

Voice Controlled Games – The approach and challenges of implementing speech recognition and voice control in games

Dominik Strzałko

Adam Mickiewicz University in Poznań

Email: dominikstrzalko@gmail.com

Abstract—The subject of voice controlled games is quite underrated and exciting. Even though there are not that many papers focusing on that particular subject, we can find many papers describing both, concept of voice recognition, and controlling the game using natural language. The most illustrative example of usage of that kind of control are games in VR/AR/MR and games for people with disabilities. Moreover, almost every game can benefit from voice commands e.g. for controlling the user interface or units in strategy game. Therefore, voice controlled games are an interesting and innovative concept for game designers and developers.

I. INTRODUCTION

THE SUBJECT of voice controlled games is quite underrated and exciting. Even though there are not that many papers focusing on that particular subject, we can find many papers describing both, concept of voice recognition, and controlling the game using natural language (e.g. a popular game “Façade” [1]). One might argue that speech control in games are even somehow part of the future of gaming industry. The most illustrative example of usage of that kind of control are games in Virtual Reality (VR)/Augmented Reality (AR)/Mixed Reality (MR) (that are getting more and more popular [2]) and games for people with disabilities. Moreover, almost every game can benefit from voice commands e.g. for controlling the user interface or units in strategy game. Therefore, voice controlled games are an interesting and innovative concept for game designers and developers. This paper will focus on different aspects of this large and broad topic.

II. THE PAPER

The whole paper will be built from 4 distinct chapters.

A. Beginning

The first part to the paper will outline the short history of voice commands in the game industry and show how this feature was used across time. From the first games that used this technology in some way (like “Command: Aces of the Deep” [3]), to the most recent productions on the market (like games built with “Just AI” tools [4]).

B. Components

The next part is going to focus on describing the main components of the speech recognition: Natural Language Understanding (NLU) and Speech-to-Text (STT). It will also

present various tools and libraries that are commonly used to implement those modules into applications. As for NLU, the description will concentrate on LUIS by Microsoft Azure, Dialogflow by Google and author’s own module created with the help of spaCy and Keras (this solution is still in the conceptual phase). As For the Speech-to-Text module, the paper will focus only on already created services like STT by Microsoft Azure, IBM and Google. It is important to mention that all selected technologies are compatible and working well with different commercial game engines. Regarding game engines, the paper will also shortly describe those engines which will later be used to implement the modules.

C. Implementation

The plans for the implementation process will take up the succeeding and largest part of this paper. A step by step explanation will be given to the process of implementing the different tools/technologies into those game engines and also comparing the accuracy and other metrics of the Speech-to-Text and Natural Language Understanding tools. It will be very important to determine the most accurate and easy to implement tools, also for the sake of future game developers that will be interested in implementing those tools into their games. The most ambitious version of the implementation plans involve the development of a plugin that will be available under an open-source license and will be easily integrated to the games in the engine of author’s choice (Where unity engine is the most probable option).

D. Demo Games

The final part will feature author’s own demo games that will be created especially for the project. The process of building them will be present in the previous chapter but the last part of the paper will sum up the technologies used for each game and describe how the voice control influences them. The planned demo games include with high probability: a VR Game, an AR Game, a game dedicated for people with disabilities, a normal game with both standard control system and voice control system, and potentially, a game for visually impaired people (idea inspired by the game “Ptolem’s Singing Catacombs” by Narayana Walters, aka Miziziziz [5]), were player uses only his voice and hearing to navigate and progress through the game. The genres of the games will vary, to

show that developer can create not enchant a simple puzzle game with voice commands, but also a more complicated and interesting products for different demographics.

III. CONCLUSION

To sum up, this paper is aimed at promoting the voice control among other game developers and designers, and helping them create more immersive games for everyone. This involves the idea of helping people with disabilities enjoy one of the largest source of entertainment, which are: video games.

REFERENCES

- [1] Michael Mateas and Andrew Stern, "Façade", <https://www.playablstudios.com/facade>
- [2] Virtual Reality (VR) - statistics & facts, <https://www.statista.com/topics/2532/virtual-reality-vr>
- [3] Teemu Kiiski, "Voice Games: The History of Voice Interaction in Digital Games"
- [4] Just-AI, <https://just-ai.com/en/voice-games>
- [5] Narayana Walters, "Ptolem's Singing Catacombs", <https://nartier.itch.io/ptolems-singing-catacombs>