

Research and Applications

A comparison of text versus audio for information comprehension with future uses for smart speakers

Gondy Leroy¹ and David Kauchak²

¹University of Arizona, Tucson, Arizona, USA and ²Pomona College, Claremont, California, USA

Corresponding Author: Gondy Leroy, MS, PhD, University of Arizona, McClelland Hall, Room 430, 1130 E. Helen Street, Tucson, AZ 85718, USA (gondyleroy@email.arizona.edu).

Received 8 February 2019; Revised 29 March 2019; Editorial Decision 3 April 2019; Accepted 6 April 2019

ABSTRACT

Objective: Audio is increasingly used to access information on the Internet through virtual assistants and smart speakers. Our objective is to evaluate the distribution of health information through audio.

Materials and Methods: We conducted 2 studies to compare comprehension after reading or listening to information using a new corpus containing short text snippets from Cochrane ($N=50$) and Wikipedia ($N=50$). In study 1, the snippets were first presented as audio or text followed by a multiple-choice question. Then, the same information was presented as text and the question was repeated in addition to questions about perceived difficulty, severity and the likelihood of encountering the disease. In study 2, the first multiple-choice question was replaced with a free recall question.

Results: Study 1 showed that information comprehension is very similar in both presentation modes (53% accuracy for text and 55% for audio). Study 2 showed that information retention is higher with text, but similar comprehension. Both studies show improvement in performance with repeated information presentation.

Discussion: Audio presentation of information is effective and the format novel. Performance was slightly lower with audio when asked to repeat information, but comparable to text for answering questions. Additional studies are needed with different types of information and presentation combinations.

Conclusion: The use of audio to provide health information is a promising field and will become increasingly important with the popularity of smart speakers and virtual assistants, particularly for consumers who do not use computers, for example minority groups, or those with limited sight or motor control.

Key words: audio, comprehension, health literacy user study, smart speakers

BACKGROUND

About 40% of the world population has internet access and there are more than 3.5 billion searches on Google daily.¹ About 77% of Americans own a smart phone² and many use this as a critical tool to access the internet with more than half of website visits served to mobile phones and other devices.³ In 2013, 59% of Americans went online for health information⁴ and this percentage has likely increased since then.

We are entering a new era where a different mode of accessing the internet is becoming increasingly used: audio using mobile devices, virtual assistants, and smart speakers. For example, in 2018,

30% of Americans used voice to find and purchase products. Smart speakers, in particular, have become an increasingly common household item. In the fourth quarter of 2016, there were 4.6 million smart speakers sold. This number increased to 19.7 million in the third quarter of 2018, reaching 57.8 million owners in the United States.⁵ Consumers are utilizing these devices for a wide range of activities including searching for health-related information. For example, as of November 2018, there were 15 sections within the Health and Fitness section of Google commands: many focus on standard activities such as tracking (exercise, medication, and sleep) or finding providers, but some focus on more involved tasks like

diagnosis. Alexa's Health and Fitness section contains over 1000 skills with 300+ receiving a customer review rating of 4 or more. Like Google, several skills focus on tracking (medication, menstruation, fertility, and calories), scheduling, or locating providers and facilities, but some also focus on providing health information, for example provided by WebMD. The popularity of these new interactive systems for health-related activities can be expected to increase. Additionally, many developments in the field will also increase use, including research into multi-model dialog,⁶ Amazon Alexa competitions for example focusing on Type 2 diabetes support,⁷ and hospitals providing Alexa skills (eg, Mayo Clinic, Boston Children's Hospital).

SIGNIFICANCE

In this article, we focus on audio for health-care information distribution. Critically, this new information access mode will bring new opportunities to reach different population groups, especially consumers in rural areas, those with impaired motor skills or vision, and nonliterate patients. Understanding how best to bring health-care information through audio will also benefit general information consumers since the use of audio is becoming increasingly popular. In addition, this new research stream will have an impact on health-care providers who wish to provide information through virtual assistants and smart speakers.

To our knowledge, we are the first to evaluate the potential of audio for accessing health information. We systematically compare text and audio presentation and lay foundations for this new type of consumer education. We conducted 2 studies that compare the use of text, as is common via browsers, and audio, as will become increasingly common through virtual assistants and smart speakers. Our goal is to evaluate the potential of this new medium. We measured comprehension in 2 ways, using multiple-choice content questions in the first study and free recall answers in the second. We found that audio presentation resulted in similar performance as text presentation when answering questions, but not when expressing the information independently. Repeat presentation of the information increases performance. These studies are a first step toward understanding the broader research question of the capacity to digest information presented through audio presentation.

OBJECTIVE

We have 2 objectives with this work. The first is to evaluate the feasibility of using audio to bring health information by comparing this mode to the current practice of using text. We accomplish this objective through 2 user studies where information is presented in audio and text format and comparing the accuracy of answering multiple-choice questions or the amount of content remembered via free recall questions. Our second objective is the creation of a corpus of text/audio information useful to other researchers.

MATERIALS AND METHODS

Corpus creation

We created a corpus containing text snippets from Cochrane (<https://www.cochranelibrary.com/>) and English Wikipedia (<https://en.wikipedia.org/>). The Cochrane library contains a range of health-related articles and summaries to "inform healthcare decision-making." We selected Cochrane and Wikipedia since they are 2

Table 1. Text snippet statistics for study corpus

	Minimum	Maximum	Mean
Cochrane snippets (<i>N</i> = 50)			
Word count	78	120	96.12
Sentence count	2	7	4.52
Wikipedia snippets (<i>N</i> = 50)			
Word count	85	109	94.36
Sentence count	3	7	4.50
Combined set (<i>N</i> = 100)			
Word count	78	120	95.24
Sentence count	2	7	4.51

common sources for obtaining accessible health-related information for patients. Articles were obtained from a previous study⁸ that contained texts from Cochrane and Wikipedia and were downloaded and processed for 60 different medical conditions. We selected 50 Cochrane and 50 Wikipedia snippets from different articles. The snippets contained 4.5 sentences on average and were approximately 95 words long (Table 1).

For each snippet, a multiple-choice content question with 3 answer choices was manually created. Each question requires information from at least 2 sentences in the snippet to be answered correctly. We created the corresponding audio version using the Microsoft Speech API (<https://azure.microsoft.com/en-us/services/cognitive-services/speech-services/>). This resulted in a corpus of 100 text snippets related to medical conditions, each with a corresponding audio version and a multiple-choice question about the content.

For example, the following is a snippet from a Cochrane article:

Children with pre-existing neurobehavioral disorders tend to be pharmacoresistant and have frequent seizures though these also remit with age. Formal neuropsychological assessment of children with Panayiotopoulos syndrome showed that these children have normal IQ and they are not on any significant risk of developing cognitive and behavioural aberrations, which when they occur they are usually mild and reversible. Prognosis of cognitive function is good even for patients with atypical evolutions. However, though Panayiotopoulos syndrome is benign in terms of its evolution, autonomic seizures are potentially life-threatening in the rare context of cardiorespiratory arrest.

And the question posed was:

Do children with Panayiotopoulos syndrome have a good prognosis of cognitive functioning?

With the following 3 options (correct answer indicated):

1. Yes/correct (*the correct answer*)
2. No/incorrect
3. Not enough information to answer the question

Study 1: Multiple-choice comprehension test

Participants and timeframe

We conducted the study using Amazon Mechanical Turk (AMT) in the summer (starting in June) of 2018. Each participant was paid \$0.50 for each text snippet that they worked on. The order of the snippets was randomized. As is the common procedure on AMT, participants could work through as many text snippets as they liked. We recruited 3 participants for each text snippet for each condition which allowed us to remove outliers from our dataset while retaining data on each snippet (ie, it is unlikely that all 3 participants for a

snippet would be an outlier). Participants were required to be in the United States and have a previous acceptance rate of 95%.

We collected demographic information from all participants before they participated in the study.

Independent variables

There were 2 independent variables in the study. The first was the source of the text: Cochrane or Wikipedia. The second was the mode of presentation of the information during the first interaction: text or audio.

Dependent variables

For each text, we measured comprehension with a multiple-choice question that asked about a particular assertion made in the text. Each question had 3 possible answers with a single correct answer, resulting in a random baseline of 33%. In addition, we also measured perceived difficulty using a 5-point Likert scale and perceived severity and perceived likelihood of encountering the condition using a 3-point Likert scale. Table 2 shows the wording used for these additional questions.

Procedures

The study was conducted in 2 phases representing 2 different conditions. The first phase was conducted using only text (Text-Text) and the second phase was conducted using first audio and then text (Audio-Text). There were several weeks between the first and second phase. This ensured that even if participants completed HITS in both phases, there would be no transfer of information based on memory of the information.

For the first phase (Text-Text), participants were first shown the text snippet. When they were done reading, they clicked a “Next” button which removed the text and presented the multiple-choice question. After answering the question, they were shown the original text together with the same multiple-choice question as well as the 3 subjective questions (see Table 2). Participants were informed they could change their answer to the multiple-choice content question. These 2 different presentations mimic common scenarios of accessing health information. The first corresponds to being presented with material to learn from (eg, a pamphlet after a doctor’s visit). The second corresponds to researching a particular topic, where there is a question to be answered and there is some resource to find the answer to that question. Both are important for health literacy and comparing the results between the 2, particularly how much improvement is seen after the second interaction with the information, can also provide useful insights.

For the second phase (Audio-Text), participants first listened to the audio version of the text. When they were done listening, they were automatically directed to a page where they answered the multiple-choice content question. After answering the question, the original information was shown as text, identical to the first phase, with the original multiple-choice content question and the 3 subjective questions.

Study 2: Free recall comprehension test

The second study was conducted in the late fall of 2018, 4 months after Study 1, and was almost identical to the first study. The only difference was the first multiple-choice question was replaced with a free recall question where participants were asked “Please write as much as you can remember of the information.” To score these text responses, we calculated 3 algorithmic scores that quantify how

Table 2. Meta-questions

Question type	Question example	Answers	Assigned score
Perceived difficulty (text)	How difficult does this text look?	Very easy	1
		Easy	2
		Neither	3
		Difficult	4
		Very difficult	5
Perceived difficulty (audio)	How difficult does this fragment sound?	Very easy	1
		Easy	2
		Neither	3
		Difficult	4
		Very difficult	5
Perceived severity	How severe does this condition seem to you?	Extremely severe	1
		Somewhat severe	2
		Not at all severe	3
Perceived likelihood	How likely are you or one of your immediate family members to develop this condition?	Extremely likely	1
		Somewhat likely	2
		Not at all likely	3

much of the information they remembered: *simple recall*, *exact recall*, and *semantic recall*. Simple recall is calculated as the number of number of unique, content-bearing terms (nouns, verbs, adjective, and adverbs). This measures how much they wrote, though it ignores the content. However, it does eliminate a few answers made by participants who did not remember anything, for example, “Don’t remember anything” does not contain content-bearing terms and so receives a score of zero. To capture actual content overlap, we measured exact recall, the number of terms in the response that were also found in the text snippet. Finally, sometimes participants recall the idea, but do not use the exact same phrasing as in the original text. To capture this, we measured semantic recall, which counted the number of terms found in the text snippet allowing for similar words based on word embeddings.

Before calculating any of the metrics, we applied an automatic spelling corrector to all free recall answers. We used the Java port (<https://github.com/Lundez/JavaSymSpell>) of JavaSymSpell (<https://github.com/wolfgarbe/SymSpell>) for spelling correction. This was done to avoid error variance due to some AMT workers using a spelling checker and others not. Using a spell checker is fairly common among workers, since they can get blocked from future tasks if they perform unsatisfactorily. We did not correct any words in the response that could be found in the original text. This avoided spurious correction for uncommon, but generally correct words, such as proper nouns, abbreviations, and acronyms.

After spelling correction, we used the Stanford CoreNLP toolkit⁹ to tokenize and part-of-speech tag the responses. Only the meaning-bearing words (nouns, verbs, adjectives, and adverbs) were counted. Simple recall and exact recall were then calculated directly using these words. To calculate semantic recall, we counted all words that were either lexically identical (ie, an exact match as used for exact recall) or were semantically similar. Two words were considered semantically similar if the cosine similarity of their word embedding vectors was greater than 0.45, a threshold determined empirically after examining the data. We used Google’s pretrained word embeddings, which contain a 300-dimensional vector for each word. The embeddings are freely available and provide large vocabulary coverage.¹⁰

To provide an additional metric for the *overall content similarity*, we also calculated the average cosine similarity of the word embeddings of all meaning-bearing words in a response compared with those in the original text snippet.

RESULTS

Study 1: Multiple-choice comprehension test

Overview

The first study focuses on information comprehension using a multiple-choice content question. The multiple-choice question is repeated twice, once after reading or listening to the text and a second time immediately following, but with the text present. Since this question is repeated, we conducted repeated measures ANOVA with the score on the multiple-choice question as the repeated measure and the text origin (Wikipedia or Cochrane) and presentation mode (Text-Text and Audio-Text) as the independent variables.

Demographic information

There were 46 participants in the first study. The data of 9 participants were removed because their accuracy on answering the multiple-choice questions when presented for a second time (ie, with text present) was at or below the random level (33%). This resulted in a total of 51 data items being removed (most participants removed had completed only 1 or 2 tasks and 1 participant had completed 42 tasks). Table 3 provides the demographic information of the remaining participants. The participants were mostly male (59%) and White (93%). One person identified as American Indian/Native Alaskan (2%), 3 as Asian (7%), and 1 as Black (2%). Participants could choose multiple races. Most participants were younger than 50 years old, with 46% less than 30 years old, 24% between 31 and 40, and 24% between 41 and 50. Three participants (7%) were between 51 and 60 years old.

A majority of the participants spoke only English at home (89%) with a few speaking mostly English at home (4%) and an even smaller group speaking English half of the time (2%) or rarely at home (4%). Note that not speaking English at home does not mean the participants were not fluent in English. All participants except 1 (2%) had at least a high school degree. A large group of participants (39%) held a bachelor's degree and half held either a high school diploma (26%) or an associate degree (24%) as their highest degree. There were 4 participants (9%) who held a doctorate degree.

Information comprehension

Figure 1 (left) shows the results for the multiple-choice question presented the first time. The overall accuracy was fairly low at 54%. There was a small difference depending on source with 50% overall accuracy for Cochrane (51% after reading and 49% after listening) and 57% for Wikipedia (55% after reading and 60% after listening). The differences in accuracy were even smaller for different presentation modes: the accuracy was overall 53% after listening to the text (regardless of source) versus 55% after reading the text. These differences were not statistically significant.

Figure 1 (right) also shows the detailed numbers for the multiple-choice question when presented a second time along with the text (regardless of the mode of presentation for the first interaction). Overall, accuracy was higher the second time with an average of 65%. The accuracy was slightly lower for Cochrane (64%) than for Wikipedia (66%), though the improvement in accuracy with the second presentation was much higher for Cochrane, that is 14% abso-

Table 3. Participant demographic information

	N	%
Total	46	
Gender		
Male	27	59
Female	19	27
Ethnicity		
Hispanic of Latino	8	17
Not Hispanic or Latino	38	83
Race (multiple choices allowed)		
American Indian/Alaska Native	1	2
Asian	3	7
Black	1	2
Native Hawaiian/Pacific Islander	0	0
White	43	93
Age		
Younger than 30 years old	21	46
31–40 years old	11	24
41–50 years old	11	24
51–60 years old	3	7
61–70 years old	0	0
71 or better	0	0
Education level		
Less than high school degree	1	2
High school diploma	12	26
Associate degree	11	24
Bachelor's degree	18	39
Master's degree	0	0
Doctoral degree	4	9
Language spoken at home		
Never English	0	0
Rarely English	2	4
Half English	1	2
Mostly English	2	4
Only English	41	89

lute versus 9% for Wikipedia. The first presentation mode of the information had a small effect on the accuracy of the question the second time around. If the participants listened to the audio originally then the accuracy was on average 62% (64% for Cochrane and 60% for Wikipedia) compared with 68% on average if they read the text (64% for Cochrane and 73% for Wikipedia).

Our statistical analysis showed that only the increase in accuracy from answering the multiple-choice question the first time to the second time was significant ($F(1, 550) = 32.672, P < .001$). There were no significant interactions with the independent variables.

Perceived difficulty, severity, and likelihood

We conducted three 2×2 ANOVAs to measure the impact of our text origin and presentation mode for perceived difficulty, severity, and likelihood. Two data points were missing for perceived difficulty. No data were missing for the other questions.

We found a significant effect of the presentation mode on the perceived difficulty of the information ($F(1, 548) = 13.605, P < .001$). The information was perceived as more difficult when it was first presented as audio (3.21) compared with text (2.88).

Information from Wikipedia was perceived as more severe (2.38) than Cochrane (2.17) ($F(1, 550) = 15.972, P < .001$). There was a second significant effect of the presentation mode: listening to the information first before reading it resulted in higher perceived

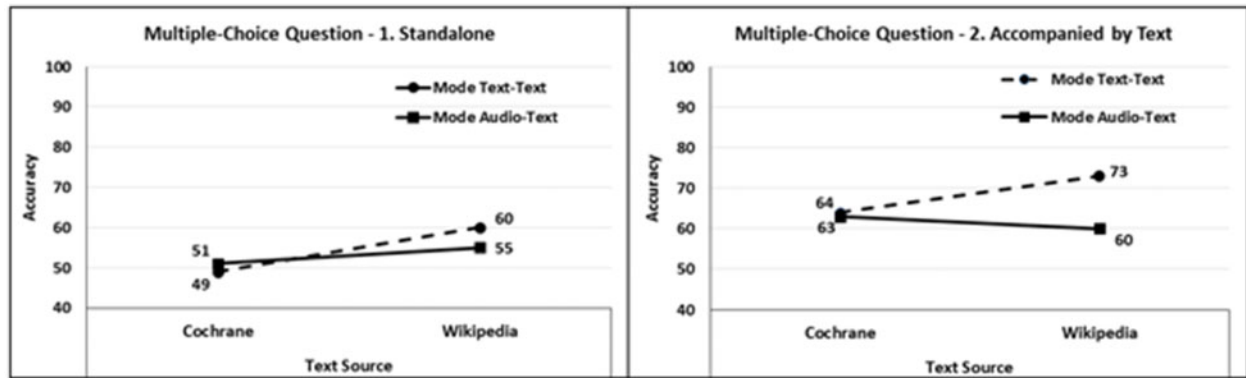


Figure 1. Study 1 with multiple-choice question.

severity (2.33) compared with reading the information twice (2.22) ($F(1, 550) = 4.644, P = .032$).

There were no significant differences in the perceived likelihood of encountering the disease and the overall perceived likelihood was 1.35.

Study 2: Free recall comprehension test

Overview

Since the first study showed very encouraging results for audio presentation of information, we conducted a second study using a different dependent variable to provide a more fine-grained and sensitive measure of information retention after the first interaction with the information. Since this study does not repeat the multiple-choice question between the 2 phases, we first conducted a 2×2 ANOVA to measure the effect of the 2 independent variables (text source and presentation mode) on free recall and a second 2×2 ANOVA to measure the effects on the multiple-choice question accuracy.

Demographic information

There were 39 AMT workers who participated in the second study. There were 7 participants in this study who had also participated in the previous study. Because there were 4 months between the studies, we retained these data. Similar to the first study, we removed workers whose score on the multiple-choice question was at or below the random level. The data of 6 workers were removed, which accounted for 82 data items being removed. Table 4 provides the demographic information of the remaining participants. A majority of the participants were female (56%) and White (97%). In this study, there were 7 American Indian/Alaska Natives (18%), 7 Asian (18%), 4 Black (10%), and 4 Native Hawaiian/Pacific Islander (10%). Most participants were younger than 50 years old, with 31% younger than 30, 21% between 31 and 40, and 38% between 41 and 50. Three participants (8%) were between 51 and 60 and 1 participant (2%) was between 61 and 70.

The majority (97%) spoke only English at home with 1 participant speaking English most of the time (3%). All participants held at least a high school diploma with the largest group holding a high school diploma (49%) as their highest degree. A smaller group held an associate degree (15%) or a bachelor's degree (31%) and there was 1 participant (3%) with a master's degree and 1 (3%) with a doctorate.

Information comprehension and retention

Figure 2 shows the results of the second study. The left side shows the simple recall, that is the number of unique words (nouns, verbs,

Table 4. Participant demographic information

	N	%
Total	39	
Gender		
Male	17	44
Female	22	56
Ethnicity		
Hispanic of Latino	3	8
Not Hispanic or Latino	36	92
Race (multiple choices allowed)		
American Indian/Alaska Native	7	18
Asian	7	18
Black	4	10
Native Hawaiian/Pacific Islander	4	10
White	38	97
Age		
Younger than 30 years old	12	31
31–40 years old	8	21
41–50 years old	15	38
51–60 years old	3	8
61–70 years old	1	3
71 or better	0	0
Education level		
Less than high school degree	0	0
High school diploma	19	49
Associate degree	6	15
Bachelor's degree	12	31
Master's degree	1	3
Doctoral degree	1	3
Language spoken at home		
Never English	0	0
Rarely English	0	0
Half English	0	0
Mostly English	1	3
Only English	38	97

adjectives, and adverbs). Overall, there were 7.94 words in the responses. The number of words was slightly higher (but not statistically significant) for Wikipedia (8.13 words) compared with Cochrane (7.76 words). The number of words was significantly ($F(1, 518) = 6.769, P = .010$) higher if participants read the text (8.84 words overall, 8.94 for Cochrane, and 8.73 for Wikipedia) compared with listening to the information (7.18 words overall, 6.74 for Cochrane, and 7.61 for Wikipedia). The interaction between the 2 variables was not significant.

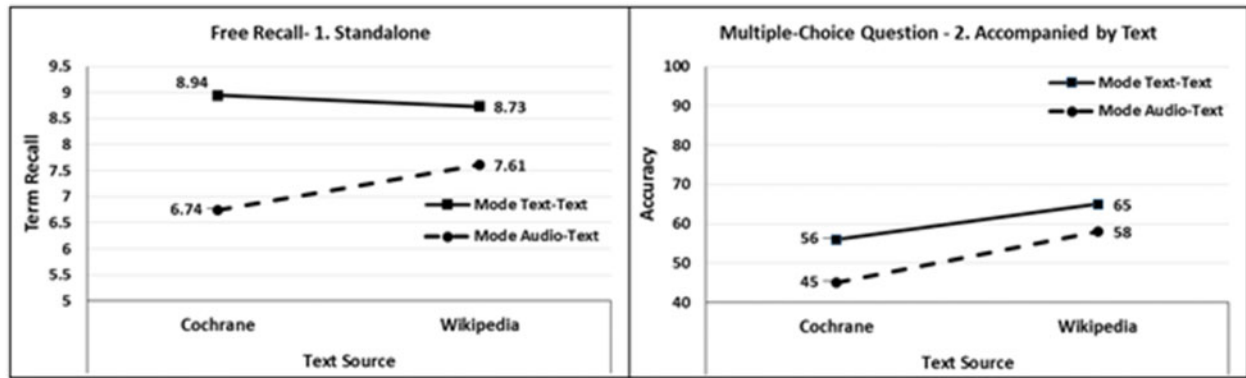


Figure 2. Study 2 with free recall (exact recall) and multiple-choice questions accuracy.

We found no significant differences between conditions for the number of words in the free response that were found in the original text (exact recall), however, the number of similar words that were found (semantic recall) showed the same trends as simple recall ($P = 0.60$) with a slightly higher number of similar terms after reading the text (6.53 words) compared with listening to the information (5.63 words).

The average cosine similarity of the words in the response compared with the original information did not differ for the audio or text presentation, however, the average similarity of terms was significantly higher ($F(1, 518) = 25.340, P < .001$) with Wikipedia ($r = 0.135$) compared with Cochrane ($r = 0.123$).

Since these analyses are based on raw counts, we calculated the *percentage* of exactly matching words and the *percentage* of similar words in the responses. In both cases, the differences between conditions were small and not statistically significant.

This study also presented the multiple-choice questions after the free recall answers (Figure 2, right side). For accuracy on the multiple-choice question, we found significant effects for both text source and presentation mode. Overall accuracy of answering the multiple-choice question (with the information present as text) was 56% (compared with 54% in the first study). Accuracy was higher for Wikipedia (61%) compared with Cochrane (50%) ($F(1, 518) = 6.437, P = .011$) and higher ($F(1, 518) = 4.640, P = .032$) when the information was first shown as text (61% overall, 56% for Cochrane, and 65% for Wikipedia) compared with audio (51% overall, 45% for Cochrane, and 58% for Wikipedia).

Perceived difficulty, severity, and likelihood

The same 3 questions about perceived difficulty, severity, and likelihood of encountering the disease were asked in the second study and we conducted a 2×2 ANOVA for each. Four data points were missing for perceived difficulty. No data were missing for the other questions.

We found significant effects for text source ($F(1, 514) = 8.389, P = .004$) with Cochrane perceived as more difficult (3.69) than Wikipedia (3.43). In addition, the condition also led to different perceived difficulty ($F(1, 514) = 9.772, P = .002$). The information was perceived as more difficult when it was first presented as audio (3.69) than as text (3.41).

We found again a significant effect of the text source on the perceived severity ($F(1, 518) = 6.748, P = .010$): Wikipedia was perceived as more severe (2.30) than Cochrane (2.46).

We also found a significant effect on perceived likelihood of encountering the conditions ($F(1, 518) = 10.328, P = .001$) with the likelihood considered higher when the information was read twice (1.35) compared with having listened then read (1.21).

DISCUSSION

Both studies showed very similar results regardless of whether information was presented as text or audio, especially when comprehension was measured with a multiple-choice question. There are 2 interesting differences. The first is related to our experimental setup. The first study presented the multiple-choice question immediately after presentation of the information and a second time with the text. The second study first collected free recall of information before presenting the multiple-choice question. When participants performed the free recall exercise, we found effects of the conditions with the following multiple-choice question. Free recall may interfere with remembering information because of the delay in answering the multiple-choice question. This might be relevant when a similar procedure is used during clinical encounters where providers might ask people to reiterate an explanation or rationale. This process may interfere with later information processing. Further investigation is required.

The second interesting difference relates to the participants. The second study was completed by a group of participants whose education level was lower overall, with a larger group of participants with only a high school diploma. In this study, the effect of the different conditions was more pronounced. Education level may play a role in comprehension when presenting information as audio or text. However, future experiments will be needed, since in our studies, both the type of recall (ie, free recall) as well as the education level were different and so these are confounded and cannot provide causal conclusions.

For both our studies, we removed workers who performed very poorly on the multiple-choice question and whose accuracy was below 33% (random). Removing participants who are outliers according to an important metric is a common approach. However, when creating multiple-choice questions, it is also possible that some questions are more difficult than others. Therefore, we reviewed our text snippets to find those where participants struggled the most. There were 26 snippets that resulted in below random (33%) accuracy in both studies. We duplicated our analysis by removing these text snippets. However, the results displayed the same trends with lower significance because of the reduced data size. Since there were no

significant differences, we have refrained from a detailed description and presented the analysis using as much data as possible in an objective manner.

Since this work addresses comprehension of information and a large body of prior work has relied on Flesch-Kincaid grade levels to estimate text difficulty and its relation to comprehension, we calculated correlations between the grade level of each text and the multiple-choice question accuracy and free recall metrics. There were no significant correlations between the grade level and the accuracy on the multiple-choice questions using all data grouped together, grouped by origin (Wikipedia vs Cochrane), or by presentation mode (audio vs text).

CONCLUSION

This study compared the effects of presenting medical information in text versus audio format. We conducted 2 studies with increasingly detailed metrics. The first study used solely multiple-choice questions while the second also included free recall of information. We found that audio presentation of information is a promising field with differences in comprehension between text and audio seemingly small.

To our knowledge, this is one of the first comprehensive studies of the effect of information comprehension when the information is presented via audio. We expect this medium to become increasingly popular with the increased use of smart speakers and virtual assistants. Additional studies are needed to evaluate the effects of these new presentation modes on different age groups, different types of information, and to optimize presentation of comprehension and retention of information.

CONTRIBUTORS

Both authors made substantial contributions to the conception and design of the study, the data acquisition, analysis and interpretation of the studies, and also write-up (draft and final version) of the manuscript.

FUNDING

The work was supported by the National Library of Medicine of the National Institutes of Health grant number R01LM011975.

Conflict of interest statement. None declared.

REFERENCES

1. Internet Live Stats. Trends and More (Statistics); 2018 [cited 2018]. <http://www.internetlivestats.com/>; last accessed December 01, 2018.
2. Pew Research Center. Mobile Fact Sheet; 2018. <http://www.pewinternet.org/fact-sheet/mobile/>; last accessed September 15, 2017.
3. Statista. Percentage of All Global Web Pages Served to Mobile Phones from 2009 to 2018; 2018 [cited 2018]. <https://www.statista.com/statistics/241462/global-mobile-phone-website-traffic-share/>; last accessed December 01, 2018.
4. Pew Research Center. Majority of Adults Look Online for Health Information; 2013. <https://www.pewresearch.org/fact-tank/2013/02/01/majority-of-adults-look-online-for-health-information/>; last accessed March 28, 2019.
5. voicebot.ai. Amazon Echo & Alexa Stats; 2018. <https://voicebot.ai/amazon-echo-alexa-stats/>; last accessed February 10, 2019.
6. K epuska V, Bohouta G. Next-generation of virtual personal assistants (Microsoft Cortana, Apple Siri, Amazon Alexa and Google Home). 2018 IEEE 8th Annual Computing and Communication Workshop and Conference (CCWC); 8–10 January; Las Vegas, NV, USA; 2018.
7. Alexa Diabetes Challenge. 2018. www.alexadiabeteschallenge.com/; last accessed September 10, 2017.
8. Revere D, Mukherjee P, Kauchak D, Leroy G. *Creating a Corpus Resource for Text Simplification Research and Development*. AMIA Fall Symposium; November, Washington DC; 2007.
9. Manning CD, Surdeanu M, Bauer J, Finkel J, Bethard SJ, McClosky D. The Stanford CoreNLP Natural Language Processing Toolkit 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations; 2014, 55–60.
10. Mikolov T, Chen K, Corrado G, Dean J. Efficient Estimation of Word Representations in Vector Space. CoRR. 2013; arXiv:1301.3781v3.