Comprehension of Synthetic and Natural Speech: Differences among Sighted and Visually Impaired Young Adults

Konstantinos Papadopoulos and Eleni Koustriava

[1]University of Macedonia, Department of Educational and Social Policy,
Thessaloniki, Greece
{kpapado, elkous}@uom.edu.gr

ABSTRACT

The present study examines the comprehension of texts presented via synthetic and natural speech in individuals with and without visual impairments. Twenty adults with visual impairments and 65 sighted adults participated in the study. Both individuals with and without visual impairments performed at a similar level in the comprehension of texts that were presented via synthetic and natural speech. The findings indicate that prospective difficulties in intelligibility do not affect comprehension. It seems that context cues provided by the text assisted the participants in identifying and comprehending the text more effectively. Moreover, the results reveal no significant differences between sighted participants and participants with visual impairments regarding the comprehension of natural and synthetic speech.

1. INTRODUCTION

Text-to-Speech (TtS) systems are often used by individuals with visual impairments to meet their daily, professional and educational needs [1, 2]. Moreover, individuals with visual impairments frequently use TtS systems and/or screen reader with synthetic speech (synthetic speech systems) as their reading medium [2, 3]. Hence, it is very important to investigate the intelligibility and comprehension of synthetic speech by individuals with visual impairments and examine whether differences in intelligibility and comprehension between synthetic and natural speech exist.

There is a great spectrum of researches examining the differences between synthetic and natural speech and how intelligibility and comprehension are affected respectively. Synthetic speech appears to be less intelligible and more difficult to comprehend [4, 5 & 6], while it engages more cognitive resources than natural speech [7, 6 for a review]. The predominance of natural speech has been proved in various experiments including segmental intelligibility, word recall, sentence transcription and comprehension of passages [8].

Although there is an abundance of research carried out regarding the intelligibility of synthetic speech systems experienced by people with no disabilities, in individuals with visual impairments, limited research is available on the perception of synthetic speech [9]. In a recent study carried out by [10], it was found that the participants, who had visual impairments, had significantly better performance when identifying words presented via natural speech than via synthetic speech; accuracy scores ranged from 89.92% for words presented via the TtS synthesizer to 99.2 % for words presented via natural speech [10]. Similarly, both groups of participants – individuals with visual impairments and sighted individuals – performed better in the task of identifying words presented via natural speech [11].

Having taken a closer look on synthetic and natural speech, the researchers shed light on variables that could affect intelligibility in a more positive or negative way. [8] refer to acoustic-phonetic differences between natural and synthetic speech, which have an impact on synthetic speech perception. Stevens, Lees, Vonwiller, and Burnham [12] found that the gender of the voice and the quality of the signal affect the intelligibility of TtS synthesis. Additionally, previous studies have indicated that synthetic speech perception in typical listeners is also dependent on listening conditions [13]. Moreover, age seems to be a critical variable affecting interaction via synthesizers. Older adults proved to have more difficulties processing synthesized speech [14], probably because with age the working memory requirements increase [15]. On the contrary, intelligibility and comprehension of synthetic speech can be unaffected by age in individuals with visual impairments [11]. One possible interpretation is that augmented experience of using TtS devices or software in older adults with visual impairments acts more as a labor-saving factor than as working memory liability.

As a matter of fact, many researches have drawn the conclusion that the ability to perceive synthetic speech improves rapidly with training and experience either in sighted individuals [8, 6] or in individuals with visual impairments [16, 17]. Thus, individuals with visual impairments who use TtS applications for educations and professional needs have an advantage over sighted peers in word perception via synthetic speech because of their experience [11]. For instance, the results for DEMOSTHéNES (TtS platform in the Greek language) in a series of psychoacoustic experiments using similar acoustic patterns ranged from 94.5% correct responses for sighted users to 96.47% correct responses for users with visual impairments in single word tasks and from 97.5% correct responses for sighted users to 98.1% correct responses for users with visual impairments in single sentence tasks [18].

Additionally, it is possible that experience interacts with Speech Presentation Rate (SPR) affecting in this way intelligibility. Thus, speaking rate is an important

variable for manipulation when attempting to maximize the comfort, acceptance, and comprehension of synthetic speech [19]. Previous studies indicated that speaking rates between 150 and 200 wpm were the most preferred when adults listening to synthetic speech [19]. Synthetic speech presented in a slow rate allows a more accurate performance on cognitive processes such as summarizing [20]. On the other hand it is known that blind persons often prefer to use synthetic speech in fast speaking rates [21]. However, fast SPR does not imply an accurate perception of words even in individuals who prefer and use indeed synthetic speech in fast presentation mode [16, 17].

Digging out the variables that could affect intelligibility of synthetic speech is, among others, very significant since intelligibility might set obstacles for comprehension. These two terms are discussed separately in the literature. Intelligibility is the listener's ability to recognize phonemes and words when they are presented in isolation [22], whereas comprehension involves the extraction of the underlying meaning from the acoustic signals of speech [23]. Higginbotham, Drazek, Kowarsky, Scally, and Segal [20] suggested that differences in perceptual level because of the quality of synthetic speech may affect the comprehension of texts presented synthetically. [24] found that there is a moderate relationship between intelligibility scores and comprehension processing measures across different speech synthesizers. On the other hand, [10] found that appropriate context cues can rupture the interaction between intelligibility and comprehension by ameliorating comprehension results, while complexity of information have a negative effect in comprehension which is exacerbated by the type of speech – synthetic or natural [6].

2. STUDY

The present study has been designed to examine the comprehension of individuals with and without visual impairments when they have "reading" texts presented via synthetic and natural speech. In particular, the study aims to compare: a) the comprehension of texts produced in natural speech and synthetic speech for both individuals with visual impairments and sighted individuals, b) the comprehension of two groups for both natural and synthetic speech. Moreover, the effect of several individual parameters (gender, age, and experience in using TtS systems) on the comprehension of synthetic speech by individuals with visual impairments was also investigated.

2.1 Participants

Twenty young adults with visual impairments and 65 sighted young adults took part in the study. These two groups were equivalent in terms of educational level. The group of sighted individuals (15 males and 50 females) ranged in age from 18 years to 30 years (M = 22.7, SD = 2.86). The group with visual impairments consisted of 13 males and 7 females. An essential requirement to include a participant in the study was not to have a hearing impairment or other disabilities, apart from visual impairments, and to speak the Greek as his/her primary language. The age range of the adults with visual impairments was from 18 years to 30 years (M = 24.5, SD = 3.39). Fourteen participants were blind or had severe visual impairments (i.e. did not read visually by using any low vision aids) and 6 had low vision. In addition, 11 of the 20 participants were congenitally visually impaired and 9 were adventitiously visually impaired.

The participants with visual impairments were asked to indicate the main reading media that they used (i.e., Braille, TtS systems, audio cassettes, lens, large print, screen magnification software), and how often they used TtS systems. The frequency of use was described using a 5-point likert scale: quite often, often, sometimes, rarely, and not at all. To determine the most precise indication of the frequency of TtS systems use, the participants stated how many years (overall) they had used TtS systems. These descriptive data are presented in Tables 1 and 2. Fifteen out of 20 participants with visual impairments used TtS systems as basic reading medium. The sighted participants did not have any previous systematic experience in the use of synthetic speech; prior exposure to synthetic speech was incidental.

	Frequency of use				
	not at all	rarely	sometimes	often	quite often
Particip.	1	2	3	6	8

Table 1. Frequency of TtS systems use by participants with visual impairments

	Years		
	0-1	2-10	>10
Participants	6	12	2

Table 2. Years of use TtS systems by participants with visual impairments

2.2 Instruments – Procedures

Before each test began, the participants were informed in detail about the procedure that would follow. They were told that they were going to listen to two texts, one produced by synthetic speech and one produced by natural speech, and that they would be asked to respond to 10 comprehension statements at the end of each text. Thus, the participant had to listen carefully to each text, without repeating what he or she heard, so that he or she would be able to answer the comprehension statements that followed. On the basis of the text that had just been presented, each participant had to answer yes or no, depending on whether he or she felt that the statement was right or wrong.

During the construction of the tests, a female voice was used to record the natural and synthetic speech. Moreover, special care was taken to ensure that the speed

of presentation was the same for both the natural and synthetic speech. The natural speech was recorded in a recording studio. For the recording of the synthetic speech, one TTS platform (in Greek) was used, together with all the appropriate recording devices.

The tests were conducted in a quiet room, to avoid the effect of background noise. The participants first listened to two texts in synthetic speech and verified or rejected 10 comprehension statements that were made up by the researcher. The same procedure was repeated with the second text, which was presented in natural speech. The texts were taken from a history book and arranged in such a way that they had the same level of difficulty (the same topic and similar vocabulary). All the texts were taken from scientific historical texts and were relatively difficult to retain because they included several historical details.

It was made sure that both texts had similar degrees of difficulty. However, to reduce further the possibility that the results would be distorted because of the different degrees of difficulty of the texts, the following procedure was used: the two texts were recorded both in synthetic and natural speech. Then, two subtests were created. In the first subtest, the first text was generated with synthetic speech and the second text was generated with natural speech. Conversely, in the second subtest, the second text was presented in synthetic speech, whereas the first text was presented in natural speech. The participants were separated into two groups. The first group consisted of 10 individuals who were given the first subtest, and the second group consisted of 10 individuals who were given the second subtest.

3. RESULTS

The minimum, maximum, mean, and standard deviation (SD) of correct answers in the comprehension test presented in Tables 3 and 4. Each correct answer was scored 1. Thus, if any participant had answered all the questions correctly, his or her score would be equal to 10.

	Natural speech			
	Min	Max	Mean	SD
Visually Impaired	4	10	7.10	1.832
Sighted	5	10	7.51	1.382

Table 3. Minimum, maximum, mean, and standard deviation (SD), of correct answers (natural speech)

	Synthetic speech			
	Min	Max	Mean	SD
Visually Impaired	5	10	7.30	1.490
Sighted	3	10	7.45	1.640

Table 4. Minimum, maximum, mean, and standard deviation (SD), of correct answers (synthetic speech)

T-tests were conducted to examine the differences between the two groups (visually impaired vs. sighted), and repeated-measures ANOVAs were conducted to examine the differences between the speech types (natural vs. synthetic). The repeated-measures ANOVAs revealed no significant differences between natural and synthetic speech comprehension for the participants with visual impairments as well as the sighted participants. These findings indicate that the participants performed equally well when listening to synthetic or natural speech. Moreover, the t-tests revealed no significant differences between sighted participants and participants with visual impairments regarding the comprehension of natural and synthetic speech.

Regarding synthetic speech, we also investigated if there was a relation between performance of individuals with visual impairments and the following variables: gender, age, frequency of TtS use, and overall duration (in years) of TtS use. The t-tests revealed no significant differences between males and females regarding the comprehension of natural and synthetic speech. Moreover, the correlation analysis showed a significant positive correlation between correct answers in comprehension test and: a) frequency of TtS use (r = .485, p < .05), and b) the overall duration (in years) of the TtS use (r = .450, p < .05). The more experience with using the TtS systems, the more correct responses that were given during the comprehension test. All the other correlations were no significant.

4. CONCLUSIONS

The results regarding the first aim of the study indicated that the participants performed at a similar level in the comprehension of texts that were presented via synthetic and natural speech. This finding is in accordance with the results of previous research with individual with visual impairments [25, 10].

Previous studies [10, 11 & 16] revealed significant differences on the intelligibility of natural and synthetic speech. However, the findings of the present study indicate that prospective difficulties in intelligibility did not affect comprehension. It seems that context cues provided by the text assisted the participants in identifying and comprehending the text more effectively. If we take into consideration that the overall purpose of reading is comprehension, this finding has a unique practical value because it indicates that the use of TtS systems by individuals with visual impairments does not affect the intended purpose of reading. Future research should experiment with various texts of increased language difficulty to verify this conclusion.

The results regarding the second aim of the study revealed no significant differences between sighted participants and participants with visual impairments regarding the comprehension of natural and synthetic speech. Moreover, the relation between comprehension of texts presented via synthetic speech and experience (years of use) as well as frequency of use TsS systems was revealed. The more the experience with using the TtS systems, the more correct the responses that were given during the comprehension test. According to Reynolds

and Jefferson [26] as well as Koul [27], the perception and comprehension of synthetic speech are improved with training and practice. They are also improved with repeated and systematic exposure to synthetic speech [28].

The findings of this study contribute to the understanding of issues that concern synthetic speech comprehension by individuals with visual impairments. Thus, the results of the study have implications for both educators and assistive technology developers.

Acknowledgements

This research has been co-financed by the European Union (European Social Fund – ESF) and Greek national funds through the Operational Program "Education and Lifelong Learning" of the National Strategic Reference Framework (NSRF) under the Research Funding Project: "THALIS - University of Macedonia - KAIKOS: Audio and Tactile Access to Knowledge for Individuals with Visual Impairments", MIS 380442 .

REFERENCES

[1] D. Freitas and G. Kouroupetroglou, "Speech technologies for blind and low vision persons," Technology and Disability, vol. 20, pp. 135–156, 2008.

[2] D. Goudiras, K. Papadopoulos, A. Koutsoklenis, V. Papageorgiou, and M. Stergiou, "Factors affecting the reading media used by visually impaired adults," British Journal of Visual Impairment, vol. 27, pp. 111–127, 2011.

[3] K. Papadopoulos and A. Koutsoklenis, "Reading media used by higher-education students and graduates with visual impairments in Greece," Journal of Visual Impairment & Blindness, vol. 103, pp. 772–779, 2009.

[4] R. Koul and J. Hanners, "Word identification and sentence verification of two synthetic speech systems by individuals with intellectual disabilities," Augmentative and Alternative Communication, vol. 13, no. 2, pp. 99-107, 1997.

[5] E. O'Bryan, "Processing differences in synthetic versus natural speech," MIT Working Papers in Linguistics, vol. 38, pp. 169-177, 2000.

[6] S. J. Winters and D. B. Pisoni, "Speech synthesis: Perception and comprehension," in K. Brown (Ed.), Encyclopedia of Language and Linguistics, vol. 12, pp. 31-49, 2005.

[7] D. B. Pisoni, L. M. Manous, and M. J. Dedina, "Comprehension of natural and synthetic speech: Effects of predictability on the verification of sentences controlled for intelligibility," Computer speech & language, vol. 2, no. 3, pp. 303-320, 1987.

[8] S. J. Winters and D. B. Pisoni, "Perception and comprehension of synthetic speech", in Research on Spoken Language Processing Progress Report no. 26, Speech Research Laboratory Psychology Department, Indiana University, Bloomington, 2004, pp. 95-138.

[9] J. Hensil and S. G. Whittaker, "Visual reading versus auditory reading by sighted persons and persons with low vision," Journal of Visual Impairment & Blindness, vol. 94, pp. 762-770, 2000.

[10] K. Papadopoulos, A. Koutsoklenis, E. Katemidou, and A. Okalidou, Perception of natural and synthetic speech by adults with visual impairments," Journal of Visual Impairment & Blindness, vol. 103, pp. 403–414, 2009.

[11] K. Papadopoulos, E. Katemidou, A. Koutsoklenis, and E. Mouratidou, "Differences amongst sighted individuals and individuals with visual impairments in word intelligibility presented via synthetic and natural speech," Augmentative and Alternative Communication, vol. 26, pp. 278–288, 2010.

[12] C. Stevens, N. Lees, J. Vonwiller, and D. Burnham, "On-line experimental methods to evaluate text-to-speech (TtS) synthesis: Effects of voice gender and signal quality on intelligibility, naturalness and preference," Computer Speech and Language, vol. 19, pp. 129–146, 1995.

[13] R. K. Koul and G. D. Allen, "Segmental intelligibility and speech interference thresholds of high quality synthetic speech in presence of noise," Journal of Speech and Hearing Research, vol. 36, pp. 790–798, 1993.

[14] J. B. Hardee and C. B. Mayhorn, "Reexamining synthetic speech: Intelligibility and the effects of age, task, and speech type on recall," in Proceedings of the Human Factors and Ergonomics Society Annual Meeting, vol. 51, no. 18, 2007, pp. 1143-1147.

[15] A. M. Sinatra, V. K. Sims, S. K. T. Bailey, and M. B. Najle, "Differences in the performance of older and younger adults in a natural vs. synthetic speech dichotic listening task," in Proceedings of the Human Factors and Ergonomics Society Annual Meeting, vol. 57, no. 1565, 2013.

[16] M. Barouti, K. Papadopoulos, and G. Kouroupetroglou, "Synthetic and natural speech intelligibility in individuals with visual impairments: effects of experience and presentation rate," in P. Encarnação et al. (Eds.) Assistive Technology: From Research to Practice, IOS press, pp. 695-701, 2013. doi:10.3233/978-1-61499-304-9-695

[17] A. Stent, A. Syrdal, and T. Mishra, "On the intelligibility of fast synthesized speech for individuals with early-onset blindness," in The proceedings of the 13th international ACM

SIGACCESS conference on Computers and accessibility, ACM, 2011, pp. 211-218.

[18] V. Argyropoulos, K. Papadopoulos, G. Kouroupetroglou, G. Xydas, and P. Katsoulis, "Discrimination and perception of the acoustic rendition of texts by blind people," Lecture Notes in Computer Science, vol. 4556, pp. 205–213, 2007.

[19] B. Sutton, J. King, K. Hux, and D. R. Beukelman, "Younger and older adults' rate performance when listening to synthetic speech," Augmentative and Alternative Communication, vol. 11, pp. 147–153, 1995.

[20] D. J. Higginbotham, A. Drazek, K. Kowarsky, C. Scally, and E. Segal, "Discourse comprehension of synthetic speech delivered at normal and slow presentation rates," Augmentative and Alternative Communication, vol. 10, no. 3, pp. 191-202, 1994.

[21] I. Hetrich, S. Dietrich, A. Moos, and J. Trouvain, "Enhanced speech perception capabilities in a blind listener are associated with activation of fusiform gyrus and primary visual cortex," Neuroscience: The Neural Basis of Cognition, vol. 15, pp. 163–170, 2009.

[22] J. V. Ralston, D. B. Pisoni, and J. W. Mullennix, "Comprehension of synthetic speech produced by rule", in Research on Speech Perception Progress Report no.15, Speech Research Laboratory, Psychology Department, Indiana University, Bloomington, 1989, pp. 77–132.

[23] S. A. Duffy and D. B. Pisoni, "Comprehension of synthetic speech produced by rule: A review and theoretical interpretation," Language and Speech, vol. 35, pp. 351–389, 1992.

[24] J. V. Ralston, D. B. Pisoni, S. E. Lively, B. G. Greene, and J. W. Mullennix, "Comprehension of synthetic speech produced by rule: Word monitoring and sentence-by-sentence listening times," Human Factors, vol. 33, pp. 471–491, 1991.

[25] K. Papadopoulos, V. Argyropoulos, and G. Kouroupetroglou, "Discrimination, perception and comprehension of synthetic speech by students with visual impairments: the case of similar acoustic patterns," *Journal of Visual Impairment and Blindness,* vol. 102, pp. 420–429, 2008.

[26] M. E. Reynolds and L. Jefferson, "Natural and synthetic speech comprehension: Comparison of children from two age groups," Augmentative and Alternative Communication, vol. 15, pp. 174-182, 1999.

[27] R. Koul, "Synthetic speech perception in individuals with and without disabilities," Augmentative and Alternative Communication, vol. 19, pp. 49–58, 2003.

[28] R. Koul and K. Hester, "Effects of repeated listening experiences on the recognition of synthetic speech by individuals with severe intellectual disabilities," Journal of Speech, Language, and Hearing Research, vol. 49, pp. 47–57, 2006