EFFECTS OF REPEATED EXPOSURE TO SYNTHETIC SPEECH BY INDIVIDUALS WITH APHASIA

by

Madeline Wollersheim

A Thesis Submitted to the Faculty of the

DEPARTMENT OF SPEECH, LANGUAGE, AND HEARING SCIENCES

In Partial Fulfillment of the Requirements

For the Degree of

MASTER OF SCIENCE

In the Graduate College

THE UNIVERSITY OF ARIZONA

2019

THE UNIVERSITY OF ARIZONA
GRADUATE COLLEGE

As members of the Master's Committee, we certify that we have read the thesis prepared by Madeline Wollersheim, titled *Effects of Repeated Exposure to Synthetic Speech by Individuals with Aphasia* and recommend that it be accepted as fulfilling the thesis requirement for the Master's Degree.

Jessica Brown — Date: 4/8/2019
Jessica Brown, PhD, CCC-SLP

Pélagie Beeson — Date: 4/8/2019
Pélagie Beeson, PhD, CCC-SLP

Kathleen Cazzato — Date: 4/8/2019
Kathleen Cazzato, MA, CCC-SLP

Final approval and acceptance of this thesis is contingent upon the candidate's submission of the final copies of the thesis to the Graduate College.

I hereby certify that I have read this thesis prepared under my direction and recommend that it be accepted as fulfilling the Master's requirement.

Jessica Brown — Date: 4/8/2019
Jessica Brown, PhD, CCC-SLP
Assistant Professor
Speech, Language, and Hearing Sciences

Acknowledgements

First and foremost, thank you to my mentor, Dr. Jessica Brown. I can genuinely say that I would not be the person I am today without the guidance and support you have provided me over the past years. Your approach to research, teaching, and mentorship has shaped my view of the field and set a gold standard of leadership that I will forever aspire to emulate.

Thank you to my husband, Paul, for your constant encouragement and strength. You are my rock. Thank you for selflessly supporting me through countless acts of kindness, words of affirmation, and delicious meals and baked goods. I am incredibly grateful for you and cherish this special chapter in our life.

Thank you also to the faculty and staff at the University of Arizona. I was constantly in awe regarding my experience in the department. To be taught by experts provided a solid foundation of clinical and research skills that I am deeply grateful for; to be supported, encouraged, and cared for by them was an invaluable experience that I will forever cherish. Your encouragement, kindness, and wisdom kept me afloat and rejuvenated throughout the program.

Finally, thank you to my participants, who completed a demanding protocol and spent energy on tasks and meetings. This study would not have been possible without you.

Table of Contents

List of Tables.. 5

List of Figures... 6

Abstract.. 8

Introduction... 11

Method.. 25

Results.. 42

Discussion... 58

Appendix A: Written Instruction for TOSREC Task Completion...................... 73

References... 77

List of Tables

Table 1. Demographic information... 28

Table 2. Performance on standardized testing measures.................................. 29

Table 3. Descriptive Data of Audio Stimuli.. 33

Table 4. TOSREC sentence stimuli across tasks.. 34

Table 5. JE | Preference ranking across time and tasks...................................... 46

Table 6. RO | Preference ranking across time and tasks..................................... 50

Table 7. NL | Preference ranking across time and tasks..................................... 54

Table 8. MA | Preference ranking across time and tasks.................................... 58

List of Figures

Figure 1. Example image of generalization task. This specific screen was shown with the auditory stimulus: "The man is carrying the child." The target image is in the upper right corner.. 36

Figure 2. Likert-Type Scale of Perceived Accuracy... 37

Figure 3. Unique pictorial representations for voice conditions......................... 38

Figure 4. Qualtrics© survey platform for baseline and independent practice tasks as presented on computer screen.. 39

Figure 5. JE | Performance on maintenance tasks over time for baseline (dark), one-week follow up (medium), and four-week follow up (light) sessions................... 43

Figure 6. JE | TOSREC independent practice accuracy over repeated exposures... 44

Figure 7. JE | Performance on generalization tasks over time for baseline (dark), one-week follow up (medium), and four-week follow up (light) sessions.................... 45

Figure 8. RO | Performance on maintenance tasks over time for baseline (dark), one-week follow up (medium), and four-week follow up (light) sessions.................... 47

Figure 9. RO | TOSREC independent practice accuracy over repeated exposures.. 48

Figure 10. RO | Performance on generalization tasks over time for baseline (dark), one-week follow up (medium), and four-week follow up (light) sessions.................... 49

Figure 11. NL | Performance on maintenance tasks over time for baseline (dark), one-week follow up (medium), and four-week follow up (light) sessions.................... 51

Figure 12. NL | TOSREC independent practice accuracy over repeated exposures 52

Figure 13. NL | Performance on generalization task over time for baseline (dark), one-week follow up (medium), and four-week follow up (light) sessions.................... 53

Figure 14. MA | Performance on maintenance tasks over time for baseline (dark), one-week follow up (medium), and four-week follow up (light) sessions.................... 55

Figure 15. MA | TOSREC independent practice accuracy over repeated exposures.56

Figure 16. MA | Performance on generalization tasks over time for baseline (dark), one-week follow up (medium), and four-week follow up (light) sessions.................... 57

Abstract

For individuals with aphasia, comprehension challenges inhibit quality of life and daily task completion. Technological supports (e.g., text-to-speech (TTS)) offer methods to compensate for comprehension challenges by providing content through multiple modalities (i.e., visual and auditory inputs), which includes the use of computer-generated speech forms (synthetic speech). Although potentially beneficial, research suggests that individuals with aphasia comprehend synthetic speech output from TTS with varying degrees of success. At this time, it is unclear whether repeated exposure to synthetic speech increase comprehension of content presented through this modality. Therefore, the purpose of this project was to evaluate the effect of repeated exposure on comprehension of recorded natural speech (digitized speech) and various synthetic speech forms by people with aphasia.

Participants included four adults, between 40 and 67 years of age, with a clinical aphasia diagnosis. Participants were exposed to one synthetic and one digitized speech output daily over a two-week period. Each day, participants listened to 30 simple, active sentences and provided a true/false response regarding the feasibility of each sentence. Participants then completed two follow-up sessions—one week and four weeks after experimental completion—to evaluate maintenance effects of repeated exposure and comprehension generalization to novel stimuli and untrained synthetic voice conditions.

The results of this study revealed four main findings: (a) comprehension variability of the presented synthetic speech across and within individuals with aphasia, (b) increased comprehension of synthetic speech and digitized natural speech following repeated exposure to stimuli, (c) unique performance patterns dependent on aphasia profile, and (d) individualized and dynamic preference selection for synthetic speech.

Findings from the study hold clinical implications when considering use of synthetic speech as a comprehension aid for individuals with aphasia. First of all, variable comprehension and preferences suggest trialing of multiple supports, and including unique patient preference when evaluating synthetic speech as a possible comprehension aid. Generally improved comprehension following repeated exposures, although not statistically significant, illustrates increased comprehension of synthetic and digitized natural speech over time among individuals with aphasia. Though improvements were not statistically significant, increases in comprehension may be clinically or personally significant for individuals using the support as a comprehension aid. Individuals with mild aphasia (n = 2) maintained high performance on comprehension tasks across time. However, performance of individuals with moderate aphasia (n = 2) fluctuated over time. This finding cautions against taking data at only a single timepoint, as it may not accurately reflect an individual's auditory comprehension potential. Finally, the variability of performance on generalization tasks was highly individualized and suggested that generalization of auditory comprehension to an untrained task may be dependent on the individual. Collectively, these findings suggest that synthetic and digitized speech may be promising supports for individuals with mild and moderate aphasia. However, conclusions may not be generalizable to a wide range of individuals due to the limited participants in the current study. Further research is warranted to better understand the effects of repeated exposure on comprehension of synthetic and digitized speech by individuals with aphasia.

INTRODUCTION

Aphasia is a language impairment secondary to neurological damage that can affect varying aspects of language including oral production, auditory comprehension, written expression, and/or reading decoding and comprehension. Profiles of strengths and deficits across language modalities vary greatly across individuals. Regardless of the modality, challenges are exacerbated with increasing complexity of the signal (e.g., increased message length or syntactic complexity; DeDe, 2013). Following occurrence of acquired aphasia, language impairments may resolve to some degree during periods of spontaneous recovery or as a result of restorative intervention (e.g., Beeson, Rising, Kim, & Rapcsak, 2010; Edmonds, Nadeau, & Kiran, 2009; Raymer, 2011).

Many individuals experience chronic symptoms. In fact, over one million individuals in the United States currently live with chronic aphasia ("Aphasia Information Page," 2019) and experience lifelong implications of communication impairment. Subsequently, individuals with aphasia may experience social isolation and decreased independence (Cruice, Worrall, & Hickson, 2006; Davidson, Howe, Worrall, Hickson, & Togher, 2008; Hilari & Northcott, 2006; Le Dorze & Brassard, 1995), both of which are important factors in overall quality of life (Cruice et al., 2006). The World Health Organization set forth the International Classification of Functioning, Disability, and Health (ICF) framework, which unified the concurrent effects of health conditions (i.e., neurological damage resulting in aphasia) with contextual factors (i.e., environmental and personal factors) in classifying the impact of disability on an individual's life (World Health Organization, 2001). Thus, both must be considered when addressing patients holistically. Deepening evidence regarding both (a) the nature of aphasia, and (b) viable options for treatment is imperative in increasing communicative success and quality of life for individuals with aphasia.

Auditory comprehension

Considering how individuals without neurological disorders understand language is a necessary first step to developing supports for those with impaired auditory comprehension. Linguists propose that successful auditory comprehension requires concurrent processing in the neurologic, linguistic, semantic, and pragmatic systems (Rost, 2013). Neurologically, an individual must (a) hear the stimulus, (b) attend to the stimulus, and (c) process the meaning of the message. Linguistic processing requires decoding of grammatical units, while semantic processing relies on word recognition and knowledge. Finally, the pragmatic system integrates suprasegmental features of the auditory signal (e.g., prosody).

Neurological damage may result in a breakdown in any of the aforementioned comprehension subsystems (Baker, Blumstein, & Goodglass, 1981; Kimelman, 1999). Aspects of the auditory signal itself such as rate, prosody (e.g., stress, pitch), as well as message vocabulary may further contribute to auditory comprehension deficits for individuals with aphasia (Baker et al., 1981; Kimelman, 1999). Additionally, higher level demands, such as interpretation of complex syntactic structures contribute to difficulties comprehending auditory information (Goh, 2000).

As with many aspects of aphasia symptoms, the extent of auditory comprehension deficits is dependent upon the location and extent of neurological damage. Thus, variable profiles of linguistic strengths and weaknesses result in a subset of individuals with relatively preserved auditory comprehension in comparison to comprehension of written text (Brown, Wallace, Knollman-Porter, & Hux, 2018; DeDe, 2012). For individuals with auditory comprehension strengths, examining how to utilize this capability to overcome for other areas of deficits is necessary in the development and selection of compensatory aids to increase participation in communication-based activities.

Reading comprehension

In addition to difficulties comprehending auditory information, chronic reading deficits are commonly experienced by individuals with aphasia (DeDe, 2013; Holland, 2007; Knollman-Porter, Wallace, Hux, Brown, & Long, 2015). Deficits can exist at the grapheme (i.e., letter), word, phrase, and/or paragraph level. Approximately 68-80% of people with aphasia experience single-word reading deficits (Brookshire, Wilson, Nadeau, Gonzalez Rothi, & Kendall, 2014; Wilson, Gonzalez Rothi, Nadeau, & Kendall, 2007). Reading deficits also exist at the sentence- and paragraph-level (Wallace, Knollman-Porter, Brown, & Hux, 2018).

As a direct consequence of language impairment, individuals with aphasia experience a change of roles in a plethora of activities which impacts how individuals interact with various materials (Parr, 1995). Changes specific to reading reported after onset of aphasia include: (a) decreased reading ability, (b) changes in motivation and decreased desire to read, and (c) changes in the reading materials of interest. Despite challenges, individuals with aphasia continue to express a desire to engage in reading material. Specifically, individuals report motivation with regards to text-based material relating to personal interests, family, friends, and community (Knollman-Porter et al., 2015).

Therapeutic approaches to remediate reading do exist in the literature (e.g., phonological treatment, multiple oral re-readings). Findings yielded individual success at single-word and passage-level tasks for individuals with a variety of reading impairment profiles (Beeson & Rapcsak, 2010; Beeson, Rising, DeMarco, Foley, & Rapcsak, 2016; Kim, Rising, Rapcsak, & Beeson, 2015). The ability to assess reading treatment success in isolation is difficult due to the concomitant treatment of reading and writing skills in many rehabilitative studies (e.g., Beeson et al., 2016; Kim et al., 2015).

Another factor which limits the validity of studies examining restorative reading treatment is the fact that many studies report outcomes in terms of decoding alone, including only reading rate and number of errors produced (e.g., Kim et al., 2015). Reporting measures of decoding alone fails to address reading comprehension. According to the simple model of reading, reading is the product of decoding and comprehension (Gough & Tunmer, 1986). Thus, effects on reading comprehension should not be assumed with measures of rate and error production alone; further, one study illustrated improved reading rate with minimal or no changes to reading comprehension (Mayer & Murray, 2002).

Finally, the time and monetary costs of extended therapy emphasizing reading at grapheme or single-word levels may not be realistic for many patients given the restrictions inherent in the provision of most speech and language services (Cameron & Wright, 2009). Whereas current literature documents success following extensive therapy of 12 to 24 sessions (e.g., Beeson, Rising, DeMarco, Foley, & Rapcsak, 2016; Kim et al., 2015), many patients are capped at as few as 20 visits per year (Cameron & Wright, 2009). Thus, many clinicians and researchers have investigated and utilized augmentative and alternative supports to aid in comprehension of both written and spoken input.

Comprehension Supports

Clinicians may suggest supports to aid in auditory or reading comprehension for individuals with aphasia. Supports are categorized as alternative or augmentative. Alternative supports transform the signal from one presentation modality into another. For example, a written passage may be presented auditorily for an individual who demonstrates superior auditory versus reading comprehension. Augmentative supports are additive in nature in that they concurrently provide multiple modalities. Unlike alternative supports which simply transform modalities, multimodal input provides a user with simultaneous presentation in multiple forms (e.g., written text with simultaneous auditory reading of the text). Use of multimodal content presentation for individuals with aphasia is advantageous because it allows the individual to capitalize on multiple subdivisions of language. If an individual has deficits in comprehending both written and spoken text, providing concurrent auditory and visual input may provide multiple opportunities for comprehension.

The concept of multimodal presentation for individuals with aphasia is supported by the resource allocation theory (McNeil, Odell, & Tseng, 1991). This theory suggests that humans possess a limited capacity of attention and therefore must allocate that attention and effort among all cognitive processes required in a given task (McNeil et al., 2004). Breakdowns occur when the demands of a given task exceed an individual's cognitive capacity. Applying this theoretical framework to individuals with aphasia who experience comprehension deficits, challenges comprehending auditory or written messages may occur due to an inability to properly allocate attention dependent on task complexity and demands (McNeil et al., 2004). In this case, multimodal input would be theorized to decrease the cognitive load required of the individual receiving the intended message, build connections between the visual and auditory modes of input, and thereby improve comprehension of material (Hux, Knollman-Porter, Brown, & Wallace, 2017; Mayer, 2002; Mayer & Moreno, 2003). The redundancy of multimodal presentation may also decrease the demands on working memory, which in turn decreases the cognitive efforts required in decoding a message (Brown, Wallace, Knollman-Porter, & Hux, 2018; Dietz, Knollman-Porter, Hux, Toth, &

Brown, 2014; Wallace, Dietz, Hux, & Weissling, 2012; Wood, Lasker, Siegel-Causey, Beukelman, & Ball, 1998).

Indeed, there is evidence to support multimodal presentation to increase comprehension among individuals with aphasia. Multimodal presentation in the forms of (a) visual (i.e., gesture) plus auditory information (Preisig et al., 2018), (b) visual written plus visual pictorial representations (Dietz et al., 2014), and (c) written plus auditory input (Knollman-Porter et al., 2015), have all been documented as effective supports. Thus, multimodal presentation of various forms is a feasible comprehension support for individuals with aphasia. A primary drawback to some multimodal options is the requirement of a third party to produce material not otherwise available. High tech options may be one solution that reduces the burden on caregivers and increase the independence of individuals with reading deficits.

Text to speech: An electronic comprehension support. Text-to-speech systems (i.e., systems that create synthetic speech from text; TTS) are one means of providing simultaneous presentation of content in multiple modalities such that information is presented visually through written text and auditorily through synthetic speech presentation of the text. Inherent advantages to TTS systems are that (a) they do not rely on human assistance to generate the auditory signal, (b) are available on a wide range of high-tech devices, and (c) may be applied to a seemingly infinite variety of electronically-based written materials (e.g., social media, PDFs, online news sources).

An electronic device's production of computer-generated (synthetic) speech for use in TTS systems involves four steps: (a) capturing text on a device, (b) converting the selected text to a digital representation, (c) transforming digital files into corresponding sounds, and (d) converting the digital representation to an analog signal. The analog signal is then transmitted to the reader/listener through the device's speaker (Beukelman & Mirenda, 2013). These TTS systems are commercially available on most high-tech devices including smart phones, tablets, and computers. Thus, they are widely accessible. Moreover, over half of the adult population in the United States owns a smartphone and approximately three quarters of the general population owns a laptop computer ("Mobile Fact Sheet," 2018), both of which are viable platforms with which to run TTS.

Digitized natural speech: An electronic comprehension support. Technology also offers the option to digitize natural speech, that is, record a human speaker and utilize the audio file. Digitized natural speech differs from synthetic speech with regard to suprasegmental features, naturalness, and pronunciation (Cowley & Jones, 1992; Luce, Feusetl, & Pisoni, 1983) and closely resembles natural, face-to-face speech because intricate aspects of human speech production, such as coarticulation and prosody, are naturally portrayed. Moreover, when manipulated to increase intelligibility (i.e., alter rate, pitch, and/or volume), digitized natural speech empirically mirrors natural speech (Delogu, Conte, & Sementina, 1998; Drager, Clark-Serpentine, Johnson, & Roeser, 2006; Koul & Allen, 1993). The ability to manipulate the signal is advantageous in that it allows individualization of the signal to maximize comprehension for individuals with aphasia. However, the primary disadvantage of digitized natural speech is that all content must be pre-recorded or recorded in time by another individuals (e.g., spouse, caretaker, clinician). Thus, signals are not readily available for many reading materials, nor are they available on commercial devices. Despite increased naturalness of digitized natural speech, the limited availability of digitized natural speech makes it a less viable option for individuals who wish to access a myriad of materials.

Comprehension of Synthetic Speech Systems

In order for a TTS system to be effective, it must be comprehensible to the user. Investigation of the comprehensibility of early synthetic systems among individuals with aphasia yielded unfavorable results illustrating variability of comprehension given differing auditory signals (Carlsen, Hux, & Beukelman, 1994; Huntress, Lee, Creaghead, Wheeler, & Braverman, 1990). In one early study, adults with mild auditory comprehension deficits secondary to aphasia performed significantly better in a natural speech condition than with synthetic speech stimuli across four different tasks (Huntress et al., 1990). Poor comprehensibility of early synthetic speech signals was also reflected in an early study in which ten out of 12 adults refused to perform a sentence comprehension task due to problems understanding the synthetic speech output of the presented system (i.e., Echo II+™; Carlsen et al., 1994).

Relatively recent technological advances have resulted in a substantially higher quality of synthetic speech production. Research of comprehension of a synthetically-generated signal by individuals with aphasia is limited (Beukelman & Mirenda, 2013; Koul, 2003). However, a growing body of evidence supports increased comprehensibility and yields promising results for use of synthetic speech as a plausible support for individuals with neurological damage (Brown et al., 2018; Wallace et al., 2018).

Software platforms provide unique speech output options varying in quality and characteristics. Apple platforms allow users to select one of approximately 25 unique voices for the English language. Voices exist in both male and female genders and vary perceptually on multiple speech characteristics (e.g., pitch; “Adjust Voices,” 2018). Microsoft platforms provide the option to choose one pre-made voice in a male or female gender and adjust features such as rate and pitch (“Chapter 7: Customizing Narrator,” 2019). Although a plethora of options exist across software platforms, electronic devices are typically pre-programmed with one default voice. The current preset synthetic speech output on Microsoft platforms is named “David,” and on Apple platforms is “Alex.” Despite customizable options for TTS voices, many individuals with aphasia and their caregivers may choose to use the preset voice conditions out of convenience. Furthermore, preset voices may be selected due to the variable rate of computer literacy in the population (OECD, 2016).

Hux and colleagues (2017) investigated sentence-level comprehension of synthetic speech voices available in Apple platforms (Alex) and Microsoft platforms (David) and digitized natural speech by individuals with aphasia. Results indicated significantly increased accuracy when comprehending sentences presented through digitized natural speech compared to either synthetic speech condition. Analysis of individual performance revealed three distinct patterns across participants: (a) relatively comparable accuracy across the three conditions (i.e., 30% of participants), (b) comparable accuracy between one synthetic speech condition and the digitized natural speech condition (i.e., 45% of participants), and (c) significantly better comprehension accuracy in the digitized natural speech condition than either synthetic speech condition (i.e., 25% of participants).

These results demonstrated comprehension variation among individuals with aphasia when accessing synthetic and digitized speech outputs (Hux et al., 2017). The researchers concluded that, on average, a group of individuals with varying aphasia types and severity comprehend at least one form synthetic speech to an equal extent as digitized natural speech. Thus, for individuals with impaired reading abilities, conversion of written text into auditory synthetic speech signals may facilitate comprehension. However, the authors stated a need to investigate the effects of repeated exposure to synthetic speech given that previous studies evaluated comprehension at only one distinct time point (Hux et al., 2017). It is possible that increased exposure to the various voice output options may influence comprehension for people with aphasia. Increased auditory comprehension following repeated exposure to synthetic speech has indeed been documented with individuals with and without intellectual disabilities (Koul & Hester, 2006; McNaughton, Fallon, Tod, Weiner, & Neisworth, 1994; Reynolds, Isaacs-Duvall, & Haddox, 2002; Reynolds, Isaacs-Duvall, Sheward, & Rotter, 2000); however, effects of repeated exposure on the comprehension efforts of individuals with aphasia with currently available synthetic speech systems has not been explored. Thus, while researchers have documented that individual differences in comprehending synthetic and digitized speech exists for individuals with aphasia following one exposure to the speech outputs, the effects of practice and repeated exposure remains unknown.

Preferences and Social Validity

Participation in reading- and auditory-based activities has a direct social impact on the lives of individuals with and without aphasia. Participation in meaningful activities provides a sense of successful living, purpose, mental stimulation, and a feeling of normalcy (Brown, Worrall, Davidson, & Howe, 2012; Knollman-Porter et al., 2015). Many daily activities are communication-based and require decoding and comprehension of written and auditory stimuli (e.g., email or text family and friends, participation on social media, or reading magazines and newspapers). Unfortunately, many individuals with aphasia express a feeling of isolation as a result of difficulties comprehending content presented across these modalities and contexts (Dalemans, De Witte, Wade, & Van Den Heuvel, 2010). Returning to communication-based activities is crucial for social participation, increased quality of life, and the ability to work and access vital information (e.g., health information, personal finance).

Taking into account user preference when exploring supportive aids is crucial. While some opinions regarding communication supports are held in relative agreement amongst individuals with aphasia, other preferences are highly individualized and are dependent on life experiences, personal background, and current life factors (Brown et al., 2018). Thus, detailed inventory of an individual's strengths, needs, and preferences are important for aid selection and long-term use (Johnson, Inglebret, Jones, & Ray, 2006; Scherer, 2005). Failure to consider and integrate a user's desires and preferences is a primary reason by which assistive technology is abandoned by individuals with disabilities (Scherer & Glueckauf, 2005). It is best clinical practice to match an individual's skills, deficits, needs, and preferences with available and effective supports and strategies.

In relation to technological supports, matching technology and user preferences may include taking into account perspectives of naturalness and ease of comprehending synthetic voices (Hux et al., 2017). Technological advances have led to widespread accessibility of high-tech supports. Emerging use of "smart" devices as communication aids is increasingly acceptable among individuals with aphasia (Moffatt, Pourshahid, & Baecker, 2017). In measuring effectiveness, it is paramount that assistive technology matches an individual's preferences and aid in functional, real-life situations (Fager, Hux, Beukelman, Karantounis, 2006). Thus, the validity of technologically-based support systems (e.g., TTS) is based both on their efficacy in increasing comprehension as well as user preference. Evaluating how comprehension and preference change over a period of repeated use will aid clinicians in interpreting baseline performance and preferences when introducing comprehension aids to individuals with aphasia.

Research purpose and questions

Evaluating how comprehension of synthetic and digitized natural speech by people with aphasia is affected following repeated exposure is a necessary first step for confirming the feasibility of electronically-based systems which utilize these speech forms as potential compensatory strategies. A model for how perception changes over time, as measured through comprehension of various voice outputs during listening tasks, would provide clinicians with an expectation regarding plausibility of TTS as a compensatory strategy based off of baseline measures. Thus, this study assessed comprehension of synthetic and digitized natural speech during and following repeated exposures by individuals with aphasia. Furthermore, we investigated the maintenance and generalization across differing synthetic speech conditions and listening tasks. The study aimed to answer the following research questions:

1. Does auditory comprehension of sentence-length material presented in digitized natural and synthetic speech forms change with repeated exposure for people with aphasia?

2. Are changes in comprehension of sentence-length material presented in digitized natural and synthetic speech forms maintained one-week and four-weeks after repeated exposure ceases?
3. Do changes in comprehension of synthetic speech generalize to different contexts (i.e., content and voices) after repeated exposure in one condition?
4. How does the perceived comprehension accuracy of digitized natural speech and synthesized speech by people with aphasia change following repeated exposure to digitized natural and synthetic speech forms?
5. Is there a relationship between injury-related characteristics and/or standardized assessment performance factors and baseline comprehension of synthetic speech?

METHOD

Participants

All participants met the following study eligibility criteria: spoke American English as a native language, had a clinical diagnosis of aphasia secondary to a stroke, were at least one-year post injury, and had reliable internet access. Furthermore, all participants successfully completed hearing, vision, and technology screenings prior to enrollment in the study. Specifically, participants (a) passed a pure-tone hearing screening at 1000Hz, 2000Hz, and 4000Hz at 40dB in at least one ear; (b) passed a vision-screening consisting of the symbol cancellation subtest of the *Cognitive Linguistic Quick Test Plus* (*CLQT+*; Helm-Estabrooks, 2017), identifying one or more correct symbols in 4/4 quadrants; and (c) passed a technology use screening in which they navigated a computer-based study-specific task through mouse movement, selection of icons through mouse click, and typing of their unique participant code.

Participant demographic information. Study participants included four adults (i.e., one female and three males) with aphasia secondary to stroke in the language dominant hemisphere. Participants ranged from 40 to 67 years old (*M* = 57.75, *SD* = 12.31) and had 13 to 16 years of education (*M* = 14.25, *SD* = 1.50). At the time of study completion, no participants reported working or volunteering full time and one participant volunteered part-time. All four participants lived with a full-time caretaker or family member. Participants ranged from 13 months to 109 months post-stroke (*M* = 62.75, *SD* = 46.01) and presented with anomic (n = 2), conduction (n = 1), and Broca's aphasia (n = 1). Demographic information for each participant appears in Table 1. Pseudonyms are used throughout the study to protect participant privacy. Participants completed the Aphasia Quotient (AQ) portion of the *Western Aphasia Battery - Revised* (*WAB-R*; Kertesz, 2006), all subtests of the *CLQT*+ (Helm-Estabrooks, 2017), and the Comprehension of Spoken Sentences and Comprehension of Spoken Paragraphs subtests from the *Comprehensive Aphasia Test* (*CAT*; Swinburn, Porter, & Howard, 2004). The researchers selected the aforementioned assessments as a means of quantifying participants' visual, cognitive, and linguistic processing skills rather than as study inclusion criteria. Standardized test performance for each individual participant appears in Table 2. Participants were also queried regarding previous exposure to technology and synthetic speech (Brown et al., 2018). One participant reported daily use of synthetic speech through a global positioning system (GPS; JE), one participant reported using digitized natural speech through a reading application twice per week (MA), and two participants did not use electronic voice outputs (i.e., RO and NL).

JE

JE was a 65-year-old, right hand dominant, white male who experienced a focal, left hemisphere stroke 35 months prior to study completion. He reported no significant medical or developmental history prior to his stroke. JE reported completing 15 years of education (i.e., three years of college) and worked as a maintenance manager pre-stroke. However, JE reported retiring post-stroke due to subsequent disability. At the time of study completion, JE lived at home with his wife. JE self-reported vision problems including double vision in the right eye and cataracts for which he was being treated at the time of study participation. He reported needing glasses to support visual acuity deficits and wore glasses when completing all experimental tasks. JE reported daily use of technology (i.e., television, phone, laptop, global positioning system or GPS) and daily exposure to speech generating devices through the use of a map/GPS. At the time of participation, he received weekly individual speech therapy with goals targeting social communication and Theory of Mind, word finding, and expression of specific language.

RO

RO was a 67-year-old, Latino male who experienced a focal, left hemisphere stroke 109 months prior to the study. He reported no significant medical or

Table 1. Demographic information.

	JE	RO	NL	MA
Gender	Male	Male	Female	Male
Aphasia type	Anomic	Conduction	Broca's	Anomic
Age (years)	65	67	40	59
Education level (years)	15	13	16	13
Time post onset (months)	35	109	13	94
Currently receiving SLP services	Yes	Yes	Yes	Yes
Living status	Spouse	Spouse	Family	Family
Employment status	Disability	Disability	Disability	Disability
Regular technology user	Yes	Yes	Yes	Yes
Regular synthetic speech user	Yes	No	No	Yes

developmental history prior to his stroke. RO reported a history of right-hand dominance prior to his stroke, and left-hand dominance post-injury. He completed 13 years of education (i.e., one year of technical school) and was employed as a mechanic prior to his stroke. RO retired due to subsequent disability following his stroke. At the time of study completion, he lived at home with his wife. He reported a history of vision problems including cataracts and laser surgery and reported needing glasses to support visual acuity deficits. RO reported use of technology (i.e., computer) approximately three times weekly and reported no previous exposure to synthetic speech. At the time of participation, RO received group speech therapy once weekly targeting implementation of multimodal communication strategies to remediate communication breakdowns.

Table 2. Performance on standardized testing measures.

Standardized Assessment		Participant			
Assessment	Subtest	J	R	N	M
Western Aphasia Battery - Revised	Spontaneous Speech (20)	20.0	14.0	11.0	19.0
	Auditory Verbal Comprehension (20)	9.5	8.5	8.5	9.2
	Repetition (20)	7.6	5.8	3.5	9.4
	Naming & Word Finding (10)	9.1	7.8	5.1	9.0
	Aphasia Quotient (100)	92.3	72.2	56.2	93.2
	Aphasia Severity	Mild	Mod	Mod	Mild
Cognitive Linguistic Quick Test Plus (CLQT+)	Attention (4)	3.0	3.0	2.0	4.0
	Memory (4)	2	2.0	1.0	2.0
	Executive Function (4)	4.0	4.0	4.0	4.0
	Language (4)	3.0	2.0	1.0	3.0
	Visuospatial Skills (4)	4.0	4.0	4.0	4.0
	Composite Severity (4)	3.2	3.0	2.4	3.4
	Linguistic/Aphasia (3)	3.0	2.0	2.0	3.0
	Non-Linguistic Cognition (4)	3.0	3.0	4.0	3.0
Comprehensive Aphasia Test	Comprehension of Spoken Sentences (32)	28.0	19.0	20.0	25.0
	Impairment level	WNL	I	I	WNL
	Comprehension of Spoken Paragraphs (4)	4.0	2.0	1.0	3.0
	Impairment level	WNL	I	I	I

NOTE: First initial used to indicate participant, such that J: JE, R: RO, N: NL, M: MA; Numbers in parentheses following subtest indicate the max possible score; Mod = Moderate impairment; WNL = within normal limits; I = impairment indicated.

NL

NL was a 40-year-old, white female who experienced a focal, left hemisphere stroke 13 months prior to study initiation. She reported no significant medical or developmental history. NL reported a history of right-hand dominance and subsequent left-hand dominance post-stroke. She completed 16 years of education worked as an industrial mechanic prior to injury; NL ceased working subsequent to her disability. At the time of study completion, NL was single and lived at home with her mother. She reported vision problems including nearsightedness for which she wore glasses. She reported daily use of technology (i.e., computer, phone, tablet); however, reported no previous exposure to synthetic speech. At the time of participation, NL received individual speech therapy once weekly in addition to weekly group speech therapy. Treatment in both contexts focused on use of multimodal communication during expressive breakdowns, reading comprehension, and phonemic self-cueing.

MA

MA was a 59-year-old, white male who experienced two focal, left hemisphere strokes, most recently 94 months prior to study participation. He reported right hand dominance prior to his first stroke and at the time of participation. MA completed 13 years of education (i.e., one year of college). Prior to his stroke, he worked in sales; however, he retired subsequent to his disability. At the time of study participation, MA volunteered weekly at his church and with animal care. MA was single and lived at home with his sister. He identified vision problems including a cataract in his right eye and a history of hemorrhaging in his left eye resolved at the time of participation. MA utilized glasses to support visual acuity and wore glasses when completing all experimental tasks. MA reported daily use of technology (i.e., phone, computer) and use of digitized natural speech while utilizing a reading application. At the time of participation, he received weekly group speech therapy targeting public speaking and use of expressive language to facilitate group conversations.

Stimuli

Audio conditions. We recorded audio stimuli in three conditions: digitized natural speech, synthesized David speech, and synthesized Alex speech. The digitized natural speech condition consisted of pre-recorded files of a male speaker of American English and Midwestern dialect. Synthetic speech conditions consisted of pre-recorded files of distinct synthetic speech settings (i.e., Alex and David). The researchers selected Alex and David due to previous work indicating that individuals with aphasia can comprehend these voices with relatively similar accuracy (Hux et al., 2017). Moreover, both synthetic speech options are widely available, as they are offered on Macintosh platforms (Alex) and Windows platforms (David).

Volunteers not previously exposed to experimental stimuli reviewed all sentences to assure that stimuli did not present with any obvious mispronunciations. Subsequently, the researchers reviewed discrepancies noted by the volunteers and altered any human error (e.g., spelling mistakes). In order to maintain ecological validity, the researchers did not alter correct spelling in an attempt to modify pronunciation. The researchers did, however, manipulate the computer-generated speech to match digitized speech recordings using Audacity© software to assure there were no significant difference in speaking rate or amplification across the three conditions.

Audio stimuli. Content of recordings varied dependent on the task. During baseline measures and daily practice, the content included sentences selected from the *Test of Silent Reading Efficiency and Comprehension* (*TOSREC*; Wagner, Torgesen, Rashotte, & Pearson, 2010). Generalization tasks were comprised of previously utilized task-specific sentences created by researchers in other studies; these stimuli are described in detail below (Hux et al., 2017).

Test of Silent Reading Efficiency and Comprehension. Stimuli for baseline testing, independent practice, and maintenance sessions included 163 simple, active sentences selected from the 1st, 2nd, and 3rd grade levels of the *TOSREC* (Wagner et al., 2010). The 163 sentences were between three and 15 words in length (M = 7.80, SD = 2.85). The researchers recorded all 163 sentences in each of the three voice conditions. Thus, independent practice stimuli included a total of 489 individual sound files of pre-recorded *TOSREC* sentences. Data regarding average speaking rate in words and syllables per minute are presented in Table 3. Computation of two one-way analyses of variance (ANOVAs) confirmed no significant differences between speaking rate of audio recordings across the three voice conditions relative to words per minute, $F(2, 486) = 0.477$, $p = 0.621$, or syllables per minute, $F(2,486) = 0.711$, $p = 0.491$.

Table 3. Descriptive Data of Audio Stimuli.

	Words per minute	Syllables per minute
Digitized natural speech	112.07 – 296.78 *M* = 195.33 *SD* = 30.29	129.59 – 342.86 *M* = 239.82 *SD* = 37.72
Alex	130.15 – 272.73 *M* = 195.87 *SD* = 28.41	144.06 – 316.34 *M* = 240.58 *SD* = 35.17
David	120.82 – 285.04 *M* = 192.79 *SD* = 30.88	140.19 – 311.47 *M* = 236.03 *SD* = 34.04

The researchers randomized stimuli into 35 total sentence sets divided into pre-testing (n = 3), practice (n = 26), and post-testing (n = 6) files using the Qualtrics© platform. One pre- and two post-test files were created for all three experimental conditions and contained 15 *TOSREC* sentences each. The 26 independent practice files each contained 30 total sentences--that is, 15 sentences presented in either the David or Alex synthetic voice followed by 15 sentences presented in the digitized natural voice. Across voice conditions, sentences appeared between zero and three times (*M* = 1.47, *SD* = 0.61). We fully randomized sentences across stimuli sets. Data regarding the number of sentence appearances across tasks are provided in Table 4.

Generalization task. Audio stimuli for the generalization tasks consisted of pre-recorded stimulus sentences across the three conditions (i.e., David, Alex, and digitized natural speech). The researchers utilized 20 main character/event sentences created for a previous study examining comprehension amongst individuals with aphasia (Wallace, Hux, Brown, & Knollman-Porter, 2017). Audio stimuli for the generalization task was paired with 132 digital images of contextually-rich Norman Rockwell drawings available from the internet. Generalization task stimuli mirrored that of the previously mentioned study conducted by Hux and colleagues (2017), such that each target image contained a main character or event matched to an auditorily presented sentence. Researchers presented four images--that is, one target and three foil images--within a 2x2 grid display on a PowerPoint slide. Each image appeared on a black background. The researchers created four PowerPoint files per voice condition (i.e., 12 PowerPoint shows total). Each file contained three practice items followed by 10 experimental stimuli. Between each stimulus item, the researchers included a "Go" slide so that participants could indicate when they were ready for exposure to the next set. Figure 1 provides an example of stimuli setup.

Table 4. TOSREC sentence stimuli across tasks.

	Range	Mean	Standard deviation
Pre-test	0-2	0.28	0.62
Independent practice	0-5	2.41	0.66
Post-test	0-3	0.56	0.85

Likert-type scales. The researchers created a Likert-type scale to quantify participants' perceived comprehension of the auditory stimuli in each condition. Pictorial representations of zero percent (thumbs down) and 100 percent (thumbs up) appeared at the ends of the scale. An image of the scale is provided in Figure 2.

Voice preference ranking. One cartoon avatar face per synthetic speech voice was created through My Blue Robot© and assigned to Alex and David. A photographic image of a male was printed out and assigned to the digitized natural speech option. Researchers printed out and laminated images onto 3.25 inch x 3.25 inch rectangles. Images were utilized to ease cognitive load and decrease language output needed to rank preference of voices following listening tasks (see Figure 3).

Procedures

All experimental procedures occurred across three phases: (1) pre-experimental tasks, (2) independent practice, and (c) two follow up sessions (i.e., maintenance and generalization tasks). One trained member of the research team facilitated completion of all research tasks. Each task is described in detail below.

Phase 1: pre-experiment tasks. The researchers administered all pre-experimental tasks across sessions lasting between one and two hours in length for each participant. Sessions included short breaks between tasks as requested by participants. Pre-experiment tasks included: completion of hearing, vision, and technology screening tasks; completion of standardized assessments; baseline generalization task; and a baseline measurement for *TOSREC* task performance.

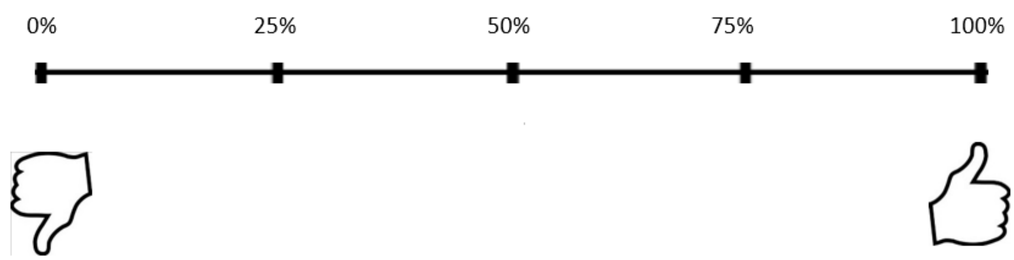
Generalization task. All participants completed the generalization task as their initial exposure to each of the three voice conditions and as a means of identifying the appropriate synthetic voice for inclusion in independent practice tasks. During this task, participants listened to 30 total sentences presented simultaneously with four images across three conditions (i.e., 10 digitized natural speech, 10 Alex synthesized speech, and 10 David synthesized speech). We randomly assigned participants to one of four PowerPoint sets in each voice condition for completion of the first generalization task.

Figure 1. Example image of generalization task. This specific screen was shown with the auditory stimulus: "The man is carrying the child." The target image is in the upper right corner.

Voice condition presentations were fully randomized within participants to control for order effects. During generalization tasks, participants viewed a 2x2 grid of four contextually rich images while simultaneously listening to a stimulus sentence. Auditory stimuli played twice for each image set. The first presentation occurred one second after the images appeared and the second presentation occurred two seconds following the first presentation. The researchers instructed participants to select the image they felt best represented the recorded sentence. Participants indicated their selection by pointing to one of the four presented images. The researchers recorded participant responses online. This task was video-recorded to calculate inter-rater reliability. Upon participant response, the researcher advanced the slideshow. The next screen contained a "go" icon. This screen allowed opportunities for participants to take breaks as requested. The researcher advanced the screen after participant selection, unless they requested a break. This procedure was replicated for three practice items and each of 10 stimuli items within each voice condition.

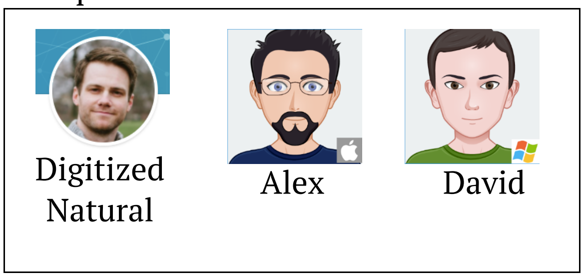
After completing the generalization task, participants reported self-perceived ease of listening through a five-point Likert-type scale by responding to the question "How much did you understand?". Additionally, participants ranked the three voices in order from most desirable to least desirable. Unique pictorial images, previously assigned to each voice, were available for participants to show preference when verbalization was difficult due to language deficits. After ranking the conditions, the researcher asked participants to further explain their preferential rankings.

Figure 2. Likert-Type Scale of Perceived Accuracy.



The researchers determined the voice condition to be used during subsequent independent practice based off of performance on the synthesized speech trials for the initial generalization task. That is, the synthesized speech condition for which the participant performed most poorly was assigned for the independent training phase (i.e., Alex or David). If participants performed equally between synthesized voices, participant preference was used to determine the practice condition. To ensure interrater reliability, the generalization task was video recorded and later analyzed by trained research assistants. The agreement between the independent coders was 98.33%.

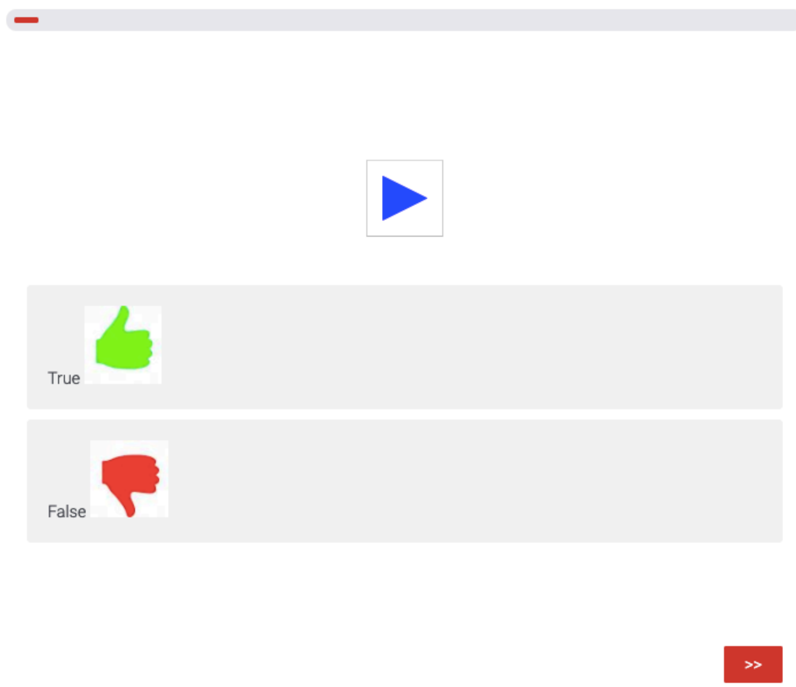
Figure 3. Unique pictorial representations for voice conditions.



TOSREC listening task. Participants also completed a baseline measure activity for the TOSREC listening task across each of the three conditions. Each Qualtrics© survey contained fifteen sentences in each condition. The researcher presented the participant with a screen containing a central blue triangle (i.e., “play”), a green thumbs up with accompanying text that read “true,” a red thumbs down with accompanying text that read “false”, and a red arrow located on the bottom right side of the screen (See Figure 4). Participants first selected the blue arrow to begin the stimulus presentation. After selection, the participant heard a single *TOSREC* sentence. Each auditory stimulus played twice. The first presentation occurred one second after selection of the “play” icon. The second presentation occurred one second following the end of the first occurrence. The researchers instructed participants to evaluate the truthfulness of the presented stimulus through selection of the thumbs-up or thumbs-down icon. Following icon selection, the participant selected the red arrow at the bottom of the screen to advance the task to the next item. Each subsequent stimulus presentation occurred in the same manner. Qualtrics© software tracked participant accuracy throughout the task.

Following completion of generalization and TOSREC baseline measures, participants completed the first day of independent practice in the presence of the

Figure 4. Qualtrics© survey platform for baseline and independent practice tasks as presented on computer screen.



researcher. Completion of the first independent practice session in the presence of a research team member allowed for provision of task instructions and provided a means for participants to troubleshoot technological issues or ask questions as needed. Prior to initiating the activity, the researcher provided the participant with written and pictorial instructions presented in aphasia-friendly format (e.g., large font, simplified language, images). The written instructions contained information regarding completion of the independent trial phase, troubleshooting instructions, and contact information for the research team (see Appendix A). Participants could reference the instructions at any time during task completion and were instructed to take the written instructions with them in order to independently practice at home.

Phase two: independent practice. We instructed participants to complete daily practice on a personal computer with internet connection. Independent practice sessions began the day after completion of pre-experimental tasks. Independent practice included the initial exposure with research team member support and 12 at-home sessions which occurred over 14 consecutive days. The researchers emailed each participant a personalized link daily. The link directed participants to an online Qualtrics© survey which included one practice stimuli set (i.e., 15 synthetic speech sentences and 15 digitized speech sentences). The synthetic speech condition, which was assigned during the initial phase, remained constant throughout practice. The researchers randomized set order to control for order effects and participants listened to each stimulus sentence set only once over the 14-day independent practice period.

Once a participant accessed the personalized Qualtrics© link, he/she recorded their previously assigned participant code. Subsequently, the first stimulus screen appeared. The procedure for completing independent practice was identical to the procedure detailed in the previous section for baseline measures of TOSREC sentence comprehension. After listening to all 30 stimuli in a set, the Qualtrics© provided an overall accuracy score (out of 30) for the participant.

Phase 3: maintenance and generalization sessions. Following independent practice, participants completed two in-person sessions to test for maintenance of practice effects across time and generalization of performance to untrained stimuli and tasks. The first session occurred between zero and seven days of completing the 13 independent practice sets (M = 2.25, SD = 3.20). The second session occurred between 26 and 35 days after completion of independent practice (M = 31.50, SD = 3.87). Sessions included three major components: (a) completion of the previously described generalization task, (b) post-exposure measures of TOSREC sentence comprehension (i.e., the maintenance task), and (c) collection of participant condition preference rankings.

Generalization task. The procedures for the generalization task mirrored the generalization task detailed in phase one. However, we randomly assigned each participant a new stimulus set such that no participant viewed the same generalization task PowerPoint twice. The order of voice presentations was systematically altered across participants to control for order effects. We queried participants regarding post-task perceived accuracy and preference rankings in an identical manner as described in phase one procedures.

TOSREC listening task. The researchers presented participants with three sets of TOSREC sentences, one per condition, during follow-up sessions. Post-test sets one and two were utilized during the first and second follow up sessions. The order of voice presentations was systematically altered across participants to control for order effects.

The procedure for post-exposure measures of TOSREC sentence comprehension mirrored that of the baseline measures and independent practice of TOSREC sentence perception. Post-task perceptual rankings were queried in an identical manner as during the pre-experimental session.

Data Analysis

Dependent variables included accuracy during baseline, independent practice, and follow up measures. We calculated a simple nonoverlap Tau-U statistic to determine whether change in performance over time on independent practice tasks was significant (Parker, Vannest, Davis, & Sauber, 2011).

RESULTS

The results section highlights individual data across time for TOSREC task performance, generalization task performance, and participant perceived accuracy. Performance-related data are organized by timepoint to demonstrate performance over time. Data are presented in a multiple-case study format as to illustrate the unique performance profiles of all four participants with aphasia.

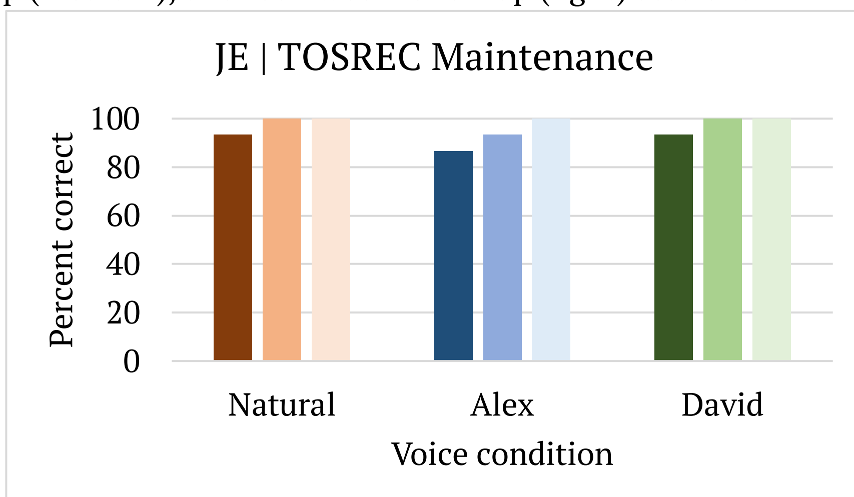
JE

All quantitative data for JE are portrayed in Figures 5, 6, and 7. Preference-related data follow. Information regarding preference across time and tasks are portrayed in Table 5.

Phase 1: pre-experimental data. JE completed the generalization task first. Across the three conditions, JE achieved an average accuracy of 86.67% (*SD* = 11.55).

Specifically, in the digitized natural condition, JE responded with 100% accuracy. In the Alex and David conditions, JE responded to 80% of stimuli accurately. Due to identical scores in the Alex and David voice conditions, the researcher assigned JE's preferred voice condition (i.e., David) for the independent practice phase. Following generalization task completion, JE predicted that he performed with 100% accuracy across the three conditions.

Figure 5. JE | Performance on maintenance tasks over time for baseline (dark), one-week follow up (medium), and four-week follow up (light) sessions.

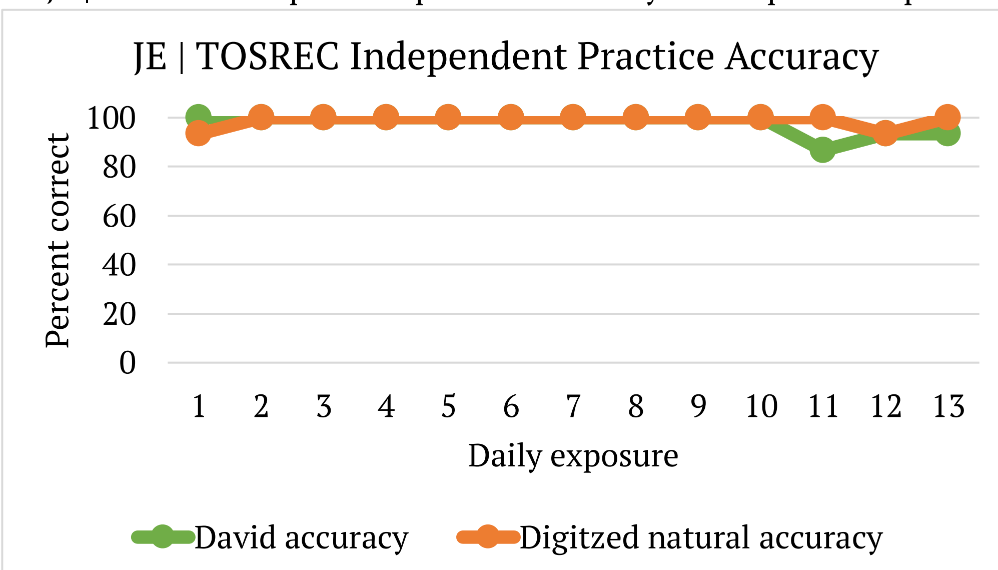


Subsequently, JE completed the pre-experimental TOSREC task. During baseline measures of the TOSREC task performance, JE performed at an average of 91.11% (SD = 3.15) across the three voice conditions. Specifically, JE responded correctly to 93.3% of stimuli in both the digitized natural speech and David conditions and to 86.6% of stimuli in the Alex condition.

Phase 2: independent practice. JE completed the independent practice phase with exposure to both the digitized natural and David synthetic conditions (Figure 6).

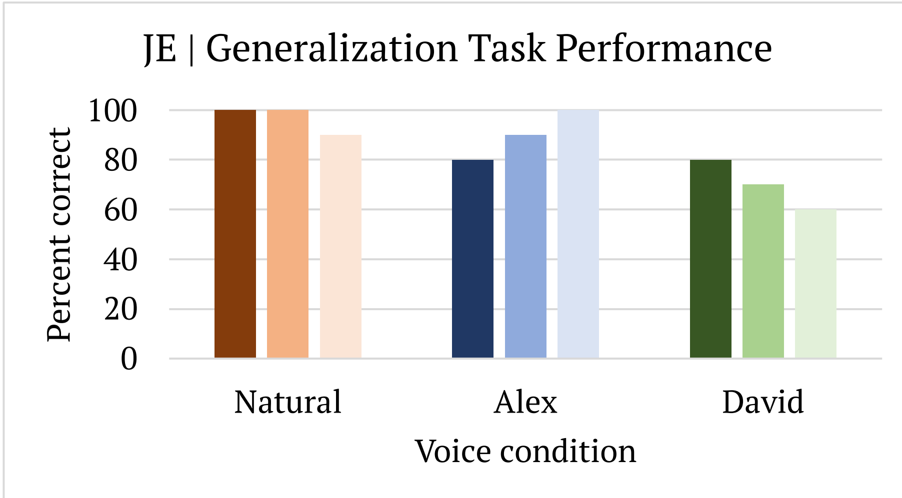
Performance across the 13 exposures to TOSREC sentences ranged from 93.33% to 100% (*M* = 98.46, *SD* = 2.59). Accuracy in the digitized natural condition increased slightly from baseline to the final independent practice exposure (baseline = 93.33%, final independent exposure = 100%) and remained constant in the David synthetic speech condition (baseline = 93.33%, final independent exposure = 93.33%). Tau-U computation revealed that performance over time in the synthetic speech condition was not significant, *U* = -0.3590, *p* = 0.0876, 90% CI [-0.750, -0.013].

Figure 6. JE | TOSREC independent practice accuracy over repeated exposures.



Phase 3: one-week follow up. JE completed the first post-experiment session one day following final independent practice. For the generalization task, he responded with an average accuracy of 86.67% (*SD* = 15.28). Specifically, he responded correctly to 100% of stimuli in the digitized natural condition. In the Alex condition, he responded with 90% accuracy; in the David condition, he responded with 70% accuracy.

Figure 7. JE | Performance on generalization tasks over time for baseline (dark), one-week follow up (medium), and four-week follow up (light) sessions.



Following generalization task completion, JE predicted that he performed with 100% accuracy across the three conditions. For maintenance measures of the TOSREC task, JE responded with an average accuracy of 97.78% (*SD* = 3.85). JE accurately responded to 100% of stimuli in the digitized natural and David conditions and 93.33% in the Alex condition.

Phase 3: one-month follow up. JE completed the second post-experiment session 33 days following the independent practice. The generalization task was completed first, and JE responded with an average accuracy of 83.33% (*SD* = 20.18). In the Alex condition, he responded with 100% accuracy; in the digitized natural condition, he responded correctly to 90% of the stimuli. JE performed with the lowest accuracy in the David condition (i.e., 60%). On maintenance measures of the TOSREC task, JE accurately responded to 100% of stimuli in all the three conditions.

Voice preference across time. Preference rankings across sessions are provided in Table 5. JE consistently preferred David as his top choice following the generalization task across all three sessions. When queried to provide rationale for his preference of David, JE stated that the condition “sounded normal” and that he felt it had an easily understood “midwestern accent.” The order of second and third choices following the generalization task changed from baseline to follow up. Whereas he initially preferred Alex over the digitized natural condition, during both follow up sessions he expressed preference for Digitized Natural Speech over Alex. During these sessions, he stated that Alex sounded “weird,” and had “a different accent,” that made it “harder to catch everything…he pronounces them different.” Interestingly, JE’s preference ranking during the TOSREC task was slightly different and more consistent than the generalization task. He preferred the digitized natural condition as his most preferred condition across all three timepoints. Regarding synthetic speech options, he preferred David over Alex following baseline measures and both follow up sessions.

Table 5. JE | Preference ranking across time and tasks.

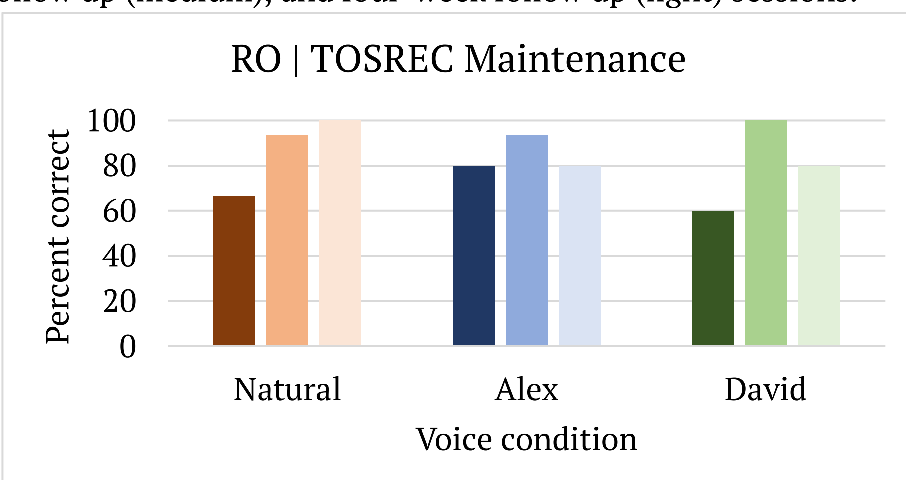
	Generalization task			TOSREC task		
Preference ranking	First	Second	Third	First	Second	Third
Baseline	David	Alex	Natural	Natural	David	Alex
1 week follow up	David	Natural	Alex	Natural	David	Alex
4 week follow up	David	Natural	Alex	Natural	David	Alex

RO

Quantitative data across experimental phases for RO are portrayed in Figures 8, 9, and 10. Condition preferences follow (see Table 6).

Phase 1: pre-experimental data. Across the three conditions, RO achieved an average of 86.67% accuracy (*SD* = 5.77) on the generalization task. He achieved 90% accuracy in both the Alex and David conditions and achieved 80% accuracy in the digitized natural condition. RO predicted that he performed with 60% accuracy across the three conditions. Due to his identical performance in the synthetic voice conditions, the researcher assigned RO's preferred voice condition (i.e., David) for independent practice. Subsequently, RO completed the pre-experimental TOSREC task. During baseline TOSREC performance, RO achieved an average accuracy of 68.90% (*SD* = 10.18). He achieved the highest accuracy in the Alex voice condition (80.00%) followed by the digitized natural (66.67%) and David voice conditions (60.00%), respectively.

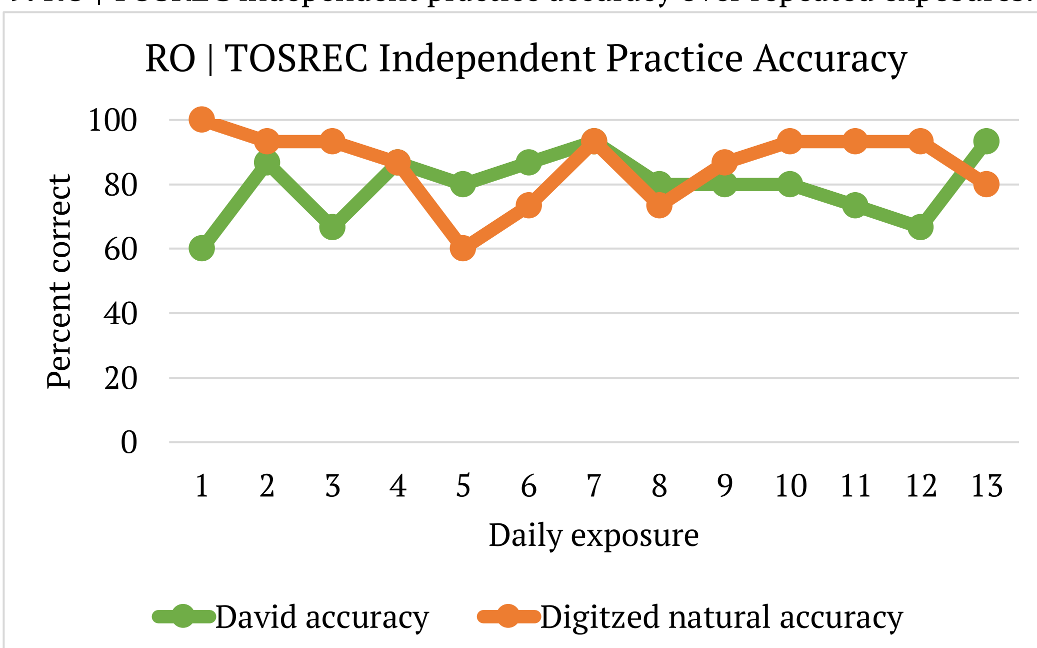
Figure 8. RO | Performance on maintenance tasks over time for baseline (dark), one-week follow up (medium), and four-week follow up (light) sessions.



Phase 2: independent practice. RO completed the independent practice phase with exposure to both the digitized natural and David synthetic conditions.

Performance across the 13 exposures to TOSREC sentences ranged from 70.00% to 93.33% (*M* = 82.82, *SD* = 6.06). Scores increased from baseline to the final independent exposure in the digitized natural condition (baseline = 66.67%, final independent exposure = 80%) and in the David synthetic speech condition (baseline = 60%, final independent exposure = 93.33%). Tau-U computation revealed that performance over time in the synthetic speech condition was not significant *U* = 0.0385, *p* = 0.8548, 90% CI [0.-0.307, 0.384]. Figure 9 represents accuracy for digitized natural and synthetic speech exposures across the two-week period.

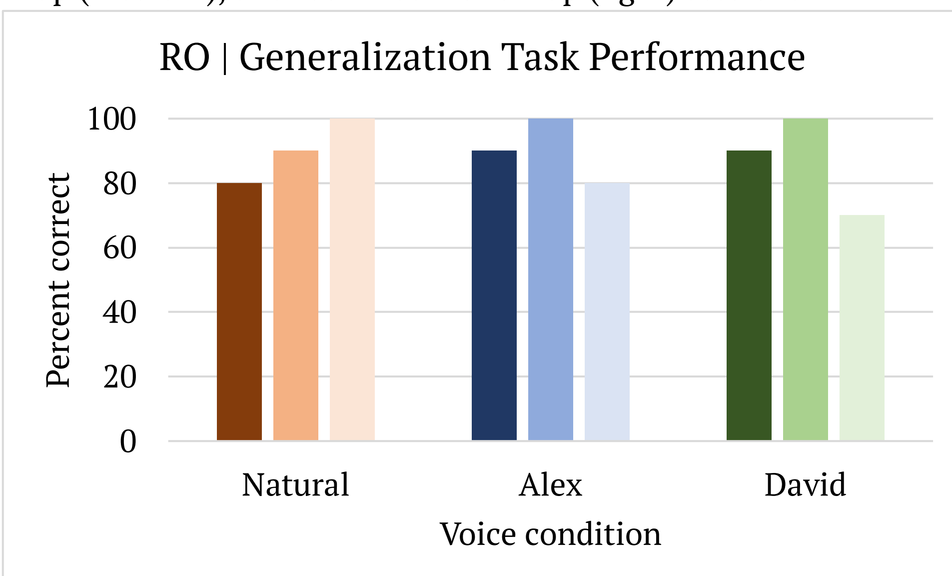
Figure 9. RO | TOSREC independent practice accuracy over repeated exposures.



Phase 3: one-week follow up. RO completed the first post-experiment session one day following final independent practice. On the generalization task, RO's average response accuracy across voices was 96.67% (*SD* = 5.77). Specifically, he responded accurately to 100% of stimuli in both synthetic voice conditions, followed by 90%

accuracy in the digitized natural condition. Following generalization task completion, RO predicted identical performance as he predicted during baseline measures (i.e., 60%) across the three conditions. On TOSREC maintenance measures, RO accurately responded to 95.56% of stimuli across voices (*SD* = 3.85). Specifically, in the David condition he responded with 100% accuracy and in the digitized natural and Alex conditions RO achieved 93.33% accuracy.

Figure 10. RO | Performance on generalization tasks over time for baseline (dark), one-week follow up (medium), and four-week follow up (light) sessions.



Phase 3: one-month follow up. RO completed the second post-experiment session 32 days following the independent practice. On generalization measures, RO's average response rate was 83.33% (*SD* = 15.28). In the digitized natural condition, he responded correctly to 100% of stimuli. In the Alex condition, he responded with 80% accuracy; in the David condition, he responded with 70% accuracy. Following generalization task completion, RO predicted that he performed with 60% accuracy

across the three conditions. On maintenance measures of the TOSREC task, RO responded with an average of 86.67% accuracy across voices. Specifically, he responded accurately to 100% of digitized natural stimuli and 80% of stimuli in the David and Alex conditions.

Voice preference across time. RO's preference changed over time and across tasks (see Table 6). During baseline measures, he stated preference for a synthetic speech condition over natural speech in both the generalization (David) and TOSREC (Alex) tasks. During both the one-week and one-month follow up sessions, RO stated preference for digitized natural speech over either synthetic speech condition following the generalization task, stating he could "hear it really good.". Regarding synthetic speech conditions, he stated "When I hear the words, I'm not sure about my hearing." Following the TOSREC task, however, he stated preference for digitized natural speech at the one-week follow up and the David condition at the one-month follow up.

Table 6. RO | Preference ranking across time and tasks.

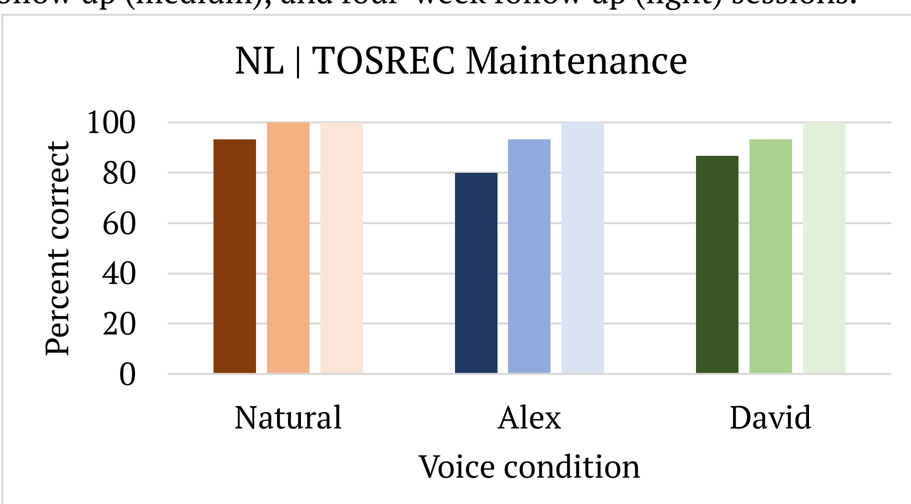
	Generalization task			TOSREC task		
Preference ranking	First	Second	Third	First	Second	Third
Baseline	David	Natural	Alex	Alex	David	Natural
1 week follow up	Natural	Alex	David	Natural	David	Alex
4 week follow up	Natural	David	Alex	David	Natural	Alex

NL

Quantitative data depicting performance for NL are portrayed in Figures 11, 12, and 13. Preference-related data are portrayed in Table 7.

Phase 1: pre-experimental data. NL completed the generalization task first. Across the three conditions, she achieved an average of 80.00% accuracy (*SD* = 10.00). She achieved the highest accuracy in the digitized natural condition (90.00%) followed by David (80.00%), and Alex (70.00%). Due to the lowest accuracy in the Alex voice condition, the researcher assigned Alex as her synthetic speech condition during independent practice. Following generalization task completion, NL predicted that she performed with 70% accuracy across the three conditions. Subsequently, NL completed the pre-experimental TOSREC task. During baseline measures of the TOSREC task performance, she achieved an average of 86.67 accuracy (*SD* = 6.67). Specifically, in the digitized natural condition, she responded with 93.33% accuracy. In the David condition she achieved 86.67% accuracy and in the Alex condition she responded with 80.00% accuracy.

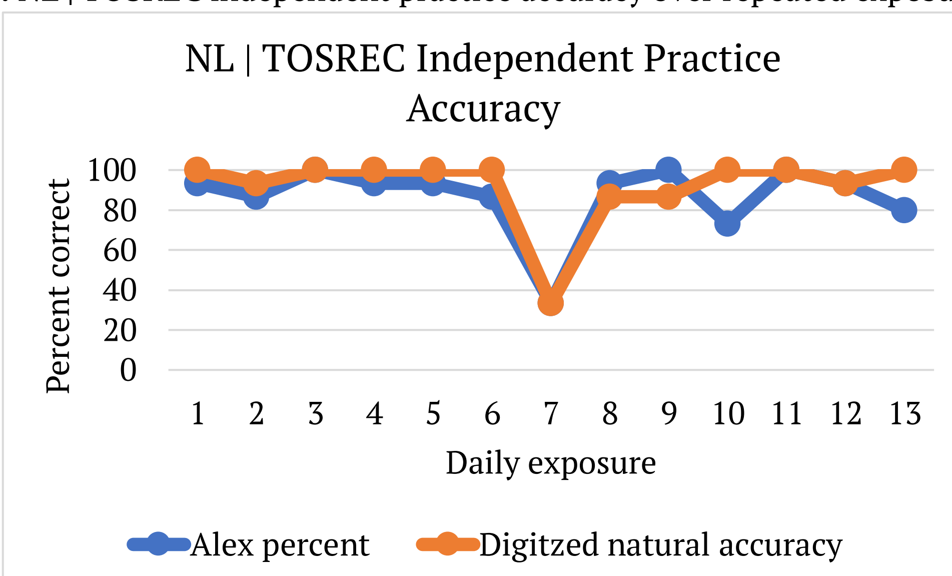
Figure 11. NL | Performance on maintenance tasks over time for baseline (dark), one-week follow up (medium), and four-week follow up (light) sessions.



Phase 2: independent practice. NL completed the independent practice phase

with exposure to both the digitized natural and Alex synthetic conditions. Performance across the 13 exposures to TOSREC sentences ranged from 33.33% to 100.00% (*M* = 89.23, *SD* = 17.28). Scores increased from baseline to the final independent exposure in the digitized natural condition (baseline = 93.33%, final independent exposure = 100%) and remained stable in the Alex synthetic speech condition (baseline = 80%; final independent exposure = 80%). Tau-U computation revealed that performance over time in the synthetic speech condition was not significant $U = -0.0769$, $p = 0.7143$, 90% CI [-0.423, 0.0269]. Figure 12 represents accuracy for digitized natural and synthetic speech exposures across the two-week period.

Figure 12. NL | TOSREC independent practice accuracy over repeated exposures.

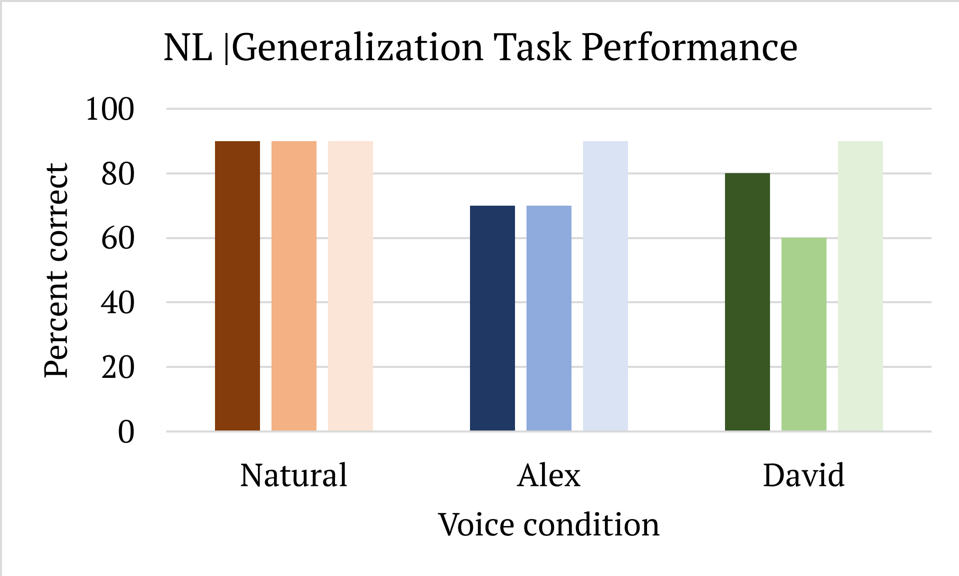


Phase 3: one-week follow up. NL completed the first post-experiment session seven days following final independent practice. On the generalization task, NL's average response accuracy across voices was 73.33% (*SD* =15.28). Specifically, in the digitized natural condition, she responded correctly to 90.00% of stimuli. In the Alex

condition, she achieved 70% accuracy, and in the David condition she achieved 60% accuracy. Following generalization task completion, NL predicted that she performed with 90% accuracy across the three conditions. For maintenance measures of the TOSREC task, NL accurately responded to an average of 95.56% of stimuli across voices (*SD* = 3.85). She achieved the highest accuracy in the digitized natural condition (100%) followed by 93.33% accuracy and in both synthetic speech conditions.

Phase 3: post-experimental session 2. NL completed the second post-experiment session 35 days following the independent practice. On generalization measures, NL responded with 90.00% accuracy across all three voices. She subsequently predicted that she performed at 90%. On maintenance measures of the TOSREC task, NL responded accurately to 100% of stimuli across all three voices.

Figure 13. NL | Performance on generalization task over time for baseline (dark), one-week follow up (medium), and four-week follow up (light) sessions.



Voice preference across time. Across time and task, NL consistently preferred the digitized natural speech condition as her top choice. Regarding synthetic speech conditions, NL preferred David over Alex during baseline and one-week follow up measures. At these times, she stated that David was “terrible” and “icky.” At these times, she expressed ease of listening in both the Alex and digitized natural conditions. However, her preference for synthetic speech changed during the four-week follow up session. At this time, she described David as “brilliant,” and Alex as “Deafen[ing].” She maintained overall preference for the digitized natural condition over either synthetic speech condition, stating it was “amazing,” and “really effective.” Data regarding preference rankings are displayed in Table 7.

Table 7. NL | Preference ranking across time and task.

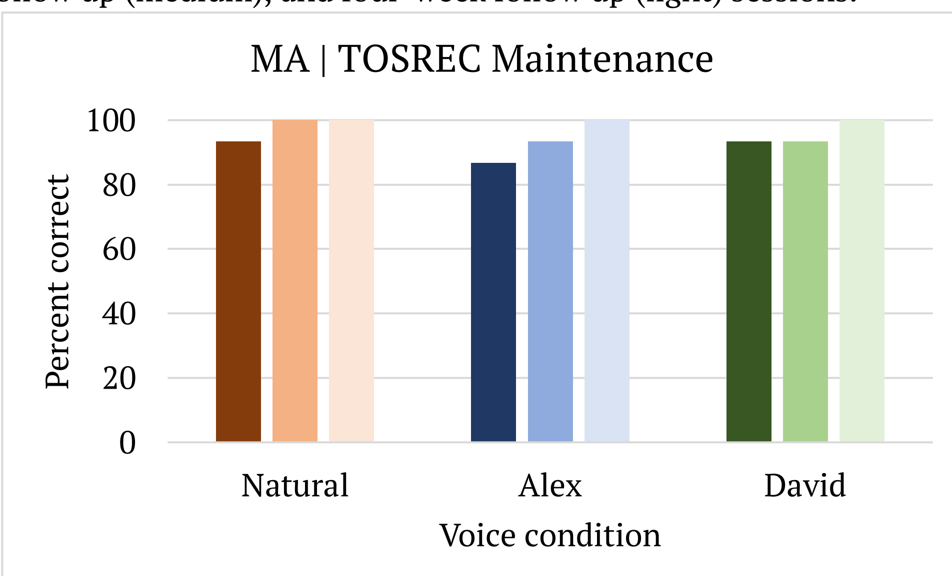
	Generalization task			TOSREC task		
Preference ranking	First	Second	Third	First	Second	Third
Baseline	Natural	Alex	David	Natural	Alex	David
1 week follow up	Natural	Alex	David	Natural	Alex	David
4 week follow up	Natural	David	Alex	Natural	David	Alex

MA

Quantitative data for MA’s performance across phases is portrayed in Figures 14, 15, and 16. Preference-related data follow (Table 8).

Phase 1: pre-experimental data. MA completed the generalization task first. Across the three conditions, he achieved an average of 93.33% accuracy (*SD* = 5.77). Specifically, in the digitized natural condition, he responded with 100% accuracy. In both synthetic speech conditions, he achieved 90%. Due to identical scores in the synthetic speech conditions, the researcher assigned MA's preferred voice condition (i.e., David) for the independent practice phase. Following generalization task completion, MA predicted that he performed with 70% accuracy across the three conditions. Subsequently, he completed the pre-experimental TOSREC task. During baseline measures of the TOSREC task performance, MA achieved an average of 91.11% accuracy (*SD* = 3.85). He performed equivocally in the digitized natural and David conditions (93.33%) and responded accurately to 86.67% of stimuli in the Alex condition.

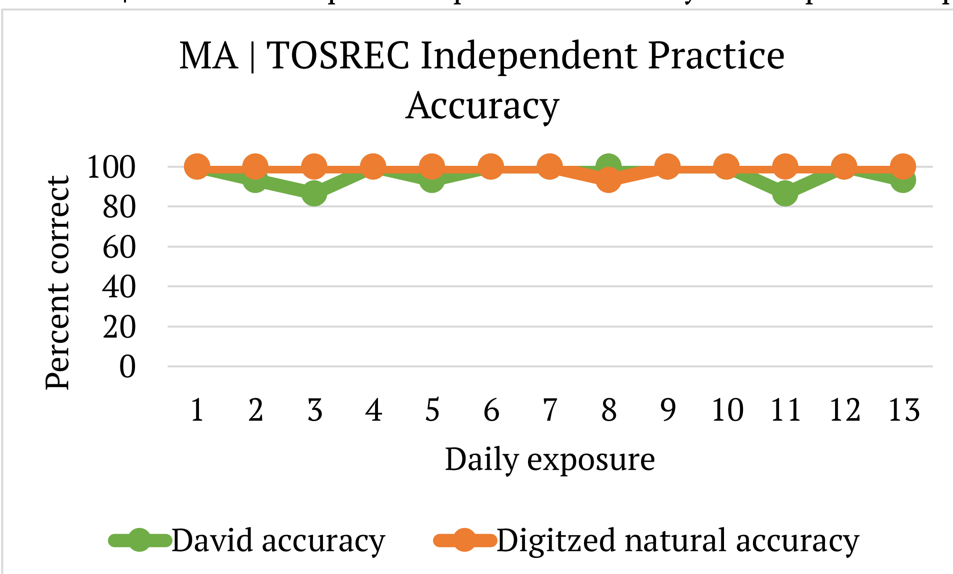
Figure 14. MA | Performance on maintenance tasks over time for baseline (dark), one-week follow up (medium), and four-week follow up (light) sessions.



Phase 2: independent practice. MA completed the independent practice phase with exposure to both the digitized natural and David synthetic conditions. Performance across the 13 exposures to TOSREC sentences ranged from 93.33% to 100% (*M* = 97.95, *SD* = 2.56). Scores remained at 100% from baseline to the final

independent exposure in the digitized natural condition and declined slightly in the David synthetic speech condition (baseline = 100%; final independent exposure = 93.33%). Tau-U computation revealed that performance over time in the synthetic speech condition was not significant $U = 0.0256$, $p = 0.9029$, 90% CI [0.-0.320, 0.371]. Figure 15 represents accuracy for digitized natural and synthetic speech exposures across the two-week period.

Figure 15. MA | TOSREC independent practice accuracy over repeated exposures.

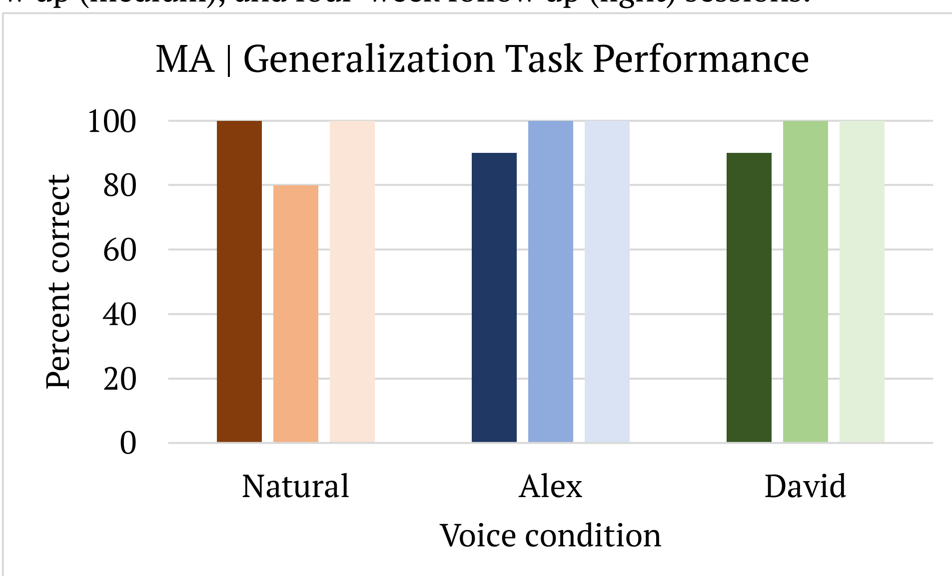


Phase 3: one-week follow up. MA completed the first post-experiment session zero days following final independent practice. On the generalization task, MA's average response accuracy across voices was 93.33% (*SD* = 11.55). Specifically, he achieved 100% accuracy in the Alex and David conditions; MA responded correctly to 80.00% of stimuli in the digitized natural condition. Following generalization task completion, MA predicted that he performed with 100% accuracy across the three

conditions. For maintenance measures of the TOSREC task, MA accurately responded to 95.56% of stimuli across voices (*SD* = 3.85). Specifically, he correctly responded to 100% of stimuli in the David condition and 93.33% in both the digitized natural and Alex conditions.

Phase 3: post-experimental session 2. MA completed the second post-experiment session 26 days following the independent practice. The generalization task was completed first. MA responded with 100% accuracy across all three voice conditions. MA accurately predicted that he performed with 100% accuracy across the three conditions. On maintenance measures of the TOSREC task, MA responded with 100% accuracy across all three voice conditions.

Figure 16. MA | Performance on generalization tasks over time for baseline (dark), one-week follow up (medium), and four-week follow up (light) sessions.



Voice preference across time. MA's preference remained the same across time and tasks. He reported consistent preference for the digitized natural condition. MA stated that the digitized natural condition sounded “more relaxed,” “personable,” and more “human sounding.” For synthetic speech conditions, he preferred David over Alex. When queried to justify this preference, he stated that, in the Alex condition, “I didn’t like the way this one sounded because I didn’t’ get all of it,” and he felt that there were “some [words] that cut off early,” resulting in a voice that was “almost not human sounding at all.” Regarding David, he stated, it was “more normal” and “more mellow” than Alex but “not as personable” as the digitized natural condition. Data regarding preference rankings are displayed in Table 8.

Table 8. MA | Preference ranking across time and task.

	Generalization task			TOSREC task		
Preference ranking	First	Second	Third	First	Second	Third
Baseline	Natural	David	Alex	Natural	David	Alex
1 week follow up	Natural	David	Alex	Natural	David	Alex
4 week follow up	Natural	David	Alex	Natural	David	Alex

DISCUSSION

To compensate for acquired language deficits, individuals with aphasia may utilize synthetic speech produced by high tech devices (e.g., TTS technology) which augments visual content (i.e., written text) with auditory content. This multimodal input provides users with multiple opportunities to process and comprehend stimuli, potentially decreasing the cognitive load required for comprehension efforts. Relatively recent technological advances further support that high-tech supplementation may be both advantageous and feasible for a subset of this population due to (a) increased quality of synthetic speech signals (Brown et al., 2018; Wallace, Knollman-Porter, Brown, & Hux, 2018), and (b) widespread accessibility and ownership of high-tech devices ("Mobile Fact Sheet," 2018; Moffatt et al., 2017). While research documents variability among individuals with aphasia in comprehension of different synthetic speech voice conditions (Hux et al., 2017), little research is available regarding how comprehension is affected by repeated exposure to these stimulus forms. Hence, the purpose of the study presented herein was to evaluate the effects of repeated exposure to synthetic speech among four individuals with mild to moderate aphasia. Moreover, the study examined maintenance of these effects across one and four-week intervals in addition to evaluating generalization to (a) a novel activity, and (b) an untrained synthetic voice. The secondary purpose of the study was to solicit information regarding the preferences of individuals with aphasia across the digitized natural and synthetic speech conditions. The following sections highlight important findings from this study and discuss potential clinical implications.

Independent Practice

Improved performance. Participants completed TOSREC listening tasks in the digitized natural condition and one assigned synthetic speech condition daily for thirteen exposures. Subsequently, they completed a TOSREC listening task during two maintenance sessions. Between baseline and both the one-week and one-month follow up sessions, all four participants demonstrated improved comprehension in the digitized natural condition. Two participants (i.e., JE, NL) also demonstrated consistent improvement comprehending both the practiced and unpracticed synthetic speech conditions at both follow up sessions. The remaining participants (i.e., RO and MA) demonstrated either equivocal or improved comprehension of both synthetic speech forms at both follow up sessions. Overall performance of the four participants demonstrated (a) improvement over time, (b) generalization of positive effects of repeated exposure to unpracticed synthetic speech conditions, and (c) maintenance of positive effects of repeated exposure for up to one month in trained and untrained synthetic speech conditions.

Maintenance and generalization of the effects of repeated exposure hold important clinical relevance. Positive effects of repeated exposure in the synthetic speech condition were maintained for up to one month across the four participants, a result that warrants consideration in creation of a practice schedule. Maintenance of improved comprehension indicates that intensive or massed practice may be advantageous when targeting comprehension of synthetic speech for individuals with aphasia. While intensive practice is supported in research relating to speech and language treatment among individuals with chronic aphasia (e.g., Brindley, Copeland, Demain, & Martyn, 1989; Poeck, Huber, & Willmes, 1989), findings from the current study suggest that short, daily exposure may only be necessary for a period of two weeks in order to improve and maintain higher levels of comprehension of synthetic speech signals. Further, initial massed practice may only be necessary for individuals with moderate aphasia, as performance across voices maintained high accuracy throughout the study for the two participants with mild aphasia (i.e., JE and MA). Generalization to untrained voices suggests possible flexibility in device use. Thus, if in an individual utilizes synthetic speech and changes devices, the distributed practice schedule may be maintained without requiring an initial period to build up comprehension in a new condition.

Performance patterns. Of note, performance of the two participants with moderate aphasia severity ratings (i.e., RO and NL) fluctuated across the independent trial phase. For example, despite average performance of 89.23% across the digitized natural and synthetic speech conditions during independent practice, NL performed at 33.33% accuracy following one exposure in both the synthetic speech and digitized natural conditions. When queried regarding performance on this day, she was unaware of any decline in performance and denied any challenges or difficult days. Though not as drastic a decline, RO averaged 82.82% accuracy across conditions during independent practice, yet his accuracy decreased to 60% in the synthetic speech condition during one exposure trial. This pattern suggests that daily performance of auditory comprehension tasks of synthetic speech may be inconsistent among individuals with moderate aphasia. Variable performance on auditory-based tasks from day to day among individuals with moderate aphasia is documented elsewhere in the literature (e.g., Smith, Nicholas, Hunsaker, & Zips, 2015) has should be taken into clinical consideration. Primarily, a single exposure to synthetic or digitized speech may not be reflective of true comprehension abilities or potential and should not be considered as such during evaluation of possible support aids. Performance over time may be a more effective measure when assessing comprehension potential for synthetic speech.

Although some variability occurred during independent practice for the participants with mild aphasia severity ratings (i.e., JE and MA), their performance remained above 90% for the entirety of the independent practice phase. At or near-ceiling performance on auditory comprehension of synthetic speech among individuals with mild aphasia is documented in the literature (Brown et al., 2018). While the study by Brown and colleagues (2018) assessed comprehension at a single timepoint, the current study expands this finding and suggests that individuals with mild aphasia perform at consistently high levels over time when presented with synthetic speech and digitized natural speech conditions. Thus, among this subgroup, high performance at a single exposure may be representative of a stable ability to understand synthetic or digitized speech signals.

The aforementioned performance patterns of individuals with moderate and mild aphasia provide preliminary data of performance over time when performing auditory comprehension tasks of synthetic and digitized speech signals. However, the small sample size of the current study result in preliminary suggestions only and requires replication with a larger, more homogeneous group, to draw conclusions regarding effects of repeated exposures on comprehension of synthetic speech signals.

Generalization Task Performance

Performance on the generalization task yielded highly variable results, which is consistent in previous work documenting generalization of auditory comprehension following intervention (e.g., Schwartz, Saffran, Fink, Myers, & Martin, 1994). In the current study, variability was documented both across and within participants such that performance increased, decreased, and stayed constant at different time points for different individuals. Specifically, two of four participants (i.e., RO and MA) increased performance on the generalization task in the practiced voice condition immediately following repeated exposures. However, one participant performed equivocally (i.e., NL) and one participant declined in accuracy (i.e., JE). Thus, effects of repeated exposures on comprehension may generalize to unpracticed tasks for some, but not all, individuals with aphasia.

Of note, the generalization task was inherently more complex than practiced tasks in that it forced participants to choose from a field of four rather than a true/false dichotomy as was practiced. Although TOSREC and generalization stimuli contained all simple-active sentence forms, the differing platforms and complexity may have contributed to performance on either task. Despite unequal task complexity, participants demonstrated improved comprehension in trained and untrained synthetic speech conditions within one-week of completing the repeated exposures (i.e., RO, MA) and four weeks after ceasing daily exposure (i.e., NL and MA). This performance suggests that effects of repeated exposure on comprehension in novel tasks in both trained and untrained synthetic speech conditions (a) may occur, and (b) be maintained for a period of four weeks post repeated exposure for some individuals with aphasia. When considering use of synthetic speech as an aid to compensate for reading deficits, this finding is hopeful given that reading patterns, responsibilities, and activities are diverse and unpredictable (Parr, 1992) resulting in varied content that an individual may wish to access.

Voice Preference

Across timepoints, all four participants expressed preference for the digitized speech condition over either synthetic speech condition at least once. Considering only synthetic speech conditions, two participants preferred David over Alex across every timepoint and task (i.e., JE and MA). NL and RO varied in preference of synthetic speech condition, each expressing preference for either David or Alex conditions at least once.

These results, including (a) preference for digitized natural speech, and (b) preference for David over Alex, are consistent with findings from Hux and colleagues (2017). It is important to note that preference changed at least once for three of the four participants (JE, RO, NL) following repeated exposure. Preference changes were noted at different time points for the three aforementioned participants. Furthermore, both JE and RO expressed differing rankings for the generalization task versus the TOSREC sentence task during baseline and both follow up measures. This demonstrates that participant preference for auditory stimuli is not static and can change with time and exposure. Thus, results of preferential data suggest (a) variability of preference across individuals with aphasia for electronic voice condition, and (b) task-dependent variability of preference within individuals with aphasia. It is essential to consider preference when evaluating effectiveness and long-term use of assistive technology (Fager, et al., 2006; Scherer, 2005). Further, the evolution of preference over time as seen in three of four participants (i.e., JE, RO, NL) implies that preference should be addressed not only during support selection but also across time.

Considering Synthetic Speech as an Aid to Comprehension

The four participants who took part in this study are only a small sample of individuals with aphasia. Two participants, JE and MA, shared similarities across standardized testing measures that were not consistent across all four participants. Specifically, they performed within normal limits for measures of comprehension of spoken sentences on the CAT (Swinburn et al., 2004) and were classified as having mild, anomic aphasia on the WAB-R (Kertesz, 2006). Unsurprisingly, both JE and MA subsequently performed consistently high within the independent practice phase and demonstrated improved comprehension from baseline to both the one- and four- week follow up sessions in all three voice conditions; both participants ceilinged out at 100% accuracy across voice conditions by the four week follow up session. Their consistently high performance suggests that individuals with similar aphasia profiles (i.e., mild, anomic aphasia) may comprehend synthetic speech with consistently high accuracy across time and thus be a good candidate for use of synthetic speech generators to aid comprehension; however, further investigation assessing a larger sample of individuals with similar profiles would be required to solidify the generalizability of these findings.

RO and NL also performed comparably on various standardized testing measures. Specifically, they each demonstrated impaired performance on both the Comprehension of Spoken Sentences and Comprehension of Spoken Paragraphs subtests of the CAT (Swinburn et al., 2004) and were classified as having moderate aphasia per the Aphasia Quotient portion of the WAB-R measure (Kertesz, 2006). Their performance over time demonstrated trends not observable in either JE nor MA's performance such that both RO and NL each demonstrated fluctuations in performance from day-to-day during independent practice. Despite inconsistent accuracy, RO and NL both demonstrated an overall increase in performance from baseline measures to one-week follow up sessions in the TOSREC task. Therefore, overall positive trends for these two participants imply that comprehension of synthetic speech may improve for individuals with moderate aphasia. However, individual data points must be interpreted with caution given variability of performance over time.

Limitations and Future Directions

Several limitations hinder the generalizability of results from the present study, which included factors related to the study participants and materials. First, while the multiple case study design allows for in-depth case by case exploration for a group of individuals with aphasia, this methodology limits the ability to generalize results to a broader audience who may exhibit varying aphasia types and severity or may be in different recovery stages (e.g., acute recovery). In fact, the participant sample in the present study represented a heterogeneous group of individuals with aphasia such that the four participants ranged in aphasia type (i.e., anomic, conduction, Broca's), severity (i.e., mild, moderate), and cognitive and linguistic capabilities as measured by the CLQT+ (Helm-Estabrooks, 2017; e.g., attention). A larger, more homogenous sample would increase the study's validity and generalizability.

Differences in comprehension of the presented synthetic speech voices is consistent with previous research by Hux and colleagues (2017). Moreover, positive trends in comprehension for a practiced task and synthetic speech condition aligns with previous research demonstrating increased comprehension of suboptimal speech among individuals with and without intellectual disabilities (Koul & Hester, 2006; McNaughton et al., 1994; Reynolds et al., 2002; Reynolds et al., 2000), suggesting that individuals with aphasia may also benefit from repeated exposure. However, the changes in this study were not statistically significant and were observed in a limited number of individuals.

Second, limitations within the study materials themselves may also have impacted performance and should be considered when interpreting the results. The independent practice task contained only two options for selection (i.e., true or false) and therefore the probability of selecting the correct response regardless of comprehension was fifty percent. The generalization task required that participants select from a field of four, therefore lowering the probability of selecting the correct response to twenty-five percent and in turn lowering the likelihood of false positive responses. In this way, the generalization task in the current study tested performance on a novel and more complex level of comprehension. It follows that participants may have demonstrated differing patterns of generalization had the two tasks been comprised of equal difficulty.

Additional consideration must be taken into account regarding length of stimulus material. While the current study assessed comprehension of synthetic speech for simple active sentences, this is not reflective of many real-life reading activities which are often longer and more complex (Berl et al., 2010). Despite chronic persistent reading difficulties, individuals with aphasia express a desire to return to reading materials enjoyed previous to their strokes (e.g., reading newspapers, books, email, text; Elman & Bernstein-Ellis, 2006; Knollman-Porter et al., 2015; Brown et al., 2012). Recent research found that multimodal input through written and synthetic speech input in the David condition improved comprehension of long narratives (i.e., 128-130 words) among individuals with chronic aphasia (Wallace et al., 2018). However, effects of repeated exposure on narrative-length texts remains unexplored.

Conclusion and Clinical Implications

The results of this study demonstrate (a) variability of comprehension of the presented synthetic speech platforms across and within individuals with aphasia, (b) generally increased comprehension of synthetic speech and digitized natural speech following repeated exposures, and (c) individualized and dynamic preference for synthetic speech condition.

The variable comprehension accuracy evident during daily independent practice for individuals with moderate aphasia (i.e., NL and RO) suggests that evaluation of synthetic speech comprehension may not be accurately represented at a single timepoint. Rather, individual data points should be interpreted with caution. Conversely, initially high comprehension demonstrated by the two participants with mild aphasia (i.e., JE and MA) remained consistent throughout the study. It follows that initially high performance of individuals with mild aphasia may closely reflect comprehension potential and performance. While currently available research documents comprehension of synthetic and digitized speech at a single timepoint (e.g., Brown et al., 2018; Wallace et al., 2018), further research is needed to evaluate the performance potential of individuals with aphasia following repeated exposure.

The positive changes in comprehension over time demonstrate that individuals with aphasia may indeed improve their comprehension of synthetic speech systems with repeated exposures. In synthesizing this result with the aforementioned consideration of variability, clinicians may recommend use of synthetic speech to aid comprehension to individuals with moderate aphasia severity ratings despite low initial comprehension provided that an individual engage in repeated practice. However, the small sample size and heterogeneity of the participants in the current study limit generalizability to the broader population of individuals with aphasia. Evaluation of a broader and more diverse sample is needed. Additionally, the effects of repeated practice on individuals with severe aphasia remains unexplored and thus unknown.

All repeated exposures were completed independently. The four participants improved their comprehension of both synthetic speech systems and digitized natural speech for the practiced task following daily independent practice. Recent research of telepractice for naming treatment for individuals with chronic aphasia documented successful technology-based home practice programs (Kurland, Liu, & Stokes, 2018). Researchers concluded that unsupervised technology-based intervention with semi-frequent conferencing may be a feasible option in treatment of individuals with chronic aphasia. They highlighted advantages such as lower cost and reducing mobility demands for the patient (Kurland et al., 2018). These advantages extend to increasing comprehension of synthetic speech via independent practice, as this preliminary data illustrates increased comprehension through repeated daily exposure alone.

Additionally, each participant demonstrated unique patterns of preference. Assessment of multiple aids and variation within support technology (e.g., differing synthetic speech platforms) is compliant with the matching person and technology assessment process, which includes person-centered measures relating to strengths, goals, and preferences, when selecting support technology (Scherer & Craddock, 2002). Incorporating user preference in evaluation and selection of aids is a central component of patient-centered care (Moffatt, McGrenere, Purves, & Klawe, 2004) and is paramount for long term acceptance and use of supports (Lasker & Bedrosian, 2001;

Scherer & Glueckauf, 2005). The differing preferences regarding the two presented synthetic speech options resulted in two participants who always preferred David over Alex (JE, MA), one participant who frequently expressed preference for David (RO), and one participant who more frequently expressed preference for Alex over David (NL). Preference for David over Alex is consistent with results of Hux and colleagues (2017). However, the variation across participants from the small sample size of the present study reiterates the importance of trialing multiple conditions and soliciting user preference prior to clinical recommendation. Moreover, the within-participant changes over time, as observed in three participants (JE, NL, RO), reiterates the dynamic nature of assessment of supplemental supports. As such, user preference should be considered not only during the initial evaluation but also throughout use of the support.

In conclusion, many individuals with aphasia experience chronic reading deficits (DeDe, 2013; Holland, 2007; Knollman-Porter et al., 2015), yet express a desire to engage in written text (Knollman-Porter et al., 2015). Multimodal input through simultaneous visual and auditory content presentation is one option to compensate for such processing and comprehension deficits. Findings in this study suggest that repeated exposure to synthetic speech may result in increased comprehension for individuals with mild and moderate aphasia. Moreover, fluctuating performance over time suggest that evaluation of synthetic speech as a compensatory aid should be a dynamic process in which single data points are interpreted with caution. Unique preferences and needs affect acceptance and use of supplemental supports and must be taken into consideration when evaluating supportive technological devices (Scherer, 2005). Despite a small sample size, study results indicated varying preference for synthetic voice condition across individuals with aphasia. Thus, both preference and performance should be taken into account when evaluating and requesting supports for individuals with aphasia.

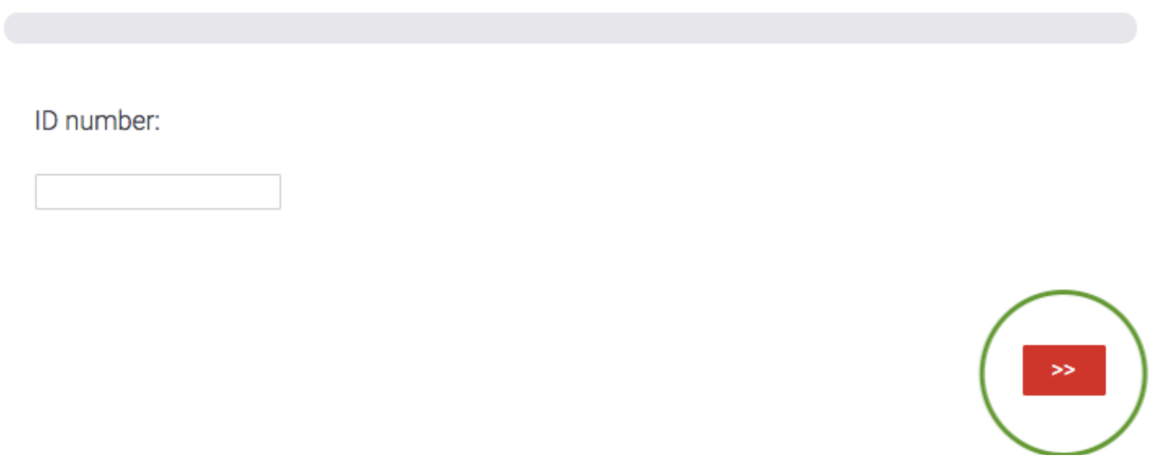
APPENDIX A. WRITTEN INSTRUCTION FOR TOSERC TASK COMPLETION.

How to Practice Listening

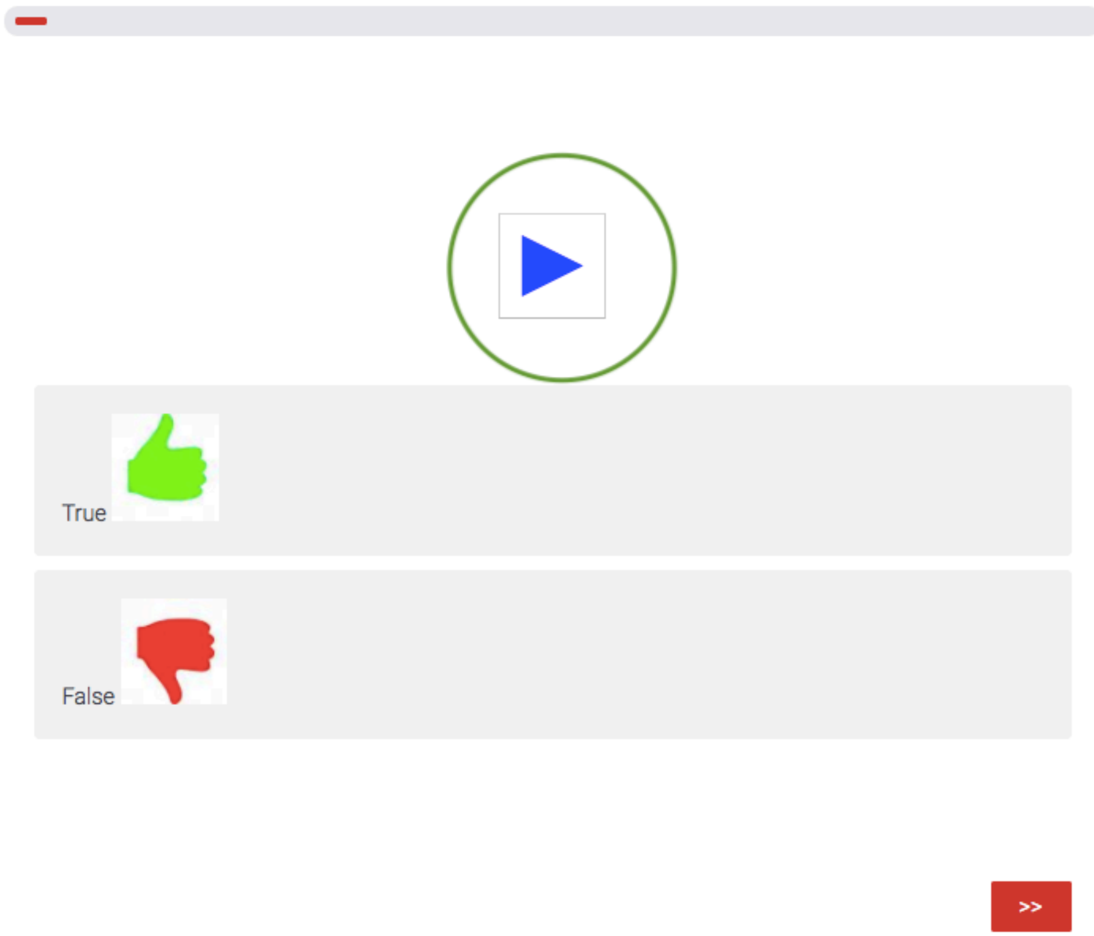
1. Open your email every day. You will see an email from [email provided]. Select the text "SELECT HERE". Your internet will open.

2. Click inside the box. Put in ____________ where it says "ID number":

3. Click on the red box.

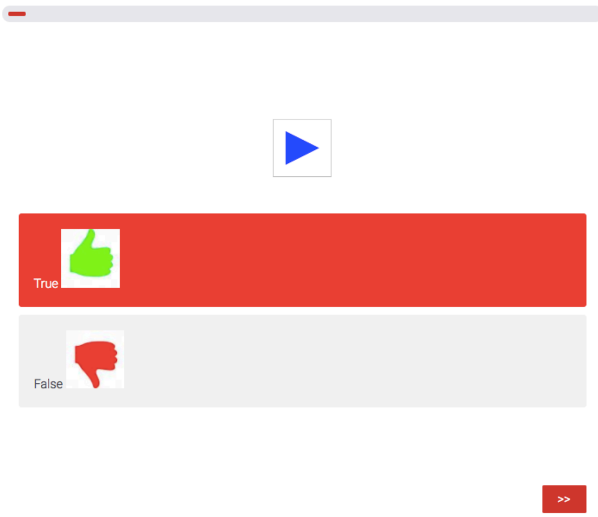


4. Click on the blue triangle in the middle of the screen.

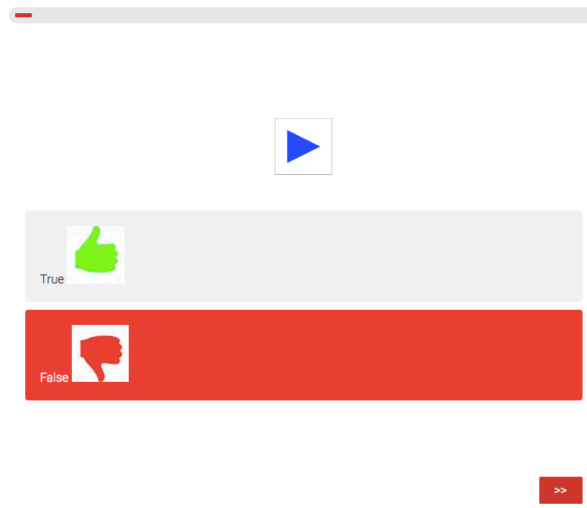


5. Listen to the sentence. It will only play two times.

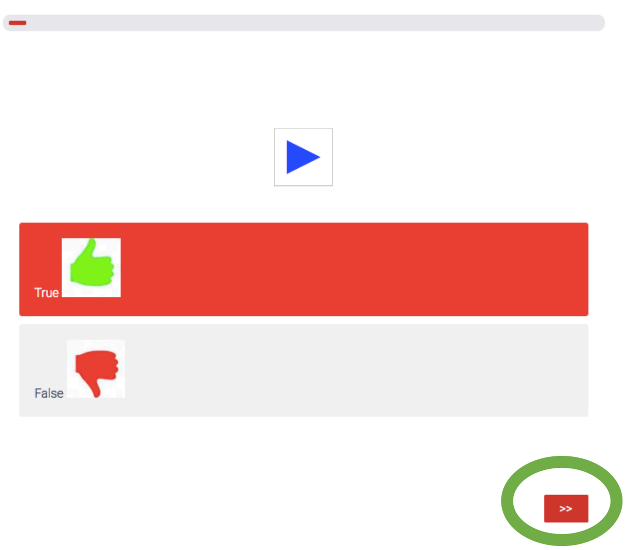
6. Select the green thumbs up if you think the sentence is true OR select the red thumbs down if you think the sentence is false.

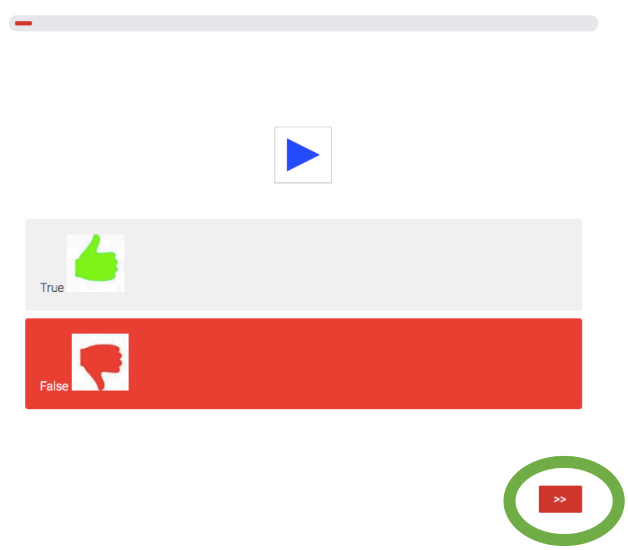


OR



7. Once you made your selection, click on the red arrow on the bottom of the screen.





8. Repeat steps 4-7. You will hear 30 sentences total.

<u>When something is wrong:</u>

- Ask a friend for help
- Make sure you have internet access
- Check the volume
- Plug in your computer, phone, or tablet
- Contact researchers [contact information provided]

References

Adjust voices for VoiceOver, Speak Screen, and Speak Selection on your iPhone, iPad, or iPod touch. (n.d.). Retrieved from https://support.apple.com/en-us/HT202362

Aphasia Information Page. (2019). Retrieved from https://www.ninds.nih.gov/disorders/all-disorders/aphasia-information-page

Baker, E., Blumstein, S. E., & Goodglass, H. (1981). Interaction between phonological and semantic factors in auditory comprehension. *Neuropsychologia, 19*, 1–15.

Beeson, P. M., & Rapcsak, S. Z. (2010). A treatment sequence for phonological alexia/agraphia. *Journal of Speech, Language, and Hearing Research, 53,* 450-468.

Beeson, P. M., Rising, K., DeMarco, A., Foley, T., & Rapcsak, S. (2016). The nature and treatment of phonological text agraphia. *Neuropsychological Rehabilitation, 28*(4), 568-588.

Beeson, P. M., Rising, K., Kim, E. S., & Rapcsak, S. Z. (2010). A treatment sequence for phonological alexia/agraphia. *Journal of Speech, Language, and Hearing Research, 53*(2), 450–468.

Berl, M. M., Duke, E. S., Mayo, J., Rosenberger, L. R., Moore, E. N., VanMeter, J., . . . Gaillard, W. D. (2010). Functional anatomy of listening and reading comprehension during development. *Brain and Language, 114*, 115–125.

Beukelman, D., & Mirenda, P. (2013). Augmentative and alternative communication: Supporting children and adults with complex communication needs (4th ed.). Baltimore, MD: Paul Brookes Publishing Co.

Brindley, P., Copeland, M., Demain, C., & Martyn, P. (1989). A comparison of the speech of ten chronic Broca's aphasics following intensive and non-intensive periods of therapy. *Aphasiology,3*(8), 695-707.

Brookshire, C. E., Wilson, J. P., Nadeau, S. E., Gonzalez Rothi, L. J., & Kendall, D. L. (2014). Frequency, nature, and predictors of alexia in a convenience sample of individual with chronic aphasia. *Aphasiology, 28,* 1464–1480.

Brown, J., Wallace, S., Knollman-Porter, K., & Hux, K. (2018). Comprehension of single versus combined modality information by people with aphasia. *American Journal of Speech-language Pathology,* 1-15.

Brown, K., Worrall, L. E., Davidson, B., & Howe, T. (2012). Living successfully with aphasia: A qualitative meta-analysis of the perspectives of individuals with aphasia, family members, and speech-language pathologists. *International Journal of Speech-Language Pathology, 14,* 141–155.

Cameron, R., & Wright, S. (2009). Options for aphasia therapy when insurance runs out. *Stroke Connection,* 16–17.

Carlsen, K., Hux, K., & Beukelman, D. (1994). Comprehension of synthetic speech by individuals with aphasia. *Journal of Medical Speech-Language Pathology, 2*, 105–111.

Chapter 7: Customizing Narrator. (2019). Retrieved from https://support.microsoft.com/en-us/help/22818/windows-10-narrator-customizing.

Cowley, C. K., & Jones, D. M. (1992). Synthesized or digitized? A guide to the use of computer speech. *Applied Ergonomics, 23,* 172–176.

Cruice, M., Worrall, L., & Hickson, L. (2006). Quantifying aphasic people's social lives in the context of non-aphasic peers. *Aphasiology, 20*, 1210–1225.

Dalemans, R., De Witte, L., Wade, D., & Van Den Heuvel, W. (2010). Social participation through the eyes of people with aphasia. *International Journal of Language and Communication Disorders, 45*(5), 537-550.

Davidson, B., Howe, T., Worrall, L., Hickson, L., & Togher, L. (2008). Social participation for older people with aphasia: The impact of communication disability on friendships. *Topics in Stroke Rehabilitation, 15*, 325 340.

DeDe, G. (2012). Effects of word frequency and modality on sentence comprehension impairments in people with aphasia. *American Journal of Speech-Language Pathology, 21*, S103–S114.

DeDe, G. (2013). Reading and listening in people with aphasia: Effects of syntactic complexity. *American Journal of Speech Language Pathology, 22*, 579–590.

Delogu, C., Conte, S., & Sementina, C. (1998). Cognitive factors in the evaluation of synthetic speech. *Speech Communication, 24,* 153–168.

Dietz, A., Knollman-Porter, K., Hux, K., Toth, K., & Brown, B. (2014). Supported reading comprehension for people with aphasia: Visual and linguistic supports. *Journal of Medical Speech-Language Pathology, 21*, 319–331.

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

Edmonds, L. A., Nadeau, S. E., & Kiran, S. (2009). Effect of verb network strengthening treatment (VNeST) on lexical retrieval of content words in sentences in persons with aphasia. *Aphasiology*, *23*(3), 402-424.

Elman, R. J., & Bernstein-Ellis, E. (2006). Aphasia book clubs: Making the connection. Stroke Connection, 32–33.

Fager, S., Hux, K., Beukelman, D. R., & Karantounis, R. (2006). Augmentative and alternative communication use by adults with traumatic brain injury. *Augmentative and Alternative Communication, 22*, 37-47.

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

Gough, P. B. and Tunmer, W. E. (1986) Decoding, reading, and reading disability. *Remedial and Special Education, 7*(1), 6–10.

Helm-Estabrooks, N. (2017). *Cognitive-Linguistic quick test–plus.* San Antonio, TX: The Psychological Corporation.

Hilari, K., & Northcott, S. (2006). Social support in people with chronic aphasia. *Aphasiology, 20*, 17 – 36.

Holland, A. L. (2007). Counseling/coaching in chronic aphasia: Getting on with life. *Topics in Language Disorders, 27*(4), 339–350.

Huntress, L. M., Lee, L., Creaghead, N. A., Wheeler, D. D., & Braverman, K. M. (1990). Aphasic subjects' comprehension of synthetic and natural speech. *Journal of Speech and Hearing Disorders, 55*,21–27.

Hux, K., Knollman-Porter, K., Brown, J., & Wallace, S. (2017). Comprehension of synthetic speech and digitized natural speech by adults with aphasia. *Journal of Communication Disorders, 69*, 15-26.

Johnson, J., Inglebret, E., Jones, C., & Ray, J. (2006). Perspectives of speech language pathologists regarding success versus abandonment of AAC. *Augmentative and Alternative Communication, 22*(2), 85-99.

Kertesz, A. (2006). *Western aphasia battery–revised*. San Antonio, TX: Pearson Education.

Kim, E. S., Rising, K., Rapcsak, S. Z., & Beeson, P. M. (2015). Treatment for alexia with agraphia following left ventral occipito-temporal damage: Strengthening orthographic representations common to reading and spelling. *Journal of Speech, Language, and Hearing Research*, 1-17.

Kimelman, M. D. (1999). Prosody, linguistic demands, and auditory comprehension in aphasia. *Brain and Language, 69*, 212–221.

Knollman-Porter, K., Wallace, S. E., Hux, K., Brown, J., & Long, C. (2015). Reading experiences and use of supports by people with chronic aphasia. *Aphasiology, 29,* 1448–1472.

Koul, R. (2003). Synthetic speech perception in individuals with and without disabilities. *Augmentative and Alternative Communication, 19,* 49–58.

Koul, R., & Allen, G. (1993). Segmental intelligibility and speech interference thresholds of high-quality synthetic speech in presence of noise. *Journal of Speech and Hearing Research, 36*, 790–798.

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

Kurland, J., Liu, A., & Stokes, P. (2018). Effects of a tablet-based home practice program with telepractice on treatment outcomes in chronic aphasia. *Journal of Speech, Language, and Hearing Research, 61*(5), 1140-1156.

Lasker, J., & Bedrosian, J. (2001). Promoting acceptance of augmentative and alternative communication by adults with acquired communication disorders. *Augmentative and Alternative Communication, 17*, 141–153.

Le Dorze, G., & Brassard, C. (1995). A description of the consequences of aphasia on aphasic persons and their relatives and friends, based on the WHO model of chronic diseases. *Aphasiology, 9*, 239–255.

Luce, P. A., Feustel, T. C., & Pisoni, D. B. (1983). Capacity demands in short-term memory for synthetic and. natural speech. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 25,* 17–32.

Mayer, R. E. (2002). Multimedia learning. *Psychology of Learning and Motivation, 41,* 85–139.

Mayer, R. E., & Moreno, R. (2003). Nine ways to reduce cognitive load in multimedia learning. *Educational Psychologist, 38*, 43–52.

Mayer, J., & Murray, L. (2002). Approaches to the treatment of alexia in chronic aphasia. *Aphasiology, 16*(7), 727-743.

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.

McNeil, M. R., Doyle, P., Hula, W., Rubinsky, H., Fossett, T., & Matthews, C. (2004). Using resource allocation theory and dual-task methods to increase the sensitivity of assessment in aphasia. *Aphasiology, 18*, 521–542.

McNeil, M.R., Odell, K., & Tseng, C. H. (1991). Toward the integration of resource allocation into a general theory of aphasia. *Clinical Aphasiology, 20*, 21–39.

Mobile Fact Sheet. (2018). Retreived from https://www.pewinternet.org/fact-sheet/mobile/

Moffatt, K., McGrenere, J., Purves, B., & Klawe, M. (2004, April). The participatory design of a sound and image enhanced daily planner for people with aphasia. In R. Grinter, T. Rodden, P. Aoki, E. Cutrell, R. Jeffries, & G. Olson (Eds.), *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 407–414). New York, NY: ACM.

Moffatt, K., Pourshahid, G., & Baecker, R. (2017). Augmentative and alternative communication devices for aphasia: The emerging role of “smart” mobile devices. *Universal Access in the Information Society, 16*(1), 115-128

OECD. (2016). Skills Matter: Further Results from the Survey of Adult Skills, Additional tables. In *Skills Matter: Further Results from the Survey of Adult Skills*. Paris: OECD Publishing.

Parker, R. I., Vannest, K. J., Davis, J. L., & Sauber, S. B. (2011). Combining non-overlap and trend for single case research: Tau-U. *Behavior Therapy, 42,* 284–299.

Parr, S. (1992). Everyday reading and writing practices of normal adults: Implications for aphasia assessment. *Aphasiology, 6*, 273–283.

Parr, S. (1995). Everyday reading and writing in aphasia: Role change and the influence of pre-morbid literacy practice. *Aphasiology, 9*, 223–238.

Poeck, K., Huber, W., & Willmes, K. (1989). Outcome of intensive language treatment in aphasia. *The Journal of Speech and Hearing Disorders, 54*(3), 471-9.

Preisig, B., Eggenberger, N., Cazzoli, D., Nyffeler, T., Gutbrod, K., Annoni, J., . . . Müri, R. (2018). Multimodal communication in aphasia: Perception and production of co-speech gestures during face-to-face conversation. *Frontiers in Human Neuroscience, 12*, 200.

Raymer, A. M. (2011). Naming and word retrieval impairments. In L.L. LaPointe (Ed.) *Aphasia and related neurogenic language disorders, 4th Edition.* (pp. 95-110). New York: Thieme.

Reynolds, M. E., Isaacs-Duvall, C., & Haddox, M. L. (2002). A comparison of learning curves in natural and synthesized speech comprehension. *Journal of Speech, Language, and Hearing Research, 45*, 802–810.

Reynolds, M., Isaacs-Duvall, C., Sheward, B., & Rotter, M. (2000). Examination of the effects of listening practice on synthesized speech comprehension. *Augmentative and Alternative Communication, 16,* 250–259.

Rost, M. (2013). *Teaching and researching: Listening.* New York, NY: Routledge.

Scherer, M. J. (2005). Assessing the benefits of using assistive technologies and other supports for thinking, remembering, and learning. *Disability and Rehabilitation, 27*, 731–739.

Scherer, M. J., & Craddock, G. (2002). Matching Person & Technology (MPT) assessment process. *Technology and Disability, 14*(3), 125-131.

Scherer, M. J., & Glueckauf, R. (2005). Assessing the benefits of assistive technology for activities and participation. *Rehabilitation Psychology, 50*, 132–141.

Schwartz, M. F., Saffran, E. M., Fink, R. B., Myers, J. L., & Martin, N. (1994). Mapping therapy: A treatment programme for agrammatism. *Aphasiology, 8*, 19–54.

Smith, Z., Nicholas, M., Eileen, H., & Lauryn, Z. (2015). *Treating auditory attention to improve auditory comprehension in persons with aphasia* (Master's Thesis). Retrieved from ProQuest.

Swinburn, K., Porter, G., & Howard, D. (2004). *Comprehensive aphasia test.* Hove, UK: PsychologyPress.

Wagner, R. K., Torgesen, J. Kj., Rashotte, C. A., & Pearson, N. A. (2010). *Test of Silent Reading Efficiency and Comprehension (TOSREC).* Austin, TX: Pro-Ed.

Wallace, S. E., Dietz, A., Hux, K., & Weissling, K. (2012). Augmented input: The effect of visuographic supports on the auditory comprehension of people with chronic aphasia. *Aphasiology, 26*, 162–176.

Wallace, S., Hux, K., Brown, J., & Knollman-Porter, K. (2014) High-context images: Comprehension of main, background, and inferential information by people with aphasia. *Aphasiology, 28*(6), 713-730.

Wallace, S., Knollman-Porter, K., Brown, J., & Hux, K. (2018). Narrative comprehension by people with aphasia given single versus combined modality presentation. *Aphasiology,* 1-24.

Wilson, J., Gonzalez Rothi, L., Nadeau, S., & Kendall, D. (2007). Prevalence and incidence of alexia in aphasia. *Poster session presented at the American speech, language, & hearing association convention.*

Wood, L., Lasker, J., Siegel-Causey, E., Beukelman, D. R., & Ball, L. J. (1998). Input framework for augmentative and alternative communication. *Augmentative and Alternative Communication, 14*, 261–267.

World Health Organization, 2001. International Classification of Functioning, Disability and Health. World Health Organization, Geneva.