Modern Psychological Studies

Volume 29 | Number 1 Article 24

2023

To Read or to Listen: The Effect of Text-to-Speech Software and Accents on Comprehension of Digital Media

Lily Kiamanesh
University of California, Los Angeles, lkiamanesh@g.ucla.edu

Dillon Murphy
University of California, Los Angeles

Alan Castel
University of California, Los Angeles

Abstract

Text-to-speech software is a popular tool for consuming information for school, work, and recreation. This study examined two aspects of text-to-speech information exposure: the effect of audio/text modality and type of accent. In Experiment one, participants received material presented in a text-only, audio-only, or dual format. Those who received the material as text-only had significantly higher scores on a comprehension test and higher score predictions than those in the audio-only condition. In Experiment two, participants were presented with audio material in a text-to-speech-generated U.S. English, Mandarin Chinese, or Italian accent. Results revealed that accent did not significantly impact retention or performance predictions. These findings call into question the use of text-to-speech software as an "easier" way to retain information.

Keywords: text-to-speech, audio, text modality, metacognition, digital media, accent

To Read or to Listen: The Effect of Text-to-Speech Software and Accents on Comprehension of Digital Media

In today's work and school environment, almost all necessary information is received digitally. Emails, textbooks, and news articles, to name a few, are frequently consumed via electronic devices. A 2018 survey of over 10,000 American college students found that about 61% used one or more e-textbooks (Abaci et al., 2019). Additionally, with the COVID-19 pandemic, the demand for e-books and electronic learning platforms has increased drastically (Kipp, 2021; Kodama et al., 2021). Many e-textbook providers (e.g., Pearson with VitalSource and Macmillan's LaunchPad) offer a "Read Aloud" function for their books that uses text-to-speech software to generate an audio recording of the material. There are also browser extensions and applications that will read text aloud, whether it is emails, articles, documents, e-books, or any other webpage (e.g., Speechify with over one million users or Microsoft Edge's built-in "Read Aloud" tool; Speechify, n.d.; Use Immersive Reader, n.d.). These services provide a dictation of the text for users to listen to or read along with. While these tools are often marketed as accessibility measures for those who have vision impairments or reading disabilities, others make broad claims that the tools decrease study time, increase learning, and boost grades. Students may find that these services increase the convenience of studying because they allow them to engage in other tasks, such as driving, or just to look away from their screens while continuing to learn (Abaci et al., 2019). Additionally, some students may believe that listening and reading together will boost recall and produce more comprehension than simply reading alone. While convenient, this format of information exposure may come with some disadvantages, which the current study seeks to examine.

Current Text-to-Speech Use

Text-to-speech software has long been used as a tool to help individuals with reading and comprehension difficulties (e.g., aphasia, dyslexia, and ADHD), focus on text, and retain more information (Bonifacci et al., 2022; Elkind et al., 1993; Grunér et al., 2018; Hecker et al., 2002; Knollman-Porter et al., 2022; Wallace et al., 2021). Bonifacci and colleagues (2022) found that with text-to-speech-assisted reading, students with dyslexia experienced reduced mind-wandering, which is defined as a shift in attention from current material and events to self-generated and task-unrelated thoughts. The researchers found that this reduction in mind-wandering may have contributed to the participants' increased comprehension. Additionally, a 2017 meta-analysis found that text-to-speech tools positively impacted comprehension among students with reading disabilities (Wood et al., 2018). However, some researchers have found that text-to-speech may have no impact or may potentially be detrimental to the comprehension of those with reading disabilities (Elkind et al., 1993; Harvey & Hux, 2015; Schmitt et al., 2011).

Along with this ambiguity, there is a lack of literature concerning the effects of text-to-speech on comprehension in the general population, specifically those without reading and comprehension difficulties. Bonifacci and colleagues (2022) found that while comprehension of students with learning disabilities strongly increased when using text-to-speech software, those with typical development experienced only marginal improvements. The current study aims to provide more insight into the general application of text-to-speech tools.

Effect of Modality on Comprehension

With the aforementioned text-to-speech functionalities, readers now have three modalities by which they can consume written information: reading alone, listening while reading, and listening alone. Experiment one explores how these methods affect learning. In

studies using audio recordings of human voices rather than text-to-speech, reading alone has typically produced the best comprehension (Daniel & Woody, 2010; Green, 1981; Kopp & D'Mello, 2016; Leahy & Sweller, 2011). For example, Daniel and Woody (2010) found that participants who listened to an audio recording of an article scored significantly worse on a comprehension quiz than those who simply read it. In another study using human narration rather than text-to-speech, Kopp and D'Mello (2016) found that participants in the audio-only condition performed significantly worse than those in the text-only and dual (concurrent audio and text) conditions. The researchers attributed this discrepancy to the executive resource hypothesis, the theory that both the text format and dual format require the brain to use more higher-order processes than the audio format. Only listening makes use of fewer systems of our executive resources, with the phonological loop (auditory component of working memory) being the primary one, whereas the text format and dual format also recruit the visuospatial sketchpad (visual and spatial component of working memory). Drawing upon this executive resource hypothesis, the researchers reasoned that tasks that demand more executive resources limit the mind's ability to wander. Therefore, the participants in the audio-only condition would be more susceptible to mind-wandering. The study concluded that there was a negative correlation between this mind-wandering and comprehension among the groups (Kopp & D'Mello, 2016).

Additionally, poor comprehension from participants who only listen to the material may be attributed to the transient information effect. As Leahy and Sweller (2011) explain, visual information is permanent. When reading, one may revisit previous information frequently without cognitive strain. However, auditory information is ephemeral, meaning if the listener wishes to return to previous material, they must memorize it, placing a burden on the working memory. This burden increases with more complex information. However, it is not harmful to

comprehension when learners consume short or familiar information in an auditory format. However, listening to lengthy and difficult material may cause comprehension to suffer (Leahy & Sweller, 2011).

While most prior work has shown that only listening to material results in less comprehension than the dual and text-only modalities, some research has produced contrasting results (Clinton-Lisell, 2022; Klein, 1989; Rogowsky et al., 2016). For example, a 2021 meta-analysis comparing participants who only read and participants who only listened determined that the two were not reliably different in terms of comprehension (Clinton-Lisell, 2022). Similarly, Rogowsky and colleagues (2016) found no significant difference in comprehension among participants in digital audiobook, e-text, and dual modality groups. In contrast, Klein (1989) found that participants who both read and listened performed better than those who solely read the material. These results highlight the uncertainty concerning the relationship between modality and comprehension.

Effect of Accent on Comprehension

Along with offering different modalities for consuming information, many text-to-speech services provide users the ability to change the accent of the dictated text. Learners may employ this tool to increase the variety of their readings or in the hope that a different accent will improve learning outcomes. Experiment two explores the potential effects text-to-speech accents have on comprehension.

Many studies have found that accented speech produces more comprehension errors and slower processing times than speech from native speakers (Adank et al., 2009; Anderson-Hsieh & Koehler, 1988; Chan et al., 2019; Clarke & Garrett, 2004; Floccia et al., 2006; Kang et al., 2019; Munro & Derwing, 1995; Ockey et al., 2016). For instance, one study found that when

native English speakers listened to passages dictated by speakers with Chinese accents of three comprehensibility levels and one native English speaker, the comprehension scores were significantly lower for the non-native passages (Anderson-Hsieh & Koehler, 1988). Furthermore, participants' scores corresponded with the degree of the speaker's accent (i.e., the stronger the accent, the lower the score). Additionally, Munro and Derwing (1995) tested participants' accuracy and response times in determining the truth and falsity of statements spoken in a native English accent and Mandarin-accented English. They found that participants correctly identified the truthfulness of statements from native English speakers more frequently than those from Mandarin speakers. Regarding response time, Floccia and colleagues (2006) conducted an experiment in which monolingual French speakers heard words spoken in familiar and unfamiliar native accents. They found that unfamiliar regional accents produced a 30-millisecond delay in word identification. Clarke and Garrett (2004) found that foreign accents resulted in a 100- to 150-millisecond delay in their task. These increases in processing time may be partially responsible for the decreased comprehension that accented speech is thought to produce.

As for the factors that contribute to differences in accent processing, there are many hypotheses. Floccia and colleagues (2006) posit that there is a time delay because short-term processing systems must normalize the accent before comprehension may occur. Munro and Derwing (1995), as well as Kang and colleagues (2019), claim that accents may be differentially processed due to their varying levels of comprehensibility. Specifically, Kang and colleagues (2019) found that as long as speakers were rated as being highly comprehensible, there was no significant difference in scores between accent groups. Many other studies have determined that social perceptions of accented speech cause listeners to regard non-native accents as less truthful, acceptable, pleasant, and credible than native accents (De Meo, 2012; Dragojevic & Giles, 2016;

Evans & Michael, 2014; Kaur & Raman, 2014; Lev-Ari & Keysar, 2010, 2012). Lev-Ari and Keysar (2012) hypothesize that listeners expect less reliability when listening to accented speech and therefore represent the language in less detail. This adjustment in language processing would decrease participants' comprehension of the material. Thus, prior research indicates that accented speech produces less comprehension than native speech; however, these studies have all employed human voices rather than computer-generated ones. Whether text-to-speech will produce any differences in comprehension of accented recordings is uncertain, and the present study will explore this question.

The Current Study

In this study, we investigated how consuming digital passages in one of three modalities or with accented text-to-speech recordings affects comprehension and metacognitive monitoring of learning. Specifically, in Experiment one, we presented participants with a passage as either only text, only text-to-speech audio, or both text and text-to-speech audio concurrently (dual modality condition). We hypothesized that those in the text-only condition would perform better on the comprehension test than those in the audio-only or dual modality conditions. As for participants' metacognitive predictions of performance, we believed the text-only condition would produce the highest score predictions.

To understand whether accented text-to-speech audio recordings serve to benefit or detriment participants' comprehension of the material, in Experiment two, we presented participants with audio passages in either a U.S. English, Mandarin Chinese, or Italian accent. Based on prior research, we hypothesized that participants would show more comprehension when the material was presented in a native U.S. English accent. We also expected participants in the U.S. English accent condition to provide the highest metacognitive predictions of their

performance. Ultimately, the results of this study will help assess the validity of claims that text-to-speech productivity tools can boost comprehension of digital information. Additionally, the results will provide insight into how text-to-speech accents may affect comprehension and predictions of performance.

Experiment 1

Most prior work has shown that learners comprehend more when consuming material in a text-only format as compared to an audio-only and a dual modality. In Experiment one, we extended this prior work by exploring the phenomenon with text-to-speech audio rather than human voices. We investigated how memory for information is affected by the method of exposure: either by reading text alone, listening to a text-to-speech audio recording only, or reading the text and listening to the text-to-speech audio at the same time. Specifically, participants read and/or listened to an article on historical and general information about Yellowstone National Park and were immediately tested on the article's content. Additionally, we solicited predictions of performance to determine whether participants were metacognitively aware of any potential effects of the method of exposure on their comprehension. Based on prior work, we expected that the text-only condition would display more comprehension than those in the audio-only and dual conditions. We also believed that the text-only group would predict their scores to be the highest.

Method

Participants

Participants consisted of 157 undergraduate students (aged 17-35: M_{age} = 20.82, SD_{age} = 2.67) recruited from the University of California, Los Angeles (UCLA) Human Subjects Pool. Informed consent was obtained, and the study was completed in accordance with the UCLA

institutional review board. Participants were tested online and received course credit for their participation. Participants were excluded from analysis if they admitted to cheating (e.g., writing down answers) in a post-task questionnaire (they were told they would still receive credit if they cheated; this exclusion process resulted in zero exclusions. We aimed to collect around 50 participants per condition. The sample size was selected based on prior exploratory research and the expectation of detecting a medium effect size. A sensitivity analysis indicated that, with the given sample size, assuming $\alpha = .05$, we had an 80% chance of detecting a medium effect size (Cohen's $f = 0.25$) concerning the relationship between modality and comprehension.

Materials

Participants listened to and/or read a passage on Yellowstone National Park taken from Little (2011). The text had 1,116 words, and the audio recording was 8 min and 52 s. To generate the audio recording, we used Google Translate's text-to-speech software with the U.S. English language selected. To measure learning, we administered a comprehension test (20 questions) consisting of multiple-choice questions with four options for participants to select from (also taken from Little (2011)). Comprehension was calculated as the proportion of questions answered correctly.

Procedure

Participants were either told that they would listen to, read, or listen to and read along with a passage and then take a comprehension test on the covered material. They were also instructed not to take any notes. Participants were then presented with the passage one paragraph at a time, and each group had the material for the same amount of time (8 min and 52 s). Participants who were reading the passage were shown one paragraph at a time, with that paragraph's duration being the same as the audio recording for that paragraph. Following

exposure to the passage, participants were asked to predict how many of the 20 questions on the comprehension test they would get correct. Finally, they completed the comprehension test.

Results

Comprehension and predictions of comprehension as a function of type of exposure are shown in Appendix A. To examine predictions of performance on the comprehension test, we conducted an ANOVA to see how the conditions of text, audio, or both affected score prediction $F(2, 154) = 3.14$, $p = .046$, $\eta^2 = .04$. Results revealed that participants in the text-only condition ($M = .52$, $SD = .17$) predicted their scores to be significantly higher than those in the audio-only condition ($M = .43$, $SD = .20$; $p_{tukey} = .040$, $d = -.48$). However, neither the text-only nor the audio-only groups' predictions were significantly different from those of participants in the dual modality condition ($M = .46$, $SD = .23$), all $ps \geq 0.21$.

To examine performance on the test, we conducted an ANOVA to see the effect of the three modality conditions on comprehension scores $F(2, 154) = 3.98$, $p = .021$, $\eta^2 = .05$. Results revealed that participants in the text-only condition ($M = .53$, $SD = .20$) scored significantly higher than those in the audio-only condition ($M = .43$, $SD = .15$; $p_{tukey} = .015$, $d = -.55$). Again, however, neither group's scores were significantly different from those of the dual modality condition ($M = .47$, $SD = .17$), all $ps \geq 0.23$.

Additionally, we wanted to analyze the effect of native speaker status on comprehension to examine whether having English as a first language would produce a difference in scores. After conducting an independent samples t-test, we found that participants who were native speakers of English performed significantly better on the comprehension test, $t = -2.43$, $p = .016$, $d = -.42$. To examine the extent of this relationship, we conducted a 3 (condition: text only, audio only, both text and audio) x 2 (native English speaker status: yes, no) between-subjects ANOVA,

which showed no interaction between native speaker status, modality, and comprehension scores $F(2, 151)= 1.02$, $p = .361$, $\eta^2 = .01$. We also conducted a 3 (condition: text only, audio only, both text and audio) x 2 (native English speaker status: yes, no) between-subjects ANOVA to investigate the relationship between native speaker status, modality, and score predictions and found no interaction $F(2, 151) = 1.59$, $p = .207$, $\eta^2 = .02$.

Experiment 2

In Experiment one, we found that the use of text-to-speech audio, either alone or in a dual modality, did not appear to enhance comprehension. In Experiment two, we investigated the effect that accents of text-to-speech recordings may have on memory for information presented in an audio passage. Specifically, participants listened to text-to-speech audio recordings of passages on the planet Saturn and stimulant drugs in either a U.S. English accent, an Italian accent, or a Mandarin Chinese accent and were tested on the passages' content immediately after listening. The U.S. English accent was selected as the neutral accent condition, while the Italian and Mandarin Chinese conditions provided intelligible yet foreign accents with which participants were not expected to be very familiar. Additionally, we solicited metacognitive predictions of performance to determine whether participants were aware of any potential effects of accents on comprehension. Based on the literature reviewed, we expected that participants would comprehend more when the text-to-speech used a native U.S. English accent. We also expected that participants in the U.S. English accent condition would have the highest predictions of their performance.

Method

Participants

Participants consisted of 150 undergraduate students (aged 18-50: M_{age} = 20.95, SD_{age} = 4.09) recruited from the UCLA Human Subjects Pool. Participants were tested online and received course credit for their participation. Informed consent was acquired, and the study was completed in accordance with the UCLA institutional review board. Participants were excluded from analysis if they admitted to cheating (e.g., writing down answers) in a post-task questionnaire (they were told they would still receive credit if they cheated; this exclusion process resulted in two exclusions. A sensitivity analysis indicated that, with the given sample size, assuming $\alpha = .05$, we had an 80% chance of detecting a medium effect size (Cohen's f = 0.26) concerning the relationship between condition and comprehension.

Materials

Participants listened to two text-to-speech audio recordings featuring the same accent (either U.S. English, Italian, or Mandarin). The passages were taken from Little (2011), and one discussed the planet Saturn (1,110 words), while the other covered stimulant drugs (628 words). The audio recordings were created using Google Translate's text-to-speech software with either the U.S. English language, Italian language, or Mandarin Chinese language selected. In the U.S. English accent, the Saturn passage was 8 min and 22 s, and the stimulants passage was 5 min and 9 s. In the Italian accent, the Saturn passage was 9 min and 40 s, and the stimulants passage was 5 min and 56 s. In the Chinese accent, the Saturn passage was 10 min and 46 s, and the stimulants passage was 6 min and 35 s. The recordings are of differing lengths because Google Translate's various accents dictate texts at different speeds (i.e., the Chinese accent "speaks" more slowly than the U.S. English). To measure learning, we used a comprehension test (20 questions) for each passage consisting of multiple-choice questions containing four options for participants to select from (also taken from Little (2011)).

Procedure

Participants were randomly assigned to listen to both passages in either the U.S. English, Italian, or Mandarin accent. Participants were told they would be listening to material and then taking a comprehension test on the passage. They were also instructed not to take any notes. The participants listened to either the Saturn or stimulants passage first (counterbalanced), made a prediction of their test performance, and took a comprehension test. When making predictions, participants were asked how many of the 20 questions they expected to get correct. Next, participants repeated this procedure for the other passage.

Results

Comprehension and predictions of comprehension as a function of accent are shown in Appendix B. To examine predictions of performance on the comprehension test, we conducted an ANOVA to see how the conditions of U.S. English accent, Mandarin Chinese accent, and Italian accent affected score prediction $F(2, 147) = 1.89$, $p = .155$, $\eta^2 = .03$. Results revealed that regardless of accent condition, score predictions were not significantly different among participants in the U.S. English accent condition ($M = .38$, $SD = .19$), Mandarin accent condition ($M = .41$, $SD = .17$), and Italian accent condition ($M = .34$, $SD = .16$).

To examine performance on the comprehension test, we conducted an ANOVA to see how the accent conditions affected score $F(2, 147) = 1.65$, $p = .196$, $\eta^2 = .02$. Again, results showed that regardless of the accent, scores were not significantly different in the U.S. English ($M = .43$, $SD = .15$), Mandarin ($M = .42$, $SD = .11$), and Italian accent conditions ($M = .39$, $SD = .11$).

Prior research indicates that in listening comprehension tests of human voices, non-native English speakers tend to achieve the best comprehension when listening to native English

speakers (Kang et al., 2019). To examine differences in predictions and scores among non-native English speakers, we conducted an independent samples t-test. Results revealed that there was no significant difference between the two groups, $t = .61$, $p = .55$. Additionally, participants who noted that they were fluent in another language did not perform significantly differently than those who did not, $t = .20$, $p = .84$.

Discussion

In today's work and school environment, digitally presented information has become the norm. From e-books to emails, we are continuously presented with information on our electronics. When tasked with reading and comprehending digital material, some people may turn toward text-to-speech services that claim to improve learning outcomes and allow for an easier way to consume this information. These services (e.g., browser extensions and players built into e-textbook readers) allow learners the opportunity to read along with the recording or simply listen to the computer-generated audio. Additionally, many text-to-speech software services allow users to manipulate the accent of the dictation, providing an array of languages from which to choose. These options may add variety to learning methods and allow for multitasking while reading. In the current study, we tested the effect of these tools on comprehension of digital media.

In Experiment one, we presented participants with a passage in one of three modalities: reading in the text-only condition, listening to the text-to-speech audio recording in the audio-only condition, or reading and listening simultaneously in the dual modality condition. Prior work has produced some contrasting results, with many studies finding that the text-only modality produces the most comprehension (Daniel & Woody, 2010; Green, 1981; Kopp & D'Mello, 2016; Leahy & Sweller, 2011), while others determined that there was no significant

difference between conditions (Clinton-Lisell, 2022; Rogowsky et al., 2016). However, almost all previous work examining modality has used human voices rather than recordings from text-to-speech software. Thus, it was largely unclear how text-to-speech and learning modality might affect comprehension of digital information.

After participants consumed the information, they made a metacognitive prediction of their comprehension score and then took a test. Results revealed that those in the text-only condition produced both the highest predictions of performance and the highest scores on the comprehension test. This finding is consistent with the body of literature that asserts that reading alone is the most effective way to consume information (e.g., Daniel & Woody, 2010). Furthermore, the significantly higher score predictions for the text-only condition as compared to the audio-only condition indicate that learners are sensitive to the impact of modality on their comprehension. Their metacognitive awareness suggests that readers felt more confident in their learning than those who used text-to-speech tools. They then justified their higher confidence levels through increased comprehension scores. However, the hierarchy of the effectiveness of learning methods in terms of comprehension remains somewhat uncertain. Our findings show that the dual modality group placed in between the text-only and audio-only conditions for both predictions and scores with there being no statistically significant difference between either one. Further investigation is required to determine how simultaneous information exposure affects learning.

Furthermore, analyses showed that native English speakers performed significantly better on the comprehension test than non-native speakers; however, there was no interaction between modality, native speaker status, and comprehension or between modality, native speaker status, and prediction. Therefore, non-native-English-speaking participants did not appear to benefit

from the use of text-to-speech tools in terms of increased comprehension or confidence in learning.

In Experiment two, we wanted to determine whether there was a relationship between the accent of a text-to-speech audio recording and comprehension of the dictated material. Previous work has shown that accented speech almost always negatively affects comprehension (Adank et al., 2009; Anderson-Hsieh & Koehler, 1988; Chan et al., 2019; Clarke & Garrett, 2004; Floccia et al., 2006; Kang et al., 2019; Munro & Derwing, 1995; Ockey et al., 2016). However, it again appears that all prior work has used human speech, so the effect of accented text-to-speech was previously unexplored.

Participants listened to two passages in one of three accents: U.S. English, Mandarin Chinese, or Italian. Again, they made metacognitive predictions and took a comprehension test. Contrary to the conclusions of prior work, our results showed that regardless of the presented accent, participants' predictions and scores were not significantly different. These findings are surprising given that the foreign language accents (Mandarin, Chinese, and Italian) were relatively strong, with the pronunciation of some English words being very different from that of the U.S. English accent. However, it is possible that the articulation of all three conditions was very intelligible, supporting the conclusion of Kang and colleagues (2019) that it is comprehensibility, not accent, that determines learning levels. In Experiment two, text-to-speech is likely responsible for this comprehensibility due to the software's clear enunciation of most words.

There were also no significant differences observed in terms of comprehension and prediction for participants who were non-native English speakers as well as for those who spoke another language fluently. This is somewhat surprising as non-native speakers typically believe

that accented English will impede their comprehension (Abeywickrama, 2013), and often non-native listeners do not comprehend non-native speech as well as they do native speech (Smith & Bisazza, 1982). Again, we suspect that this discrepancy between prior work and current results may be attributed to text-to-speech's ability to make accents equally comprehensible.

Limitations and Future Directions

Overall, it is important to consider the potential limitations of the current study that may affect the generalizability of the findings. In Experiment one, it is possible that participants could have prior knowledge that would enhance their scores on the comprehension test. Additionally, native U.S. English speakers could be more knowledgeable about Yellowstone National Park. This familiarity could explain why native English speakers scored significantly better than non-native speakers.

Potential limitations for Experiment two also include the unknown influence of prior knowledge. Participants who were familiar with Saturn and/or stimulant drugs could have scored more highly regardless of their accent condition. Another limitation is that, when generating accented text-to-speech recordings, audio lengths varied, with the Chinese and Italian accent recordings being longer than those of the English. Although there were no significant differences in the results between the conditions, these discrepancies in audio lengths may have affected participants' predictions and/or performance. The increased audio length would require longer sustained attention from participants, potentially resulting in decreased effort and comprehension. Additionally, another limitation may stem from there being only three accent conditions whereas other accents may produce different results.

While the present study demonstrated that there is no significant difference between text-to-speech accents, future work could investigate how text-to-speech accents compare with

human accents in terms of comprehension. This research could help determine if accented text-to-speech is more comprehensible than accented human speech as we suspected in the current study. Additionally, previous work has examined how text-to-speech software benefits individuals with learning or reading disabilities, but there is no research that shows the effect that accented text-to-speech produces. Future researchers could examine whether accented text-to-speech recordings have any effect on the metacognitive predictions or comprehension of this population.

Conclusion

The results of the current study provide insight to those who must comprehend digital information and seek an easier way to consume it. Text-to-speech software, such as third-party extensions or built-in players on websites and e-textbooks, may be appealing due to claims that it can improve productivity, make studying easier, and decrease the amount of time needed to absorb information. However, the present study shows that these tools do not benefit comprehension. Additionally, the results suggest that manipulating the accent of the text-to-speech audio recording has no significant effect on comprehension. Thus, despite the increasing prevalence of text-to-speech tools to dictate text aloud, simply reading and forgoing assistance from various software may be the most effective way to comprehend material.

References

Abaci, S., BrckaLorenz, A., & Quick, J. (2019, April 8). *Examining students' use of, preferences for, and learning with e-textbooks* [Paper presentation]. American Educational Research Association (AERA) 2019 Annual Meeting, Toronto, Canada.

Abeywickrama, P. (2013). Why not non-native varieties of English as listening comprehension test input? *RELC Journal*, *44*(1), 59–74. https://doi.org/10.1177/0033688212473270

Adank, P., Evans, B. G., Stuart-Smith, J., & Scott, S. K. (2009). Comprehension of familiar and unfamiliar native accents under adverse listening conditions. *Journal of Experimental Psychology: Human Perception and Performance*, *35*(2), 520–529. https://doi.org/10.1037/a0013552

Anderson-Hsieh, J., & Koehler, K. (1988). The effect of foreign accent and speaking rate on native speaker comprehension. *Language Learning*, *38*(4), 561–613. https://doi.org/10.1111/j.1467-1770.1988.tb00167.x

Bonifacci, P., Colombini, E., Marzocchi, M., Tobia, V., & Desideri, L. (2022). Text-to-speech applications to reduce mind wandering in students with dyslexia. *Journal of Computer Assisted Learning*, *38*(2), 440–454. https://doi.org/10.1111/jcal.12624

Chan, K. Y., Chiu, M. M., Dailey, B. A., & Jalil, D. M. (2019). Effect of foreign accent on immediate serial recall. *Experimental Psychology*, *66*(1), 40–57. https://doi.org/10.1027/1618-3169/a000430

Clarke, C. M., & Garrett, M. F. (2004). Rapid adaptation to foreign-accented English. *The Journal of the Acoustical Society of America*, *116*(6), 3647. https://doi.org/10.1121/1.1815131

Clinton-Lisell, V. (2022). Listening ears or reading eyes: A meta-analysis of reading and listening comprehension comparisons. *Review of Educational Research*, *92*(4), 543–582. https://doi.org/10.3102/00346543211060871

Daniel, D. B., & Woody, W. D. (2010). They hear, but do not listen: Retention for podcasted material in a classroom context. *Teaching of Psychology*, *37*(3), 199–203. https://doi.org/10.1080/00986283.2010.488542

De Meo, A. (2012). How credible is a non-native speaker? Prosody and surroundings. In M. Grazia Busà & A. Stella (Eds.), *Methodological perspectives on second language prosody* (pp. 3–9). Padova: CLEUP.

Dragojevic, M., & Giles, H. (2016). I don't like you because you're hard to understand: The role of processing fluency in the language attitudes process. *Human Communication Research*, *42*(3), 396-420. https://doi.org/10.1111/hcre.12079

Evans, J. R., & Michael, S. W. (2014). Detecting deception in non-native English speakers. *Applied Cognitive Psychology*, *28*(2), 226–237. https://doi.org/10.1002/acp.2990

Elkind, J., Cohen, K., & Murray, C. (1993). Using computer-based readers to improve reading comprehension of students with dyslexia. *Annals of dyslexia*, *43*(1), 238–259. https://doi.org/10.1007/BF02928184

Floccia, C., Goslin, J., Girard, F., & Konopczynski, G. (2006). Does a regional accent perturb speech processing? *Journal of Experimental Psychology: Human Perception and Performance*, *32*(5), 1276–1293. https://doi.org/10.1037/0096-1523.32.5.1276

Green, R. (1981). Remembering ideas from text: The effect of modality of presentation. *British Journal of Educational Psychology*, *51*(1), 83-89. https://doi.org/10.1111/j.2044-8279.1981.tb02458.x

Grunér, S., Östberg, P., & Hedenius, M. (2018). The compensatory effect of text-to-speech technology on reading comprehension and reading rate in Swedish schoolchildren with reading disability: The moderating effect of inattention and hyperactivity symptoms differs by grade groups. *Journal of Special Education Technology*, *33*(2), 98–110. https://doi.org/10.1177/0162643417742898

Harvey, J., & Hux, K. (2015). Text-to-speech accommodations for the reading challenges of adults with traumatic brain injury. *Brain Injury*, *29*(7–8), 888–897. https://doi.org/10.3109/02699052.2015.1022878

Hecker, L., Burns, L., Elkind, J., Elkind, K., & Katz, L. (2002). Benefits of assistive reading software for students with attention disorders. *Annals of Dyslexia, 52,* 243–272. https://doi.org/10.1007/s11881-002-0015-8

Kang, O., Thomson, R., & Moran, M. (2019). The effects of international accents and shared first language on listening comprehension tests. *TESOL Q*, *53*(1), 56–81. https://doi.org/10.1002/tesq.463

Kaur, P., & Raman, A. (2014). Exploring native speaker and non-native speaker accents: The English as a lingua franca perspective. *Procedia - Social and Behavioral Sciences*, *155*(6), 253–259. https://doi.org/10.1016/j.sbspro.2014.10.288

Kipp, M. (2021). Impact of the COVID-19 pandemic on the acceptance and use of an e-learning platform. *International journal of environmental research and public health*, *18*(21), 11372. https://doi.org/10.3390/ijerph182111372

Klein, H. A. (1989). The effect three testing modes - Reading while listening, listening and silent reading - Have on sixth grade boys and girls. *Reading Improvement, 26*(4), 298.

https://www.proquest.com/scholarly-journals/effect-three-testing-modes-reading-while/docview/1994304035/se-2

Knollman-Porter, K., Brown, J. A., Hux, K., Wallace, S. E., & Crittenden, A. (2022). Reading comprehension and processing time when people with aphasia use text-to-speech technology with personalized supports and features. *American Journal of Speech-Language Pathology*, *31*(1), 342–358. https://doi.org/10.1044/2021_AJSLP-21-00182

Kodama, M., Ishita, E., Watanabe, Y., & Tomiura, Y. (2021). Usage of e-books during the COVID-19 pandemic: A case study of Kyushu University Library, Japan. In K. Toeppe, H. Yan, & S.K.W. Chu (Eds.), *Diversity, Divergence, Dialogue* (pp. 475–483). Springer. https://doi.org/10.1007/978-3-030-71305-8_40

Kopp, K., & D'Mello, S. (2016). The impact of modality on mind wandering during comprehension. *Applied Cognitive Psychology*, *30*(1), 29–40. https://doi.org/10.1002/acp.3163

Leahy, W., & Sweller, J. (2011). Cognitive load theory, modality of presentation and the transient information effect. *Applied Cognitive Psychology*, *25*(6), 943–951. https://doi.org/10.1002/acp.1787

Lev-Ari, S., & Keysar, B. (2010). Why don't we believe non-native speakers? The influence of accent on credibility. *Journal of Experimental Social Psychology*, *46*(6), 1093–1096. https://doi.org/10.1016/j.jesp.2010.05.025

Lev-Ari, S., & Keysar, B. (2012). Less-detailed representation of non-native language: Why non-native speakers' stories seem more vague. *Discourse Processes*, *49*(7), 523–538. https://doi.org/10.1080/0163853X.2012.698493

Little, J. L. (2011). Optimizing multiple-choice tests as learning events. [Doctoral dissertation, University of California, Los Angeles]. https://eric.ed.gov/?id=ED539534

Munro, M. J., & Derwing, T. M. (1995). Processing time, accent, and comprehensibility in the perception of native and foreign-accented speech. *Language and Speech*, *38*(3), 289–306. https://doi.org/10.1177/002383099503800305

Ockey, G. J., Papageorgiou, S., & French, R. (2016). Effects of strength of accent on an L2 interactive lecture listening comprehension test. *International Journal of Listening*, *30*(1-2), 84–98. https://doi.org/10.1080/10904018.2015.1056877

Rogowsky, B. A., Calhoun, B. M., & Tallal, P. (2016). Does modality matter? The effects of reading, listening, and dual modality on comprehension. *SAGE Open*, *6*(3). https://doi.org/10.1177/2158244016669550

Schmitt, A. J., Hale, A. D., McCallum, E., & Mauck, B. (2011). Accommodating remedial readers in the general education setting: Is listening-while-reading sufficient to improve factual and inferential comprehension? *Psychology in the Schools*, *48*(1), 37–45. https://doi.org/10.1002/pits.20540

Smith, L. E., & Bisazza, J. A. (1982). The comprehensibility of three varieties of English for college students in seven countries. *Language Learning*, *32*(2), 259–269. https://doi.org/10.1111/j.1467-1770.1982.tb00971.x

Speechify [@speechifyapp]. (n.d.). *Reels* [Instagram profile]. Retrieved October 21, 2022, from https://www.instagram.com/speechifyapp/reels/

Use Immersive Reader in Microsoft Edge. (n.d.). Retrieved October 21, 2022, from https://support.microsoft.com/en-us/topic/use-immersive-reader-in-microsoft-edge-78a7a17d-52e1-47ee-b0ac-eff8539015e1

Wallace, S. E., Hux, K., Knollman-Porter, K., Brown, J. A., Parisi, E., & Cain, R. (2021). Reading behaviors and text-to-speech technology perceptions of people with aphasia. *Assistive Technology*. https://doi.org/10.1080/10400435.2021.1904306

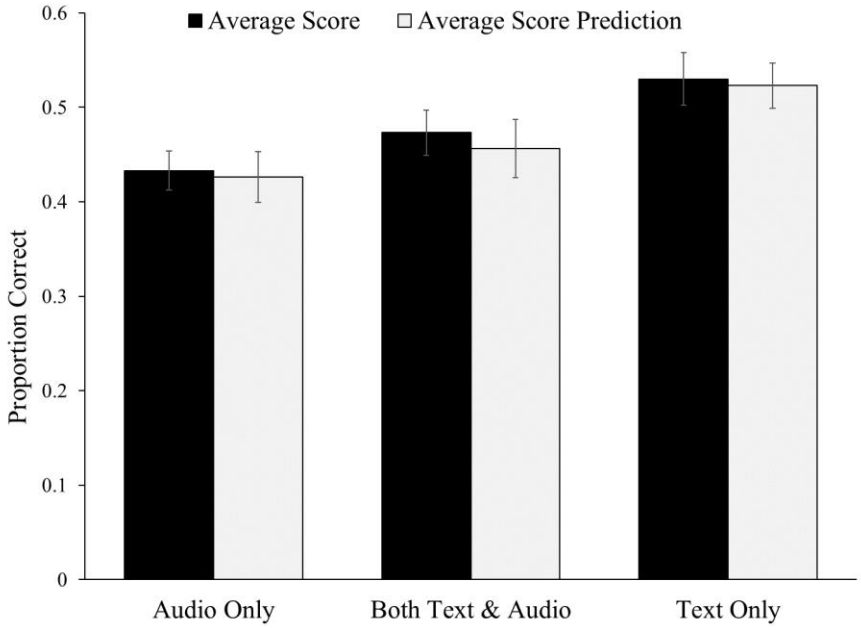
Wood, S. G., Moxley, J. H., Tighe, E. L., & Wagner, R. K. (2018). Does use of text-to-speech and related read-aloud tools improve reading comprehension for students with reading disabilities? A meta-analysis. *Journal of learning disabilities, 51*(1), 73–84. https://doi.org/10.1177/0022219416688170

Conflicts of Interest

The authors certify that they have no affiliations with or involvement in any organization or entity with any financial or non-financial interest in the subject matter or materials discussed in this manuscript.

Appendix A

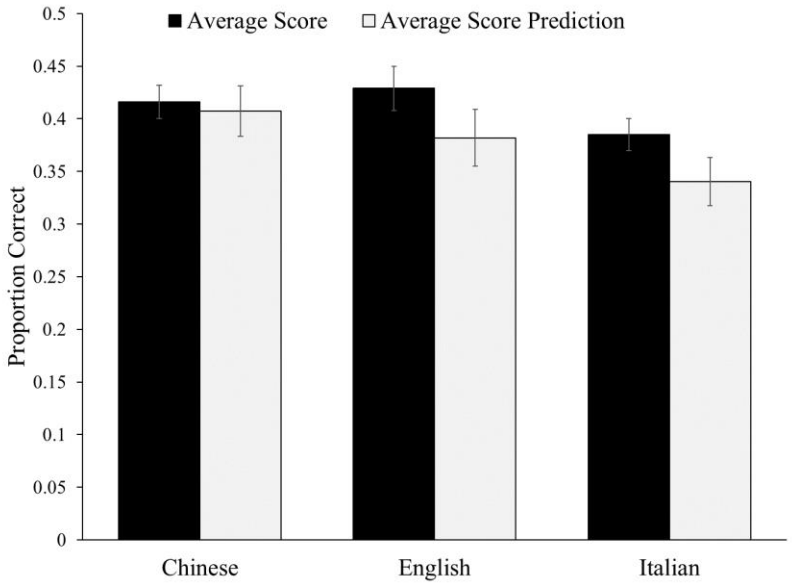
Average score and average predicted score as a function of method of exposure in Experiment 1



Note. Error bars reflect the standard error of the mean.

Appendix B

Average score and average predicted score as a function of accent in Experiment 2



Note. Error bars reflect the standard error of the mean.