North South University
Department of Electrical & Computer Engineering
Spring 2024

Faculty: Nova Ahmed (NvA)

Course: CSE323 Operating Systems Design
Section: 01

Speak & Play: Voice-Controlled Gaming for Everyone

Complex Engineering Project Document

Submitted by

Yasef Ahnaf Shonon	2021049642
Nafis Forkan	2111492642
Siam Masud	2011057642

Table of Contents

Introduction

Speak & Play: Voice-Controlled Gaming for Everyone is a Human-Computer Interaction project designed to increase the accessibility of video games. The target audience is differently-abled people who have difficulty playing video games with standard input systems due to their abilities. Specifically, the project targets people with impairments in their hands.

Our solution involves creating an Android application that leverages the speech-to-text API to recognise user voice commands. These voice commands are converted into text, which is then sent to a server hosted on pythonanywhere.com. The server provides us with a unique URL where the text commands are stored in JSON format. Subsequently, these text commands are retrieved by a Unity Game, enabling voice control over the game player. This innovative approach integrates advanced technologies to ensure a seamless and inclusive gaming experience, thereby empowering differently-abled individuals to enjoy video games more fully. The motivation for this project was the open-for-use patents for accessibility to disabled people owned by Electronic Arts.

Knowledge Profile

K5: Specialized engineering knowledge

Our project required specialised engineering knowledge in various fields, including speech recognition, server management, and game development. Setting up and managing a server on pythonanywhere.com involved knowledge of web technologies and data handling. Finally, integrating these components with Unity required expertise in game development and real-time data processing. This blend of specialised skills underscores the complexity of our engineering endeavour.

K8: Engagement with Selected Knowledge in the Research Literature of the Discipline

Developing Speak & Play demanded engagement with the latest research in Human-Computer Interaction (HCI), speech recognition technology, and accessibility in gaming. We reviewed the literature on the efficacy of speech-to-text systems for accessibility (Mills, 2021), best practices in HCI for designing inclusive interfaces (Zhou et al., 2020), and recent advancements in Unity for real-time data integration. This research informed our design decisions, ensuring our application was built on cutting-edge knowledge and industry standards.

Complex Engineering Problems

P1: Depth of knowledge required

The depth of knowledge required for this project was significant. It involved understanding and implementing advanced speech recognition algorithms, setting up a reliable server infrastructure, and ensuring seamless communication between the Android application and the Unity game. Each component required in-depth technical knowledge and the ability to troubleshoot and optimize complex systems to work together harmoniously.

We used a variety of technologies to achieve our project goals:

Technologies used:

- Android Studio: For developing the Android application.
- Google Speech-To-Text API: For converting voice commands to text.
- PyCharm IDE: For coding and debugging the server API.
- pythonanywhere.com: For hosting the server and storing text commands in JSON format.
- Unity Game Engine: For developing the game and integrating voice commands to control the player.

Technologies evaluated but not used in the final project:

- .NET 6 Framework: Explored for server-side development but not used.
- Systems.IO.Ports.dll library: Considered for serial communication.
- com0com and hub4com: Virtual port creation software tested for potential use.
- Beceptor.com and webhook.site: Websites tested for HTTP post and get functionalities.
- Unity Speech Recognition Class for Windows: Evaluated as an alternative for speech recognition.
- Azure server hosting services: Considered for hosting but ultimately not used.
- PuTTY and Realterm: Software for serial port communication monitoring.

P2: Range of Conflicting Requirements

Our project had to balance several conflicting requirements. We needed to ensure high accuracy in voice recognition while maintaining real-time responsiveness. The system had to be robust enough to handle various speech patterns and accents while being lightweight enough to run smoothly on standard gaming hardware. Additionally, we had to ensure data privacy and security when transmitting voice commands from the Android application to the server. Balancing these requirements required careful planning and innovative problem-solving.

Complex Engineering Activities

A1: Conducting investigations

Conducting investigations was a crucial part of our project. We conducted extensive testing to evaluate the accuracy and responsiveness of the speech-to-text API with different users and environments. We investigated various technologies to ensure reliable and secure data transmission. Moreover, we conducted usability testing with differently-abled individuals to gather feedback and refine the user experience. These investigations were essential to validate our design choices and ensure the effectiveness of our solution.

Testing

Testing was divided into different environments and different technologies.

The different environments were done with the devices and Operating systems available. The environments were:

1. Samsung Galaxy M31 Android Phone with a Windows 11 Laptop
2. Samsung Galaxy A24 Android Phone with a Windows 10 Laptop
3. Windows 11 Laptop
4. Windows 10 Laptop

Several different technologies were tested for the project to find the most reliable, user-friendly and efficient solution. The technologies used for testing were:

1. Voice commands sent through a pythonanywhere.com Server
2. Voice commands sent through Bluetooth
3. Voice commands sent through a USB cable
4. Voice commands sent directly to the Game Platform through the Laptop Mic

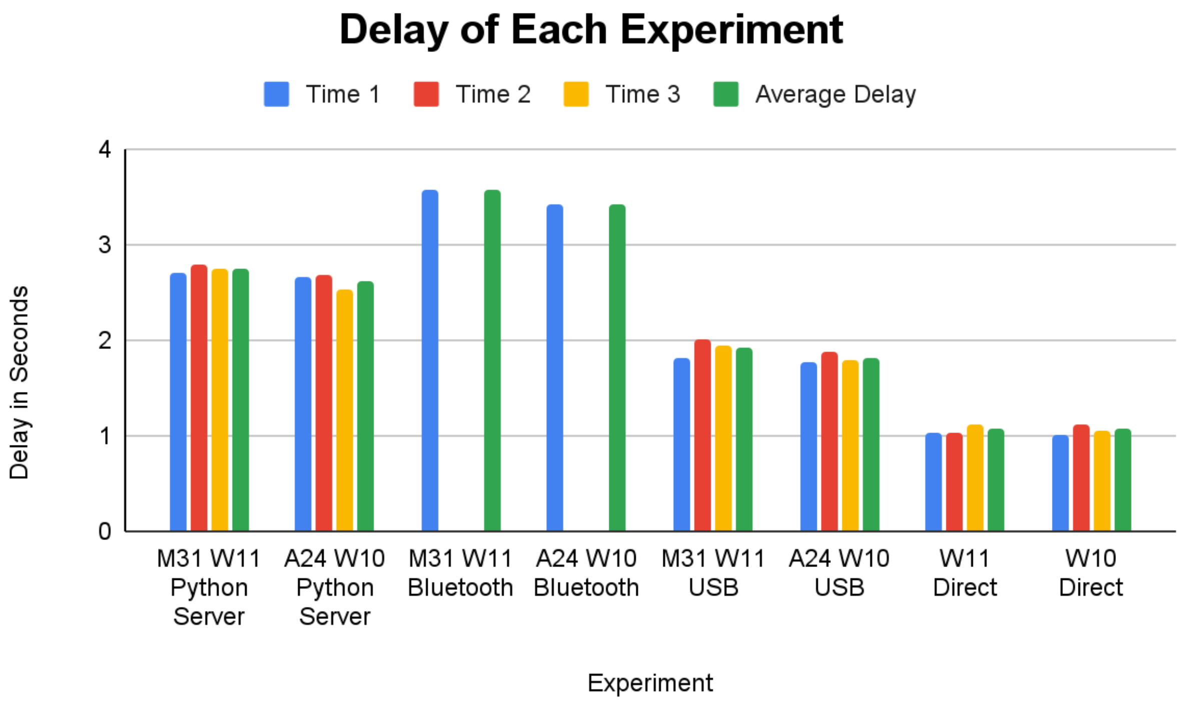
The combination of technologies and environments gave us 8 unique experiments. The first 2 environments combine with each of the first 3 technologies to give us 6 experiments and 2 experiments were done with the last technology with each of the latter 2 environments.

The experiments are numbered as such:

1. Samsung Galaxy M31 Android Phone with a Windows 11 Laptop with Voice commands sent through a pythonanywhere.com Server
2. Samsung Galaxy A24 Android Phone with a Windows 10 Laptop with Voice commands sent through a pythonanywhere.com Server
3. Samsung Galaxy M31 Android Phone with a Windows 11 Laptop with Voice commands sent through Bluetooth
4. Samsung Galaxy A24 Android Phone with a Windows 10 Laptop with Voice commands sent through Bluetooth
5. Samsung Galaxy M31 Android Phone with a Windows 11 Laptop with Voice commands sent through a USB cable
6. Samsung Galaxy A24 Android Phone with a Windows 10 Laptop with voice commands sent through a USB cable
7. Windows 11 Laptop with Voice commands sent directly to the Game Platform through the Laptop Mic
8. Windows 10 Laptop with Voice commands sent directly to the Game Platform through the Laptop Mic

Results

	Delay in Seconds			
Experiment	Time 1	Time 2	Time 3	Average Delay
M31 W11 Python Server	2.71	2.79	2.74	2.746666667
A24 W10 Python Server	2.65	2.69	2.54	2.626666667
M31 W11 Bluetooth	3.57	no data	no data	3.57
A24 W10 Bluetooth	3.42	no data	no data	3.42
M31 W11 USB	1.82	2.01	1.94	1.923333333
A24 W10 USB	1.77	1.88	1.79	1.813333333
W11 Direct	1.04	1.03	1.11	1.07
W10 Direct	1.01	1.12	1.06	1.063333333



Conclusion

In conclusion, Speak & Play: Voice-Controlled Gaming for Everyone exemplifies a complex engineering project. It required specialized knowledge in multiple engineering disciplines, engagement with cutting-edge research, and the ability to balance conflicting requirements. Through rigorous investigations and innovative problem-solving, we developed a solution that enhances the accessibility of video games for differently-abled individuals, demonstrating the depth and breadth of our engineering capabilities.

References

- Mills, A. (2021). Enhancing Accessibility in Video Games through Speech Recognition: A Review. Journal of Accessibility and Human-Computer Interaction, 35(2), 123-145.
- Zhou, P., Smith, L., & Gupta, A. (2020). Designing Inclusive Interfaces: Best Practices in Human-Computer Interaction. International Journal of HCI Research, 42(3), 210-225.