

Using Voice Input to Control and Interact With a Narrative Video Game

Andrei Copaceanu
School of Computing and
Mathematical Sciences
University of Greenwich
London, UK
a.copaceanu@gre.ac.uk

Jonathan Weinel
School of Computing and
Mathematical Sciences
University of Greenwich
London, UK
j.r.weinel@greenwich.ac.uk

Stuart Cunningham
School of Computer and
Engineering Sciences
University of Chester
Chester, UK
s.cunningham@chester.ac.uk

1. INTRODUCTION

With the advancement of artificial intelligence (AI) over recent years, especially the breakthrough in technology that OpenAI achieved with the natural language generative model of ChatGPT (Roumeliotis & Tselikas 2023), virtual assistants and voice interactive devices such as Amazon's Alexa or Apple's Siri, have become popular with the general public. This is due to their ease of use, accessibility, and ability to be used without physical interaction (Lopatovska et al. 2018).

When it comes to the video games industry, there have been attempts to implement voice input as a core mechanic, with various levels of success. Ultimately, voice input has been mostly used as a separate mechanic or as an alternative to traditional input methods (Allison et al. 2017).

This project will investigate different methods of using voice input to control and interact with a narrative video game. The research will analyse which method is most effective in facilitating player control of the game and identify challenges related to implementation. This paper also includes a work-in-progress demonstration of a voice-activated game made in Unreal Engine.

2. BACKGROUND

2.1 Voice interaction in video games

Throughout video game history, with each technological breakthrough, there has been a rise in interest from developers to produce games that use voice interaction (Allison et al. 2018).

One major contributor in this area is Nintendo, which has been at the forefront of innovation when it comes to voice interaction. For example, the Famicom had a microphone incorporated in the

controller as an on-off switch for certain interactions such as attempting to negotiate a lower price for items with the shopkeeper in *Kid Icarus* (Nintendo R&D1 1986). Later, the Nintendo 64 had a voice recognition add-on that introduced and popularised the virtual pets with *Hey You, Pikachu!* (Ambrella 1998) by using keywords as voice input. Along similar lines, the Nintendo DS also successfully popularised voice interaction with games such as *Nintendogs* (Nintendo EAD 2005), that took the same concepts for virtual pets from earlier games, and improved on them, increasing the number of interactions, improving the voice recognition and making it more accessible by being implemented on a portable device (Kiiski 2020).

With the increase in popularity of voice chats in video games, a new wave of voice input games has arisen in North America. In Western media, most games that made use of voice input were strategy games in which the player could give orders to multiple units. The most well-received voice input system was present in *Tom Clancy's EndWar* (Ubisoft Shanghai 2008), a real-time strategy game that was released for consoles and features a Who-What-Where pattern to control the units. The system was praised for its elegant and reliable solution for the ergonomic constraints of the console format (Allison et al. 2017).

More recently, following the improvement of AI and language processing, *Phasmophobia* (Kinetic Games 2020), is a four-player co-op virtual reality (VR) game, which showed what can be done when voice interaction is incorporated in a relevant context. In this case, players are tasked to investigate haunted locations and determine which type of ghost is present in that location. The players are encouraged to utilise microphones to communicate between themselves, but also to interact with the ghost. The interaction with the ghost is done through recognition of key phrases.

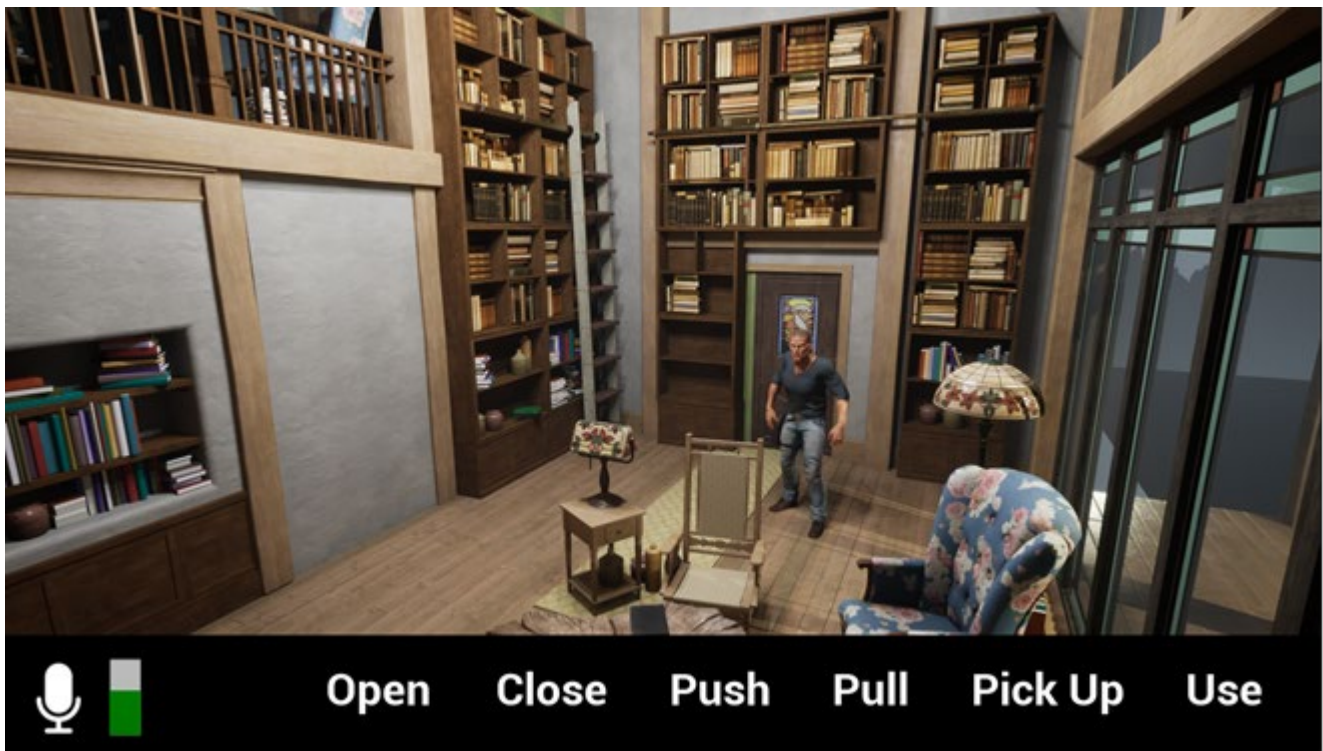


Figure 1: Screenshot from the prototype showing the voice input system listening for the keywords on display

The players can use items such as a 'spirit box', which allows them to put questions such as "Is this your home?", in the hope of a response that would help the players identify the type of the ghost. The ghost would also listen constantly to the player's communications, and would change its behaviour based on words such as "Scared" or "Hide". This not only makes the experience more immersive, but it introduces gameplay elements that otherwise wouldn't be achievable (Stanton 2020).

2.2 Narrative video games

Narrative video games feature their story as a core element. Within this broad umbrella term, the genres that put the most accent on the story, characters, and word-building are adventure games and their sub-genres, such as visual novels (Mallon & Webb 2005).

Interaction with narrative video games is often achieved through natural language, with early examples such as *King's Quest I* (Sierra On-Line, 1984) requiring the player to type text: a verb representing the action and a noun representing the target, to interact with the game.

3. A VOICE ACTIVATED GAME PROTOTYPE

The goal of this project was to investigate and analyse different methods of implementing existing

technologies for voice recognition control in a game created in Unreal Engine 5.

The game is set in a small room where the player must escape by investigating, finding clues, and solving puzzles. Interaction is solely through voice input, drawing from past adventure games like *King's Quest I* (Sierra On-Line 1984), with verbs such as "Open" or "Pick up" to trigger actions. Inspired by *Phasmophobia* (Kinetic Games 2020), players can also engage with the main character by asking questions, adding depth and personality.

Two voice input methods are employed:

- The first system contains a library of predefined keywords that represent actions or objects. The game listens to the player's voice input using voice processing and identifies the keywords to trigger actions. This system ensures reliable responses but requires extensive synonym registration to avoid player frustration.
- The second method transcribes the player's input and feeds the text to an AI for natural conversation. While this promotes fluid interaction without specific word reliance, AI responses may be less dependable, necessitating constraints and safeguards against revealing answers.

Early tests showed that the first method proved to be more reliable and efficient for actions such as investigating the area or giving orders to the character, while the second method was more engaging when interacting with the character, creating more diversity in replies and making the conversation more believable. As both systems proved to work better where the other one had flaws, a combination of the two systems was implemented in the final prototype.

4. SUMMARY

The outcome of this project is a prototype that utilises existing technologies and applies them in a narrative video game, making use of the natural language interaction that corresponds with this type of video game. The project has its flaws, especially when it comes to recognising different player accents, but it still is in its early prototyping stage.

For future development, more accent should be put on accessibility, by increasing the number of accents that the language processing can recognise, on the user interface as right now is not very aesthetic and gives away too much information, a more diegetic user interface would be ideal, and more improvements on the voice processing system as sometimes it does misunderstand the player and activates the wrong action.

5. REFERENCES

Allison, F., Carter, M. and Gibbs, M. (2017) 'Word play: A history of voice interaction in Digital Games', *Games and Culture*, 15(2), pp. 91–113. doi:10.1177/1555412017746305.

Allison, F., Carter, M., Gibbs, M. and Smith W. (2018) 'Design patterns for voice interaction in games', *Proceedings of the 2018 Annual Symposium on Computer-Human Interaction in Play* [Preprint]. doi:10.1145/3242671.3242712.

Ambrella (1998) *Hey You, Pikachu!* [video game]. Nintendo 64.

Kiiski, T. (2020) *Voice Games: The History of Voice Interaction in Digital Games*. thesis.

Lopatovska, I., Rink, K., Martinez, A. et al. (2018) 'Talk to me: Exploring user interactions with the Amazon Alexa', *Journal of Librarianship and Information Science*, 51(4), pp.984–997. doi:10.1177/0961000618759414.

Mallon, B. and Webb, B. (2005) 'Stand up and take your place', *Computers in Entertainment*, 3(1), pp. 6–6. doi:10.1145/1057270.1057285.

Nintendo EAD (2005) *Nintendogs* [video game]. Nintendo DS.

Nintendo R&D1 (1986) *Kid Icarus* [video game]. Famicom.

Phasmophobia (2020). Kinetic Games.

Roumeliotis, K.I. and Tselikas, N.D. (2023) 'ChatGPT and open-AI models: A preliminary review', *Future Internet*, 15(6), p.192. doi:10.3390/fi15060192.

Sierra On-Line (1984) *King's Quest I*.

Stanton, R. (2020) 'Phasmophobia is the best ghost game ever made', *PC Gamer*, 12 October. Available at: <https://www.pcgamer.com/phantophobia-is-the-best-ghost-game-ever-made/> (Accessed: 9 March 2024).

Ubisoft Shanghai (2008) *Tom Clancy's EndWar* [video game]. Various platforms.