10-2022

Audio Description in Video Games Research in Progress

Stevie C. Sansalone

Claire M. Culver

Mahadeo A. Sukhai

7. Audio Description in Video Games Research in Progress

Stevie C. F. Sansalone
Ontario Tech University
stevie.sansalone@ontariotechu.ca

Claire M. Culver
Ontario Tech University
claire.culver@ontariotechu.ca

Mahadeo A. Sukhai
Canadian National Institute for the Blind
mahadeo.sukhai@cnib.ca

Abstract

As video games continue to grow in popularity, accessibility is a key concern which developers must consider to ensure the most people possible can enjoy the games they create (Cairns et al., 2019; Nova et al., 2021). With approximately 500,000 blind and partially sighted people in Canada alone, visual accessibility is a central concern of game accessibility. Visual accessibility has developed for decades with one of the most popular and effective methods of this being audio description (AD) (Fryer, 2016). Audio description comes in different styles depending on its use, with standard and extended AD being 2 of the most common types (Canadian National Institute for the Blind, 2019). Despite the success of this option in film and television, AD has not caught on in the game industry (SightlessKombat, 2020). This research looks to investigate AD as a method for visual accessibility in video games with a focus on determining the advantages and disadvantages of both standard and extended AD in this medium.

Keywords: Visual Accessibility Technologies, Accessible Game Development, Audio Description

1. Introduction

Video games are an increasingly popular pastime and serve both as a source of entertainment and a method for connecting with other people through shared interests (Cairns et al., 2019; Nova et al., 2021). The social aspect of games has only grown in recent years with multiplayer centered games such as Fortnite and Apex Legends captivating audiences around the globe (Cairns et al., 2019; PlayerCounter, 2022a; PlayerCounter, 2022b). With this increased interest, it is vital that games are accessible to all communities so that everyone has the option to participate in this hobby and the social connections associated with it (Kulik et al., 2021). To ensure that as many users as possible are able to participate in gaming, accessibility options are necessary. In the case of visual accessibility, audio description (AD) has long been used as an accessibility tool in film and television, but, despite its success in these formats, has not caught on as an option in video games (Fryer, 2016; SightlessKombat, 2020; Bodi et al., 2021). AD is the practice of narrating visually displayed information to ensure that the full meaning, or narrative information, of a scene is conveyed in an accessible manner (Fryer, 2016; Wang et al., 2021; Bodi et al., 2021). In this paper, narrative information will refer to the intended message of a given scene. AD serves to impart this narrative information and provide context for scenes as they occur (Canadian National Institute for the Blind, 2019; SightlessKombat, 2020).

As an interactive medium, games use a variety of methods with varying levels of required user input to convey narrative information. These methods can loosely be split into 3 main categories no input scenes, low input scenes, and high input scenes. No input scenes require no interaction from users and will play regardless of what the player does once triggered. These sequences are independent of player input and will often cause the game to no longer receive input or will be unaffected by user input while they are playing. Cutscenes are the most common example of this type as they are effectively mini movies which are meant to be observed by the player and, once triggered, will play through without user input. Low input scenes require some simple interactions from users and may require occasional input to continue playing once they are triggered. These sequences are dependent on player input, but require such input infrequently. Finally, high input scenes require frequent and complex interactions from the user to continue playing once they are triggered. These sequences are dependent on player input and include gameplay sequences in which the player is required to navigate obstacles, complete puzzles, or complete some other form of challenge to receive narrative information. Given that no input scenes are the most similar to movies and television shows, where audio description has already been successfully implemented, these scenes are a logical place to start for testing audio description as a tool for visual accessibility in video games (SightlessKombat, 2020). The first step for implementing audio description is to create the transcripts which will then be narrated during the scene (Bodi et al., 2021; SightlessKombat, 2020; Wang et al., 2021). Multiple styles exist for writing and implementing audio description transcripts with the 2 main methods being standard audio description and extended audio description (University of South Carolina, 2020). This research seeks to confirm whether AD is an effective tool for visual accessibility in video games, and, if so, what style of AD is most suitable.

2. Types of Audio Description

Audio Description, sometimes referred to as described audio, has been used for many years in the medium of film and television to provide audible descriptions of visual elements for the benefit of blind and partially sighted audiences (Bodi et al., 2021; Fryer, 2016). Ideally, descriptions are provided during pauses in dialogue, but this is not always possible; some scenes may either be too short or too dialogue heavy to fit an audio description that is long enough to adequately convey all important narrative information. The necessary solutions to these issues are to either provide descriptions while dialogue is playing, or to pause a scene in order to provide the necessary information without interrupting the dialogue (3PlayMedia, 2022; University of South Carolina, 2020). These 2 methods are known as standard and extended audio descriptions respectively (3PlayMedia, 2022; University of South Carolina, 2020).

2.1 Standard

For standard AD, scenes are played normally, with an additional audio track providing a description for any elements in the scene that are only portrayed visually (3PlayMedia, 2022). These audio tracks are inserted between lines of dialogue that occur naturally in the scene (3PlayMedia, 2022; University of South Carolina, 2020). There are limits to this when a scene is filled with a lot of dialogue with few spaces to insert an audio description. In these cases, the content covered in the AD must be limited to avoid delivering AD at the same time as dialogue. While some information may be lost, it is necessary to ensure that the AD does not detract from the experience and make users miss important information from dialogue (University of South Carolina, 2020). In these cases, an alternative is to use extended AD to ensure that dialogue is delivered and the AD is still comprehensive (University of South Carolina, 2020).

2.2 Extended

Extended AD functions similarly to standard AD, but implements pausing in the scene to provide more time for AD (University of South Carolina, 2020). In practice, this means that scenes will run as usual until they hit a point where AD is required, but more time is needed for the delivery. When this occurs, the scene will pause so AD can be provided and will automatically resume once the AD is finished (3PlayMedia, 2022). This ensures that users are provided the full range of information in a scene, but does so at the cost of lengthening scenes, and, by extension, increases the overall length of the viewed media. This pausing can disrupt narrative tension and so extended AD is generally only used when required (3PlayMedia, 2022). Cases where standard AD may not suffice include scenes that are heavy with dialogue and offer few pauses as well as scenes that are complex and require more time to describe than is offered in existing pauses. In situations such as these, extended AD is a better option for providing users with the full range of narrative information in a scene (3PlayMedia, 2022; University of South Carolina, 2020).

3. Audio Description in Video Games

3.1 Current Projects

While audio descriptions have been present for television and film for many decades, audio description has only recently begun to gain traction as a tool for visual accessibility in video games (Fryer, 2016; SightlessKombat, 2020). One notable effort to introduce AD to video games is the Transcribing Games Project, a community–driven effort to create AD transcripts for commercially available games (SightlessKombat, 2020). The initiative looks to raise awareness about visual accessibility in video games and create AD transcripts as a proof of concept for the benefit of AD in video games (SightlessKombat, 2020). Some of these transcripts have additionally been implemented on existing cutscenes which have then been posted on YouTube as a demo for what this would look like in game (SightlessKombat, 2020). While the AD transcripts created by members of the initiative are at this time only for internal use, some companies such as Sony have taken an interest in collaborating with the Transcribing Games Project to develop AD transcripts for some of their games (SightlessKombat, 2020). While this is still in early stages, the fact that there are large studios which are interested in pursuing this option is a good sign for its future inclusion as an option for visual accessibility.

3.2 Challenges of Adding AD to Video Games

Unlike film and television which rely solely on pre-recorded video sequences, most games have content that is dynamic and require user interaction. This makes it more difficult to implement AD as developers need to find suitable times during gameplay to play AD transcripts rather than just looking for pauses in dialogue. However, this problem is not present in cutscenes as these sequences do not rely on player input and function the same way as film and television. When applying AD to cutscenes, developers can rely on established techniques of finding pauses in dialogue or implementing pauses when necessary to deliver AD narration (3PlayMedia, 2022; University of South Carolina, 2020). The issue of pausing scenes becomes more complicated in the case of multiplayer games where truly pausing the game would result in pausing the game for all other players. For this reason multiplayer games typically do not give players the ability to pause gameplay which may create issues in the case of extended AD where pauses are necessary (Feuchtner et al., 2016). Additionally, implementing AD requires more file space as additional narration must be recorded and downloaded along with other game files to make this option possible. In games that are especially heavy on cutscenes, the additional space required for AD will become more noticeable which may present a potential barrier to accessing the game for some users.

4. Research Proposal

4.1 Problem Statement

While a variety of options exist for visual accessibility in video games ranging from colour filters to spatial audio, these options are only able to assist some users in receiving narrative information in scenes heavily reliant on visuals (Yuan et al., 2011). Despite the success of AD in films and television as a tool for providing narrative information, AD has not similarly been pursued as a tool for visual accessibility in video games (Bodi et al. 2021; Fryer, 2016; SightlessKombat, 2020). Without AD, narrative information contained in cutscenes is often not effectively communicated to blind and partially sighted users, resulting in the meaning of a scene being lost (Bodi et al., 2021; Fryer, 2016; Wang et al., 2021). This affects both the user's experience of the cutscene they are experiencing as well as future cutscenes given that later scenes will often build upon information presented earlier in the plot, thus creating an increasingly disjoint narrative as more pieces of information are lost in each scene due to lack of description and lack of context from previous scenes.

4.2 Solution

This research aims to fill this gap and strengthen the communication of narrative information in cutscenes by investigating the use of audio description in video games with the goal of developing guidelines for effectively writing audio description transcripts to be implemented in cutscenes. Both standard and extended audio description formats will be tested with a pair of demo scenes intended to both deliver narrative information and get participants emotionally invested in the stories they are experiencing. The primary focus of this research will be on judging the effectiveness of each style of audio description in clearly delivering narrative information and generating an emotional response. Scenes will focus on 2 salient emotional states, sadness and excitement, but the specific content contained in each will evolve over the course of the study as demo scenes are tested with participants to ensure they are effective at eliciting the desired emotional states.

4.3 Research Plan

Research for this project will be split into 2 stages representing the development of the scenes to be analyzed with audio description and the eventual audio description study using these scenes.

4.3.1. Phase 1 Developing and Refining Scenes

To begin, the 2 scenes being used in the study will be written and built using Unity, a video game engine that is commonly used to create novel simulations and experiences for research of this kind. The first scene will be a science-fiction themed action scene with a focus on eliciting excitement in participants, while the second will be a fantasy themed drama scene with a focus on eliciting sadness in participants. Each scene will be designed to use a mixture of visual information, such as silent character actions, and audio information, such as dialogue, to convey narrative information in the same way as cutscenes. Following this, a small study will be conducted to test how well these scenes convey information and elicit emotional responses in sighted audiences when no audio description is used. This study will perform similarly to the eventual audio description study with 10 sighted participants first being given a survey to collect information regarding demographics, experience with gaming, and preference for science fiction and fantasy as genres of fiction. This information will be used to give context to collected data and control for the effects of differing levels of gaming experience, narrative preferences, and individual backgrounds. After this, participants will be split into 2 groups of 5 with each group being shown the scenes in opposite order from each other (i.e. group A will start with scene 1 and move to scene 2 while group B will start with scene 2 and move to scene 1) to control for ordering effects in scene presentation. Once the first scene has concluded for each group, participants in both conditions will be interviewed individually. Interviews will begin with having participants describe the events of the scene to determine how well the scene communicated the intended narrative information and will

then have participants answer a series of questions intended to gauge their emotional response to the scene. This interview process will then be repeated after the second scene is shown and results collected from participants in both groups will be analyzed to determine if scenes were effective in both clearly presenting narrative information and eliciting the desired emotional responses from participants. Depending on the results of this study, scenes may be adjusted before proceeding to the next phase of research.

4.3.2. Phase 2 Audio Description Study

The second phase of the study will begin with the creation of audio description transcripts for the refined scenes. Standard and extended transcripts will be written to narrate the visual information present in each of these scenes and these transcripts will then be implemented. With audio description in place, participants will then be gathered and split into groups for the audio description study. This study will function with 3 groups of 10 then each split into smaller subgroups of 5 in the same way as was done with the previous study. The first group of participants will include blind or partially sighted users who are reliant exclusively or nearly exclusively on audio for the presentation of information, the second group will include partially sighted users who rely on a mixture of audio and visual information, and the third group will include fully sighted users. These groups will then be divided in half so that scenes can be presented with counterbalancing for each group. Participants will then be given a survey for baseline data collection as was done in the scene analysis study before being shown the sequence of scenes assigned to their group.

On each of the tracks, participants will start by viewing a scene with no AD, then that scene will be viewed again with standard AD. Following this, participants will be shown the remaining scene in no AD format and then again with extended AD format. Similar to the previous study, participants will be interviewed individually after each of the viewings with questions focusing on how clear the scene was in providing narrative information, how emotionally engaged participants were, and, in the case of scenes with AD, participant preferences regarding the form of AD they experienced. Additionally, blind, and partially sighted participants from groups 2 and 3 will be asked whether they have any additional comments on what they would like to see implemented to improve their experience of playing video games. While this exploratory inquiry is not the direct focus of this study, it is imperative to listen to members of blind and partially sighted communities when determining how best to improve visual accessibility and answers given may inspire other lines of inquiry. Responses will then be compared across groups to see how audio description influenced the experience of each scene for participants in each of the conditions. Questions related to AD will focus on learning whether participants preferred it to no AD and what they felt the style of AD they were shown contributed. Once all scenes have been viewed and participants have been interviewed about the individual scenes, participants will then be asked questions comparing their experiences of standard and extended AD to determine the advantages and disadvantages of each format. The qualities which will be highlighted in the comparison will be clarity of the AD, emotional engagement, and seamlessness of implementation, or whether the AD was disruptive to the experience. Following these questions, participants will be asked for any further comments they have regarding audio description and how it contributed to their experience of each scene.

5. Conclusion

This research looks to fill the gap of audio description in video games by analyzing AD as it is practiced currently in film and television to determine how it may be applied to video games and whether it is an effective tool for visual accessibility in this medium. If this research determines that AD is well-suited for games, the results will then be used to develop guidelines for effectively applying AD in video game cutscenes. Using the results of the interviews, the benefits and drawbacks of each form of AD will be

determined so that developers can make informed decisions about which style of AD to apply in their cutscenes. This will help developers to ensure they are not over-relying on visually presenting narrative information, and, if they are, that they understand the associated costs of scenes requiring pausing for extended AD to fully describe everything that is occurring. These guidelines will assist in applying AD to video games and getting developers to think critically about how they are choosing to display narrative information throughout the games they are creating. This research also seeks to amplify the voices of blind and partially sighted gamers by drawing attention to the importance of presenting narrative information in visually accessible ways and demonstrating to developers that, without steps such as AD, they will be unable to effectively communicate their narrative ideas to members of their player base.

References

3PlayMedia (2022). *The Ultimate Guide to Audio Description.*

Bodi, A., Fazli, P., Ihorn, S., Siu, Y., Scott, A. T., Narins, L., Kant, Y., Das, A., Yoon, I. (2021). "Automated Video Description for Blind and Low Vision Users," in *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*, CHI EA '21. Yokohama, Japan: Association for Computing Machinery

Cairns, P. A., Power, C., Barlet, M., and Haynes, G. (2019). "Future design of accessibility in games: A design vocabulary," *Int. J. Hum. Comput. Stud.* (131), pp. 64–71.

Canadian National Institute for the Blind (2019). "The case for described video," (available online at https: //cnib.ca/en/news/case-described-video?region=on/; accessed Apr. 18, 2022).

Fryer, L. (2016). *An Introduction to Audio Description*, Englewood Cliffs, NJ: Routledge.

Feuchtner, T., Walter, R., & Muller, J. (2016). "Interruption and Pausing of Public Display Games," in *Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services*, MobileHCI '16. Florence, Italy: Association for Computing Machinery, pp. 306-317.

Kulik, J., Beeston, J., and Cairns, P. (2021). "Grounded Theory of Accessible Game Development," in *The 16th International Conference on the Foundations of Digital Games (FDG) 2021*, FDG'21. Montreal, QC, Canada: Association for Computing Machinery.

Nova, A. N., Sansalone, S. C. F., and Mirza-Babaei, P. (2021). "PathOS+: A New Realm in Expert Evaluation," in *Extended Abstracts of the 2021 Annual Symposium on Computer-Human Interaction in Play*, CHI PLAY '21. Virtual Event, Austria: Association for Computing Machinery, pp. 122–127.

PlayerCounter (2022a). "Apex Legends Live Player Count - How Many People Are Playing Now?," (available online at https://playercounter.com/apex-legends/; accessed Apr. 18, 2022).

PlayerCounter (2022b). "Fortnite Live Player Count - How Many People Are Playing Now?," (available online at https://playercounter.com/fortnite/; accessed Apr. 18, 2022).

SightlessKombat (2020). *Audio Description: Seeing Videogame Narrative Through Sound*. IGDA GASIG.

University of South Carolina (2020). "Writing Audio Description - Digital Accessibility," (available online at https : / / sc . edu / about / offices _ and _ divisions / digital - accessibility / guides _ tutorials / audio _ descriptions/writing-audio-descriptions/index.php#:~:text=Standard/; accessed

Apr. 18, 2022).

Wang, Y., Liang, W., Huang, H., Zhang, Y., Li, D., & Yu, L. (2021). “Toward Automatic Audio Description Generation for Accessible Videos,” in *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, CHI ’21. Yokohama, Japan: Association for Computing Machinery, pp. 1-12.

Yuan, B., Folker, E., and Jr., F. C. H. (2011). “Game accessibility: a survey,” *Universal Access in the Information Society* (10), pp. 81–100.