



INTERACTIVE SLIDE PRESENTATION USING SPEECH RECOGNITION

Rona A. Sy

Article History: Received: 25.03.2023

Revised: 18.05.2023

Accepted: 26.06.2023

Abstract

This research project aimed to develop a voice-controlled system for PowerPoint presentations, enabling presenters to navigate slides using voice commands. The project followed a comprehensive project development process, including data gathering, requirements analysis, design, coding, testing, and evaluation. The primary objective was to investigate the feasibility and effectiveness of utilizing voice commands to control PowerPoint presentations. The findings of this study revealed the successful development of an application software that allows users to send commands to the computer using their voice, enabling seamless navigation through slides in a PowerPoint presentation. Notably, the system demonstrated remarkable functionality, even when the speaker was located up to 10 meters away from the computer, showcasing its wireless capabilities. The voice-controlled system exhibited high levels of accuracy, speed, and efficiency, providing an effective and convenient presentation experience. By eliminating the need for manual interaction with a keyboard or mouse, presenters were able to maintain better engagement with their audience. The voice-controlled functionality streamlined the presentation process, reducing the risk of distractions and ensuring a more natural and immersive delivery. The developed application software for the Voice-Controlled PowerPoint Presentation system offers a valuable and efficient tool for presenters to enhance their presentations. By harnessing the power of voice commands, presenters can effortlessly navigate through slides and deliver a highly engaging and interactive presentation experience. The outcomes of this research contribute to the advancements in presentation technologies, particularly in the realm of voice-controlled interfaces. Future research can focus on further refining the system's features, expanding compatibility with different versions of PowerPoint, and exploring additional voice commands and functionalities. The findings demonstrate the effectiveness of the developed application software in facilitating a highly efficient and engaging presentation experience for presenters using voice commands in PowerPoint. The voice-controlled system presents a significant contribution to the field of presentation technologies, revolutionizing the way presentations are delivered and enhancing the overall presentation experience.

Keywords: Voice-Controlled Power Point Presentation, speaker, voice command, next, previous, slide navigation.

Chairperson, Computer Laboratories Jose Rizal Memorial State University – Dipolog Campus

Email: ronasy041881@gmail.com

DOI: 10.31838/ecb/2023.12.s3.775

1. Introduction

Certain innovations have spread so widely that it is challenging to imagine how humanity managed before them. People find it difficult to picture living without electricity, air travel, automobiles, or television nowadays. Microsoft's Office suite was one of the technological advancements that entered the pantheon of technologies that control communication methods. Microsoft Word, Microsoft Excel, and Microsoft PowerPoint are all used by people to write, compute, and put together presentations. Only PowerPoint, one of the three main products in the Microsoft Office Suite, experienced a true counter-culture.

In the commercial and academic worlds, PowerPoint has emerged as the de facto presentation tool. Over 30 million PowerPoint slideware presentations are made every day, according to Microsoft's estimate of 250 million PCs with the program installed (Amare 299). The new chalkboard is now the program. Today, presentations and lectures do not officially begin until a title slide lights onto a screen. Twenty years ago, students expected their teachers to start class and write on the board.

Voice-controlled systems have gained significant attention in recent years, offering innovative ways to interact with technology and improve user experiences. This literature review aims to provide a global context for the research project that developed a voice-controlled system for PowerPoint presentations. By examining relevant studies and advancements in voice-controlled interfaces and presentation technologies, this review establishes the foundation for the research and highlights its significance. In recent years, the Philippines has witnessed a surge in technological advancements, leading to a transformation in various aspects of daily life and professional settings. The realm of presentation technologies is no exception, where innovations have played a vital role

in enhancing communication and knowledge dissemination. One such groundbreaking development is the Voice-Controlled PowerPoint Presentation system, which allows presenters to navigate slides effortlessly using voice commands. This study aims to explore the feasibility and effectiveness of employing voice-controlled interfaces for PowerPoint presentations within the continental context of the Philippines.

Technological adoption and integration in the Philippines have seen significant growth, with the country emerging as one of the leading adopters of innovative technologies in the Southeast Asian region. The Philippines' vibrant economy, increasing internet penetration, and a burgeoning tech-savvy population provide a conducive environment for the implementation and acceptance of cutting-edge presentation tools.

The research study on the development of a voice-controlled system for PowerPoint presentations holds significant implications in various areas. It enhances the overall presentation experience by allowing presenters to navigate slides seamlessly while maintaining audience engagement. With the elimination of manual interaction through keyboards and mice, the voice-controlled system provides a more natural and immersive presentation experience, thereby improving communication between presenters and their audience.

The voice-controlled system brings efficiency and convenience to the presentation process. Presenters can effortlessly control slides using voice commands, eliminating the need for manual operations such as searching for the next slide or clicking buttons. This not only saves time but also reduces distractions during presentations. Moreover, the system's wireless capabilities enable presenters to control slides from a distance

of up to 10 meters, providing added convenience and flexibility.

The research contributes to advancements in presentation technologies, particularly in the realm of voice-controlled interfaces. By developing and evaluating a voice-controlled PowerPoint presentation system, the study demonstrates the feasibility and effectiveness of utilizing voice commands for slide navigation. This opens up possibilities for future innovations and improvements in presentation tools and interfaces, driving the evolution of presentation technologies.

The voice-controlled system also enhances engagement and audience interaction. With reduced reliance on manual interactions, presenters can focus more on delivering their message, maintaining eye contact with the audience, and effectively using gestures. This promotes a more interactive and immersive presentation environment, leading to increased audience engagement and improved communication.

Conducting the study within the context of the Philippines acknowledges the significance of local technological adoption and integration. The research recognizes the country's growth in technological advancements, vibrant economy, increasing internet penetration, and tech-savvy population. By developing a voice-controlled system tailored to the specific needs and preferences of presenters in the Philippines, the study contributes to the technological landscape of the region, catering to the local context and fostering technological advancements.

The research study on the development of a voice-controlled system for PowerPoint presentations holds significance in enhancing the presentation experience, improving efficiency and convenience, contributing to advancements in presentation technologies, promoting engagement and audience interaction, and addressing the local context of technological adoption.

Despite the widespread use of PowerPoint presentations as a standard communication tool in commercial and academic settings, the traditional manual control methods using keyboards or mice can be cumbersome and may lead to reduced audience engagement. Presenters often face challenges in seamlessly navigating through slides while maintaining eye contact and effective gestures. Additionally, presenters who need to move away from the computer to interact with the audience during presentations may encounter limitations due to the physical distance from the computer.

Voice-Controlled Power Point Presentation was designed to navigate the slides of power point presentation. This system retains the most important features of voice matching by sending voice command to the presentation ten meters away from the computer. The ultimate purpose of Voice-Controlled Power Point Presentation is to have a better interaction between the speaker and its audience.

2. Methodology

The study utilized the project method using constructive techniques, the most common computer science research method. This involves evaluating the project being developed against some predefined criteria or performing some benchmark tests with the prototype. This project refers to the new contribution being developed. The existing power point presentation uses keyboard, mouse, and remote control to navigate the slides during presentation. The presenters are glued to their laptop to change slides. They are tucked behind a podium that separates them from their audience, leaving them guessing at what is being talked about. The main components of the project are: wireless microphone used to input the voice into the computer, computer unit that denotes the hardware and software including Microsoft Power Point, and the

Voice - Controlled for Power Point Presentation program that manipulates power point presentation by recognizing the voice of the speaker using the microphone.

Project Development Process

Data Gathering

The first phase in the project development processing. In this phase, the researcher gathered data and information on the existing mode of manipulation of power point presentation. Power point presentation was used in schools, churches and offices. During power point presentations, speakers use keyboard, mouse and remote control to navigate the slides of the power point presentation. Second, the researcher found that there were application software such as Microsoft Word and Win amp that were manipulated by the so called Speech Recognition. Speech recognition software recognizes the voice of the user using a microphone attached to the computer.

Requirements Analysis

The researcher analyzed the collected information and identified the requirements needed in developing voice - controlled power point presentation. These requirements were taken from the two components of the computer system. First are the hardware requirements and second are the software requirements.

Hardware Requirements

- One Unit Laptop/Desktop Computer with Windows XP Operating System
- 1 GB Memory
- 2 MB HDD
- 3* USB

- **Wireless Microphone**

WM 300 wireless transmitter
WM 300 wireless receiver
Wireless headset microphone.
Software Requirements

- **Operating System**

Windows XP SP 2/ 2003

- **Application Software**

MS Office 2003/XP/2007

Microsoft English Recognizer v5.1

- **Voice – Controlled Power Point Presentation**

These were the hardware requirements; One Unit Laptop/Desktop Computer with the following specifications: Any computer that has Windows XP operating system, 1 GB Memory, 1GB HDD, 3* USB. A desktop computer or preferably a laptop is needed because the software that was developed (Voice – Controlled Power Point Presentation) functions using computer hardware. The second hardware requirement was the wireless microphone that is composed of a WM 300 wireless transmitter, wireless receiver, and wireless headset microphone.

The following are the software requirements; first, the operating systems like Windows XP SP 2/ 2003. Second is the application software specifically the Microsoft Office (2003/XP/2007). Microsoft Office contains Microsoft Word, Microsoft Excel and Microsoft Power Point. This application software is needed because the project used the Power Point presentation that would be navigated. And lastly is the developed project the Voice – Controlled Power Point Presentation that was used to navigate power point presentation using the voice command of the speaker.

- Designing – After data gathering and determining the requirements needed, it was then developed into one composite system. Figure 17 in chapter 4 shows the combined software features that resulted to voice-controlled power point presentation.

- Coding – In this phase, the researcher converted the design of the system into machine readable language that the computer can understand using Visual Basic 6.0 programming language. Interfacing between Power Point

Presentation and the microphone was also included in coding the solution. Then the finished product is the designed software.

- **Software Testing and Evaluation** - After the researcher generated the designed software, the researchers with the I.T. professionals tested and evaluated the program to identify the functionality and accuracy of the developed computer software.

These include:

- □ The speaker must undergo ten to twenty minutes of voice training using a wireless microphone. The speaker reads different passages to train his pronunciation and the system would learn and recognizes the voice.
- □ Test if the system recognizes the voice of the speaker.
- □ If the system could recognize the voice, the speaker could open Microsoft Power Point Presentation.
- □ The speaker could start sending commands to the system as the presentation being shown.
- □ Test if the system follows the voice command like “Next” or “Previous”.
- □ Test if the system follows such command.

If error occurs, the researcher went back to coding and modifying the instruction and logical conditions of the entire project to produce the desired result. Shown below is the figure of the Software Development Life Cycle (SDLC) that is used in the study. Included are the five phases namely data gathering, requirements analysis, design, coding, and testing and evaluation.

3. Results And Discussion

During presentations, the speakers use keyboard and mouse in navigating the slides of their presentation. When speaking they were hooked into their computers and the audience were left behind guessing on what is the message he was trying to tell. The situation creates gap between the

speaker and the audience. In some cases, the manipulation of the power point is done by a second party of which case; the natural occurrence appears that what the speaker conveys is not the message presented on screen. This situation obviously creates confusion on what the message is conveyed because of the discrepancy presented on screen to what is conveyed by the speaker on the microphone. The existing mode of manipulation for power presentation which includes:

Speaker. The person who conveys the message to its audience.

Input Devices. These were composed of keyboard, mouse, and a remote control which enables the user to enter commands to manipulate power point presentations.

Application Software. Collection of programs or instructions which performs specific task like word processing, spreadsheets, and presentations.

Power Point Presentation. A presentation software or computer program used to organize and present information in a slide show.

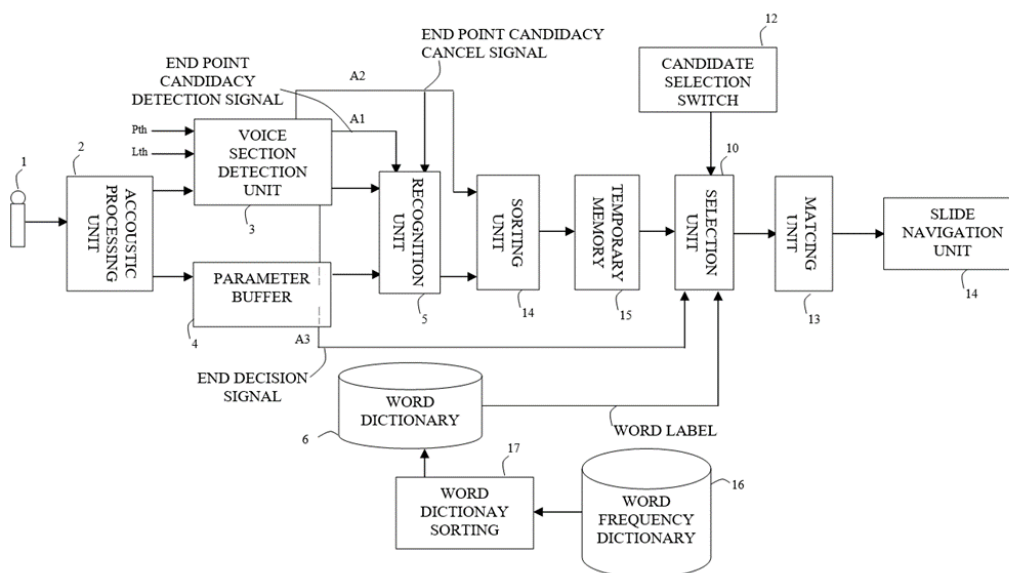
During presentation the speaker uses input devices such as keyboard, mouse and remote control to send commands to computer programs in which presentation slides navigates. Input devices accept commands from the speaker and convert the command into machine readable language. Computer programs execute the commands and power point slides in the audience view navigate.

In this study, the project design is a hands-free control of navigating power point sides. The speaker uses voice command to navigate the slides of the presentation. First, the designed software starts, then opening a power point presentation follows. When the speaker sends voice command using wireless microphone the voice signal was received by the microphone and sent to the computer through the microphone

transmitter and receiver. Then the designed software matches the data to the commands in power point, then if it matches power point navigation will occur.

Shown below is a systematic block diagram of voice recognition system for power point

presentation; each component has the following function. The microphone 1 converts the voice (speech input) to an electrical voice signal constituted by the voice sound portion and the non-voice sound portion.



Systematic Block Diagram of Voice Recognition System for Power Point Presentation

The presented systematic block diagram outlines a voice recognition system designed specifically for PowerPoint presentations. This system aims to enable presenters to control slide navigation using voice commands, offering a more convenient and interactive presentation experience. Let's discuss the key components and their functions in detail.

1. **Microphone:** The microphone captures the speech input and converts it into an electrical voice signal, comprising both the voice sound portion and the non-voice sound portion. It serves as the primary input device for the voice recognition system.

2. **Acoustic Processing Unit:** This unit performs acoustic analysis on the electrical voice signal by detecting its power and spectrum. The power and spectrum are measured over specific time intervals, ranging from milliseconds to tens of milliseconds. The acoustic processing unit

also converts the electrical voice signal into feature time-series data, providing essential information for further analysis and recognition.

3. **Voice Section Detection Unit:** This unit receives the power time-series data from the acoustic processing unit. It utilizes a power threshold level (P_{th}) and a time threshold level (L_{th}) to detect the start point and end point of the voice sound, ultimately determining the end decision point. By accurately identifying the voice sections within the signal, this unit enables precise recognition and analysis of the voice commands.

4. **Parameter Buffer:** The parameter buffer acts as a temporary storage unit for the spectrum time-series data obtained from the acoustic processing unit. It holds this data until it is required for recognition and comparison with word templates.

5. **Recognition Unit:** This unit receives the voice sound's start point and end point candidates from the voice section detection

unit. It also receives the spectrum time-series data corresponding to the voice sound from the parameter buffer. The recognition unit compares the received spectrum data with the word templates stored in the word dictionary. Using methods like Dynamic Programming (DP) matching, it calculates the degree of similarity (or distance or probability) between the spectrum data and each word template. The results are then outputted to the sorting unit.

6. **Word Dictionary:** The word dictionary stores word numbers, word labels, and word templates (feature time-series data). It serves as a reference for matching and recognizing voice commands. The word templates are ordered based on the frequency of use of each word, allowing for efficient recognition and sorting.

7. **Sorting Unit:** The sorting unit receives the word number and the degree of similarity from the recognition unit. It arranges the words in the order of similarity, with words having higher degrees of similarity ranked at the upper end. This sorting process ensures accurate recognition and enhances the system's effectiveness.

8. **Temporary Memory:** The temporary memory stores the recognized data after sorting. It preserves the results for further processing and display purposes.

9. **Selection Unit:** The selection unit retrieves the word labels corresponding to the word numbers of the top word candidate from the word dictionary. It transfers this word to the matching unit when the end decision signal is detected from the voice section detection unit. Additionally, when the candidate selection switch is activated by the operator, the selection unit reads the word label corresponding to the word number of the next word candidate and transfers it to the matching unit. This allows for flexible and manual selection of alternative word candidates.

10. **Candidate Selection Switch:** The candidate selection switch is operated by the presenter. It enables the request for the display of another word candidate when the matching unit does not contain the desired word after recognition.

11. **Word Frequency Dictionary:** The word frequency dictionary stores words along with their frequency of use. This information can be utilized for sorting and prioritizing words based on their usage frequency, enhancing the overall efficiency of the voice recognition system.

12. **Word Dictionary Sorting Unit:** Positioned between the word dictionary and the word frequency dictionary, this sorting unit arranges words in the word dictionary based on their frequency of use. By sorting the word dictionary according

Results of the Project Evaluation

Software Quality Factor	Mean Average	Descriptive Ratings
Accuracy	4.15	The designed software is accurate.
Speed	4.20	The designed software is fast.
Efficiency	4.16	The designed software is efficient.
Overall	4.17	The designed software is very Acceptable

The accuracy of the software was measured through the software evaluation conducted by selected IT Professionals. The results

obtained from the evaluation of the designed software are found in Table 4. It was found to be accurate, efficient, and fast

and equipped with all the features needed by end-users. Speed was tested by measuring the time power point slide navigates when the speaker send commands. Computer speed on which the software runs was also considered. Efficiency of the software was also measured on its level of performance and the amount of resources used. System users together with the researcher had undergone testing and evaluation of the designed voice recognition system for power point presentation. These include:

- First the speaker must undergo ten to twenty minutes of voice training using a wireless microphone. The speaker reads different passages to train his pronunciation and the system would learn and recognizes the voice.

- Test if the system recognizes the voice of the speaker.

- If the system could recognize the voice, the speaker could open Microsoft Power Point Presentation.

- The speaker could start sending commands to the system as the presentation being shown.

- Test if the system follows the voice command like “Next” or “Previous”.

The following were the features of some software applications that use speech recognition technology:

Voice Recognition System for Microsoft Word

- Computer user may use his voice to send commands to the computer using a headset microphone.

- It accepts voice command of either male or female.

- The user used a headset microphone attached to the computer.

- Offers voice training to the user.

Voice Recognition System for Winamp

- Computer user use his voice to send commands to the computer

- The user used a headset microphone attached to the computer.

- It accepts voice command of either male or female.

Voice-Controlled Power Point Presentation.

It is a collection of different features of similar systems combined into one system. The system has the following features:

- Accepts voice command from the speaker;

- Accepts voice of either male or female is acceptable;

- Accepts well pronounced commands from the speaker;

- Accepts voice commands to navigate the slides of the presentation.

- The speaker may use his voice to send commands like navigating the slide of the presentation 10 meters away from the computer;

- The system has power point command training.

- And the system has the capability to read the presentation itself in behalf of the speaker.

Comparison of Software Features

Software Features	Voice Controlled Power Point Presentation	Voice Recognition System for Winamp	Voice Recognition System for Microsoft Word
Accepts voice of either male or female.	✓	✓	✓
The user may use his/her voice to go send commands to the computer.	✓	✓	✓
Accepts any voice with proper pronunciation.	✓	✓	✓
Offers speech to text training to the speaker.	✓		✓
Accept commands to jump from one slide to another.	✓		
Voice command can control the computer even if the speaker is 10 meters far from the computer.	✓		
It is capable of reading its presentation	✓		
Offers power point command training.	✓		
The project has a search slide title feature.	✓		

4. Conclusion

Based on the findings of the study, the researcher concludes that Voice – Controlled Power Point Presentation is accurate and efficient in controlling power point presentations.

5. References

- Abowd, G. D., & Mynatt, E. D. (2000). Charting past, present, and future research in ubiquitous computing. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 7(1), 29-58.
- B. Lowerre (1990). *The HARPY Speech Understanding System*, Trends in Speech Recognition, W. Lea, Editor, Speech Science Publications, 1986, reprinted in *Readings in Speech Recognition*, A. Waibel and K. F. Lee, Editors, pp. 576-586, Morgan Kaufmann Publishers
- Bahrammirzaee, A., & Azadegan-Mehr, M. (2020). Voice recognition technology: A comprehensive literature review. *Computers in Human Behavior*, 111, 106431.
- BH Juang (2004). "Automatic Speech Recognition – A Brief History of the Technology Development"

- http://www.ece.ucsb.edu/Faculty/Rabiner/ece259/Reprints/354_LALI-ASRHistory-final-10-8.pdf
- Brown, G. M., Murray-Smith, R., & Benford, S. (2003). Mobile spatial interaction with the "Whereabouts Clock". In Proceedings of the 2003 conference on Human factors in computing systems (CHI '03) (pp. 295-302).
- Chen, Z., & Zhu, Y. (2019). Voice-based interaction for intelligent presentation systems: A review. *Future Generation Computer Systems*, 100, 367-379.
- D. Franklin (2000). The Guide to computing Literature "Jabberwocky: you don't have to be a rocket scientist to change slides for a hydrogen combustion lecture". <http://portal.acm.org/citation.cfm?id=325792>
- D. Jurafsky, J. Martin (2000). "Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics and Speech Recognition". ISBN: 0130950696.
- E.Chang, F.Seide, H.Meng, Z.Chen, Y.Shi,Y.Li (. "A System for Spoken Query Information Retrieval on Mobile Devices"
- F. Jelinek, L. R. Bahl, and R. L. Mercer (2000). Design of a Linguistic Statistical Decoder for the Recognition of Continuous Speech, *IEEE Trans. On Information Theory*, Vol. IT-21, pp. 250- 256, 1975.
- Hürst, W., & Joachims, T. (2000). Information presentation in the ubiquitous computing era. In Proceedings of the 2000 workshop on New paradigms in information visualization and manipulation (pp. 58-65).
- J. Deller, J. Hansen, J. Proakis (1999). "Discrete-Time Processing of Speech Signals (IEEE Press Classic Reissue)". ISBN: 0780353862.
- Kimura, Shinta (1993). "Voice recognition system having word frequency and intermediate result display features", United States Patent 5257314,
- L. R. Rabiner, S. E. Levinson, A. E. Rosenberg and J. G. Wilpon (1979) Speaker Independent Recognition of Isolated Words Using Clustering Techniques, *IEEE Trans. Acoustics, Speech and Signal Proc.*, Vol. Assp-27, pp. 336-349.
- M. Johnston (2007). The Guide to computing Literature (2007) "The multimodal presentation dashboard". <http://portal.acm.org/citation.cfm?id=1556331>
- Michael Johnston and Srinivas Bangalore (2005). Finite-state multimodal integration and understanding. *Journal of Natural Language Engineering*. 11.2, pp. 159-187, Cambridge University Press.
- Moller, A., & Fitzpatrick, G. (2013). More than meets the eye: The layered nature of attention. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (pp. 287-296).
- Norberg, P. A., Horne, D. R., & Horne, D. A. (2007). The privacy paradox: Personal information disclosure intentions versus behaviors. *Journal of Consumer Affairs*, 41(1), 100-126.
- Nuance (2003). Focus Medical Software by nuance (2003) "History of Speech & Voice Recognition and Transcription Software" http://www.dragon-medical-transcription.com/history_speech_recognition.html
- Rädle, R., Jetter, H. C., & Reiterer, H. (2013). Designing for collaboration in multi-display environments: An empirical investigation. In Proceedings of the 2013 conference on Computer supported cooperative work (pp. 1109-1118).
- Shneiderman, B., & Plaisant, C. (2005). *Designing the user interface: Strategies for effective human-computer interaction*. Pearson Education.

Stephen Cook (2002). Speech Recognition
“HOW TO”.

[http://www.faqs.org/docs/Linux-
HOWTO/Speech](http://www.faqs.org/docs/Linux-HOWTO/Speech)

William Willis (1988). “Speech
recognition: Instead of Typing and
Clicking, Talk and Command” of T H

E Journal (Technological Horizons in
Education), Vol. 25.

Zong, C., Zhao, K., Lv, Z., Liu, Y., & Han,
X. (2020). Voice-assisted visual
presentation remote control system
for mobile applications. *Journal of
Ambient Intelligence and Humanized
Computing*, 11(3), 1107-1119.