Journal of Communication Disorders 69 (2017) 15–26

Contents lists available at ScienceDirect

Journal of Communication Disorders

journal homepage: www.elsevier.com/locate/jcomdis

Comprehension of synthetic speech and digitized natural speech by adults with aphasia

Karen Hux[a,*], Kelly Knollman-Porter[b], Jessica Brown[c], Sarah E. Wallace[d]

[a] University of Nebraska–Lincoln, United States
[b] Miami University, United States
[c] University of Minnesota, United States
[d] Duquesne University, United States

ARTICLE INFO

Keywords:
Synthetic speech
Computer-generated speech
Auditory comprehension
Aphasia

ABSTRACT

Using text-to-speech technology to provide simultaneous written and auditory content presentation may help compensate for chronic reading challenges if people with aphasia can understand synthetic speech output; however, inherent auditory comprehension challenges experienced by people with aphasia may make understanding synthetic speech difficult. This study's purpose was to compare the preferences and auditory comprehension accuracy of people with aphasia when listening to sentences generated with digitized natural speech, Alex synthetic speech (i.e., Macintosh platform), or David synthetic speech (i.e., Windows platform). The methodology required each of 20 participants with aphasia to select one of four images corresponding in meaning to each of 60 sentences comprising three stimulus sets. Results revealed significantly better accuracy given digitized natural speech than either synthetic speech option; however, individual participant performance analyses revealed three patterns: (a) comparable accuracy regardless of speech condition for 30% of participants, (b) comparable accuracy between digitized natural speech and one, but not both, synthetic speech option for 45% of participants, and (c) greater accuracy with digitized natural speech than with either synthetic speech option for remaining participants. Ranking and Likert-scale rating data revealed a preference for digitized natural speech and David synthetic speech over Alex synthetic speech. Results suggest many individuals with aphasia can comprehend synthetic speech options available on popular operating systems. Further examination of synthetic speech use to support reading comprehension through text-to-speech technology is thus warranted.

1. Introduction

Single word reading challenges occur in 68% to 80% of people following the acquisition of aphasia (Brookshire, Wilson, Nadeau, Gonzalez Rothi, & Kendall, 2014; Wilson, 2008; Wilson, Gonzalez Rothi, Nadeau, & Kendall, 2007); difficulty reading sentence and paragraph-length texts is likely even more common, but statistics are unavailable about the prevalence of such challenges. What is known is that reading problems associated with aphasia often persist as a chronic condition (DeDe, 2013; Holland, 2007; Knollman-Porter, Wallace, Hux, Brown, & Long, 2015; Parr, 1995; Pedersen, Vinter, & Olsen, 2004). In part, this may reflect the tendency of professionals to prioritize spoken expression over reading as the focus of formal intervention efforts (Conkyln, Novak, Boissy, Bethoux, & Chemali, 2012; Lynch, Damico, Abendroth, & Nelson, 2013). Furthermore, even when people receive restorative reading

* Corresponding author at: 351 Barkley Memorial Center, University of Nebraska–Lincoln, Lincoln, NE 68583-0738, United States.
E-mail address: khux1@unl.edu (K. Hux).

http://dx.doi.org/10.1016/j.jcomdis.2017.06.006
Received 6 July 2016; Received in revised form 12 June 2017; Accepted 13 June 2017
Available online 23 June 2017
0021-9924/

interventions, functional deficits often remain and prevent engagement in literary activities commensurate with premorbid levels (Knollman-Porter et al., 2015). The result may be long-term reliance on others to decipher and interpret written materials, thus causing people with aphasia to feel frustrated and dependent (Knollman-Porter et al., 2015; Parr, 2007; Worrall et al., 2011).

A possible solution to this problematic scenario is using assistive technology applications with text-to-speech (TTS) conversion capabilities and computer-generated (i.e., synthetic) speech output to provide simultaneous bimodal (i.e., written and auditory) presentation of content. Synthetic speech generation involves using a device to (a) capture text appearing in digital or analog format; (b) convert it to a digital representation, as needed; (c) transform it into corresponding phonemes and allophones; and (d) convert it from a digital signal to analog speech waveforms (Beukelman & Mirenda, 2013). It typically is modifiable with regard to the rate and volume of speech output. However, using a TTS system to generate synthetic speech production of a written text will only be effective as a compensatory reading strategy if two conditions are met: (a) the generated speech output is comprehensible to a person with aphasia and (b) the individual has sufficient auditory—or combined auditory and reading—comprehension skills to understand the presented message. The focus of the study reported herein was to establish the plausibility of the first of these conditions—that is, the assumption that people with a variety of aphasia types and severities can comprehend current renditions of synthetic speech with accuracy levels comparable to their comprehension of natural speech presented in a digitized format.

1.1. Synthetic speech comprehensibility

The comprehensibility of synthetic speech has been a focus of research for several decades. At a basic level, speech perception theories linking the processing of acoustic signals with the production of motor movements to articulate phonemes underlie much of this research (Galantucci, Fowler, & Turvey, 2006). However, modeling synthetic speech comprehension also requires incorporation of information processing frameworks that consider the cumulative effects of factors such as acoustic signal degradation, task complexity, and the presence or absence of contextual supports; further complications arise when listeners have impairments limiting their accessing or application of semantic or syntactic knowledge to facilitate the assignment of meaning to incoming messages (Koul, 2003). The consequence is greater taxing of cognitive resources than occurs when hearing and processing natural speech that, in turn, leads to lower levels of comprehension.

Research published in the 1990s provided documentation about the difficulty people with aphasia had comprehending the output of at least some early synthetic speech systems (Carlsen, Hux, & Beukelman, 1994; Huntress, Lee, Creaghead, Wheeler, & Braverman, 1990). In particular, Carlsen et al. (1994) found that ten of 12 adult participants with aphasia refused to perform a sentence comprehension task using the Echo II+™ because of problems comprehending the synthetic speech output; the remaining two participants performed the experimental task but achieved scores significantly lower than those they achieved when performing the same task with recorded natural speech or two other synthetic speech generators popular at the time (i.e., Smoothtalker™ 3.0 and RealVoice™). Similarly, Huntress et al. (1990) had eight adults with mild auditory comprehension problems secondary to aphasia perform four language tasks (i.e., matching spoken words to pictures, following oral commands, responding to single-sentence yes/no questions, and responding to yes/no questions about paragraph-length material presented orally) in response both to natural speech and synthetic speech generated by the Votrax Personal Speech System. They found significant accuracy differences between the two speech output conditions on all of the presented language tasks, with the natural speech condition consistently yielding better comprehension scores. However, these researchers also found that four repeated exposures to the synthetic speech output improved participants' comprehension to the point that the difference from natural speech no longer reached the criterion for statistical significance.

The quality of synthetic speech production has improved substantially in recent years. However, research about the effects of these improvements on the processing and comprehension of synthetic speech by people with aphasia is lacking (Beukelman & Mirenda, 2013; Koul, 2003). Such research is necessary because, despite substantial improvements, differences continue to exist between synthetic and natural speech, particularly with regard to prosodic elements such as timing, stress, and inflection (Koul, 2003). Furthermore, research with normal listeners as well as those with developmental disabilities suggests that these suprasegmental differences make single sentences, as well as extended discourse generated with synthetic speech, slower and more cognitively demanding to process than comparable speech samples generated with natural speech (Koul, 2003; Koul & Dembowski, 2010). Understanding whether similar or exacerbated challenges occur when people with aphasia attempt to process currently available renditions of synthetic speech is an important clinical question with regard to the feasibility of using TTS systems as devices to present self-administered, computerized aphasia treatments or to compensate for persistent reading impairments.

Another important question regarding the use of TTS systems as compensatory devices for people with aphasia relates to the opinions these individuals have about the quality and acceptability of listening to the generated synthetic speech output. Researchers have documented that the foremost reason people with disabilities abandon assistive technology devices relates to the failure of practitioners to consider a user's desires and preferences (Scherer & Glueckauf, 2005). Furthermore, such abandonment occurs after only brief use in up to 90% of cases (Scherer, 2005). Hence, regardless of the sophistication or applicability of an external aid, a vital component contributing to its long-term use is the matching of features to a person's preferences and needs. With regard to TTS system use by people with aphasia, this means taking into consideration opinions about the naturalness, ease of understanding, and clarity of the generated speech.

1.2. Digitized natural speech

Creating digital recordings of natural speech production is another method of capturing and reproducing speech for use in