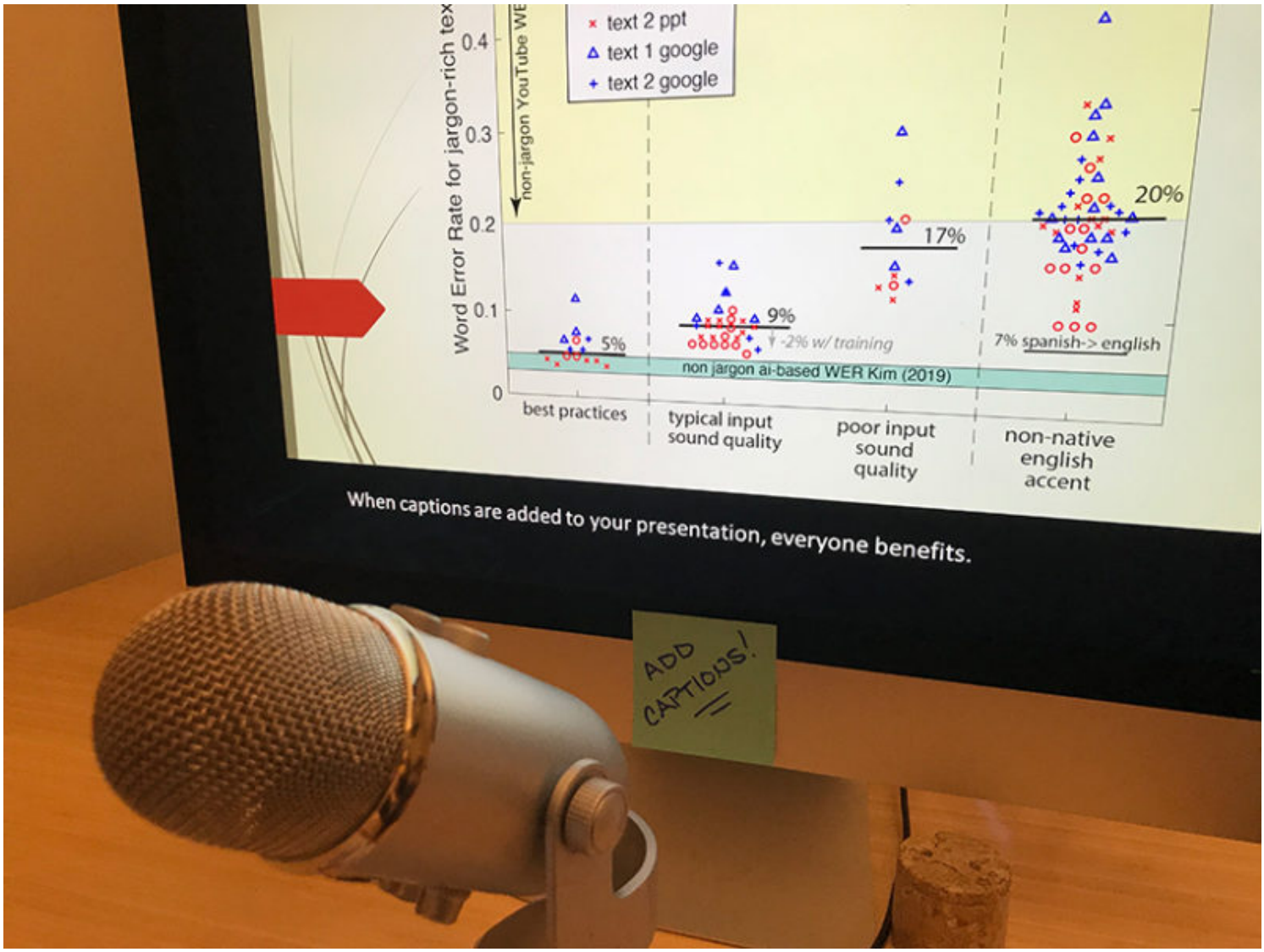
Caption This! Best Practices for Live Captioning Presentations

We demonstrate the effectiveness of straightforward strategies using widely available auto-captioning tools to greatly improve accessibility of jargon-rich content.



Credit: Michele L. Cooke

By Michele Cooke, Celia R. Child, Elizabeth C. Sibert, Christoph von Hagke, and S. G. Zihms 9 October 2020

Presentations that have captions are better understood, whether they are in-person or remote.

Captions make verbal material more accessible to a wider variety of people. A study of BBC television viewers (https://www.ofcom.org.uk/consultations-and-statements/category-1/accessservs) reported that 80% of caption users are not deaf or hard of hearing. During English-spoken scientific presentations, not-yet-fluent English speakers, people who are deaf or hard of hearing, and people who have auditory processing disorder (https://www.understood.org/en/learning-thinking-differences/child-learning-disabilities/auditory-processing-disorder/understanding-auditory-processing-disorder) develop listening fatigue that can inhibit their understanding and limit their participation in discussions.

Increasing the accessibility of presentations and improving inclusivity of discussions provide a path toward increasing diversity within the sciences. Studies have shown that subtitles or captions improve both English language skills [e.g., *Vanderplank* (https://onlinelibrary.wiley.com/doi/abs/10.1002/tesq.407), 2016; *Wang and Liu* (http://cscanada.net/index.php/sll/article/view/j.sll.1923156320110303.1200), 2011] and accessibility of science for deaf and hard of hearing participants [e.g., *Kawas et al.*, 2016; *Vanderplank* (https://onlinelibrary.wiley.com/doi/abs/10.1002/tesq.407), 2016]. Furthermore, for remote presentations, audio may not be accessible in all shared workspaces.

A myriad of tools and platforms can provide captioning for live presentations. Why then don't we regularly caption geoscience presentations? Our resistance may be due to such factors as not knowing or believing that captioning is needed, not knowing how to use these tools, and/or believing that the resulting captioning will be inadequate. However, presenters should make their talks accessible without requiring participants to request captions each time.

This article outlines different strategies for providing effective captions using widely available captioning tools and presents results of our performance assessment of artificial intelligence (AI)–based auto-captioning of jargon-rich geological passages. Because most scientific presentations are delivered using either Microsoft PowerPoint or Google Slides presentation software, we focus our performance assessment on the built-in auto-captioning provided by these platforms.

Our evidence supports five best practices and key takeaways:

Implement AI-based auto-captioning directly within the presentation software.
Use an external microphone.
Speak deliberately and clearly.
Practice with the presentation software beforehand and add to text of the slides words that are typically missed with your accent.
Always accommodate requests for human captionists.

In-Person Presentations

For in-person presentations, either trained human captionists or AI-based auto-caption or transcription software can provide live captioning (Figure 1). Captionists use stenography tools to provide accurate transcriptions. For everyone to access the captions, the captionist's transcriptions can be projected onto a separate screen near the presentation slides.

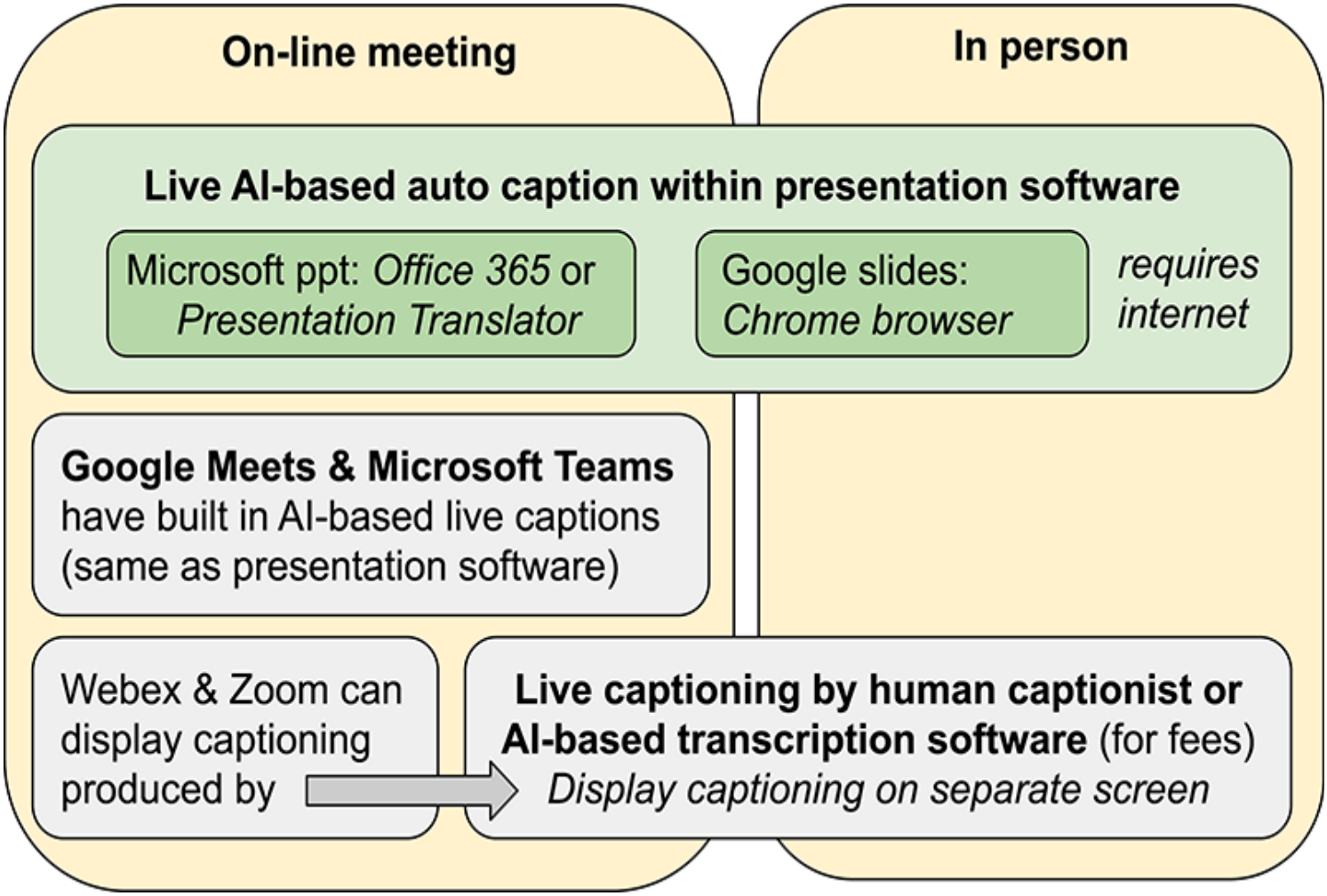


Fig. 1. Microsoft PowerPoint and Google Slides artificial intelligence (AI)–based auto-captioning can work for both online and in-person meetings. Google Meet and Microsoft Teams provide captioning for online meetings. Webex and Zoom can display captions if the host has either hired a human captionist or paid for third-party transcription services. Credit: Michele L. Cooke

Both Microsoft PowerPoint (with Office 365 (https://support.microsoft.com/en-us/office/present-with-real-time-automatic-captions-or-subtitles-in-powerpoint-68d20e49-aec3-456a-939d-34a79e8ddd5f) or Presentation Translator (https://www.microsoft.com/en-us/translator/help/presentation-translator/)) and Google Slides (https://support.google.com/docs/answer/9109474?hl=en) (with the Chrome browser) provide built-in AI-based auto-captioning directly onto the presented slides that can be used by anyone (instructions here (https://doi.org/10.6084/m9.figshare.12996719.v1)). Third-party software, such as Ava (https://www.ava.me/), Rev (https://www.rev.com/), and Otter.ai (https://otter.ai/login), can also provide AI-based transcriptions. In addition to their wide availability, an advantage of Slides and PowerPoint auto-captions over third-party

transcription software is that the captioning is projected onto the same screen as the presentation. Having captions within the presented slides frees the audience from repeatedly having to shift its focus from the presentation material to a separate caption screen.

Online Presentations

For remote online presentations, any of the in-person strategies can also work. Human captionists anywhere in the world can join remote meetings. In addition, the online meeting platforms Google Meet and Microsoft Teams offer built-in live auto-captioning that uses the same AI-based transcription tools as their presentation software. With the Webex (https://www.webex.com/) and Zoom (https://zoom.us/about) platforms, captioning can be available to everyone if the host appoints the captionist within the meeting software. Zoom and Webex also allow for third-party auto-captions if the host has paid for those services.

The benefits of providing captioning directly within Microsoft PowerPoint and Google Slides is that the built-in AI-based captioning means you don't need to add another tool and pay for that service. Many online presentations are also recorded. While a variety of tools can add carefully edited captions to recorded lectures that didn't have live captioning, offering a transcript after a live presentation is not a suitable solution to improving participation.

How Accurate Are Captions for Scientific Talks?

If you have watched auto-captions provided by YouTube, then you have seen low-quality captions, sometimes called craptions.

If you have watched auto-captions provided by YouTube, then you have seen low-quality captions, sometimes called craptions (https://www.theatlantic.com/health/archive/2019/08/youtube-captions/595831/). The word error rate (WER) of YouTube's non-AI-based auto-captioning is 20%–50%, (https://www.3playmedia.com/2019/02/04/the-difference-between-youtubes-automatic-captions-diy-captions-and-3play-media-captions/) which renders it practically useless unless creators manually edit the autogenerated transcript. Typical word errors include split or blended words, incorrect spelling, and incorrect guesses. For both AI-based and human captioning, WER is affected by microphone quality, Internet quality, accent and style of the speaker, and advance access of the captionist to the presentation content.

Jargon, such as is often encountered in geoscience presentations, can be particularly challenging for accurate captioning. To challenge the performance of live auto-captioning software to capture scientific presentations, we chose two passages rich with geological jargon taken from *Van der Pluijm and Marshak* [2004] and *Weil* (https://agupubs.onlinelibrary.wiley.com/doi/full/10.1029/2005TC001861) [2006]. Both passages have complex words that are rarely used outside the discipline as well as common English words that are used differently by experts. For example, "thrust" is typically a verb, but geologists use it

as an adjective for a type of fault. The second passage also tests the recognition of acronyms. Prior to testing the auto-caption performance, we identified words that we expected to be challenging (Table 1).

Table 1. Words Missed with Captioning of American-Accented English and Standard Sound Quality

Words That We Expected AI-Based Captions to Miss	Words That Captions Missed Much of the Time	Words That Captions Missed Consistently
nappes, substratum, lithosphere, vergent, accretionary, nonsubductable, radiogenic, Barrovian, metamorphism, paleomagnetic, Variscan, Western European Variscan belt (WEVB), Carboniferous, Permian, orocline, kinematic	nappes, lithosphere, nonsubductable, Barrovian, Variscan, WEVB, orocline, granitic, phases; blended words: thrusts and, hinge zone, WEVB's core	nappes, nonsubductable, Barrovian, Variscan,* WEVB,* orocline

*Captioned correctly under best practices and after some training.

We measured the WER of Microsoft PowerPoint and Google Slides AI-based live auto-captioning for both passages under a variety of conditions. WER indicates occurrence of error, so if the captioning never caught the acronym WEVB (https://www.abbreviations.com/term/289466) (Western European Variscan belt), for example, this would count as four mistakes in the second passage.

With a recording of an American-accented English female voice, we repeatedly tested the caption performance of both PowerPoint and Slides. For some tests, we decreased the sound quality by adding background noise and lowering input volume. In another set of tests, we assessed the WER of recordings of nonnative English-speaking geologists reading the two passages. The accents (Chinese, Mexican, Spanish, and German) are not meant to provide a complete accounting of the potential WER of nonnative English speakers but instead to show the relative performance of the AI-based auto-captioning for native and nonnative speakers.

Surprisingly, many technical words that we expected to be missed were accurately captioned (Table 1). Some words and phrases were missed in some, but not all, of the repeated tests. For example, while the phrase "hinge zone" comprises common English words, the captioning sometimes made this unfamiliar phrase into a single word. Repeating each recording at least three times allowed us to assess the variability of performance due to Internet quality and other fluctuations. Only six words from the two

passages were never correctly captioned with the AI-based auto-captioning using the American English voice recorded under typical sound conditions (Table 1). Words that were missed much of the time for American-accented English were missed more often with non-American-accented English recordings.

When flummoxed, Google Slides captioning, at the time of our testing, would sometimes omit parts of the passage, whereas Microsoft PowerPoint misguessed a few words. This difference accounts for the larger range of WER for Slides captions in Figure 2. Otherwise, the performance of Microsoft PowerPoint and Google Slides AI-based captioning was similar under most of the scenarios tested. While analyzing recordings of different accents, we noticed that some words, such as Variscan (https://en.wikipedia.org/wiki/Variscan_orogeny), were learned by the AI-based captioning and later recognized by the English recording, yielding a 2% improvement in WER.

Our experience suggests that jargon may be learned if the AI-based software hears the word in different ways. These codes are updated all the time and might in the future also yield improved caption performance with consistent recognition of jargon placed within the slides or notes.

We tested the effect of audio quality by adding background noise and reducing the sound level of the American-accented English. The tests showed that poor sound quality has a dramatic impact on the quality of the captions (Figure 2). The WER with poor sound quality reached the error levels of auto-captions, exceeding 20% in some cases.

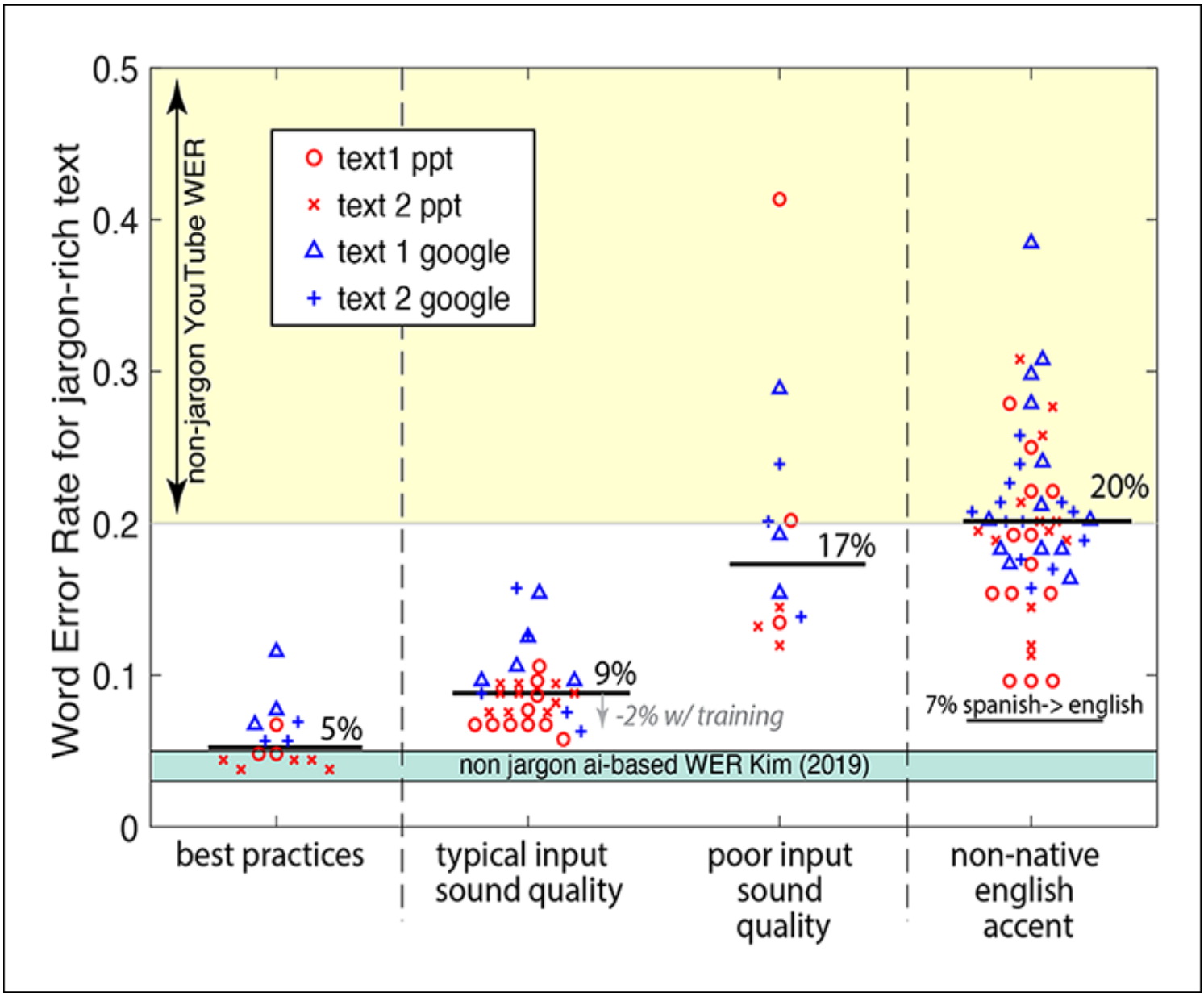


Fig. 2. Word error rate (WER) varies for auto-captioning for different settings. The best performance is with a lapel microphone. Under these conditions, the WER approaches that of nonjargon text. Poor sound quality and nonnative English accents decrease the quality of the AI-based auto-captions for both Microsoft PowerPoint and Google Slides. Credit: Michele L. Cooke

The WER from recordings of several different people with nonnative English accents showed that accents strongly decrease the quality of captioning. Microsoft PowerPoint allows the user to choose among several variants on English accents, such as British and Australian, that were not tested in this investigation. Presumably, if one spoke with an Australian accent with this accent setting chosen, the performance would be similar to that presented here of American-accented English (Figure 2). PowerPoint also provides captioning of an extensive set of languages. In a limited test, we found that spoken Spanish to Spanish captions performed as well as spoken American English to English. PowerPoint also provides translation from one spoken language to another captioned language. We

found that the WER for captioning of spoken Spanish to captioned English (~7%) was less than most of the nonnative English recordings tested here, and the resulting captions missed much of the same jargon presented in Table 1. Some nonnative English speakers may find a reasonable WER if they use the PowerPoint translation feature and speak in their native language, allowing the software to translate the captions into another language.

Best Practices

Implementing AI-based auto-captioning in live presentations using Microsoft PowerPoint or Google Slides is straightforward and can yield acceptable quality captioning.
Implementing AI-based auto-captioning in live presentations using Microsoft PowerPoint or Google Slides is straightforward and can yield acceptable quality captioning. Our findings highlighted the following best practices.

Implement AI-based auto-captioning directly within the presentation software. Your audience or meeting participants won't have to run a separate transcription service and switch attention between the presentation and the transcription.
Speak deliberately and clearly. The tests in Figure 2 for American-accented English were from recordings spoken at a conversational pace (average WER of 7.5%). When the same speaker spoke more intentionally, the WER dropped to less than 6%. The geological jargon was still missed, but the captioning caught nearly all of the nonjargon words when the speaker pace was slowed.
Practice with the presentation software beforehand and see which words are typically missed with your accent. Adding that missed jargon within the text of the slide ensures that the audience can see what the word should be and understand your message. As you repeat jargon in different ways, the AI-based captioning may learn this new word.
In our tests, having the presenter use a lapel microphone produced the greatest improvement to caption quality regardless of other variables.
Use an external microphone to improve audio quality. In our tests, having the presenter use a lapel microphone produced the greatest improvement to caption quality regardless of other variables.

Following these best practices of speaking intentionally with a good quality microphone decreased the WER for the two passages to approximately 5% over several recordings, a reasonable rate for jargon-rich material (Figure 2). Some jargon that was often missed in early tests using the built-in microphone and conversational pace was captured accurately using these best practices, which also eliminated other errors from blended and missed words.

Finally, a deaf or hard of hearing person may specifically request a human captionist for live presentations, because captionists provide more accurate captions. Accommodation requests should

always be honored. Captionists are expected to have a word error rate of 1% for nonjargon (https://www.theatlantic.com/health/archive/2019/08/youtube-captions/595831/) speech. While this level of accuracy is required for some participants, many of us can benefit greatly from captioning with an error rate of up to 5% such as provided with AI-based live auto-captioning.

Always include captioning in your live meetings, workshops, webinars, and presentations.

Acknowledgments

The authors thank Alina Valop, Xiaotao Yang, David Fernández-Blanco, and Kevin A. Frings for recording their readings of the two passages and David Fernández-Blanco for reviewing this article.

References

Kawas, S., et al. (2016), Improving real-time captioning experiences for deaf and hard of hearing students, in *Assets'16: Proceedings of the 18th International ACM SIGACCESS Conference on Computers and Accessibility*, pp. 15–23, Assoc. for Comput. Mach., New York.

Kim, J. Y., et al. (2019), A comparison of online automatic speech recognition systems and the nonverbal responses to unintelligible speech, preprint, *arXiv:1904.12403*.

Vanderplank, R. (2016), *Captioned Media in Foreign Language Learning and Teaching: Subtitles for the Deaf and Hard-of-Hearing As Tools for Language Learning*, xiii + 269 pp., Springer, New York, https://doi.org/10.1002/tesq.407 (https://doi.org/10.1002/tesq.407).

Van der Pluijm, B. A., and S. Marshak (2004), *Earth Structure: An Introduction to Structural Geology and Tectonics*, 2nd ed., Norton, New York.

Wang, K., and H. Liu (2011), Language acquisition with the help of captions, *Stud. Lit. Lang., 3*(3), 41–45, https://doi.org/10.3968/j.sll.1923156320110303.1200 (https://doi.org/10.3968/j.sll.1923156320110303.1200).

Weil, A. B. (2006), Kinematics of orocline tightening in the core of an arc: Paleomagnetic analysis of the Ponga Unit, Cantabrian Arc, northern Spain, *Tectonics*, *25*(3), TC3012, https://doi.org/10.1029/2005TC001861 (https://doi.org/10.1029/2005TC001861).

Supplementary Materials

We used two passages to test the AI-based live auto-captioning.

Passage 1, from *Van der Pluijm and Marshak* [2004]:

"Since the Alpine nappes exclusively consist of thin slices of upper crustal basement and/or its cover, detached from their lower crustal and mantle substratum, all European lower crust, including parts of the upper crust, must have been subducted together with the mantle lithosphere. Hence, north vergent nappe stacking during this collisional stage took place within an accretionary wedge that starts to grow as more nonsubductable upper crustal granitic material of the European margin enters the subduction zone. Radiogenic heat production within this granitic basement, perhaps in combination with slab break-off, leads to a change in the thermal regime and to Barrovian type metamorphism."

Passage 2, from *Weil* (https://doi.org/10.1029/2005TC001861) [2006]:

"Paleomagnetic and structural analyses of the Western European Variscan Belt (WEVB) suggest that the most viable kinematic model for Variscan deformation in northern Iberia is oroclinal bending of an originally linear belt in a two-stage tectonic history. This history represents two regional compression phases (East West in the Late Carboniferous and North South in the Permian, both in present day coordinates), which resulted in the refolding (about steeply plunging axes) of initially north south trending thrusts and folds in the hinge zone, and oroclinal tightening due to vertical axis rotation of the belt's limbs. However, the orocline model has yet to be critically tested in the WEVB's core. This study reports new paleomagnetic, rock magnetic, and structural data from the inner core of the WEVB in order to test opposing kinematic models for the well documented fault and fold interference structures formed by late stage Variscan deformation and to better understand the overall development of the WEVB arc."

Author Information

Michele Cooke (@geomechCooke (https://twitter.com/geomechcooke)), Department of Geosciences, University of Massachusetts Amherst; Celia R. Child, Department of Geology, Bryn Mawr College, Bryn Mawr, Pa.; Elizabeth C. Sibert (@elizabethsibert (https://twitter.com/elizabethsibert)), Department of Earth and Planetary Sciences, Yale University, New Haven, Conn.; Christoph von Hagke (@StrucGeology (https://twitter.com/StrucGeology)), Department of Geography and Geology, University of Salzburg, Salzburg, Austria; and S. G. Zihms (@geomechSteph (https://twitter.com/GeomechSteph)), University of the West of Scotland, Paisley, U.K.

Citation: Cooke, M., C. R. Child, E. C. Sibert, C. von Hagke, and S. G. Zihms (2020), Caption this! Best practices for live captioning presentations, *Eos, 101,* https://doi.org/10.1029/2020EO150246. Published on 09 October 2020.

This article does not represent the opinion of AGU, *Eos,* or any of its affiliates. It is solely the opinion of the author.
