

Generating real-time audio description for live streamed gaming

David Libeau
phdproposal@davidlibeau.fr

Abstract

In films and TV shows a verbal commentary is included in order to help visually impaired people follow the plot, access information on set design and costumes as well as other visual aspects of the production. This verbal commentary is known as Audio Description (AD). Newer media experiences are still to catch up in terms of access services, and live streamed gaming is an example of that. Live streamed gaming is an emerging form of entertainment where video games being played are streamed in real time to thousands of spectators. Visually impaired people are already using text-to-speech software on the text chat of live streamed services; however, this technology does not exist on the video stream yet. This PhD project takes a user-centered approach to the study of the application of audio description to live streamed video games, by analysing its use from both theoretical and technological standpoints.

I. INTRODUCTION

Visually impaired people are often excluded from playing video games as very few have been designed or adapted to allow people with vision impairments to play [1, 2, 3]. Except for text games where blind players can use screen readers, audio games are the only ones fully accessible to blind people [4]. This PhD will focus on live streamed games. Before choosing to play a game, people typically watch someone else playing it in a direct real-life situation or through Let's Play videos, and nowadays increasingly through live streaming. Glas (2015) conceptualises this as *vicarious play* [5]. Live streamed games are part of this field of study. It is a new audiovisual phenomenon that lacks accessibility features for visually impaired people. In traditional media, such as cinema and television, visually impaired audiences have access to audio description which helps them understand the action [6]. But it is in theatre that we can mainly find live AD, [7] an approach that could be applied to gaming experiences. Although there might be similarities between live AD for theatre and its application to game streaming, it is important to note the specific challenges involved in the latter. For example, a live streamed game only exports the video stream of the game but game events are needed to generate AD. Exporting them in real time is possible, although the latency needs to be taken into consideration as the video stream is already delayed by a couple of seconds. The proposed PhD project will explore the differences between a variety of media as well as reflect on potential solutions.

II. RESEARCH PROPOSAL

The main research question is:

To what extent could generated audio description be used to make live streamed games accessible to visually impaired people while ensuring a balance between information and entertainment?

Firstly, the use of both games and live streamed gaming by visually impaired people will be researched in order to determine the features currently available as well as the use of those features by visually impaired gamers [8].

The technological advancement of audio description is perpetual [9]. When looking into accessibility for video games, we can cite Microsoft, who is offering in-game text-to-speech of the community chat in two Xbox games [10]. For live streaming, live AD could be used [7]. The Xbox added AD on their E3 events live streamed on the website Mixer [11]. For live streamed games, as it is user-generated content, having a describer for hours is impossible as it may be costly and pre-recorded AD may be irrelevant as the player defines complex actions in real-time while playing. As a solution, AD could be generated thanks to game events. Exporting the data out of the game will permit a text-to-speech software to generate AD [12]. The important and relevant visual information needs to be delivered at a high quality for the user [13, 14], in order to give them the key information to understand the game session. It is crucial to consider that some information may be already delivered through in-game sound. For example, if the player uses a gun in the game, it will surely produce a recognisable sound: a gunshot. As a result, the AD does not need to describe this action as it is already understandable through an in-game sound. On the contrary, describing a place (like a room and the objects present in it) or a non-playable character (like a monster) may be key as this information is usually only available through graphics. It is also important to notice that different type of games may present different challenges for AD. A theoretical study can be performed in the game design field in order to better understand the uses and the needs of visually impaired people. Consulting visually impaired gamers on what their needs and expectations are is crucial to this research. Nevertheless, AD of game sessions could also be inspired by AD in cinema, TV and theatre as many studies explore key issues of AD in mainstream audiovisual media [13, 15, 9]. For example, AD is often included between dialogue lines [6], live streamed game AD could also be triggered when the streamer is not talking.

A technological study of game data collection and export will be useful for the proposed research project. Nowadays, video games collect data for user analytics [16, 17]. Exporting data for AD could be similar to user analytics data collection. Game analytics software could be asynchronous (export data at the end of the game, for example) or synchronous (export data in real time). Studying how the synchronous software are coded and implemented in games will help build a real time data exporter for AD.

At a CHI 2019 workshop on live streaming [18], I conducted a taxonomy of live streaming integration in games. Sometimes video games include special features which pertain to the live streaming context [19]. Adding a real-time exporter of data for generating AD could be a new feature for games. In order to test this feature, a simple prototype will be created. The video game will send data over the internet to a web server that will distribute it to users. A prototype will be easy to do but implementing this feature into commercial games may require a partnership with a game publisher. However, some games are including tools that let developers modify the game. Then, *modding* could be used to test the technology further in commercial games.

The main goal of the prototype will be to explore solutions for transforming raw data into understandable sentences for the text-to-speech software. These sentences need specifications such as being descriptive, not redundant, and fairly short, among others. Furthermore, as different types of games may present different challenges, it is proposed to focus on only one type of game as a starting point. For example, it will be valuable to make a prototype on a first-person shooter game as

it is one of the most popular game genre nowadays.

A key issue for this PhD will be to merge the audio streams in a pleasant and entertaining way for the user. Having the audio from the live stream, the AD of the game and the text-to-speech of the community chat at the same level may require high levels of concentration for visually impaired gamers, which may result in loss of enjoyment. Different solutions could be explored such as user-control of audio levels, panning (including binaural panning) and others [20, 15]. The user could personalise the level of each audio source and place them around them in a 3d virtual space [21]. As well as high quality AD [22], this will positively impact the sense of presence [23] and, as a result, the user's immersion. In addition, previous studies have shown that text-to-speech is not as engaging as human voiced AD [24]. In order to provide an engaging experience, different solutions will have to be tested with visually impaired gamers.

III. MOTIVATION & PERSONAL STATEMENT

This PhD project will contribute to the increase of accessibility of live streamed games for visually impaired people. I choose to start to research in the live streaming field as this new medium is increasingly popular. There is an average of more than 1 million people constantly watching live streams on Twitch [25]. Including visually impaired gamers in these communities by improving accessibility is only a first step towards increasing gaming accessibility.

The field of live streamed gaming is one I am familiar with as I conducted research on this topic during my Masters studies [19]. This new medium involves numerous technical challenges related to web technologies (real-time and latency, among others). My background in IT and web development as well as my previous research experience makes me an excellent candidate to conduct the research needed to advance the field.

REFERENCES

- [1] D. Archambault, T. Gaudy, K. Miesenberger, S. Natkin, and R. Ossmann. Towards generalised accessibility of computer games. In *Edutainment*, 2008.
- [2] F. C. Harris, E. Folmer, and B. Yuan. Towards generalized accessibility of video games for the visually impaired. 2009.
- [3] B. Yuan and E. Folmer. Blind hero: Enabling guitar hero for the visually impaired. In *Proceedings of the 10th International ACM SIGACCESS Conference on Computers and Accessibility, Assets '08*, pages 169–176, New York, NY, USA, 2008. ACM.
- [4] G. R. White, G. Fitzpatrick, and G. McAllister. Toward accessible 3d virtual environments for the blind and visually impaired. In *Proceedings of the 3rd International Conference on Digital Interactive Media in Entertainment and Arts, DIMEA '08*, pages 134–141, New York, NY, USA, 2008. ACM.
- [5] R. Glas. Vicarious play: Engaging the viewer in let's play videos. *Empedocles: European Journal for the Philosophy of Communication*, 5(1-2):81–86, 2015.
- [6] L. Fryer. An introduction to audio description: A practical guide. *Routledge*, 2016.
- [7] J.-P. Udo and D. I. Fels. Universal design on stage: live audio description for theatrical performances. *Perspectives*, 18(3):189–203, 2010.
- [8] J. Beeston, C. Power, P. A. Cairns, and M. Barlet. Accessible player experiences (apx): The players. In *ICCHP*, 2018.

- [9] A. Walczak. Audio description on smartphones: making cinema accessible for visually impaired audiences. *Universal Access in the Information Society*, 17(4):833–840, Nov 2018.
- [10] Xbox.com. (n.d.). Use game transcription on Xbox One and Windows 10. [online] Available at: <https://beta.support.xbox.com/help/account-profile/accessibility/use-game-chat-transcription> [Accessed 1 Jan. 2020].
- [11] American Council of the Blind. (2019). Microsoft/Xbox Receives ACB's 2019 Achievement Award in Audio Description-Media. [online] Available at: <https://acb.org/microsoftxbox-receives-acb%E2%80%99s-2019-achievement-award-audio-description-media> [Accessed 1 Jan. 2020].
- [12] A. Szarkowska. (2011). Text-to-speech audio description: towards wider availability of AD. *The Journal of Specialised Translation*, 15, 142-162.
- [13] L. Fryer and J. Freeman. Cinematic language and the description of film: keeping ad users in the frame. *Perspectives*, 21(3):412–426, 2013.
- [14] M. D. Naraine, D. I. Fels, and M. Whitfield. Impacts on quality: Enjoyment factors in blind and low vision audience entertainment ratings: A qualitative study. In *PloS one*, 2018.
- [15] Lopez, M., Kearney, G., & Hofstädter, K. (2018). Audio Description in the UK: What works, what doesn't, and understanding the need for personalising access. *British Journal of Visual Impairment*, 36(3), 274–291.
- [16] K. Hullett, N. Nagappan, E. Schuh, and J. Hopson. Empirical analysis of user data in game software development. In *Proceedings of the 2012 ACM-IEEE International Symposium on Empirical Software Engineering and Measurement*, pages 89–98, Sep. 2012.
- [17] G. Wallner, S. Kriglstein, F. Gnadlinger, M. Heiml, and J. Kranzer. Game user telemetry in practice: A case study. In *Proceedings of the 11th Conference on Advances in Computer Entertainment Technology, ACE '14*, pages 45:1–45:4, New York, NY, USA, 2014. ACM.
- [18] R. Robinson, J. Hammer, and K. Isbister. All the world (wide web)'s a stage: A workshop on live streaming. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems, CHI EA '19*, New York, NY, USA, 2019. ACM.
- [19] D. Libeau. Audience participation games: Streamer and viewers' engagement in the audience participation features of choice chamber and dead cells. Master's thesis, 2019.
- [20] K. Drossos, N. Zormpas, G. Giannakopoulos, and A. Floros. Accessible games for blind children, empowered by binaural sound. In *Proceedings of the 8th ACM International Conference on Pervasive Technologies Related to Assistive Environments, PETRA '15*, New York, NY, USA, 2015. ACM.
- [21] F. Ribeiro, D. Florêncio, P. A. Chou, and Z. Zhang. Auditory augmented reality: Object sonification for the visually impaired. In *2012 IEEE 14th International Workshop on Multimedia Signal Processing (MMSP)*, pages 319–324, Sep. 2012.
- [22] A. Walczak and L. Fryer. Vocal delivery of audio description by genre: measuring users' presence. *Perspectives*, 26(1):69–83, 2018.
- [23] T. B. Sheridan. Musings on telepresence and virtual presence. *Presence: Teleoperators and Virtual Environments*, 1(1):120–126, 1992.
- [24] A. Szarkowska and A. Jankowska. Text-to-speech audio description of voiced-over films. a case study of audio described volver in polish. 2012.
- [25] Twitch.tv. (n.d.). Press center. [online] Available at: <https://www.twitch.tv/p/press-center/> [Accessed 1 Jan. 2020].