A DESKTOP APPLICATION TO HELP SPEAKERS SWITCH SLIDES BY USING AI AND VOICE RECOGNITION

Yixin Liang[1], Marisabel Chang[2]

[1]Portola High School, Portola High School, 1001 Cadence, Irvine, CA 92618

[2]Computer Science Department, California State Polytechnic University, Pomona, CA 91768

ABSTRACT

Presentation is a skill that everyone has, and it is very commonly seen in companies, schools, conferences, etc [1]. And the purpose of a slide is to give the audience a better understanding of the topic and to add ideas that they forgot to mention [2]. It also adds visual support to the speaker's discussion. Usually the presenter held a slide remote or just used their computer to control the slide pace while presenting. However, the slide remote can often be unstable due to battery switching. Even those who do not have a slide remote are unable to ensure a smooth presentation because they need to constantly switch back and forth on the computer screen with the mouse, which not only makes the speaker more nervous but also likely to skip the slide. Slidecontroller uses existing AI technology,voice recognition, as a medium to allow users to enter the transition word used to switch slides [3]. For example, when the user enters "Now I am going to talk about" when this word is spoken the Slidecontroller will receive the voice and match the speaker's turn to the next slide. The user can be creative with the keyword selection that best fits their presentation vibe. Or the user could use the Slidecontroller default option which controls the slide by simply saying "Next" to go to the next slide, "Previous" to go to the previous slide, and "Thank you" to stop the App to prevent from catching a similar keyword that accidentally switches the slide [4].

KEYWORDS

AI, voice recognition, Slide controller

1. INTRODUCTION

In modern society, almost all successful entrepreneurs are presenters [5]. Whether it is entrepreneurs or the workplace daily work report, results can not be separated from speech, over time has become a necessary skill for everyone. And the speech itself is a skill to help put this realization of self-worth and create value for others to open up the road, able to enhance the connection between each other and maximize the output of their own experience. And to further assist in improving people's presentation skills, the first thing to address is technology [6]. Especially on some formal occasions with the mouse, slide remote, or keyboard constantly switched or very distracting behavior. To allow users to reduce the anxiety of the presentation, and in the case of not considering the use of slide remote and keyboard, the speaker's hands can be free to do gestures to increase the interest of the speech. At the same time, there is no need to worry about the system short circuit because only a computer and the presenter's "voice" is needed. This not only improves the efficiency of the presentation but also makes the transition between slides more smooth [7]. Slidecontroller also takes into account the inconvenience of the

David C. Wyld et al. (Eds): CCNET, AIMLA, CICS, IOTBS, NLTM, COIT - 2023
pp. 151-160, 2023. CS & IT - CSCP 2023 DOI: 10.5121/csit.2023.130412

disability group, so it uses only the user's voice to make this desktop application more convenient and comprehensive.

Currently, two technologies can control slide conversion, one is the traditional hardware configuration of slide remote, and the other is the new google slide presentation remote on cell phones launched by google in 2018. First of all, let's analyze the traditional slide remote, let'stake the Canon PR10-G wireless presentation remote as an example to analyze the object. Same as with other slide remotes, the speaker needs to hold it in his/her hands during the presentation. The difference is that the Canon PR10-G wireless presentation remote is designed with injection-molded plastic for easy grip, but the location of the buttons will inevitably be worn and cause the remote to malfunction under prolonged use. And some speakers are not comfortable with holding something in their hands while presenting, and this only makes them more nervous. Second thing to analyze is the core of each presentation remote control, the control that allows the user to control the slide presentation at any time. On top of the Canon PR10-G, the wireless presentation remote is the typical forward and backward buttons, and below it is the "present" button. This can be used with PowerPoint or Keynote to enlarge images, videos, or charts to full size. The problem is that the sensitivity of the controls is unquestionable, so users are likely to press the forward or back button more than once in the most stressful situations, causing the slide show to miss. The last is its battery life, Canon PR10-G wireless presentation remote control requires users to replace the battery, and can not be recharged, and every time to replace the battery is also very environmentally unfriendly behavior, especially teachers in the use of 7 days a week its battery life is only 7-9 months, so the average consumption of the battery is at least two per year. The next existing technology to analyze is google's new cell phone presentation remote control, the basic function is almost the same as the traditional slide remote, just replaced by holding the phone in your hand only. Google developed the cell phone presentation remote control's biggest drawback is that it only applies to slides made with Google Slide and others like PowerPoint. can't be used. And it becomes very inconvenient and unattractive on serious occasions or when the use of cell phones is not allowed.

Slidecontroller is a desktop application based on AI voice recognition technology [8]. By accurately receiving the user's voice, it collects data and analyzes the hidden keywords, and connects them to the "up arrow" and "down arrow" on the computer keyboard to make relative commands. The goal of this application is to allow all users to speak without worrying about any technical problems. Even if the speaker can not pay the price of a traditional slide remote, then this app must be the most affordable and accurate replacement. In contrast to existing remote control technology, Slidecontroller allows users to customize their Keywords and works with all Slide-creating apps, not just Google slide or PowerPoint. The Slidecontroller is also equipped with a default option if the user just wants a small presentation. The technology used for this option is Hotword detection, similar to the technology used by Siri or Alexa, which is a miniature algorithm that monitors the audio stream of special hotwords [9]. Slidecontroller is very easy to use and does not require the user to have anything in their hands while presenting, so anyone can use it.

The best proof of the usefulness of Slidecontroller desktop applications on the spot is its use in everyday presentations. As the developer of the desktop application myself, I use Slidecontroller for every presentation opportunity at school, and the actual results are very good, helping me to relieve my nervousness to a greater extent and adding more hand movements to make the presentation more vivid. The data proves that Slidecontroller has completely replaced the existing slide remote technology with efficiency and accuracy. The number of seconds it takes to rotate the slides after each keyword is detected by the record. The accuracy of Voice recognition was also checked to ensure that the correct keywords were collected. After nearly two thousand

tests, the Response time was only increased to about 2.3 seconds when the user Customized more than 5 keywords, but all other functions were stable and accurate.

The rest of the paper is organized as follows: Section 2 gives the details on the challenges that were found while developing the Software and finding the most accurate Voice Recognition AI library; Section 3 focuses on the component that were used to solve the challenges as mentioned before and will present part of the code to shows the details of how the AI technology was used; Section 4 presents the relevant details about the experiment we did, following by presenting the related work in Section 5; Section 6 gives the concluding remarks, as well as pointing out the future work of this project.

2. CHALLENGES

In order to build the project, a few challenges have been identified as follows.

2.1. Efficiency & Accuracy

The core of Slidecontroller, a desktop application, is the clever use of existing AI voice recognition technology to connect to the front and back keys of the keyboard for control. Therefore, the voice recognition library needs to be carefully selected, and the most basic requirement is that it should be able to run on both Windows and IOS platforms without any problems. Then is the accuracy of the received speech. In the Slidecontroller system accuracy is very important, because the computer needs to store the correct data to output the right command, but the existing AI speech recognition technology is limited in inclusiveness [10]. For example, when the user's voice input with another country's accent, voice recognition will recognize the wrong word. The accuracy of speech recognition is also affected when the user maintains a slight pause of more than six or seven seconds. Only when the user speaks perfect English and there is no pause, the speech recognition can be 100% error-free. But because the current AI speech recognition technology has not been able to expand its inclusiveness, the space available is limited. The next problem is the efficiency affected by the accuracy of speech recognition. Without the correct recognition of the user's voice output, Slidecontroller is unable to complete the command of switching slides the first time. The existence of Slidecontroller is meaningless if it cannot respond to the user's needs in the first place, and this is the biggest challenge of this desktop application.

2.2. Voice Receiving

As mentioned earlier, Slidecontroller has very high requirements for the speech recognition library, so the reception of the user's voice needs to be very subtle and not affected by outside influences. In most cases, there will be an audience on the stage during the speech, and if there isa little noise from the audience, speech recognition is likely to be affected, causing the stored content to be different from what the speaker describes. Or when the speaker is at a certain distance from the computer, speech recognition is difficult to capture complete sentences, and all these factors can directly affect the operation of the Slidecontroller. But Hotword Detection, which is used for the default keyword option of Slidecontroller, achieves the maximum absolute sensitivity of speech recognition [11]. But unlike normal speech recognition libraries, Hotword detection is limited to one or two keywords first, to improve the detection speed. But to maintain the diversity of Slidecontroller, for users who want to create their keywords, then the regular speech recognition library is inevitable. Therefore, the speech recognition library needs to be sensitive and unaffected by any external factors to maximize the benefits of Slidecontroller

3. SOLUTION

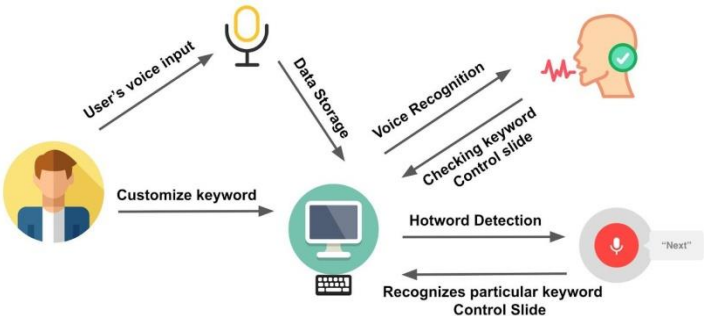

Figure 1. Overview of the solution

As the above prototype shows, the main source of data for Slidecontroller is the user's audio source. The computer's microphone picks up the audio and then the computer stores the data, which is analyzed by different algorithms depending on the keyword option selected by the user. The first important step is the computer keyboard control, which converts the control keys that can control the slide transition into commands. Then is the Hotword detection, where the user will use the default keyword, and the last one is if the user chooses to customize his keyword, where the Slidecontroller will automatically switch to Voice Recognition mode. Combining the above three main sections with the post-processed GUI screen, the user can operate more conveniently and smoothly. Figure 2 shows the complete Slidecontroller in the user's view. If the user chooses the default option, then first chooses whether the system is Windows or IOS, Picovice will change the Hotword Detection template according to the computer platforms but the function is still the same [12]. Then the user needs to copy the link at the top of the slide to the text entry of "Insert Link". Then if the user wants to save the link, they can click "Save" and name the link to make it easier to find. Finally, the user can control the slideshow by saying "Next" and "Previous". There is also a detailed tutorial in the lower right corner, which can be accessed by the single question mark icon. If the user wants to change the mode, click the arrow below to switch to Customize keyword, the procedure is similar to Default Keyword mode, you don't need to worry about your computer system, just copy the slideshow link to the corresponding Text entry and enter the corresponding Customize keyword to control the slideshow. After all the selections, the user can then click "Start Presentation" to begin!

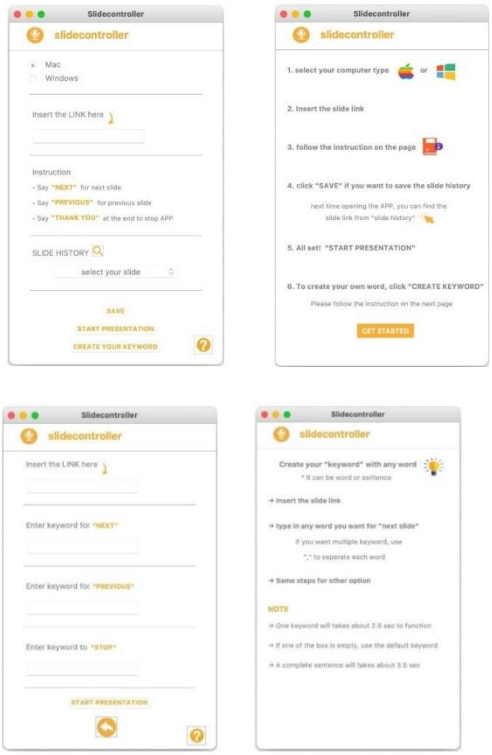

Figure 2. Overall looking of the desktop application

```
pageNumber = 1 # keep track on which slide
def nextPage():
    global pageNumber
    keyboard.press(Key.page_down)
    keyboard.release(Key.page_down)
    pageNumber += 1
    print("next function is working properly")
```

Figure 3. Page number code

As mentioned earlier Slidecontroller is composed of three main components. And the very first step is to control the slides using the up and down keys of the computer keyboard. Generally, in the absence of a slide remote, the presenters can only use the up and down keys of the computer for control, and the slide remote itself uses this method. So, as shown in the following code, the "Next" function for switching to the next slide is used as an example, and the first step is to define the “pageNumber” accumulator to keep track of the slide number. An accumulator is being defined, and then using the keyboard press function in the keyboard module is to let the keyboard autonomously click the forward button. The keyboard simulator is the basis of the whole Slidecontroller, and the final product is made up by adding voice control to the keyboard control.

```
def open_slide():
    global pv
    if urlEntry.get() == "":
        url = slide_history[value_inside.get()]
        webbrowser.open(url)
        time.sleep(3.5)
        present()
    else:
        webbrowser.open(urlEntry.get())
        time.sleep(3.5)
        present()
    if pv == True:
        voice3()
```

Figure 4. Default option code

The second component is the Default option, which is created using Hotword detection technology and covers the "Next" keyword to turn to the next slide, the “Previous”keyword togo back a slide, and the “ Thank you ” keyword at the end of the presentation to stop the program. Hotword detection is very similar to the Always-Listening Commands technique,which executes commands by using multiple words with the help of Always-Listening Commands. The difference is that Hotword Detection relies on hotwords, trigger words, keywords, or wake-up words to activate the dormant software. It is a combination of voice activation and Always Listening Commands. Also, voice activation is a key point of Hotword Detection, which activates the application by voice and then works with Always-Listening Commands to control the slideshow. Then by combining the previously set keyboard controls with Hotword Detection, the slideshow can be controlled by touching the "Next", "Previous",and "Thank you" keywords to complete the slideshow control commands in time. As shown in the code above, a web browser opening technique was added to allow users to copy the slide URL and then the computer would automatically convert it to present mode, making it more convenient without the need for users to do it by hand and allowing the audience to better understand the content of the slide after zooming in.

```
def verify_keywords(self,text):
    text = text.lower().replace('.', '').replace(',', '')
    print(self.next)
    if text == self.next:
        print('next')
        nextPage()
    elif text == self.previous:
        print('previous')
        previousPage()

    elif text == self.thankyou:# stop the program
        print(self.thankyou)
        sys.exit(0)
```

Figure 5. Customize keyword code

The last component is the creative Customize Keyword, where users can use their imagination to create a keyword that fits the speech. Speech recognition technology uses Always-Listening technology, which always listens to the voice used and activates Speech-to-Text to transcribe the

surrounding dialogue when speech is detected. In more detail, the system first analyzes the audio, then breaks it down into parts, digitizes it into a computer-readable format, and finally uses an algorithm to match the audio with the most appropriate text representation, and then looks for the specified keyword in the matched text [13]. As the above coding shows, it first confirms what the user specified as the keyword and then works with the keyboard to the command is completed with the keyboard control.

4. EXPERIMENT

4.1. Experiment 1

To test the efficiency and accuracy of speech recognition libraries supported by existing AI technology, I compare two speech recognition libraries, Google Voice Recognition and Assembly AI, using speech recognition filtered from 15 different companies before. I will apply each of the two voice recognition libraries to the Slidecontroller system, where the keywords will be the same for each voice input, and then record how many seconds it takes to respond and implement the instructions to switch slides.

Two separate computers with the same platforms(either both windows or ios). The speaker should stand at equal distance to both computers to ensure the computer receives the voice data equally. Run both Slidecontroller with different voice recognition libraries at the same time, and speak the same amount of words (constant variable). Different amounts of keywords will be tested 3 times, the maximum keywords included will be 5, so the data will take the average of 15 times. Then record the number of seconds it takes to process the keywords and respond to turnthe slide.

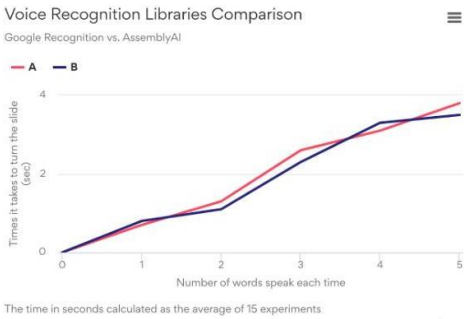

Figure 6. Voice recognition libraries comparison

As shown from the statistical chart above, the difference between Google Recognition and AssemblyAI is only a fraction of a second. But this fraction of a few seconds difference will directly affect the efficiency of the Slidecontroller. In the 15 tests, AssemblyAI is more stable and accurate than Google Recognition in terms of general trends. Google Recognition is a popular speech recognition system that does not require payment and is not used for software development, considering the target audience. Assembly AI, on the other hand, is a paid speech recognition library, but for a fee of ¥5, you can use it for an unlimited time. However, there is no difference between the two speech recognition libraries in general, but Slidecontroller requires higher sensitivity, so AssemblyAI is better and the cost is reasonable, so users do not have to worry about the financial burden.

Experiment 2

In this experiment, two algorithms that were used in the Slidecontroller are being compared based on time efficiency, which will show how many seconds it takes for the keyword to react. Hotword detection is provided by the picovoice platform, and speech recognition is provided by Assembly AI. The purpose of this test is to allow users to better visualize the operation and response time of the two different programs so that they can be more secure when choosing one. It is also to ensure that the user selects the appropriate function for the presentation. Since Picovoice provides hotword detection only with the Default keyword option provided by Slidecontroller (all the default keyword that use to control slide is one-word “Previous” and “Next”, except the ending keyword that is used to stop the App which is two words “Thank you”), it does not provide the ability for users to create their keywords. Assembly AI, however, can fulfill this need by using the freedom of Voice Recognition combined with the Keyword Capture feature to allow users to create keywords that match the atmosphere and fluency of the presentation.

Two separate computers with the same platforms (either both windows or ios). The speaker should stand at equal distance to both computers to ensure the computer receives the voice data equally. The Hotword detection will run the same keyword for 15 times, and the Assembly will run different keywords based on user customize keywords. Then record the number of seconds it takes to process the keywords and respond to turn the slide.

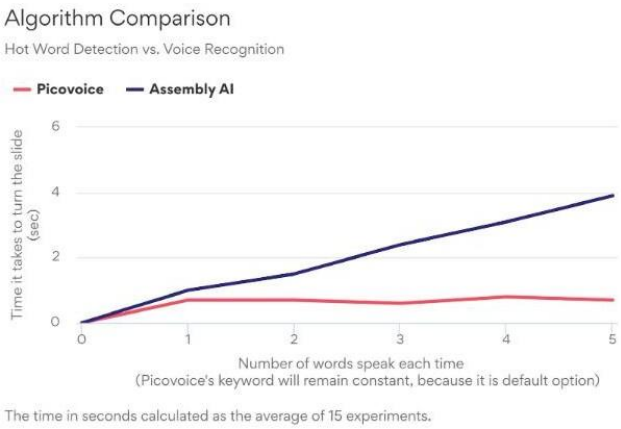

Figure 7. Algorithm comparison

As shown in the line data graph, the hotword detection provided by Picovoice has a response time of basically no more than one second when the user speaks the keyword of default and is very stable. Because of its special voice monitoring system, "Always-listening Commands" allows it to always be listening, knowing that the user speaks a specific keyword to activate the corresponding command. On the other hand, Assembly AI wins the comparison with Google Recognition, but the algorithm of detection is very different from Hotwork's, and the response time increases according to the number of keywords. Therefore, the results of this experiment are

intended to show users the advantages and disadvantages of the two different algorithms and to give them a better reference to evaluate which one to use. If the user is not overly concerned about the overall smoothness or does not have any idea to customize the keyword, it is recommended to use the default option keyword already provided by Slidecontroller, which not only guarantees overall accuracy and fast response. But if the user wants to incorporate more personal ideas or wants to fit the speech scene, you can use Assembly AI, although the responsiveness will be slightly inferior to Hotword detection, but comparable to other Voice Recognition with the same algorithm, and its accuracy can also be guaranteed.

In the existing AI speech recognition technology, Hotword Detection, and Assembly AI are used in much the same way. For Assembly AI, the more words the user input, the longer it will take. But according to the graph above, it shows the maximum amount of time for 5+ keywords is 3 seconds which doesn ’ t affect the overall presentation. However, Hotword Detection is recommended for less formal situations because the keywords that can be used are very limited. For example, students using Slidecontroller's default keyword when the teacher does not provide a Slide Remote is a good choice and will respond promptly. Both speech recognition libraries have very good performance and accuracy, so no matter which one you use, it will help you to complete a perfect speech.

5. RELATED WORK

In 2018, Google released "Remote for slides," a slide remote that could be operated by phone [14]. Each operation required a 6-digit unique number that was linked to the relevant slide that the user was presenting. Users can utilize this free web application by following a few straightforward steps. On the smartphone UI, there are simply two enormous buttons that read "Next Slide" and "Previous Slide," respectively. The Back and Forward buttons that come with Google Slide are combined in this app. In the first week after its introduction, it attracted a lot of attention, but its flaws soon became apparent. Users must have both a cell phone and a computer, without either one then the slide will not be able to turn into the present mode, this step must be operated on the computer. Also, if the user's phone is blacked out during the presentation, then it will directly lead to the APP not working. And on some formal occasions, holding a cell phone in your hand is not suitable and affects the audience's perception. Slidecontroller can easily solve all the above three points. First of all, users don't need to do anything, when users click "Start Presentation", the computer will automatically turn into Presentation Mode, secondly, when the desktop application is running, unless it is forced to shut down, basically it will not black screen. Finally, the speaker does not need to hold anything in his hands when he is in Slidecontroller, so it does not affect the overall appearance.

6. CONCLUSIONS

Slide controller is a desktop application designed to assist presenters in becoming more confident in their presentations. Implementing "Hands-Free ", allows speakers to use their hand gestures to make their presentations more interesting and to reduce their worries about slide presentations. Users are not limited to a single keyword option, they can create their Transition Keyword to suit the atmosphere of their presentation, and if they don't need it, then Slide controller's Default Keywords will work just as well. Anyone can use this software, and the ultimate goal of Slide controller is to help you deliver a better presentation [15]. Because of the limitations of the current AI voice recognition technology, there is no way to make greater use of Slidecontroller's advantage is to switch slides in one to two seconds or less. To complement this shortcoming, I asa developer will try to train my speech recognition system, and refine it to be sensitive and accurate.

REFERENCES

[1] Pittenger, Khushwant KS, Mary C. Miller, and Joshua Mott. "Using real-world standards to enhance students' presentation skills." Business Communication Quarterly 67.3 (2004): 327-336.

[2] Hayama, Tessai, Hidetsugu Nanba, and Susumu Kunifuji. "Structure extraction from presentation slide information." Pacific Rim International Conference on Artificial Intelligence. Springer, Berlin, Heidelberg, 2008.

[3] McGehee, Frances. "An experimental study of voice recognition." The Journal of General Psychology 31.1 (1944): 53-65.

[4] Turban, Georg, and Max Mühlhäuser. "A uniform way to handle any slide-based presentation: the universal presentation controller." International Conference on Multimedia Modeling. Springer, Berlin, Heidelberg, 2006.

[5] Hornaday, John A., and John Aboud. "Characteristics of successful entrepreneurs." Personnel psychology (1971).

[6] Haber, Richard J., and Lorelei A. Lingard. "Learning oral presentation skills." Journal of general internal medicine 16.5 (2001): 308-314.

[7] Steinman, Ralph M. "Dendritic cells and the control of immunity: enhancing the efficiency of antigen presentation." The Mount Sinai journal of medicine, New York 68.3 (2001): 160-166.

[8] Extance, Andy. "How AI technology can tame the scientific literature." Nature 561.7722 (2018): 273-275.

[9] Zhang, Li, et al. "Accelword: Energy efficient hotword detection through accelerometer." Proceedings of the 13th Annual International Conference on Mobile Systems, Applications, and Services. 2015.

[10] Ran, Duan, Wang Yingli, and Qin Haoxin. "Artificial intelligence speech recognition model for correcting spoken English teaching." Journal of Intelligent & Fuzzy Systems 40.2 (2021): 3513-3524.

[11] Huang, Yiteng, et al. "Multi-microphone adaptive noise cancellation for robust hotword detection." (2019).

[12] Grønli, Tor-Morten, et al. "Mobile application platform heterogeneity: Android vs Windows Phone vs iOS vs Firefox OS." 2014 IEEE 28th International Conference on Advanced Information Networking and Applications. IEEE, 2014.

[13] McDonald, Robert S., and Paul A. Wilks. "JCAMP-DX: A standard form for exchange of infrared spectra in computer readable form." Applied Spectroscopy 42.1 (1988): 151-162.

[14] Haque, Md, Abu Raihan, and Mohd Khalidi. "BTpower: An Application for Remote Controlling PowerPoint Presentation Through Smartphone." Advances in Computer and Computational Sciences. Springer, Singapore, 2017. 13-20.

[15] Alley, Michael, et al. "How the design of headlines in presentation slides affects audience retention." Technical communication 53.2 (2006): 225-234.