In summary, children as young as ages 2 years 7 months to 3 years 6 months were able to understand synthesized speech in typical listening environments such as a preschool classroom. DECtalk ™ intelligibility was significantly lower than human recorded speech and AT&T Voice ™ for the participants. Participants significantly improved their performance accuracy over repeated exposures to DECtalk ™, but never attained the same level of accuracy that was attained with AT&T Voice ™ or human recorded speech. Although it did not reach statistical significance, intelligibility accuracy improved for both novel and repeated single word vocabulary items, indicating that participants' gains may be due to practice effect generalized from trained to untrained single words presented in the same synthetic speech voice.

References

American Speech-Language-Hearing Association. (2005). *Guidelines for Addressing Acoustics in Educational Settings* [Guidelines]. Available from www.asha.org/policy.

Axmear, E., Reichle, J., Alamsaputra, M., Kohnert, K., Drager, K., & Sellnow, K. (2005) Synthesized speech intelligibility in sentences: A comparison of monolingual English speaking and bilingual children. *Language, Speech, and Hearing Services in Schools, 36,* 244-250.

Baayen, R.H., Davidson, D.J., and Bates, D.M. (2008) Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language 59,* 390 – 412.

Beukelman, D. & Mirenda, P. (2005) Educational inclusion of students who use AAC. In D. Beukelman & P. Mirenda (eds.) Augmentative and Alternative Communication (p.392). Baltimore: Paul H. Brookes Publishing Co.

Binger, C. & Light, J. (2006) Demographics of preschoolers who require AAC. *Language, Speech, and Hearing Services in Schools, 37,* 200-208.

Case, R. (1985) *Intellectual development: Birth to adulthood.* Toronto, Ontario, Canada: Academic Press.

Dempster, F.N. (1981) Memory span: Sources of individual and developmental differences. *Psychological Bulletin, 89,* 63-100.

DiCarlo, C.F. & BanaJee, M. (2000). Using voice output devices to increase initiations of young children with disabilities. *Journal of Early Intervention, 23,* 191 – 199.

Drager, K. D. R., Reichle, J. and Pinkoski, C. (in press) Synthesized speech output and children: a scoping review.

Drager, K.D.R., & Reichle, J.E. (2001a) Effects of age and divided attention on listeners' comprehension of synthesized speech. *Augmentative and Alternative Communication, 17,* 109-119.

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

Drager, K.D.R., Clark-Serpentine, E.A., Johnson, K.E. & Roeser, J.L. (2006) Accuracy of repetition of digitized and synthesized speech for young children in background noise. *American Journal of Speech-Language Pathology, 15,* 155-164.

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

Dunn, L., & Dunn, L. (2007) Peabody Picture Vocabulary Test. Circle Pines, MN: American Guidance Service, Inc. Publishing.

Fenson, L., Dale, P.S., Reznick, J.S., Bates, E., Pethick, S.J., Hartung, J & Reilly, J. (1993). *MacArthur-Bates Communicative Development Inventories.* San Diego, CA: Singular Publishing Group (Thomson Learning).

Foulds, R. (1980). Communication rates of nonspeech expressions a function of manual tasks and linguistic constraints. In *Proceedings of the International Conference on Rehabilitation Engineering* (pp. 83 – 87). Toronto: RESNA Press.

Foulds, R. (1987) Guest editorial. *Augmentative and Alternative Communication, 3,* 169.

Fucci, D., Reynolds, M.E. Bettagere, R., & Gonzales, M.D. (1995) Synthetic speech intelligibility under several experimental listening conditions. *Augmentative and Alternative Communication, 11,* 113-117.

Goldman- Eisler, (1986). *Cycle linguistics: Experiments in spontaneous speech.* London: Speechmark Publishing.

Gierut, J.A., Morrisette, M.L., & Champion, A.H. (1999) Lexical constraints in phonological acquisition. *Journal of Child Language, 26 no. 2,* 261-94.

Greene, B.G. (1983) Perception of synthetic speech by children. *Journal of the Acoustical Society of America, 73,* S28-29.

Greene, B. G., Logan J.S., & Pisoni, D.B. (1986). Perception of synthetic speech produced automatically by rule: Intelligibility of eight text-to-speech systems. *Behavior Research Methods, Instruments, Computers, 18* 100 – 107.

Greenspan, S.L., Nusbaum, H.C. & Pisoni, D.B. (1985) *Perception of synthetic speech generated by rule: Effects of training and attentional limitations* (Progress Report no. 11). Bloomington, IN: Indiana University, Department of Psychology

Hale, S. (1990) A global developmental trend in cognitive processing speed. *Child Development, 61,* 653-663.

Hale, S., Fry, A. F., & Jessie, K.A. (1993). Effects of practice on speed of information processing in children and adults: Age sensitivity and age invariance. *Developmental Psychology, 29,* 880 – 892.

Halle, J., Brady, N.C., & Drasgow, E. (2004) Enhancing Socially Adaptive

Communicative Repairs of beginning communicators with disabilities. *American Journal of Speech-Language Pathology, 13*, 43 – 54.

Helsel – Dewert, M., & van den Meiracker, M. (1987) The intelligibility of synthetic speech to learning handicapped children. *Journal of Special Education Technology, 9,* 38-44.

Higginbotham, D. J., Drazek, A. L., Kowarsky, K., Scally, C., & Segal, E. (1994). Discourse comprehension of synthetic speech delivered at normal and slow presentation rates. *Augmentative and Alternative Communication, 10*, 191-202.

Higginbotham, D.J., Scally, C.A., Lundy, D.C., & Kowarsky, K. (1995). Discourse comprehension of synthetic speech across three augmentative and alternative communication (AAC) output methods. *Journal of Speech, Language, and Hearing Research, 38,* 889-901.

Hustad, K., Kent., R.D., & Beukelman, D.R. (1998). DECtalk™ and MacinTalk speech synthesizers: Intelligibility for three listener groups. *Journal of Speech, Language, and Hearing Research, 41,* 744 – 752.

Jaeger, T. Florian, (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language, 59,* 434-446.

Kalikow, D.N., & Stevens, K.N., (1977) Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability. *Journal of the Acoustical Society of America, 61,*1337 – 1351.

Kangas, K.A., & Lloyd, L. L. (1988) Early cognitive skills as prerequisites to augmentative and alternative communication use: What are we waiting for? *Augmentative and Alternative Communication, 4,* 211 – 221.

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* 751-755.

Knecht, H., Nelson, P., Whitelaw, G., & Feth, L. (2002). Background noise levels and reverberation times in unoccupied classrooms: predictions and measurements. *American Journal of Audiology, 11,* 65-71.

Koul, R., & Clapsaddle, K.C. (2006). Effects of repeated listening experiences on the perception of synthetic speech by individuals with mild-to-moderate disabilities. *Augmentative and Alternative Communication, 22,* 112-122.

Koul, R., & Hanners, J. (1997). Word identification and sentence verification of two synthetic speech systems by individuals with mental retardation. *Augmentative and Alternative Communication, 13,* 99-107.

Koul, R., & Hester, K. (2006). Effects of listening experiences on the recognition of synthetic speech by individuals with severe intellectual disabilities. *Journal of Speech, Language, and Hearing Research, 49,* 47-57.

Kucera, H., & Francis, W. (1967) *Computational Analysis of Present-Day American English.* Providence: Brown University Press.

Light, J., Drager, K., Hayes, E. Kristiansen, L., May, H., Page, R. et al. (2004, November). *AAC interventions to maximize language development for young* *children.* Seminar presented at the Annual Convention of the American Speech-Language-Hearing Association, Philadelphia.

McNaughton, D., Fallon, K., Tod, J., Weiner, F., & Neisworth, J. (1994). Effect of repeated listening experiences on the intelligibility of synthesized speech. *Augmentative and alternative communication, 10,* 161-168.

Miller, L.T., & Vernon, P.A. (1997) Developmental changes in speed of information processing in young children. *Developmental Psychology, 33,* 549-554.

Mirenda, P., (2009) Introduction to AAC for individuals with Autism Spectrum Disorders. In Mirenda, P. & Iacono, T. eds. Autism Spectrum Disorders and AAC (pp 13 – 14). Baltimore: Paul H. Brookes Publishing Co.

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

Mirenda, P., & Beukelman, D.R. (1990). A comparison of intelligibility among natural speech and seven speech synthesizers with listeners from three age groups. *Augmentative and Alternative Communication, 6,* 61-68.

Munson, B. (2001). Relations between vocabulary size and spoken word recognition in children aged 3 to 7. *Contemporary Issues in Communication Science and Disorders, 28,* 20-29.

National Research Council, Committee on Educational Interventions for Children with Autism, Division of Behavioral and Social Sciences and Education. (2001). *Educating children with autism.* Washington, DC: National Academies Press.

Nelson, P., Soli, S., & Seitz, A. (2002). *Acoustical barriers to learning.* Melville, NY: Technical Committee on Speech Communication of the Acoustical Society of America.

Panorska, A., Uken, D., & Qeaden, F. (2009) DECtalk™ and VeriVox™: Intelligibility, Likeability, and Rate Preference Differences for Four Listener Groups. *Augmentative and Alternative Communication, 25 no. 1,* 7-18.

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. Bloomington, IN: Indiana University, Department of Psychology, Research Laboratory

Pisoni, D., Nusbaum,H., Luce, P., & Slowiaczek, L. (1985) Speech perception, word recogniton, and the structure of the lexicon. *Speech Communication, 4,* 75-95.

Reynolds, M.E., & Fucci, D. (1998). Synthetic speech comprehension: A comparison of children with normal and impaired language skills. *Journal of Speech, Language, and Hearing Research, 41,* 458-466.

Reynolds, M. E., Isaacs – Duvall, C., Sheward, B., & Rotter, M. (2000). Examination of the effects of listening practice on synthesized speech comprehension. *Augmentative and Alternative Communication, 16,* 250 – 258.

Reynolds, M., & Jefferson, L. (1999). Natural and synthetic speech comprehension: Comparison of children from two age groups. *Augmentative and Alternative Communication, 15,* 174-182.

Romski, M., & Sevcik, R. (2005) Augmentative communication and early intervention: Myths and realities. *Infants and Young Children, 18,* 174 – 185.

Schum, D.J. & Matthews, L.J. (1992) SPIN test performance of elderly hearing impaired listeners. *Journal of the American Academy of Audiology, 3* 303 – 307.

Segers, E., & Verhoeven, L. (2005). Effects of lengthening the speech signal on auditory word discrimination in kindergartners with SLI. *Journal of Communication Disorders, 38,* 499-514.

Von Berg, S., Panorska, A., Uken, D., & Qeadan, F. (2009). DECtalk^{TM} and VeriVox ^{TM}: Intelligibility, likeability, and rate preference differences for four listener groups. *Augmentative and Alternative Communication, 25,* 7-18.

Weitz, C., Dexter, M., & Moore, J. (1997). AAC and children with developmental disabilities. In S.L. Glennen & D.C. DeCoste (Eds.), *Handbook of augmentative and alternative communication* (pp. 395 – 431). San Diego, CA: Singular.

Wetherby, A.M. & Prizant, B.M. (2004). Introduction to Autism Spectrum Disorders. In A.M. Wetherby & B.M. Prizant (Eds.), *Autism Spectrum Disorder: A transactional developmental perspective* (pp. 11 – 30). Baltimore: Paul H. Brookes Publishing Co.

Tables

Table 1: Group means for participants percent and standard deviations of participants’ correct imitations of repeated and novel single words across each of five sessions for DECtalk™, AT&T™ synthesized speech and human recorded speech.

			Session			
Condition	1	2	3	4	5	Total (%)
DECtalk™ novel						
Mean %	80.5	82.7	86.6	83.8	88.8	84.5
SD	3.9	3.7	3.4	3.6	3.1	3.6
DECtalk™ repeated						
Mean %	74.4	85.0	85.5	89.4	93.3	85.0
SD	4.3	3.5	3.5	3.0	2.5	3.5
AT & T Voice ™ novel						
Mean %	92.2	92.7	90.0	91.1	89.4	91.0
SD	2.6	2.5	3.0	2.8	3.0	2.8
AT & T Voice ™ repeated						
Mean %	87.7	92.7	93.3	92.2	93.8	92.0
SD	3.2	2.5	2.5	2.6	2.4	2.7
Human Speech novel						
Mean %	96.1	95.5	96.1	98.8	97.7	96.8
SD	1.9	2.0	1.9	1.0	1.4	1.7
Human Speech repeated						
Mean %	95.5	9.83	98.3	98.3	98.8	97.8
SD	2.0	1.2	1.2	1.2	1.0	1.4

Table 2: Group mean percentages of correct single word imitations with increase/decrease in mean accuracies across each of five sessions for DECtalk™, AT&T synthesized speech and human recorded speech for repeated and novel single word vocabulary items.

			Session				
Conditions	1	2	3	4	5	Total (%)	1-5 total
DECtalk™ novel							
Mean %	80.5	82.7	86.6	83.8	88.8	84.5	+8.3
growth		+2.2	+3.9	-2.8	+5		
DECtalk™ repeated							
Mean %	74.4	85.0	85.5	89.4	93.3	85.0	+10.6
growth		+10.5	+.5	+3.9	+3.9		
AT & T Voice ™ novel							
Mean %	92.2	92.7	90.0	91.1	89.4	91.0	-1.2
growth		+.5	-2.7	+1.1	-1.7	+1.6	
AT & T Voice ™ repeated							
Mean %	87.7	92.7	93.3	92.2	93.8	92.0	+4.3
growth		+5	+.6	-1.1	+1.6	-1.8	
Human Speech novel							
Mean %	96.1	95.5	96.1	98.8	97.7	96.8	+.7
growth		-.6	+.6	+2.7	-1.1	-.9	
Human Speech repeated							
Mean %	95.5	98.3	98.3	98.3	98.8	97.8	+2.3
Growth		+2.8	=	=	+.5	-1	

Table 3. Estimates for coefficients, standard error, t-tests and p – values for session and speech output condition effect on human recorded speech, DECtalk™ and AT&T™ synthesized speech.

Factor	Coefficient	St. Err.	t(df=1)	p-value
(Intercept)	2.420866	.309789	7.815	<0.0001
Session	0.175832	.092666	1.897	0.0578
Voice: DECtalk™ (D)	-1.402149	0.359044	-3.905	0.0001
Voice: Human recorded (H)	.996887	.577042	1.728	0.0841
Novel/repeated (N/R)	.647071	.417468	1.550	0.1211
Session *x* Voice: D	0.227869	0.119438	1.908	0.564
Session *x* Voice: H	0.173794	0.205570	0.845	0.3979
Session *x* N/R	-0.269789	0.128045	-2.107	0.0351
Voice: D *x* N/R	0.007911	0.519333	0.015	0.9878
Voice: H *x* N/R	-0.830946	0.783349	-1.061	0.2888
Session *x* Voice: D *x* N/R	0.005634	0.164122	0.034	0.9726
Session *x* Voice: H *x* N/R	0.174773	0.268459	0.651	0.5150

Appendix A

Participant characteristics, results of PPVT vocabulary comprehension screening and speech output condition assignment (D = DECtalk™ , A = AT & T Voice ™ , H = Human recorded speech).

Participant Code	Gender	Age	PPVT score	Speech output condition order
SSA	F	2 YR 9 M	114	DAH
SSB	F	3 YR 0 M	110	DAH
SSC	M	3 YR 0 M	108	DAH
SSD	F	3 YR 2 M	113	LDA
SSE	F	3 YR 2 M	110	AHD
SSF	F	3 YR 6 M	101	AHD
SSG	F	3 YR 3 M	112	AHD
SSH	M	3 YR 0 M	112	HDA
SSI	F	3 YR 4 M	106	HDA
SSJ	M	2 YR 9 M	111	DAH
SSK	M	3 YR 4 M	108	DAH
SSL	M	3 YR 4 M	114	DAH
SSM	M	3 YR 5 M	113	AHD
SSN	F	2 YR 11 M	113	AHD
SSO	M	3 YR 0 M	111	AHD
SSP	M	2 YR 7 M	102	HDA
SSQ	F	3 YR 4 M	112	HDA
SSR	M	3 YR 3 M	113	HDA

Appendix B Verbal Imitation screening word list

Pull
Smile
Nail
Bring
Quiet
White
Splash
Today
Big
Hello

Appendix C

Single word stimuli vocabulary list with lexical characteristics of frequency of word recognition percentage for 30-month year olds on the MacArthur-Bates CDIs (Fenson, et al.,1993) phonological neighborhood density(Hoosier Mental Lexicon Database. Pisoni, Nusbaum, Luce, & Slowiaczek 1985), frequency of occurrence in English language (Kucera and Francis 1967), phoneme length and syllable length.

word	McArthur %	Number of phonological Neighbors	Frequency	Phoneme Length	Syllable Length
carrots	88.6	0	1	5	2
napkin	92.9	0	3	6	2
balloon	100	2	10	5	2
broom	94.3	10	2	4	1
tape	91.4	16	35	3	1
pig	92.9	19	8	3	1
duck	95.7	25	9	3	1
boat	95.7	32	72	3	1
me	95.7	32	1181	2	1
boy	94.3	13	242	2	1
orange	95.7	2	23	5	2
crayon	97.1	2	1	5	2
pancake	90	0	1	6	2
swing	94.3	12	24	4	1
table	95.7	9	198	4	2
bus	94.3	20	35	3	1
kiss	98.6	13	17	3	1
play	95.7	16	200	3	1
girl	92.9	16	220	3	1
door	100	13	312	2	1
airplane	97.1	0	1	5	2
banana	98.6	1	4	6	3
chicken	94.3	0	1	5	2
tickle	90	10	1	4	2
potty	95.7	12	1	4	2
open	95.7	4	319	4	2
dinner	91.4	13	91	4	2
keys	100	0	1	3	1
go	100	26	626	2	1
mouth	97.1	7	103	3	1
breakfast	90	0	53	8	2

button	97.1	6	10	5	2
cracker	98.6	2	1	5	2
stick	88.6	18	39	4	1
bottle	95.7	8	76	4	2
tongue	92.9	16	35	3	1
corn	90	20	34	3	1
hug	98.6	21	3	3	1
you	92.9	29	3287	1	1
what	91.4	7	1908	3	1
present	91.4	3	377	7	2
doctor	91.4	0	100	5	2
clock	88.6	15	20	4	1
cat	92.9	35	23	3	1
sit	92.9	36	67	3	1
hat	94.3	34	56	3	1
shirt	97.1	15	27	3	1
bib	88.6	13	2	3	1
fork	95.7	13	14	3	1
book	98.6	18	193	3	1
cereal	94.3	1	24	5	3
popcorn	94.3	0	1	6	2
toothbrush	98.6	0	1	7	2
brush	92.9	5	44	4	1
bread	92.9	15	42	4	1
pizza	98.6	0	3	4	2
block	94.3	9	76	4	1
toy	97.1	12	4	2	1
face	92.9	21	371	3	1
foot	100	10	70	3	1
camera	88.6	0	36	5	3
noodles	90	7	1	5	2
plate	90	0	22	4	1
milk	100	3	49	4	1
butter	91.4	13	27	4	2
doll	90	16	10	3	1
eye	98.6	2	5296	1	1
bed	98.6	26	127	3	1
house	94.3	7	591	3	1
bear	95.7	14	86	2	1
blanket	97.1	0	30	7	2

elephant	95.7	1	7	7	3
outside	98.6	2	210	5	2
clap	91.4	16	1	4	1
drive	90	9	105	4	1
cry	95.7	10	48	3	1
look	95.7	17	399	3	1
home	94.3	20	547	3	1
dog	97.1	8	75	3	1
walk	98.6	15	100	3	1
closet	88.6	0	16	6	2
baby	100	5	62	4	2
dance	91.4	6	9	4	1
puzzle	90	5	10	4	2
turtle	90	8	8	4	2
nose	100	18	60	3	1
ear	98.6	31	29	2	1
coat	91.4	31	43	3	1
bite	92.9	13	10	3	1
horse	98.6	11	122	3	1
pajamas	90	1	3	7	3
candy	94.3	10	16	4	2
pillow	98.6	6	8	4	2
bird	98.6	22	31	3	1
room	90	29	1	3	1
towel	94.3	9	6	3	2
rock	94.3	23	75	3	1
read	94.3	28	178	3	1
work	95.7	20	760	3	1
yes	100	9	144	3	1
pencil	91.4	2	34	6	2
stroller	94.3	1	1	6	2
raisin	92.9	7	1	5	2
kitchen	97.1	0	90	5	2
grass	94.3	13	53	4	1
mine	91.4	28	59	3	1
blow	92.9	15	33	3	1
toe	98.6	32	10	2	1
comb	91.4	24	6	3	1
watch	91.4	5	81	3	1
bathroom	95.7	0	1	6	2

grapes	97.1	16	3	5	1
shower	90	6	15	4	2
jump	92.9	8	24	4	1
puppy	95.7	6	2	4	2
bunny	92.9	14	1	4	2
coffee	88.6	2	78	4	2
ball	100	24	4	3	1
see	88.6	40	997	2	1
good	92.9	12	807	3	1
street	94.3	6	244	5	1
slide	90	10	20	4	1
break	92.9	13	90	4	1
hold	92.9	13	169	4	1
beach	88.6	18	67	3	1
nap	92.9	20	4	3	1
shoe	98.6	24	14	2	1
daddy	98.6	7	4	3	2
rain	94.3	38	80	3	1
hi	95.7	9	503	2	1
finger	94.3	1	40	5	2
frog	90	4	1	4	1
kitty	97.1	7	7	4	2
pen	94.3	29	18	3	1
eat	100	24	61	2	1
cup	98.6	19	45	3	1
tree	95.7	16	59	3	1
dry	88.6	12	68	3	1
no	100	27	2084	2	1
run	92.9	26	212	3	1
chocolate	88.6	0	9	7	3
sleep	94.3	13	65	4	1
mommy	98.6	0	1	4	2
dirty	98.6	4	1	4	2
heavy	92.9	3	110	4	2
cow	94.3	19	29	2	1
ice	94.3	16	45	2	1
cake	95.7	26	13	3	1
box	91.4	5	70	3	1
sun	92.9	0	278	3	1
grandma	98.6	0	1	7	2

picture	94.3	2	162	5	2
please	100	6	62	4	1
swim	88.6	11	15	4	1
spill	88.6	13	1	4	1
knee	95.7	31	35	2	1
hair	98.6	24	149	2	1
store	92.9	18	74	3	1
light	100	35	333	3	1
man	88.6	26	1207	3	1
medicine	91.4	0	30	7	3
diaper	95.7	4	1	4	2
water	100	3	442	4	2
hand	95.7	13	431	4	1
help	91.4	11	311	4	1
bowl	97.1	33	23	3	1
bath	100	17	26	3	1
tooth	95.7	14	20	3	1
fall	98.6	27	147	3	1
bee	90	35	6505	2	1
hungry	91.4	0	1	6	2
bubbles	100	10	12	6	2
tissue	90	2	41	4	2
lunch	91.4	7	33	4	1
cookie	100	6	2	4	2
little	91.4	6	831	4	2
sky	88.6	6	58	3	1
wet	98.6	29	53	3	1
leg	95.7	15	58	3	1
sock	100	26	4	3	1
hamburger	90	0	6	7	3
bicycle	98.6	1	5	5	3
tummy	97.1	7	1	4	2
paper	97.1	9	157	4	2
ride	90	30	49	3	1
TV	98.6	3	1	2	2
cheese	100	13	9	3	1
apple	98.6	8	9	3	2
head	97.1	25	424	3	1
hit	92.9	33	115	3	1
scissors	88.6	0	1	5	2

truck	98.6	9	57	4	1
toast	94.3	11	19	4	1
jelly	88.6	5	3	4	2
stop	94.3	11	120	4	1
clean	92.9	11	70	4	1
cold	100	15	171	4	1
soup	91.4	22	16	3	1
fish	95.7	13	35	3	1
car	98.6	24	274	2	1
asleep	88.6	1	29	5	2
monkey	91.4	1	9	5	2
pants	100	11	9	5	1
window	95.7	5	119	5	2
flower	97.1	5	23	4	2
spoon	97.1	11	6	4	1
money	95.7	10	265	4	2
kick	90	28	16	3	1
soap	100	21	22	3	1
hot	98.6	28	130	3	1

Appendix D
Sample Experimental session wordlist: Participant SSA session 5

DECtalk™ repeated	DECtalk™ novel	AT&T repeated	AT & T novel	Human repeated	Human novel
Cry	Noodles	Eat	Sock	Grass	TV
Dog	House	Frog	Bubbles	Blow	Apple
Home	Milk	Finger	Leg	Stroller	Hamburger
Drive	Plate	Run	Hungry	Mine	Cheese
Look	Eye	Dry	Little	Raisin	Bicycle
Outside	Bear	Kitty	Cookie	Toe	Paper
Walk	Camera	Pen	Wet	Watch	Hit
Blanket	Butter	Cup	Sky	Kitchen	Head
Clap	Bed	No	Tissue	Comb	Tummy
Elephant	Doll	Tree	Lunch	Pencil	Ride