HIERARCHICAL AUDIO STRUCTURE FOR ONLINE COLLABORATION

A Thesis
by
BRODERICK M. DESANTIS

Submitted to the School of Graduate Studies
at Appalachian State University
in partial fulfillment of the requirements for the degree of
MASTERS OF SCIENCE

AUGUST 2019
Department of Computer Science

HIERARCHICAL AUDIO STRUCTURE FOR ONLINE COLLABORATION

A Thesis
by
BRODERICK M. DESANTIS
August 2019

APPROVED BY:

Raghuveer Mohan, Ph.D.
Chairperson, Thesis Committee

Mohammed Mohebbi, Ph.D.
Member, Thesis Committee

Rahman Tashakkori, Ph.D.
Member, Thesis Committee

Rahman Tashakkori, Ph.D.
Chairperson, Department of Computer Science

Mike McKenzie, Ph.D.
Dean, Cratis D. Williams School of Graduate Studies

Abstract

HIERARCHICAL AUDIO STRUCTURE FOR ONLINE COLLABORATION

Broderick DeSantis
B.A., Appalachian State University

Chairperson: Raghuveer Mohan, Ph.D.

As online communication is increasingly used for collaborative purposes, it is important to rethink some of the audio and text-based chat environments currently in use today. Popular online video conferencing tools like Zoom and Skype, and text-based tools like Discord offer a sleuth of features and allow users to interact and exchange ideas freely. However, they may not be particularly well suited for certain types of tasks, namely a virtual classroom with hierarchical breakrooms for focused discussions and other forms of interaction in virtual spaces. For example, a teacher who wishes to teach a course entirely in a virtual classroom can give a problem for group discussion. The students may need to form hierarchically smaller groups to discuss their solutions. This increases participation among students and allows them to delegate responsibilities more easily. In this thesis, we develop a hierarchical-based audio and text system that is conducive for the aforementioned tasks.

Our audio structure is built within a virtual environment and implemented using the Unity game engine and the Dissonance Voice Chat package. Users can navigate in the virtual space as avatars and have access to increasingly private audio and text chat channels called audio regions. Users can communicate exclusively on these audio regions but may hear

communication at dimmed volumes on parent channels. This allows them to hear announcements from public channels while still retaining focus of discussion on private channels.

We performed a comparative analysis of our audio structure with popular online communication tools like Zoom, Skype and Discord on two types of activities or games -- the Pirates game and the Post Office of the Future experiment. Each of these games provide different qualities to a group discussion ([1], [2]) and are based on similar experiments conducted by Lober et al. [3]. The mediums were compared based on enjoyment, usability and feasibility, and data from participants were collected through a survey and analyzed. We performed a power analysis, correlation analysis and descriptive analysis using ANOVAs and t-tests. Due to low statistical power, we could not detect significant differences between our audio structure and other online tools. We, however, hope that this thesis contains the experimental methodology and data analysis for a large-scale study in comparing various online communication tools.

Acknowledgments

To Dr. Patricia Johann. In any journey, it is important to start at the beginning, and to lay a solid foundation for what is to come. Thank you for giving me that foundation, it serves me well every single day.

To Dr. Mohammed Mohebbi, thank you for your assistance serving on my thesis committee, and for taking the time to offer your insight and wisdom in its refinement.

To Dr. Rahman Tashakkori, my graduate advisor. A single conversation one sunny afternoon in your office was all it took to change the course of my life forever. I cannot thank you enough for suggesting the Master's track to me. It has been challenging, and worth it 100% of the trip. I cannot thank you enough for your continual belief and support.

To Dr. Raghuveer Mohan, you have been a true friend and mentor to me for the duration of my time at Appalachian State University. Your courses challenged me, our friendship has been invaluable, and your tutelage has changed the way that I approach my very profession. No matter where life takes you, the students you influence are blessed to have you.

To Dr. Twila Wingrove, your assistance with the data analysis was invaluable and served as the backbone on which this thesis could be supported. Thank you for all the work you have done.

To Professor Larry Bridges, thank you for providing your students to be participants in this thesis. Their contribution was invaluable in this project's completion.

To my parents Carl and Debra, thank you for your love and support. Not a day goes by that I am not grateful for all you have given and all you have done. I work every day to return that kindness and will continue to do so. This is something we can celebrate together.

To my Grandfather, you have served as a role model for me for as long as I can remember. The lessons you have taught I will carry with me all my life. Thank you for your love and support when pursuing my academic career.

To Lori Dean, this project started as a joint project under your guidance and direction. This Thesis was only possible due to your influence and direction. Thank you for the many conversations and instruction that helped my thesis take form.

Dedication

This thesis is dedicated to my Grandmother, Barbara DeSantis. She supported my academic endeavors when I did not support my own, and encouraged me to never stop moving forward, no matter how incremental. A truly incredible person and a brilliant academic, she understood the value of a strong education and did her best to impart that knowledge to her family. None of this would have been possible without her influence. I love you and wish you could be here to see it.

Table of Contents

Chapter 1

Introduction

In this thesis, we present an alternative form of audio technology to better address some of the current issues pressing group communication in online collaborative environments. Presently, online audio environments have a difficult time preserving the flow of conversation as group sizes expand. This can be caused by lack of visual cues, inability to communicate after splitting into smaller break rooms, or difficulty in directing focus. Using the Theory of Media Richness [1] and the Theory of Media Synchronicity [2], we evaluate the structures of various audio technologies and propose a new hierarchical channel-based audio structure that provides a richer audio experience for online collaboration.

1.1 Media Richness

According to the Media richness theory [1], a medium's richness is the ability of a media to enable users to communicate and convey information effectively. Media that can overcome ambiguity and direct clarity are said to be more rich as they are better at conveying information. This means that the most productive environment is one in which sensory cues are heavily backed by rapid feedback and clearly identifiable focus points. The highest tier in the media richness theory is face-to-face interaction due to the immediate feedback and clear focus on an individual. Many current Voice Over IP (VoIP) clients such as Discord place all users in a single chat room and have little or no visual cues to help direct the flow of conversation. Other conferencing environments like Zoom provide fully functioning audio-visual communication but do not allow for an effective means of conversation overlap without distraction, like an intercom form of broadcasting. There are also stand-alone products like YouTubeTV, which dims audio slightly while browsing. This allows for increased focus while browsing without losing information of the media stream. All these tools however, were not designed with the intention of online collaborative environments.

1.2 The Audio Structure and UNITY

The proposed audio structure uses a series of audio channels and audio regions within the UNITY game engine to help divide group conversations into easily joinable breakrooms. This allows concurrent conversation with as little information loss as possible. That is, we allow for conversation overlap at reduced volumes based on the hierarchy of the user within the group.

Photon is a UNITY based development company that focuses on developing multiplayer assets and hosting consumers and developers using its cloud hosting service the PhotonCloud. Photon offers a VoIP plus text solution called Photon Voice that exhibits the same audio concepts discussed in this thesis but was not discovered until much later in development.

The highest level in the hierarchy is the public channel – an audio channel that all users subscribe to within the virtual lobby. The second component to the structure is the group layer. The group layer consists of several disjointed breakrooms with a smaller subset of subscribers.

Finally, the private layer is reserved for communication between two members and holds authority over the other two layers. The volume of the public channel is dimmed when listening from within a breakroom on the group or private level. This is to allow increased focus within the smaller breakrooms that still overhear important conversations from higher levels in the hierarchy, such as deadline announcements.

Figure 1.1 is a representation of the proposed audio structure, upon loading into the virtual classroom, users are placed into an overarching public channel, as seen in light blue. From the public channel, users have the option of breaking off into smaller group channels. Conversations contained by other groups hear what is being said in all parent classes, but parents cannot hear their children. This means that all private groups contained by the public chat are exclusive – they cannot hear one another, they cannot hear the conversations within groups one to three but can hear the public channel.

We chose UNITY for its simplicity of interface and low barrier to entry on learning its environment. Used primarily as a virtual game engine, UNITY provided an ideal environment to develop a virtual classroom that allows for both audio and visual focus cues. This makes it a competitive option against other audio-visual conferencing environments. The ability to jump in and out of smaller private breakrooms allows for the larger group conversation to retain focus without disruption.

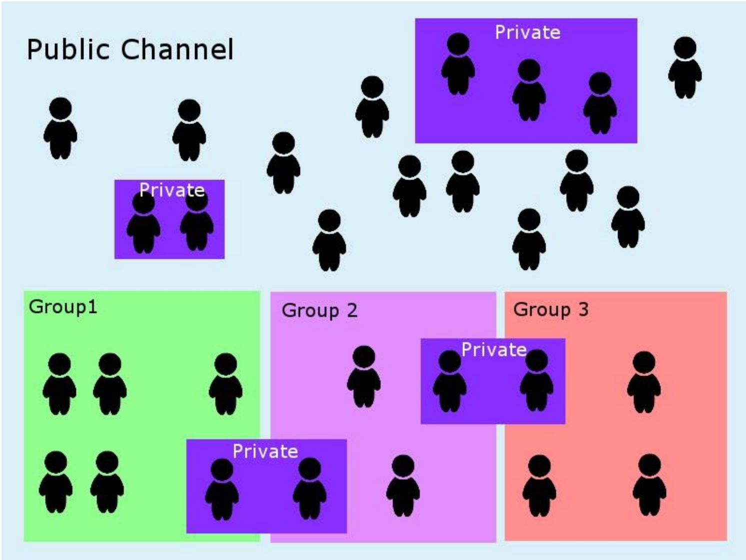

Figure 1.1: The Hierarchy of Audio Channels

UNITY also allows for individuals to share information openly that may not be urgent to the larger group but necessary to share between smaller group members. Using visual cues in UNITY such as a raised hand or a virtual exclamation point could further help facilitate conversation flow without direct interruption, and the audio channels are ideal for splitting into smaller breakrooms and returning into larger groups in seamless fashion. Using this audio structure serves to circumvent several of the issues identified by the theories of Media Richness [1] and Media Synchronicity [2]; the effect of an increasing group size discouraging personal contribution is one such issue.

1.3 Results

This thesis makes the following contributions.

1. We implemented a hierarchical-based audio structure that will allow to more effectively communicate in online collaborative environments. We implemented using

the UNITY game engine and the Dissonance Voice Chat system. Online users can have access to several audio regions in the virtual space, a text chat and private audio channels.

2. We compared our audio structure with other popular online collaborative environments like Zoom, Skype and Discord. We used two activities or games, each providing different scenarios to compare the mediums. Using a survey on eight groups of four, we analyzed based on three objectives – enjoyment, usability and feasibility.
3. Our audio structure performs best in environments with a low degree of uncertainty and a high degree of equivocality. Our audio structure performs reasonably well in environments with a high degree of uncertainty and a low degree of equivocality. Although our experiments have low statistical power, these experiments can be easily scaled to get statistically significant results.

1.4 Roadmap

In Chapter 2 we will cover the significant literature that serves as the backbone of this thesis. We will start by identifying the Theory of Media Richness and Media Synchronicity and follow by supporting their place in this thesis and how they helped to bring about the proposed audio structure. Chapter 3 will outline the architecture of the proposed audio structure, breaking down and explaining at length how it was constructed and what functions it serves within the UNITY game environment. Chapter 4 is a comparative analysis that compares the proposed audio structure to other modern and current VoIP solutions such as Skype, Discord, and Zoom. We describe the methodology used in our experiments and the results obtained. We conclude in Chapter 5 with a summary of the thesis and identify a niche market where the audio structure will excel.

Chapter 2

Literature Review

2.1 Organizational Information Requirements

Introduced in 1986, the Theory of Media Richness [1] discussed the ideals of uncertainty and equivocality in an organizational environment. They proposed that both equivocality and uncertainty, branches of ambiguity were present in all social problem-solving settings. Uncertainty is defined as the absence of information. Equivocality is a measure of a group's ability to use personal experiences with other collaborator's knowledge to draw together a unified conclusion. In settings of high equivocality, an increase of information may not help to better consolidate an answer. For example, during a round table discussion or debate, a clear-cut single answer may not be identifiable, or even possible to attain. In settings of high uncertainty, a group consensus may already have been acquired but external information is required to further develop an answer. These two forms of ambiguity identify two issues encountered in any social collaborative environment.

A larger pool of information is not necessarily beneficial in situations involving equivocality. More information can in many ways be worse as it adds to the amount of information needed to be processed. For example, ideally all forms of language cues, audio and visual should be present when there is a large amount of information or strongly opposing ideals and practices. In a virtual environment, visual cues and rapid feedback of those cues is often not present. Even in online video conferencing environments like Zoom, there will always be a slight delay between responses, due to the network travel time and bandwidth between end users. Using breakrooms to go from a larger group to a smaller group removes a large portion of the information flow, allowing smaller groups to work on removing ambiguity and forming a consolidated answer more effectively. To further hone the focus of information, breakrooms will often dim surrounding conversation volume, highlighting the breakroom members voices and causing them to stand out. This will encourage group collaboration without missing out on important information that could be conveyed in the public channel.

A Medium's richness is the ability of a media to enable users to communicate and convey information effectively. Mediums high in richness such as face-to-face interaction, have audio and visual cues, including tone, atmosphere, posture and gesture. Mediums of low richness, such as an email, have a potentially low response time, and neither audio nor visual cues to stimulate understanding. Daft et al.[1] proposed that mediums of high richness are ideal for environments of high equivocality such as group discussions, where audio visual cues are present to help direct the flow of conversation and assist in alerting members to the present atmosphere of the discussion. Also, mediums of low richness are ideal for situations of low uncertainty, where it is easy to clearly communicate non-ambiguous answers. Text Chat, email, and even telephone calls sit comfortably on the lower end of the richness spectrum and remain ideal for communicating previously established ideas or serve as a starting point for a group project.

Using the Theory of Media Richness [1] and the Theory of Media Synchronicity [2] as guidelines, we propose an audio structure which allows group members to break off into multiple smaller groups, also called breakrooms, when in environments of high equivocality. This allows for more clearly defined focus and clarity between group members. The breakroom structure facilitates an ideal environment for problems involving equivocality, where outside information is not necessarily helpful, and group consensus is the current focus. Upon entering the breakroom, surrounding volume begins to dim and creates focus on the members, developing a productive work space without drowning out surrounding public announcements or information that may still be critical to the conversation.

2.2 Rethinking Media Richness: Towards a Theory of Media Synchronicity

Developed not long after the Theory of Media Richness [1], the Theory of Media Synchronicity [2] states conveyance and convergence occur in every group communication process. Conveyance is the distribution of information and the following consolidation of its meaning between group members. In conveyance, multiple viewpoints may be present and the group need not decide on one meaning for all members, such as a round table discussion or debate. Media of low synchronicity are ideal for tasks that require conveyance. Convergence follows up conveyance and expresses the single developed meaning of information between a group. This is good for mediums of high synchronicity, as conversation and ambiguity are deliberated and filtered out through the process of Convergence.

Like the Theory of Media Richness, the Theory of Media Synchronicity proposes that each form of communication medium has a set of capabilities that, when assigned to a task appropriate for the given media significantly boosts its performance. This set of capabilities can be judged on five criteria.

- Immediacy of Feedback
- Symbol Variety
- Parallelism
- Rehearsability
- Reprocessability

Immediacy of feedback is the ability of a receiver to respond quickly to communication they receive. Symbol variety is the number of ways that information can be conveyed. It is further broken down into four categories – number of formats in which information may be conveyed, the verbal and nonverbal communication, cost of composition, and social presence. Parallelism is the number of conversations that can effectively be simultaneously transmitted across a medium. Rehearsability is the ability to edit a message before transmission, allowing for the meaning to be more clearly expressed. Finally, reprocessability is the ability for a transmission to be reexamined and altered after transmission to allow for development of its meaning.

Dennis and Valacich [2] evaluate the Theory of Media Richness, stating that due to these 5 criteria, no single medium could be identified as the richest since each has varying strengths and weaknesses. The most important factor is that of context, since it allows for the most appropriate media to be attached to the task, making it the most rich for that situation.
Dennis and Valacich touch on the idea of unnecessary present symbol variety. Their example uses the analogy of a Philips head screwdriver being present when a flat head is needed, explaining the necessity of appropriate information suited to each task. If too much information is provided, specifically of the wrong type, then the medium can become cluttered and the extra information can become more of a hindrance than help.

Parallelism becomes a hindrance in groups with many participants. This is because matters of convergence require time to identify each member's stance on a topic, so group with more members require a longer period of time to convey those views.

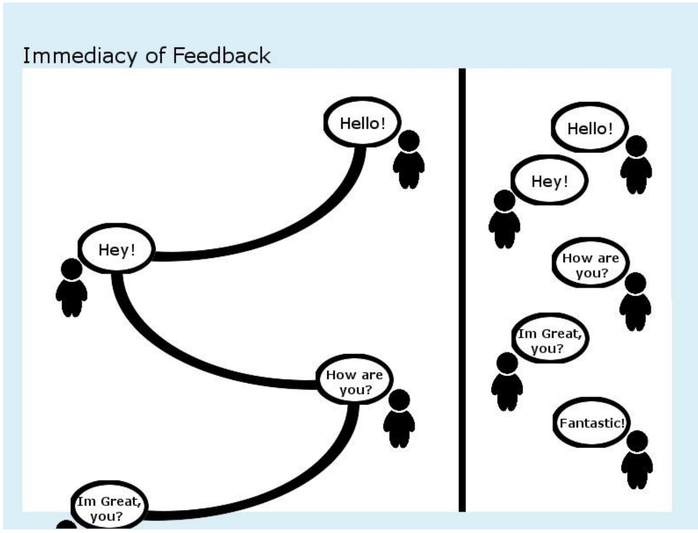

Figure 2.1: Immediacy of feedback is the ability of a receiver to respond quickly to communication they receive.

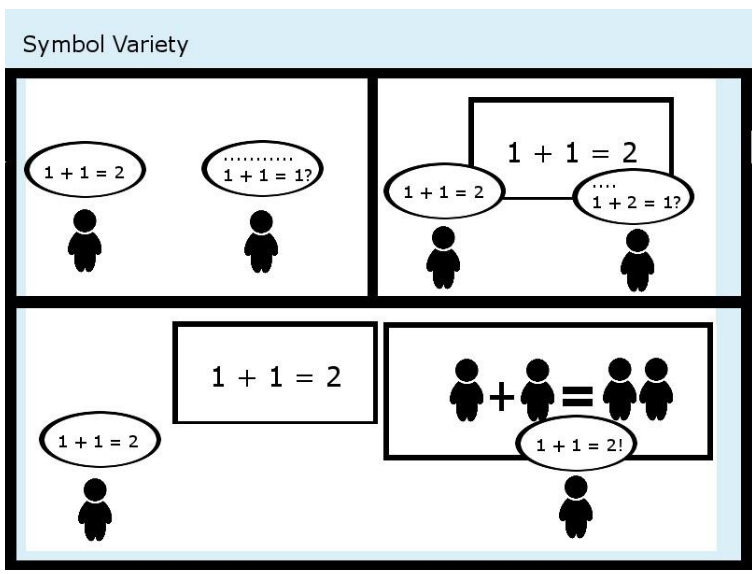

Figure 2.2: Symbol Variety is the number of ways that information can be conveyed

By using breakrooms, larger groups can more quickly and simultaneously convey information between members, and upon reconvening more effectively distribute those ideas to the group at large. By using breakrooms to modularize activities, it becomes easier to build larger group sizes without losing viewpoints of individual members, which helps to mitigate issues caused by parallelism.

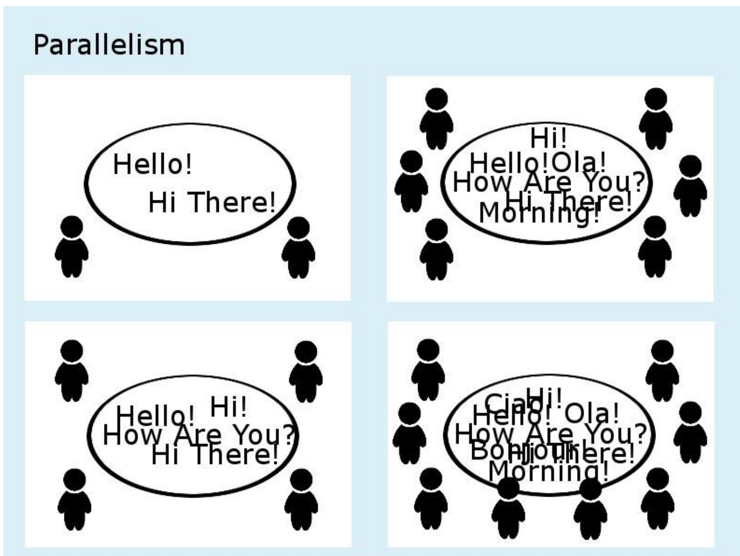

Figure 2.3: Parallelism is the number of conversations that can effectively be simultaneously transmitted across a medium.

Immediacy of Feedback places emphasis on the ability to immediately correct and fine tune a transmission, allowing for the quick eradication of unnecessary or misinterpreted information to a group. Using a system like a text chat, feedback has a delay that allows for receivers to form their own opinions about potentially wrong or misunderstood information, while taking advantage of an audio environment significantly reduces the feedback time between messages. This allow for members to more quickly dispute points of contradiction, and smooth out what may have been misunderstanding.

In Conclusion, social environments that support high immediacy of feedback and low parallelism encourage synchronicity that is key to the convergence process. Social environments that have a low immediacy of feedback and high parallelism are more suited to tasks of conveyance, as they allow for more time to deliberate and deliver carefully processed messages.

When the goal of a social environment is convergence, media of high synchronicity will lead to better performance. When the goal of a social environment is conveyance, media of low synchronicity is ideal. Media symbol variety will only negatively impact information distribution when the need of a symbol is present, but the symbol itself is unavailable. Usage of media with higher rehersability leads to a better performance under all circumstances.

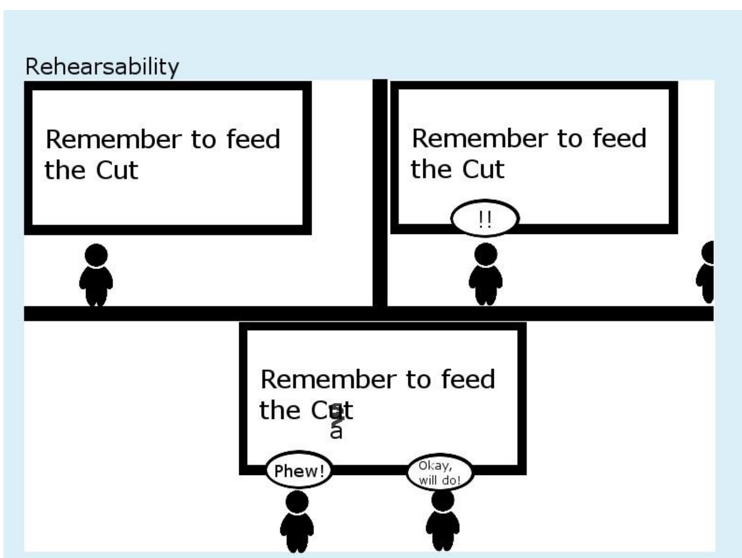

Figure 2.4: Rehearsability is the ability to edit a message before transmission.

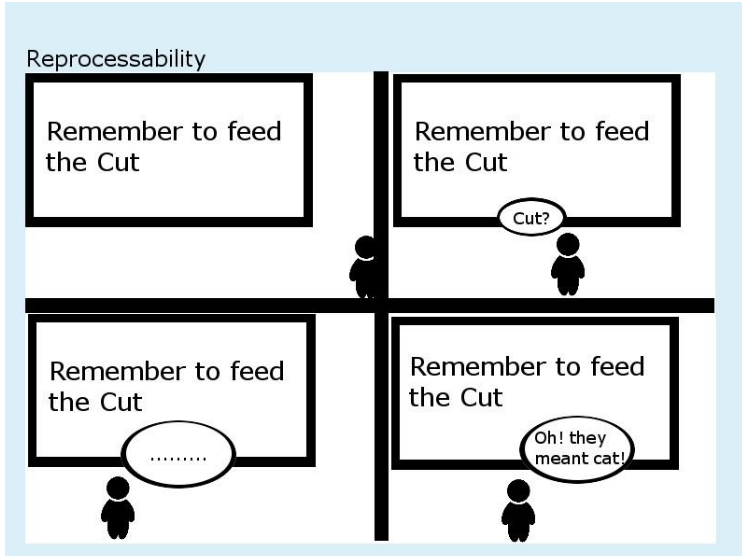

Figure 2.5: Reprocessability is the ability for a transmission to be reexamined and altered after transmission to allow for a development of its meaning

2.3 Audio VS. Chat: Can Media Speed Explain the Differences in Productivity?

Calling upon both the Theory of Media Richness and the Theory of Media Synchronicity, Lober et al. [3] evaluated the productivity of groups working towards both tasks of equivocality and uncertainty in varying group sizes and mediums, comparing VoIP and chat environments. They proposed that audio communication is a faster form of communication, but that chat is more effective when judging the productivity of groups working towards a given task.

Two experiments were designed to test the productivity of chat VS. audio, each on groups of four. The first was a murder mystery experiment. Designed to be a task of high uncertainty and low equivocality, participating groups had to identify a guilty murderer from a group of three suspects. Each member received a different set of information, requiring them to discuss and come to a consolidated consensus on both the information they possessed and the group decision as a whole. The second experiment was designed with a low degree of uncertainty but a high degree of equivocality. The goal was to develop an ideal model for a futuristic post office. This experiment had little instruction and encouraged group conversation due to its high ambiguity. All groups consisted of four-member teams and had ten groups for every combination of task and media.

The following conclusions were drawn from their experiments: audio groups transfer more useful information than chat groups in the same amount of time. This is due largely to the immediacy of feedback and the ability to speak much faster than typing. Audio groups transfer more critical information than chat groups in the same amount of time. Much like the first conclusion, this is largely due to speed of communication, which was sufficient enough so that the amount of information conveyed outperforms the rehearsability to convey critical information iconized by chat clients. Finally, chat groups have a higher ratio of critical information to audio groups in terms of overall information. This is due to the ability to retain focus due to lack of personal interaction.

Chat groups were more efficient than audio groups at transferring critical information in the Murder Mystery experiment performed by Lober et al. [3], but the ratio of critical information to time spent communicating could not hold up when compared to the overall speed of communication possessed by the audio groups. While chat groups are a highly efficient means of conveying information, they do not outperform the speed of audio interaction between groups. Media speed is a high identifier of group productivity, even more so than the type of media. This contradicts both the Theories of Media Richness and Media Synchronicity by placing an emphasis on the vehicle speed rather than its richness or appropriateness to the task at hand. It does, however, closely follow the idea of increased parallelism on chat messaging VS. low parallelism in audio chat, suggesting that such groups be kept to a maximum of four individuals per group.

2.4 Audio vs. Chat: The Effects of Group Size on Media Choice

Lober et al. [4] proposed that chat-based interaction scale up better than audio groups as group size increases, properly demonstrating the behavior predicted by the Theory of Media Richness [1] in small groups and the theory of Media Synchronicity [2] for larger groups [4]. Lober mentions that the Theory of Media Richness theory, while claiming to be applicable to larger groups, only used smaller testing groups to draw their conclusions.

As group size grows, its members become subject to productivity blocking. Productivity blocking occurs when members are unable to voice their options and views due to overlapping sound with another participant. This is an issue in audio-based chat systems and is addressed by the concept of parallelism in the Theory of Media Synchronicity. Chat based social systems can help to mitigate production blocking, as multiple concurrent users may give input without having to wait for another to finish. It is noted that the Media Richness Theory does not take such matters into consideration. However, personalized chat media had not been commercially developed at that time.

When using an audio-based chat system, parallelism becomes a matter of increasing importance. Recall that parallelism is the number of simultaneous senders that a medium can handle. With a large number of users, a medium with low parallelism causes the immediacy of feedback to be drastically increased in time, as users must wait longer to convey their information. This results in productivity blocking and can be severely detrimental to the productivity of a group.

Using the Post Office of the Future experiment [3] for its testing environment, audio groups were tested using various VoIP clients. Teamspeak was used for groups of seven, and Skype for groups of four since Skype could not host more than five users at that time. After the experiment, all users were given a survey to express their satisfaction with the communication medium.

Results of the experiment indicated that audio groups are significantly more productive than chat groups and held a higher degree of satisfaction when the group size was limited to four. When reviewing the result of groups size being seven, there was no significant difference found in the productivity of the audio groups over chat groups. The experiments did find that in groups of seven, chat groups felt more satisfied with their experience. Their results suggest that text chat scales better than audio chat as group size increases. While in small work groups, audio is superior in terms of both satisfaction and productivity. When considering groups between those margins, audio groups should still be considered as the amount of satisfaction and productivity change between four to seven does not merit a change in medium.

Our audio hierarchy structure is implemented based on these ideas. It includes a text chat to help convey important information. It includes audio chat to encourage satisfaction of participation, and it includes breakrooms to help mitigate the dissatisfaction caused by having too many overlapping voices on an audio channel. The interface was kept to a minimum to help stay within the bounds of symbol variety, and the audio regions help to mitigate scaling parallelism.

Chapter 3

Implementation

In this chapter we start with an overview of the implementation of the audio structure followed with the mechanics of the UNITY game engine and the various audio and video packages that are available in UNITY. We will focus on the Dissonance audio chat system, and then discuss our audio structure, the different audio channels and regions, as well as how they work together to produce the desired effect.

3.1 Implementation Overview

We use a client-server architecture where the user can login to the virtual environment which is entirely built in UNITY. Once the user logs in, he is given an avatar that allows for navigation of the environment and automatically tunes into the public audio channel. When the user moves into a designated audio region within the virtual space, an event is triggered, and the user moves into a separate group channel. The user can also privately communicate with others. Both the audio region and the private chat use separate audio channels which is built using the Dissonance Voice Chat package. The control of these channels is managed using tokens. These are described in more detail in the rest of the chapter.

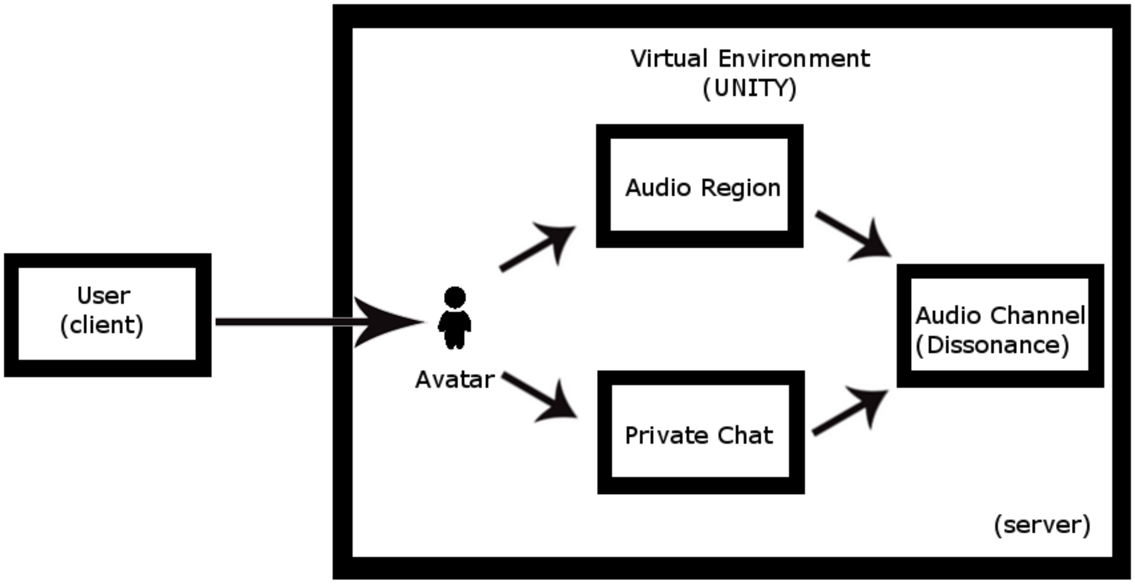

Figure 3.1: Implementation Flow Chart

Some of the challenges in the project was to learn programing in the UNITY game engine, its underlying C# implementation, and the various game design techniques. The UNITY game ending adds another dimension of coding in the form of game objects and their components and how they interact with the underlying C# code. Learning the best practices of programming in UNITY and the unique relationship between the game environment and the coding was very crucial to the development of the audio structure.

3.2 UNITY

UNITY is a cross platform game engine servicing 25 different platforms, from iOS and Android to Sony's Playstation and Microsofts Xbox series. First released in 2005, UNITY Technologies boasts that over half of the world's released games have been developed using their game engine. We chose to implement our project in UNITY due to its low barrier to entry on learning and its easily understandable interface. With a large marketplace of extensions for any game to accompany the C driven platform, UNITY helped to streamline the process as it made connecting in-game objects with underlying scripts smooth and straight forward. The proposed audio structure has several components that make up its foundation and allow it to function. The Dissonance Voice Chat is one of the external plug-ins that handles the basic Audio chat portion of our Audio Structure.

3.3 Dissonance, Photon, and DFVoice

The Dissonance voice Chat package is a VoIP solution developed by Placeholder Software as a paid plug-in for the UNITY environment. Placeholder software is a UNITY based development company with two products -- Dissonance Voice Chat and Wet Stuff, which is an add on to make any surface in the UNITY Game Engine appear wet. We chose Dissonance due to its team geared implementation and easy adaptability to existing environments. Dissonance relies on the Opus Interactive Audio Codec as its underlying coding format. Opus is the current highest rated audio transmission format for interactive speech, music transmission, storage and streaming application and contains support for all platforms used today. The Dissonance Voice Chat UNITY package allows audio transmission of data by attaching audio channels, which are audio frequencies determined by access tokens, to a local game object called the AudioManager. These tokens are used to manipulate membership to different audio channels.

Originally DFVoice was the audio implementation chosen due to its basic simplicity of communicating audio between clients. DFVoice was an ideal candidate due to its low cost and skeleton framework. However, after further investigation, we note that the concept was independently developed and distributed by PhotonVoice, a package supplied by Photon. Photon is an independent industry-solution of a networking engine geared towards multiplayer online gaming. PhotonVoice was possibly developed by a team of programmers, software developers and game designers.

This thesis stemmed from a previous project geared towards making an online classroom environment at Appalachian State University where the research was conducted. Looking for possible Audio implementations to help create the proposed audio structure, Photon was not initially considered due to its high price and was therefore overlooked to be used for the audio structure implementation. When this research grew and became independent from the virtual classroom implementation, I revisited Photon looking for inspiration and found that its audio implementation was the exact concept originally proposed when using DFVoice. The original proposal was to add a frequency array to every player, and by entering and exiting different audio regions, frequency tokens would be added to and taken from that array. By broadcasting all audio data to every player, the frequency arrays would allow only subscribed audio channel data to pass through to the player speakers, preventing data they did not have access to from being heard. So, we chose Dissonance to reimplement our concept and rather than storing this data with every player. We relied on the Dissonance Audio manager to keep track of all players and their subscriptions to different audio frequencies.

3.4 The Environment

Initially we intended to integrate our proposed audio structure with a virtual classroom. However, this thesis focuses purely on the audio and text environments. We modeled a simple box structure with four audio channels at each of the structure's corners to focus purely on the communication portion of the audio structure. We created new basic avatar models for users to maneuver from region to region, and implemented the environment, the audio and text chat structures. However, it should be noted that we borrowed code for the movement and camera settings of the avatar, which allowed the avatar to move freely in our virtual environment.

3.5 Controlling the Avatar

The user in our environment controls an avatar, which is a visual representation of the person in a virtual space represented, see Figure 3.2. This avatar can move using the arrow keys or the (w)forward, (a)left, (s)backward, (d)right, movement scheme often seen in first-person games today. The user in our virtual environment has several functions: movement, text chat, and audio chat. By moving the character into an audio region, which is a designated area represented by a game object, a method is called that adds a token to an AudioManager game object. The AudioManager is a stock game object provided by the Dissonance Voice Chat extension that keeps track of player IDs and tokens that are attached to them.

```
public class PlayerController : NetworkBehaviour
{
    public ControlStrategy movement;
    private ControlStrategy storedMovement;
    private CharacterController cc;
    private CameraController cam;
    public DissonanceComms comms;
    private HlapiPlayer hlapi;
    private const string PUBLIC_CHAT = "PublicChat";
    private const string PUBLIC_HEAR = "PublicHear";
    private const string AUDIO1 = "AudioRegion1";
    private const string AUDIO2 = "AudioRegion2";
    private const string AUDIO3 = "AudioRegion3";
    private const string AUDIO4 = "AudioRegion4";
    ArrayList PlayerList = new ArrayList<PlayerNode>();
    [SyncVar]
    public string currentChannel;
    [SyncVar]
    public string PlayerName;
    [SyncVar]
    public string PlayerID;
```

Figure 3.2: The PlayerController script, a component attached to the Avatar that controls most of its functionality

Once in an audio region, the AudioManager then allows the user to broadcast or receive audio data from others in the same region while continuing to overhear outside conversation at a decreased volume. The text chat is directly connected to the audio chat and functions in the same way. Text conversations held within a region are exclusive to that region, but public text may be broadcast to all regions without interference. The avatar's main function in this audio structure is to act as the vehicle that moves user's region to region. This allows users to interact with various virtual environments, like accessing syllabi and other class/course information in a virtual classroom or viewing digital displays from the avatar's camera.

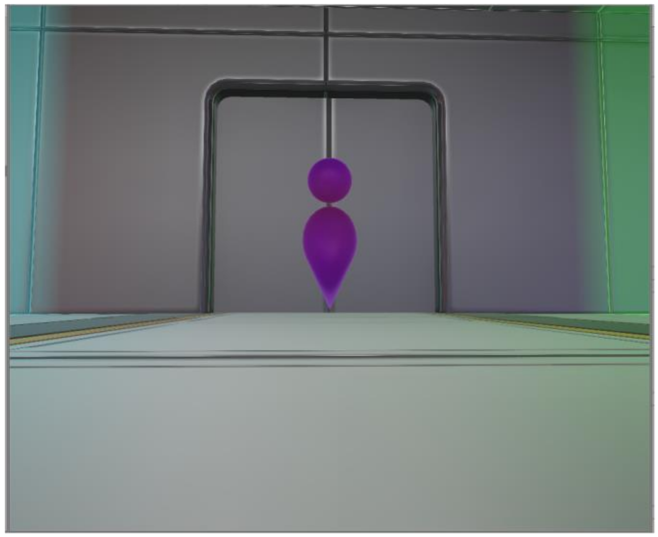

Figure 3.3: This floating character is the avatar, a virtual representation of the user and a vehicle by which they move around the virtual environment.

3.6 Audio Channels

The purpose of an audio channel is to transmit audio data and allow users to hold overlapping conversation without interruption. The audio channel is comprised of three objects – a Broadcast Trigger, a Receive Trigger, and a token. The Broadcast and Receive Triggers are components that allows users to transmit and receive audio data when attached to the AudioManager. A token is an identifier that can be added to the Broadcast and Receive triggers that allows membership to those channels. By adding multiple pairs of triggers to the AudioManager, team style multi-player mechanics can be achieved. Within this environment, each of the audio regions has a single token shared between sets of audio and receive triggers.

This allows easy manipulation of the player membership to those audio channels. The Public channel is a set of two tokens: PublicChat and PublicHear that manage the Broadcast and Receive Triggers, respectively. By manipulating those tokens, with a series of action triggers based on the users location in the virtual space, we can simulate the exclusive audio relationship as proposed in earlier chapters.

The audio structure is divided into three levels – the group, public, and private levels. Upon entering into the virtual environment, all players have the PublicChat and PublicHear tokens added to the local AudioManager. The PublicHear token is a permanent addition to the AudioManager and is not removed at any time. This allows the user to constantly hear what is being broadcasted from the Public channel. The PublicChat token is removed when a player enters an audio region and is added when they leave, allowing them to participate in public conversations exclusively when they are in the Public channel. When a user enters an audio region, an audio region token is added to the local AudioManager, and when they leave, that token is removed. These channels work in tandem with the AudioManager and audio region game objects to obtain channel-based interaction to specific locations, thereby effectively creating a private breakdroom to hold conversations. The final tier of the audio structure is the private conversation. This is a special function that came prepackaged with Dissonance that allows two players to converse from any distance without being heard by anyone else, even if their memberships to audio channels are entirely different. This is achieved by entering the PlayerID you would like to speak with into a whisper box above the chat box. The AudioManager searches for a player with that ID, and if found, creates a set of private tokens exclusive to those players, thereby allowing them to communicate, which ignores the rules of both the group and public channels.

3.7 Audio Regions

Audio Regions are empty game objects within UNITY. An empty game object is a placeholder with an affixed name and Identifier tag. An identified tag is a string value attached to a game object that can be searched using one of UNITY's search functions. This identifier tag is assigned a unique name that will serve as the token sent to the AudioManager that will allow users to receive audio data within the audio region. After adding a name and Identifier tag, an object mesh and collision collider are attached to the game object.

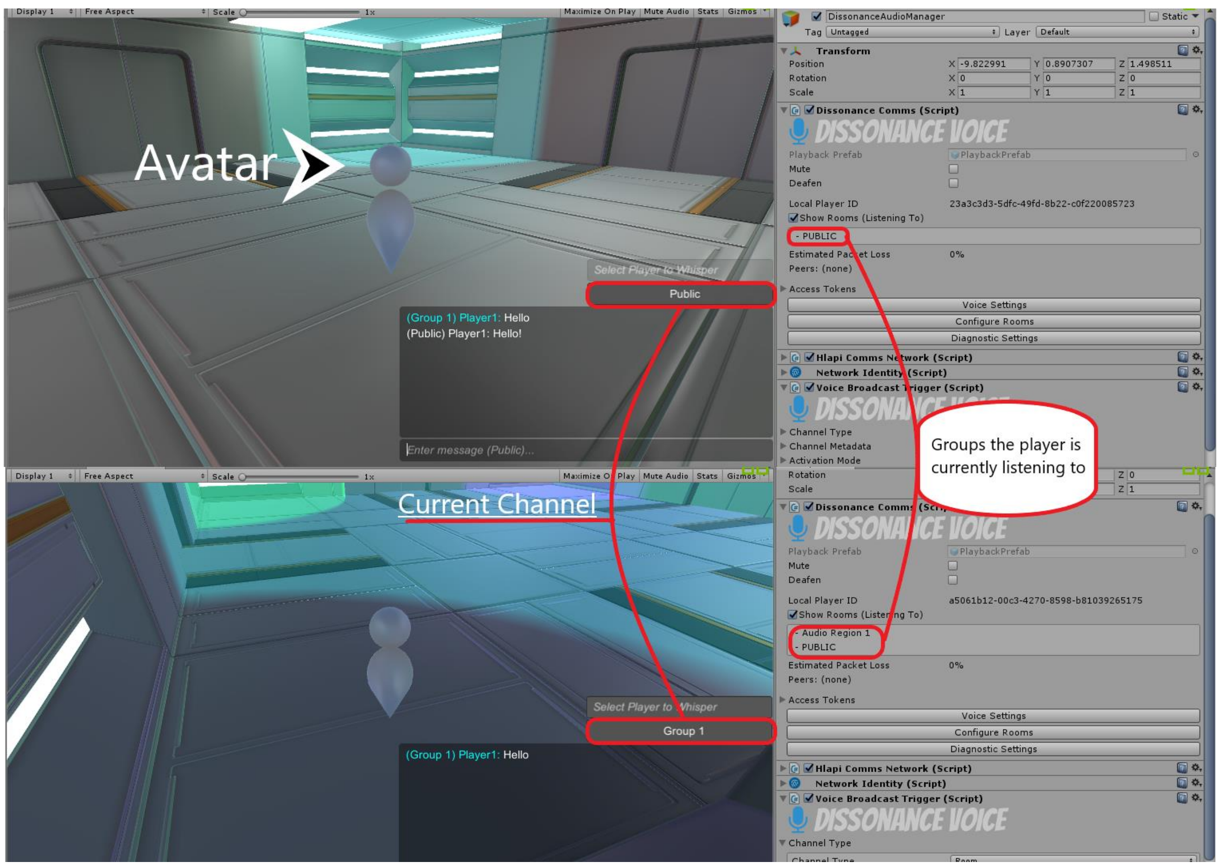

Figure 3.4: The region of blue light shown in the image depicts the boundary of an audio region, by entering or leaving that space, tokens are added to the Comms script shown on the right that show membership to which audio channels the user is a member.

An object mesh is an invisible boundary that exerts influence within the UNITY environment. By adjusting various settings on the game object, the boundary becomes both transparent and un-intractable, meaning a user may pass through that object as if it was not there. This boundary is the ”physical” presence of the audio region. Attached to the object, now called a region is a set of triggers, instantiated by the collision collider. When a user passes through the collision collider from the outside of the boundary, the method OnTriggerEnter is

```
void OnTriggerEnter(Collider region)
{
    if (!isLocalPlayer) { return; }
    switch (region.tag)
    {
        case AUDIO1:
            AddChannel(region.tag);
            RemoveChannel(PUBLIC_CHAT);
            CmdSetCurrentChannel(AUDIO1);
            setTeamIndex(1);
            break;
        case AUDIO2:
            AddChannel(region.tag);
            RemoveChannel(PUBLIC_CHAT);
            CmdSetCurrentChannel(AUDIO2);
            setTeamIndex(2);
            break;
        case AUDIO3:
            AddChannel(region.tag);
            RemoveChannel(PUBLIC_CHAT);
            CmdSetCurrentChannel(AUDIO3);
            setTeamIndex(3);
            break;
```

Figure 3.5: OnTriggerEnter, Part of the PlayerController Script, are what communicates with the AudioManager and facilitate the exchange of tokens when moving between regions.

called, and the token of that region tag is added to the AudioManager. As long as the player is within the region, they are now eligible to receive audio data with the ”PUBLIC” tag or audio region to which they are subscribed. Upon leaving the boundary of the audio region, the method OnTriggerExit is called, and that same token is removed. This completes the illusion of a private region where users can interact without being overheard. The Audio Manager transmits

```
void OnTriggerExit(Collider region)
{
    if (!isLocalPlayer) { return; }
    switch (region.tag)
    {
        case AUDIO1:
            CmdSetCurrentChannel(PUBLIC_CHAT);
            RemoveChannel(region.tag);
            AddChannel(PUBLIC_CHAT);
            setTeamIndex(ChatSystem.PUBLIC);
            break;
        case AUDIO2:
            CmdSetCurrentChannel(PUBLIC_CHAT);
            RemoveChannel(region.tag);
            AddChannel(PUBLIC_CHAT);
            setTeamIndex(ChatSystem.PUBLIC);
            break;
        case AUDIO3:
            CmdSetCurrentChannel(PUBLIC_CHAT);
            RemoveChannel(region.tag);
            AddChannel(PUBLIC_CHAT);
            setTeamIndex(ChatSystem.PUBLIC);
            break;
        case AUDIO4:
            CmdSetCurrentChannel(PUBLIC_CHAT);
            RemoveChannel(region.tag);
            AddChannel(PUBLIC_CHAT);
            setTeamIndex(ChatSystem.PUBLIC);
            break;
    }
}
```

Figure 3.6: OnTriggerExit, Part of the PlayerController Script, are what communicates with the AudioManager and facilitate the exchange of tokens when moving between regions.

the audio data to all players within its drop off distance, but due to a function called when players enter or exit a region, only they may hear people speaking within. The interaction of the player and Audio Region Object are what defines this audio structure and creates its channel-based hierarchy.

3.8 How it Works

When a player logs into the environment, the PublicHear and PublicChat tokens are immediately added to the AudioManager as they are spawn into the starting area. This area

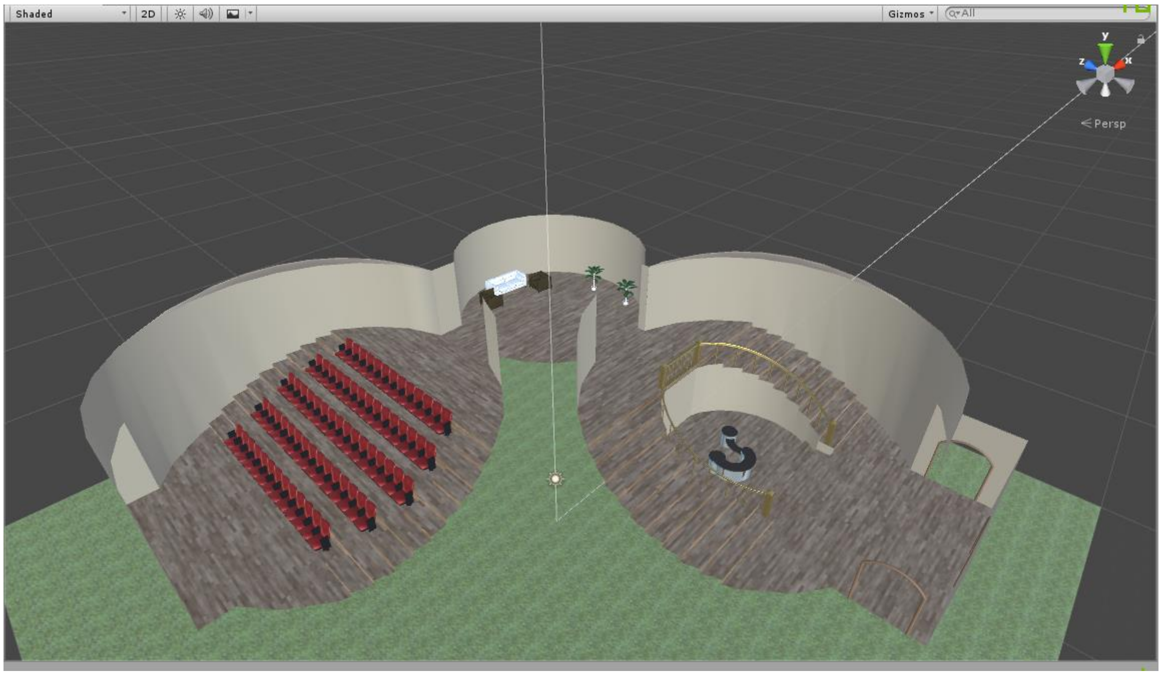

Figure 3.7: This is the design for the lobby region of the classroom

could be a lobby, or a classroom, or even a forest within the virtual environment.

A virtual classroom may take on artistic visualizations based on the teacher or creator. Players may move around and talk openly within this space. All users within this space will hear each other. There can be many predefined audio regions within the public space. These regions are more localized and may be represented by any number of things, for example audio regions within the virtual environment can be attached to an in-game object, such as a tree, bench, couch, or a table, which allows the audio regions to be more easily identifiable. Upon passing the boundary of the audio region space, the region calls the method *OnTriggerEnter* on the user and the token for that region is then added to the AudioManager, and the PublicChat token is removed, which therefore removes the user's ability to broadcast to the public channel. Now that the player is a member of that region, they have open access to all communications on that frequency, while surrounding users passing around the region do not have access and therefore cannot have conversations in that channel. Should the user then choose to leave the

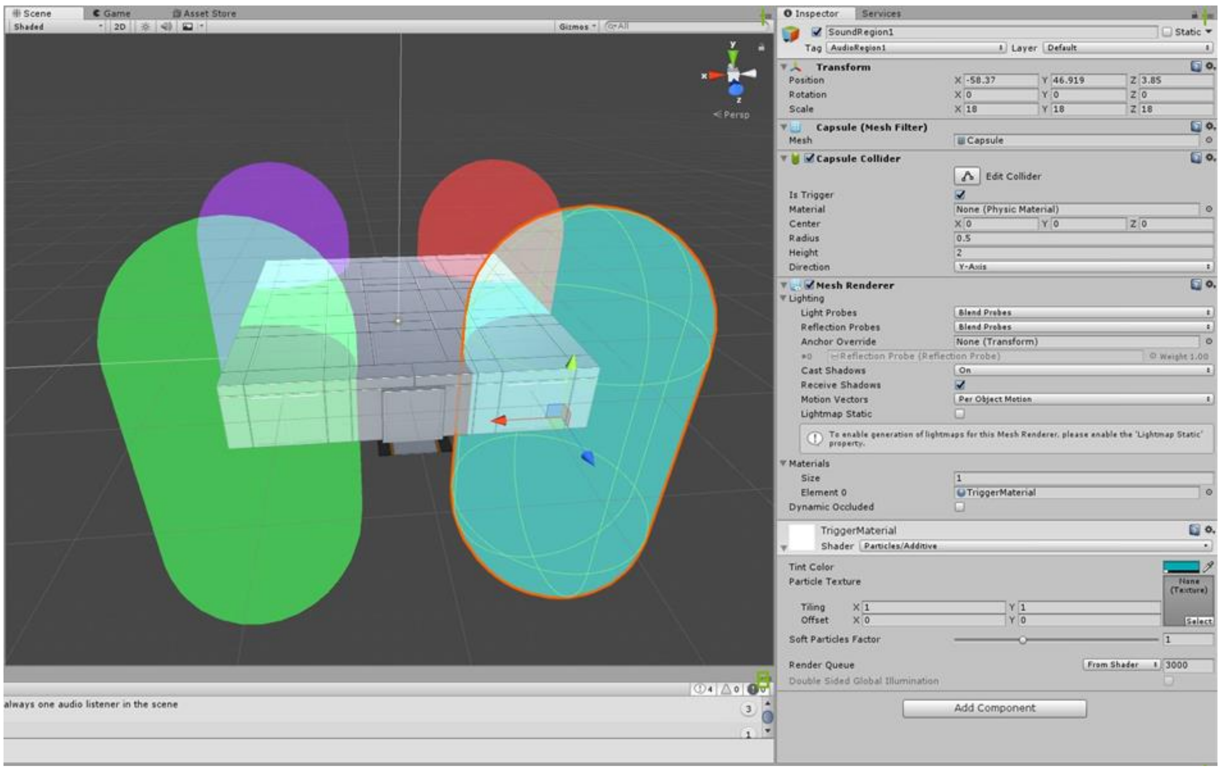

Figure 3.8: The highlighted capsules shown are the representative audio regions. They are colored for the sake of demonstration but would normally be invisible and tied to a more natural object.

region perimeter, the region token is removed from the AudioManager and the user is no longer able to participate on that region, but the PublicChat token is restored and they may continue conversing on the public channel.

There is another part to the audio structure. Users may also communicate amongst themselves privately. Using a text box positioned above the chat box on the screen, a player may input the name of another user. By entering another players name, a private channel is created between the users, this means that those two players may communicate with one another regardless of tokens or channels, while still having full access to all listening channels around them. This is because a private token is created that only those two players share. This form of communication has the highest authority, and cuts across all channels and broadcasts. Using private communication, two players may communicate from within two separate audio regions, even though their communication would normally be restricted. This form of communication also removes all other broadcast tokens, rendering them incapable of speaking over those other channels while engaged in private conversation

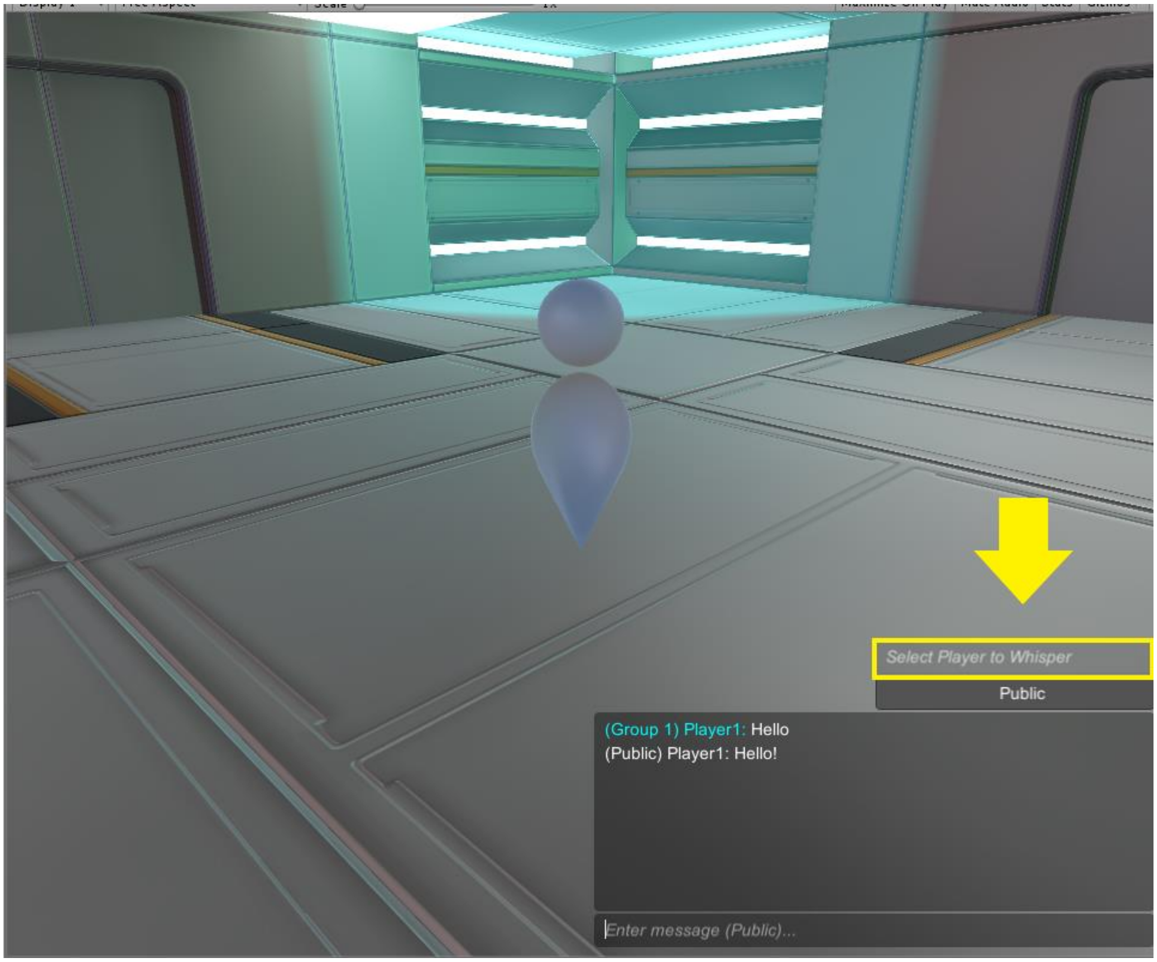

Figure 3.9: The highlighted box allows players to communicate privately. By inputting the name of another player into the box, it will turn green, signifying that all spoken communication is now directly to that player. If an invalid name is placed into the box, it will turn red and cut off all outward communication until it has been cleared or a correct name has been added.

Chapter 4

Evaluation

4.1 Experiment setup

We compare our audio structure with popular online communication mediums, namely Zoom, Skype and Discord. These form our independent variable "Medium." Each of these software mediums have a suite of features and would be expected to vary in their effectiveness at conveying information. We use two forms of activities, which forms another independent variable called "Game." Each of these games depict properties described by the theories in Chapter 2. The two games we use are the Pirates game and the Post Office of the Future experiment. These games were taken from similar experiments conducted by Lober [3]. We have also identified three dependent variables – Enjoyment, Usability, and Feasibility. We describe all these in this section below.

4.1.1 Mediums

Our Audio Structure:

As mentioned in Chapter 3, our proposed Audio structure is a UNITY developed 3D environment meant to focus on the audio and chat aspects of online communication. By operating visual representations of the users, called Avatars, the users may freely move from location to location within the virtual space to communicate using audio and text chat mediums. By entering a virtual boundary called an audio region, users are able to communicate to others exclusively within that region. This environment was created in hopes of generating an educationally geared virtual environment, where students could engage with each other in a visual way by utilizing the audio regions to communicate more effectively.

Figure 4.1: Features of Skype

Skype:

Skype is a telecommunications software used by businesses around the world to make free audio and video calls internationally. Skype was one of the first major audio telecommunications software to enter public use and is still used heavily in both the business and non-enterprise worlds.

Skype is a multiplatform telecommunications software that utilizes audio, video, and text chat services to connect people around the world. It has support for almost every platform -- Windows, Android, iOS, and Linux, and includes the file sharing of images, audio, and video files. It allows conferencing between multiple groups and is used by businesses internationally. Skype was introduced in August of 2003 [6] and is one of the leading audio and video conferencing software in use today.

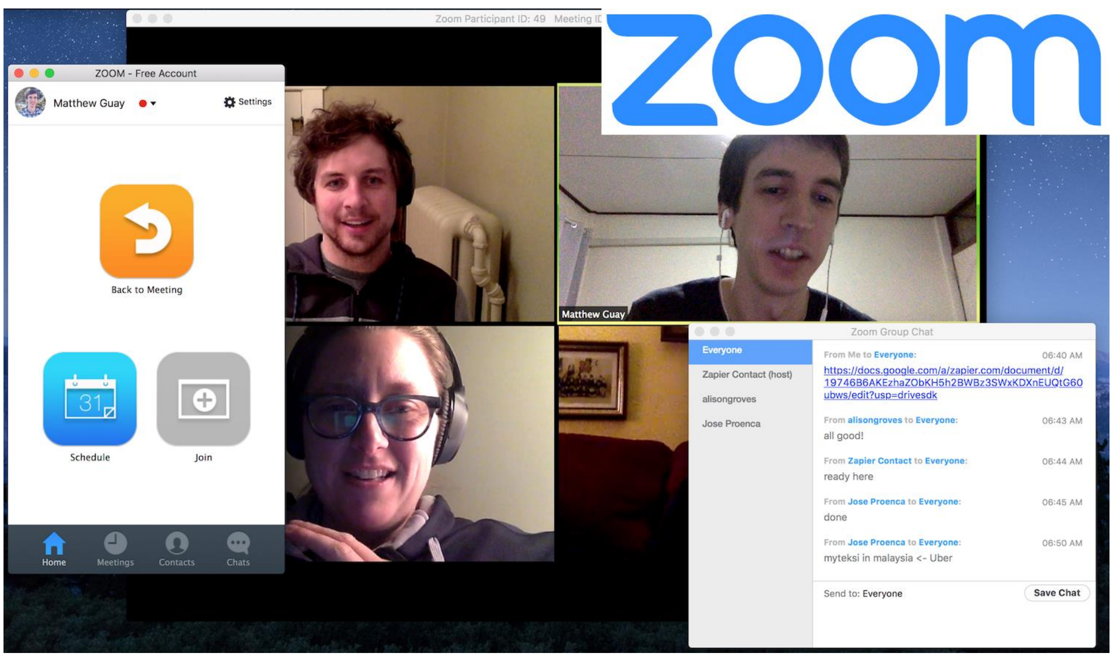

Figure 4.2: Features of Zoom

Zoom:

Heavily business focused, Zoom is the most professional of the audio structures covered in this thesis. It is designed to support the business culture and create a professional business environment with an easy and reliable cloud platform. It offers video and audio conferencing, and a variety of group-based software options to make communications seamless.

Zoom is another audio-video conferencing software oriented around business and large-scale conferencing environments. Like Skype, it offers audio and video conferencing, file sharing and multiplatform support. But it also boasts itself as being the leader in modern enterprise video communications (see https://zoom.us/about). One of Zoom's unique functionality is the incorporation of Zoom Rooms. Zoom Rooms are software-oriented virtual meeting spaces similar to the audio regions for the audio structure.

With exclusive membership to invited members. the Zoom rooms are designed to augment a physical space or spaces using cameras and microphones, to allow persons in various locations to experience a single room style conference. This is achieved through using additional equipment, cameras and audio technology, to shift focus to speakers and hone in on them when they are speaking.

Designed with gamers in mind, Discord was developed with the goal of easy connection, and drop in drop out capability. It has support for specific games, and is designed to draw the gaming community closer, making communication easier and more available through its free communication software.

Figure 4.3: Features of Discord

Discord:

Discord is a free VoIP software aimed at the computer gaming population. It combines audio, video, text, and image sharing technologies in chatroom style group settings. Discord is comprised of channels, which are created and maintained by users acting as administrators. Administrators are responsible for public and private management of the groups and their members. A channel is comprised of Internet chatroom style text boards, where members may post files and talk through audio or text. Groups, like their mother channels, may be locked with a password that prevents certain users from accessing it. Discord supports all commonly used platforms. First released in 2015, Discord has overtaken the gaming world as one of the most popular means of communication.

4.1.2 Games

We discuss the two types of activities or games used in our experiments below. In Section 4.1.3 we provide the experimental setup and the methodologies used in evaluating our audio structure with other online tools mentioned above.

Pirates Game:

The Pirates game is a question that is ideal for environments of high uncertainty and low equivocality. As defined in Chapter 2, uncertainty is the absence of information. For the Pirates puzzle, all information is available to solve the problem, and it is intended that participants work to find this unique solution. Equivocality is a measure of a group's ability to use personal experiences with other collaborators to draw together a unified conclusion. As stated above, an increase in information would not serve to better elicit an answer, so providing the bare minimum requirements to focus on the problem is ideal. Bogging down the participants with symbol variety can serve to hinder more than help. The Pirates problem below is similar to the Murder Mystery used by Lober [3]. A description of the problem is given below.

Pirates Game:

There are 5 pirates with a bag of 100 gold coins. The pirates are ranked 1 through 5 based on seniority (1 being the oldest and 5 being the youngest). The most senior pirate proposes a strategy to divide the gold, which is then voted by everyone (himself included). If he fails to receive at least 50% of the votes, he is thrown into the sea and the next senior pirate makes a proposal. This process is repeated until only one pirate is left standing, or if the proposed pirate receives at least 50% of the votes. You may assume that each pirate votes *primarily for their own survival, then having survived, tries to get as many gold coins as possible for themselves. Being the most senior pirate, how many gold coins can you get? Note that you want to ensure your own survival first, then try to maximize the amount of gold coins you can get. You have 30 minutes, you'll know when the test starts.*

Post Office Game:

The Post Office of the Future Experiment developed by Lober [3] was designed to test a low degree of uncertainty and a high degree of equivocality. Environments of high equivocality thrive when presented with multiple symbols and interpretations between answers, because there is no clear solution or definitive correct method to answer. In the Post office of the Future, the participants were given little to no direction and are left that way intentionally to pave a unique solution derived by the group for themselves. This game was designed to encourage round table discussions and debates, making it a perfect example of a high equivocality environment. A description is provided below.

Post Office Game:

Your team has been tasked with developing an initial requirement for a stand-alone post office of the future. This schematic needs to function much like an ATM in terms of usability. For this task, please discuss with your team and assemble the following:

- *A list of core services that post office would offer*
- *A list of the equipment required for full functionality*
- *An explanation of how it would work*
- *A rough cost/benefit Analysis*
- *A list of things to investigate before your next (imaginary) meeting.*

4.1.3 Methods

As mentioned above, the two independent variables “medium” and “Game” provide 8 unique combinations for our experiments. From the Theories of Media Richness [1] and Media Synchronicity [2], we identify three dependent variables – Enjoyment, Usability, and Feasibility. Enjoyment identifies the overall satisfaction of using a medium when completing the experiment. Usability is a personal measure of difficulty when operating the medium over the course of the experiment, and Feasibility is the ability of the medium used to assist in creating a solution from the game.

Using a combination of each game and medium, the groups were each given a mode and a game, designed to compare the effectiveness of our audio structure against leading communications software described in Section 4.1.1. Each group consists of 4 users. They were given a total of 10 minutes to adjust to the software and understand its basic features. They were then assigned a game and given 30 minutes to come up with a consolidated solution with little to no outside interaction. A series of hints were devised for the Pirates experiment, and no interaction was permitted with the Post Office of the Future game. At the end of 30 minutes, the users were thanked for their participation and given a short 9 question survey: Question 1 was given a rating of 1 through 5. While others were rated from Strongly Disagree (1) to Strongly Agree (5) The survey is given below.

- ***Enjoyment***

  1. I enjoyed the experience of using this virtual environment

  2. Did you feel that your voice was heard, and that you could clearly communicate your ideas?

  3. How enthusiastic were you about this virtual environment?

- ***Feasibility***

  1. Would you use this virtual environment for similar group discussions?

  2. How easy did you feel using the environment to solve the given task?

  3. Did the environment provide you with the necessary tools to communicate?

- ***Usability***

  1. How effective was the audio in conveying information?

  2. How effective was the text chat in conveying information?

  3. How easy was it to use this environment for group discussions?

Using these questions, we gathered the information used in the data analysis segment of this chapter. In total there were 32 students that were tested across 4 mediums.

We chose a group size of four to better align ourselves with the predictions and conclusions drawn from our literature review detailed in Chapter 2. Similar experiments concluded previously also used groups of four [3]. Lober et al. [2] discovered that keeping audio groups down to a size of four promoted the most optimal satisfaction from the testing group when communicating through an audio medium. Having a smaller group size removes the amount of parallelism that can occur on a channel. With groups less than four, there is less of a reason for members to use multiple means of communications, deferring to audio chat since there is little overlap between members and audio was determined to be the most effective with small group sizes [3].

Finally, due to the constraints of time and availability, we kept the group sizes to four to best make use of the total number of participants present for the experiment. However small the sample size, this thesis provides an evaluation model for a large scale experiment to compare different online communication software.

4.2 Data Analysis

Covered in the previous section, our experiment used two independent random variables – Medium and Game. We have four mediums – our audio structure, Zoom, Skype, and Discord, and two types of games – Pirates and Post Office of the Future. This provides a total of eight combinations between four mediums and two games. We had eight groups to represent one of these combinations. Each group had four members, therefore there were a total of thirty-two subjects. With a small sample size, we first performed a power analysis to see if we could get significant differences between groups, and to detect an interaction between mode and game. Using G*Power of Faul et al. [9] we needed a very large effect ($F = .58$ where .40 is considered “large”) to have even an 80% chance of detecting it. Here, the F value is defined as the ratio of explained variance and unexplained variance or the ratio of between group variability and the within group variability (see [10]). Given the small power of this study, it’s best considered exploratory and interpretation is mostly focused on describing the patterns. However, we feel that this thesis provides a prototype to perform a much large-scale analysis if need be by increasing the sample size.

We compared the mediums based on three qualities, which defines our dependent variables – Enjoyment, Feasibility, and Usability. For each dependent variable, we asked three items measured on an x-point scale. For each set of three items, we first examined correlations to make sure they were strongly related as expected. For each variable, the items were significantly intercorrelated at the $p < 0.05$ level, therefore we created a scale for each variable by averaging the three items together. Section 4.2.1 shows the results of the correlation-analysis of the dependent variables. We then perform two types of descriptive statistics based on ANOVAs and t-tests. All our analysis was performed with Software Package for Statistical Analysis (SPSS).

4.2.1 Correlations

Table 1: Correlation Analysis of Enjoyment

Correlations

		enj1	enj2	enj3
enj1	Pearson Correlation	1	0.471**	0.716**
	Sig. (2-tailed)		0.006	0.000
	N	32	32	32
enj2	Pearson Correlation	0.471**	1	0.416*
	Sig. (2-tailed)	0.006		0.018
	N	32	32	32
enj3	Pearson Correlation	0.716**	0.416*	1
	Sig. (2-tailed)	0.000	0.018	
	N	32	32	32

As can be seen in Table 1, there is a strong correlation between the variables enj1, enj2, and enj3 which is also statistically significant as seen by very low standard deviations. Therefore, we can combine the three into one variable avgenj by averaging the responses for these questions for each of the 32 subjects. We see similar correlations for variables use1, use2, and use3, and feas1, feas2, and feas3, and therefore averaging them into variables avguse and avgfeas respectively.

Table 2: Correlation Analysis for Usability

		us1	us2	us3
us1	Pearson Correlation	1	0.525**	0.538**
	Sig. (2-tailed)		0.002	0.001
	N	32	32	32
us2	Pearson Correlation	0.525**	1	0.836**
	Sig. (2-tailed)	0.002		0.000
	N	32	32	32
us3	Pearson Correlation	0.538**	0.836**	1
	Sig. (2-tailed)	0.001	0.000	
	N	32	32	32

Table 3: Correlation Analysis for Feasibility

		feas1	feas2	feas3
feas1	Pearson Correlation	1	0.378*	0.433*
	Sig. (2-tailed)		0.033	0.013
	N	32	32	32
feas2	Pearson Correlation	0.378*	1	0.579**
	Sig. (2-tailed)	0.033		0.001
	N	32	32	32
feas3	Pearson Correlation	0.433*	0.579**	1
	Sig. (2-tailed)	0.013	0.001	
	N	32	32	32

Table 4: Summarizing the data

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
Avgenj	32	2.33	5.00	4.1354	.75662
Avgeuse	32	1.00	5.00	3.7500	.95789
Avgefeas	32	2.00	5.00	3.9688	.83969
Valid N (listwise)	32				

Table 4 shows simple statistics of the three variables avgenj, avguse and avgfeas. These are taken from Table 5 that shows the average responses of each of the thirty-two subjects.

Table 5: Consolidated survey data for avgenj, avguse and avgfeas

avgenj	avgeuse	avgefeas
4	3.333333333	4
4.333333333	4	3
4	3.666666667	4
4.666666667	3.666666667	4
3.333333333	2	2
4	4	3.666666667
4.333333333	4.666666667	4.333333333
5	4.333333333	4.333333333
4.666666667	3.666666667	3.666666667
3	3.333333333	3.333333333
3.333333333	3	3.666666667
4	3.666666667	4.666666667
3	2	3.333333333
4.666666667	4	4.666666667
2.333333333	3.333333333	2
4.666666667	3.666666667	3.666666667
3.666666667	3.666666667	5
4.666666667	4.333333333	5
4.666666667	4.666666667	5
4.666666667	4.666666667	4.666666667
5	4.333333333	4

4	3.666666667	3.666666667
4.666666667	3.666666667	4
2.666666667	2	2
4	4.333333333	4.333333333
5	5	5
3.333333333	4	4.333333333
4.333333333	5	4.333333333
5	5	4
3.333333333	5	4.666666667
5	1	5
5	3.333333333	3.666666667

4.2.2 Descriptive Statistics

Now we provide two types of descriptive analysis -- UNIANOVA and ANOVA. UNIANOVA or Univariate Analysis of Variance, is a statistical comparison of groups using multiple models and their estimations. UNIANOVA provides regression analysis and analysis of variance with focus on a singular variable compared against several others. ANOVA is a statistical procedure that compares multiple models and their estimations to determine a correlation between two groups. Whereas UNIANOVA assumes the presence of two or more independent variables, in our case, Medium and Game, ANOVA performs the same analysis with only one independent variable, the Medium. This allows us to compare the different mediums independent of the games.

Section 1: UNIANOVA

The Estimated marginal means is the average scores from a group or subgroup in an experiment. We use the estimated marginal means from the following models to better see how our audio structure compares with other telecommunications software discussed.

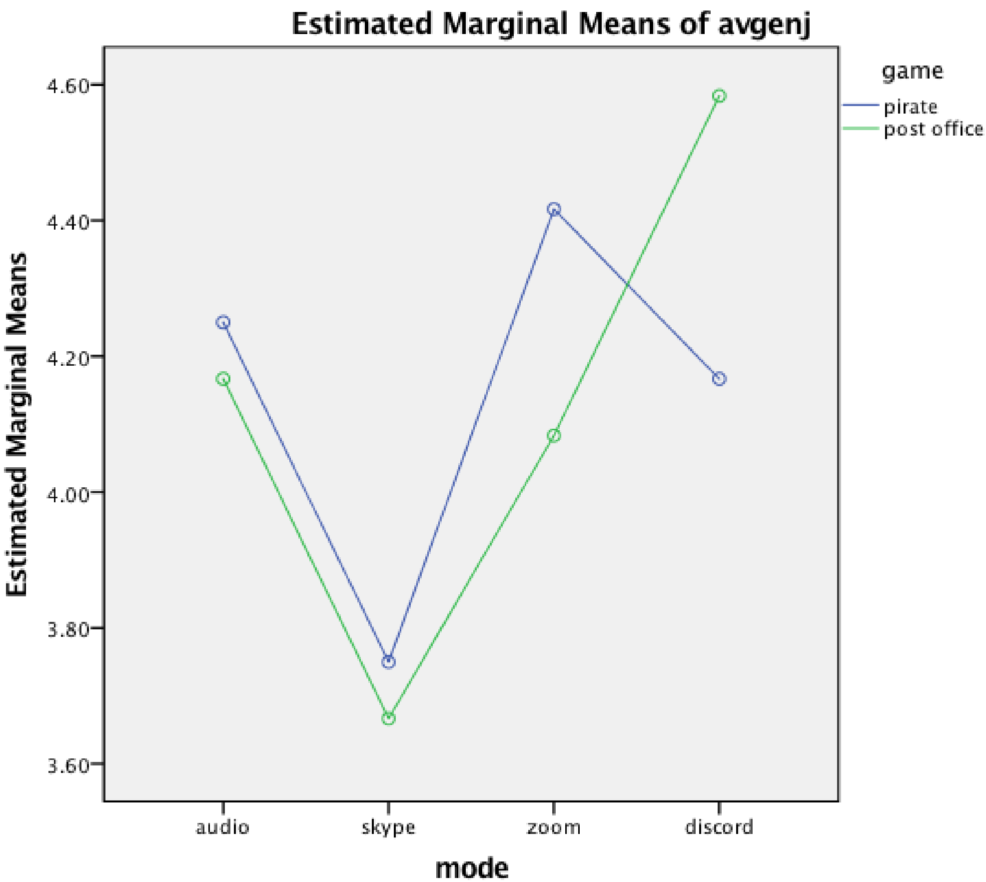

Figure 4.4: Estimated Marginal Means of avgenj

In terms of average enjoyment, our audio structure performed better than Skype and Zoom for the Post Office of the future game, and better than Skype and Discord for the Pirate game. Figure 4.7 compares the average enjoyment of our audio with other mediums combined.

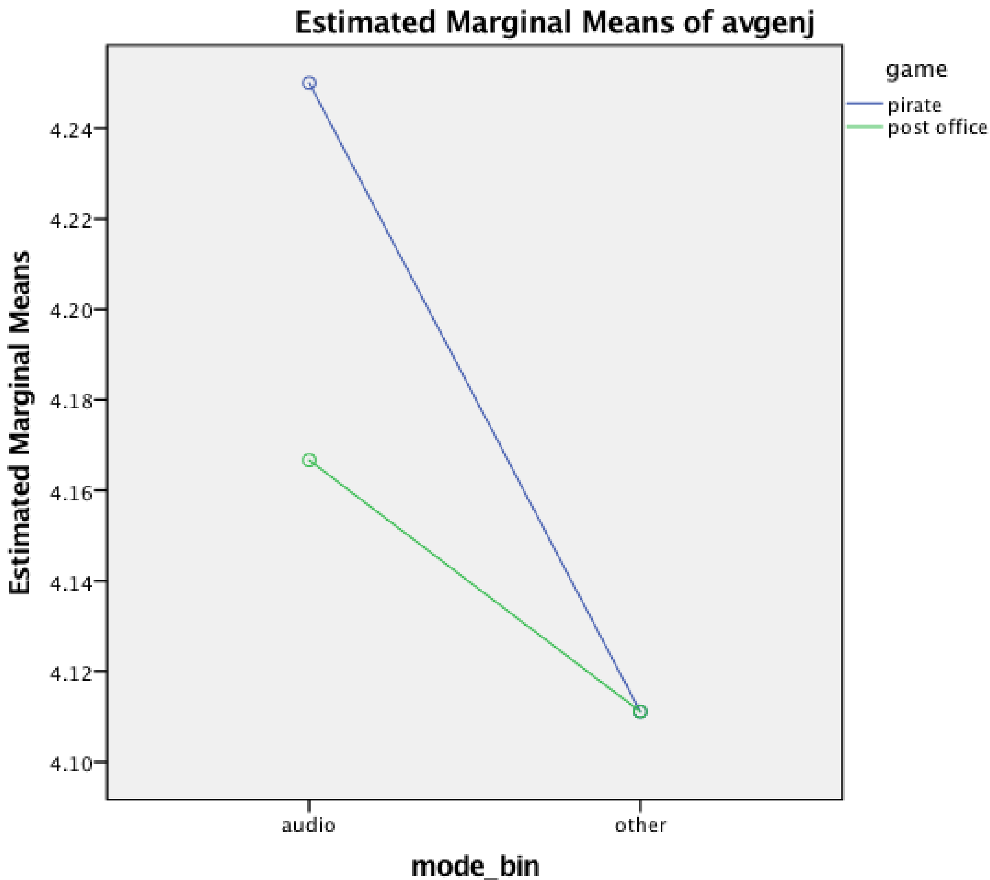

Figure 4.5: Estimated Marginal Means of avgenj compared between audio and all other mediums combined

In terms of enjoyability participants in the experiment tended to favor the Pirates experiment more than the Post Office of the Future. The audio structure had a high enjoyment rating than the combined marginal means of the other mediums for both games.

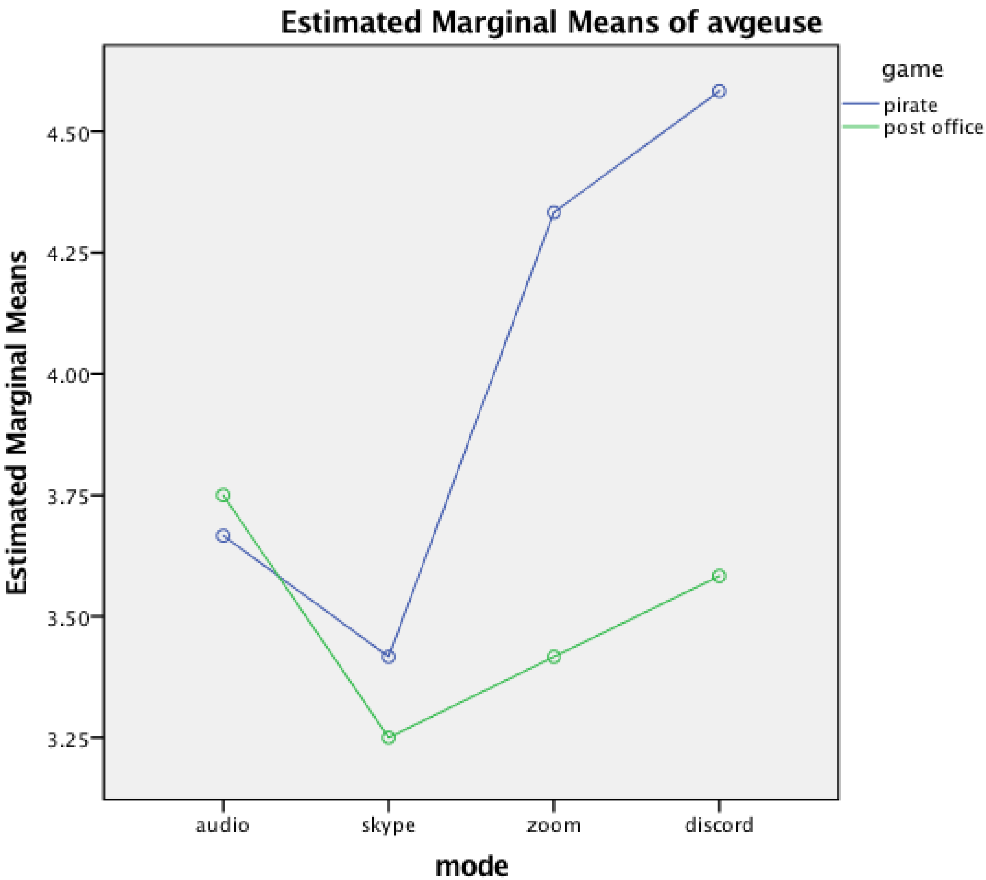

Figure 4.6: Estimated Marginal Means of avgeuse

Figure 4.6 shows the estimated marginal means of average usability for each medium and game pair. Our audio structure fell behind both Zoom and Discord for the Pirates game but was rated the overall highest in terms of usability for the Post Office game.

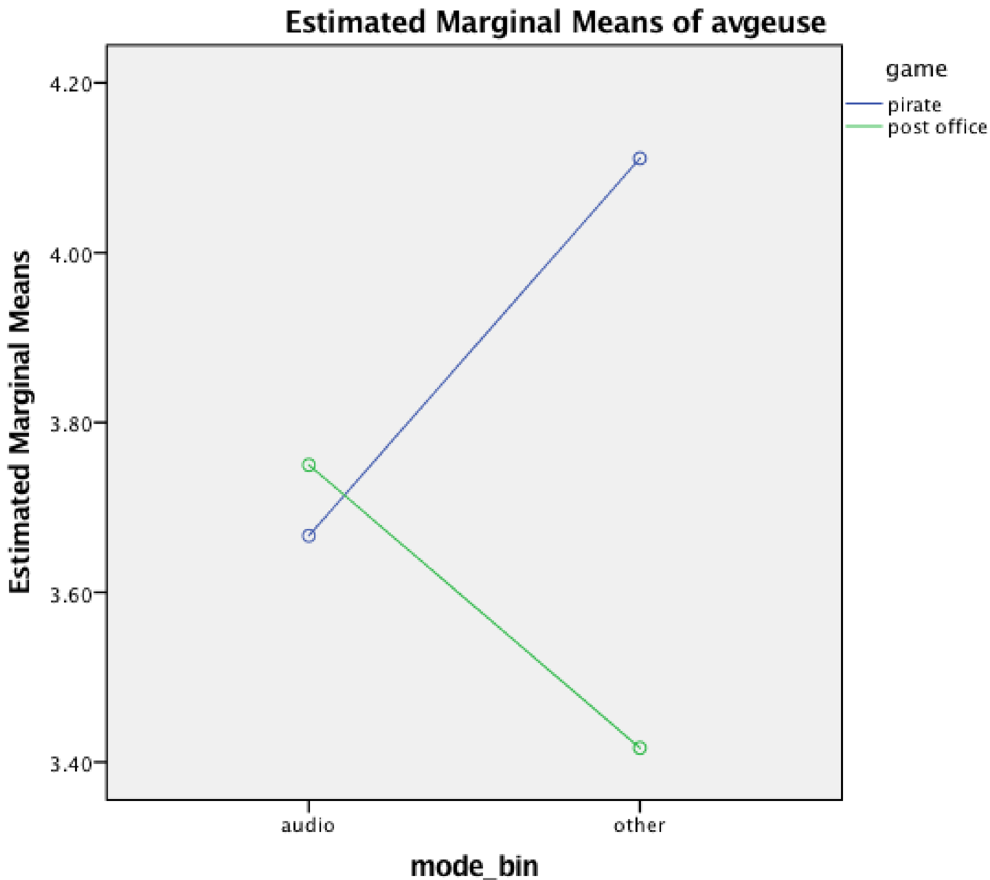

Figure 4.7: Estimated Marginal Means of avgeuse compared between audio and all other mediums combined

This graph shows that our audio structure had an overall lower average usability for the Pirates game, but higher for the Post Office Game. Our audio region performs well in tests with a low degree of uncertainty and a high degree of equivocality because of the users access to audio regions. The ability to breakdown round table discussions allows for a continuity of conversation. Lober et al. [3] discuss the idea of Parallelism: the ability of a medium to handle multiple consecutive conversations at once. Our proposed audio structure can break down conversations into smaller overlaying parts, increasing the parallelism of the medium.

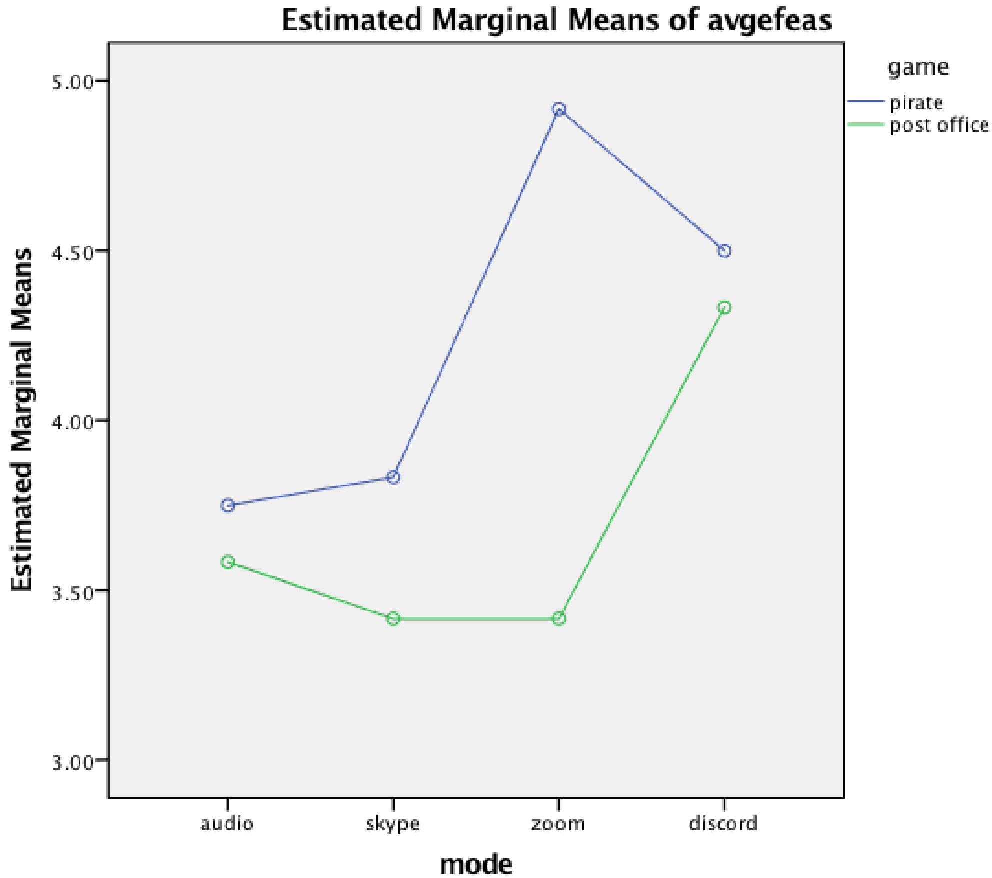

Figure 4.8: Estimated Marginal Means of avgefeas

Feasibility is the ability of a medium to assist in the convergence of an answer. Our audio structure performed better than Skype and Zoom in the Post Office game and fell behind all other mediums with the Pirates game. This may be largely in part due to the unsophisticated nature of the audio structure. Skype, Zoom, and Discord, were all created and maintained by teams of people. With a smoother interface, the score for our audio structure may have come in line with the other mediums.

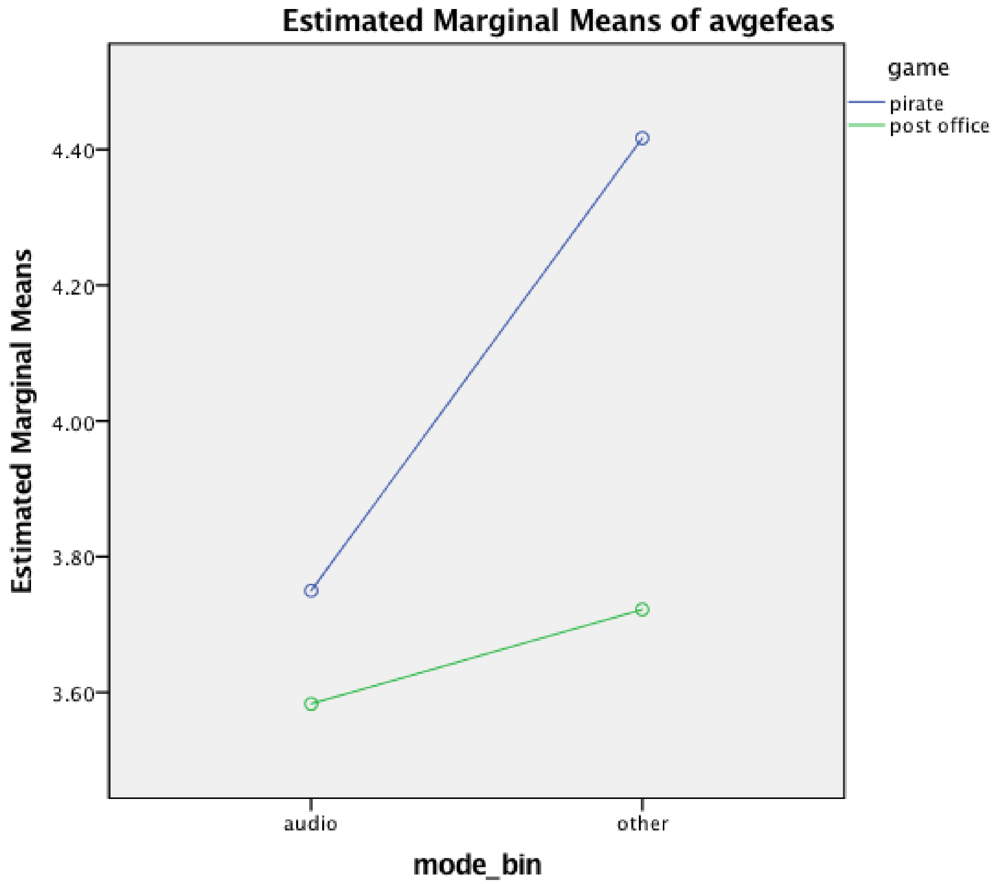

Figure 4.9: Estimated Marginal Means of avgfeas compared between audio and all other mediums combined

Much like the previous graphs, this model depicts the average feasibility of the audio structure compared to the average of all other mediums. It is in this category that it falls behind all other mediums.

Section 2: ANOVA

ANOVA is a collection of statistical models and their associated estimation procedures used to analyze the differences among group means in a sample. Using ANOVA a specific variable is pulled aside and performed a t-test to examine the equality of several groups.

Table 6: Between-Subjects Factors

Between-Subjects Factors

		Value Label	N
mode	1.00	audio	8
	2.00	skype	8
	3.00	zoom	8
	4.00	discord	8
game	1.00	pirate	16
	2.00	post office	16

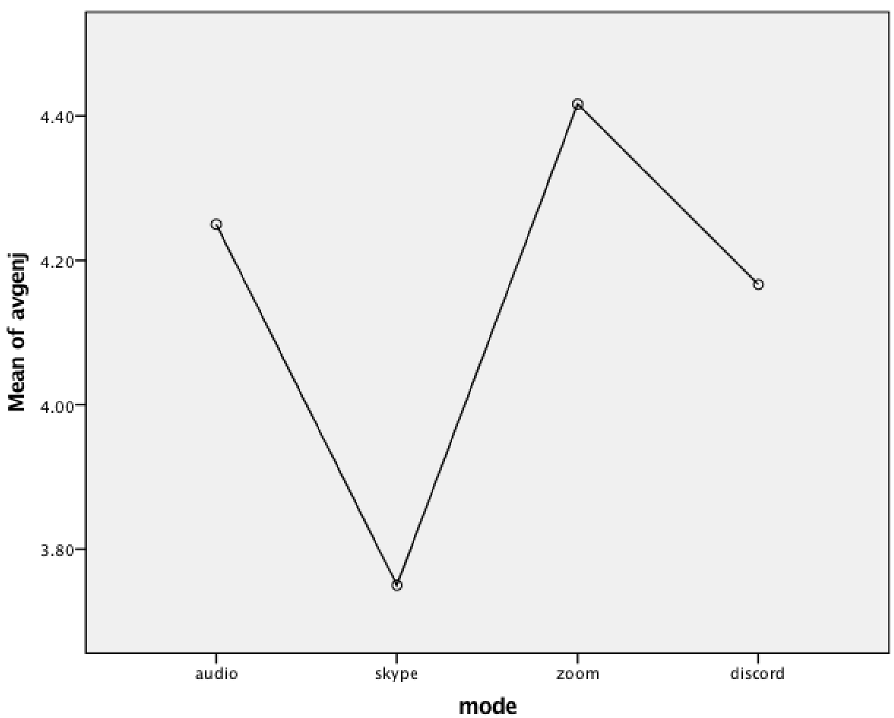

Figure 4.10: Estimated Marginal Means of avgenj

For a total of 32 surveys, we had 16 surveys per game, coming out as 8 surveys per medium. We got statistically significant results even from small sample sizes, we shown in table 6

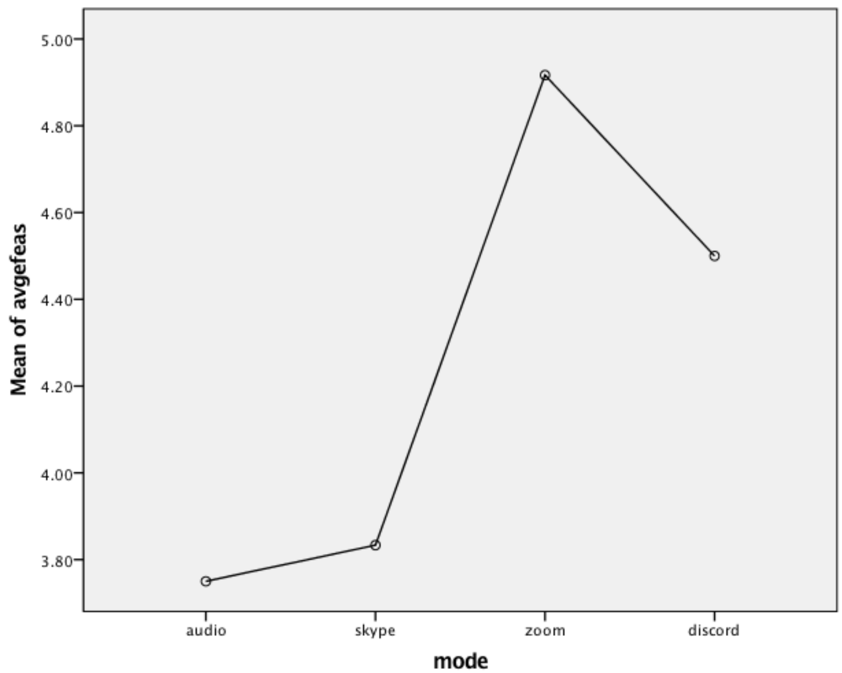

Figure 4.11: Estimated Marginal Means of avgfeas

Figure 4.11 depicts the mean of average feasibility across the 32 participants for four mediums. The audio structure seems to be comparable with other mediums. Much like the graph Figure 4.11, Figure 4.12 compares the mean of average usability for each medium.

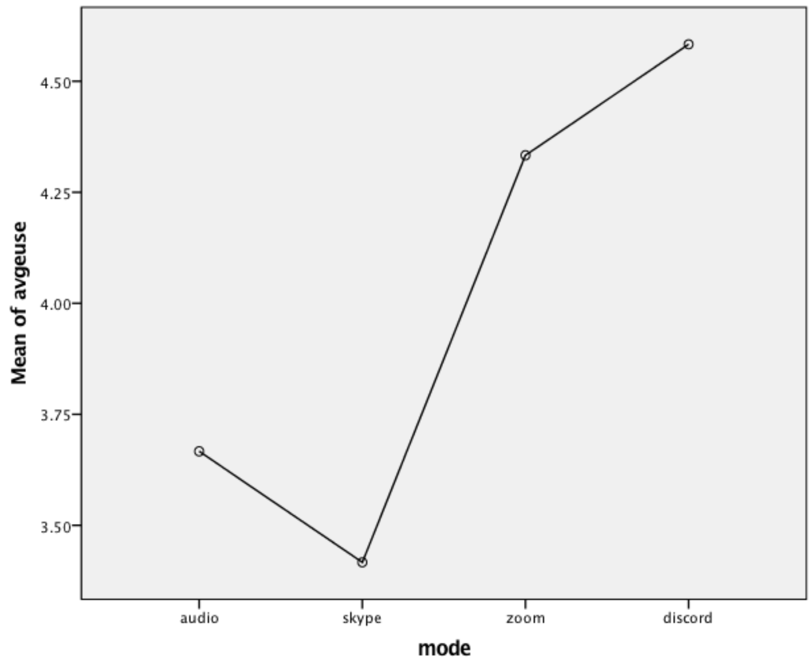

Figure 4.12: Estimated Marginal Means of aveguse

It is important to note that the due to low statistical power, the results obtained above may not be significant. However, this chapter can serve as a prototype for a large-scale experiment, and we hope to get statistically significant results if sample size is increased and by performing the same analysis.

Chapter 5

Summary and Future Work

This thesis seeks to identify the positive potential of our proposed audio structure in a group environment. Using the Theories of Media Richness [1] and Media Synchronicity [2], we created a multi-tier audio environment to segment communication with increasing group size. We composed a series of experiments using two games which provide different scenarios based on qualities identified by the above theories. We performed a comparative analysis of our audio structure with other popular audio and chat communications.

Our audio structure was developed using the UNITY game engine. It features a small avatar the user may use as a vehicle to move around an open environment. A voice chat implementation allows users to communicate using audio data, and a chat box system allows users to communicate via text chat. Within the environment is a series of audio regions. These audio regions house their own audio and text channels. When an avatar enters one of these regions they are subscribed to respective audio and text channels, gaining permission to communicate with others within that region. Players outside of the audio region may broadcast audio and text information to players within the region, but players inside the audio region cannot be heard by others outside of the region. The final tier of the audio structure is a private chat. Users may enter a name of another user into a name box and connect with them privately through text or audio.

To test the effectiveness of our audio structure, we constructed experiments and compared with other popular audio and chat communications like Zoom, Skype, and Discord. As mentioned before, we used two types of activities or games based on the theories described above. The Pirates game is a logical puzzle that tested a group's ability to logically arrive at a consolidated solution, and a Post office game allowed users to engage in a round table type discussion. By pairing each medium with a game, we constructed 8 experiments to test a total of 32 participants. The users had 15 minutes to familiarize themselves with the audio environment, and then 30 minutes to participate in the game. Afterwards the participants were given a 9-question survey to evaluate based on 3 dependent variables – enjoyment, usability, and feasibility.

The post office of the future game tests a low degree of uncertainty and a high degree of equivocality. From our results we found that the audio structure had an overall higher enjoyment rating than the combined average of all other mediums, though it fell behind discord when compared individually. In terms of usability the audio structure performed better both individually and at large than the other audio structures. On terms of feasibility the audio structure outperformed both Skype and Discord but fell short of the combined average of the other mediums.

The pirates game tests a high degree of uncertainty and a low degree of equivocality. The results of our analysis found that our audio structure outperformed both Skype and Discord and had a higher overall enjoyment rating against the combined average of the other mediums. In terms of usability, our audio structure fell behind Zoom and discord, and fell below the average of the other combined mediums. In terms of feasibility our audio structure fell below all other mediums in all other categories.

Our initial conclusion is that our audio structure performs best in environments with a low degree of uncertainty and a high degree of equivocality, such as our Post Office experiment. This supports the original design of the audio structure to reflect a classroom-oriented environment. Where problems are often given way as a round table discussion. However, our audio structure performed well in terms of enjoyment during situations of high uncertainty and low equivocality, such as in the Pirates game. This shows that even during unideal environmental situations, our audio structure was still enjoyable to use.

In terms of feasibility our audio structure fell short. This could likely be due to the prototype nature of the audio structure. With a more polished interface and more intuitive key mapping, it is possible this could be re-evaluated to be more effective at solving the problem at hand. It is also important to note that the experiments conducted above had a small sample size, which did not give us statistically significant results. But we believe this analysis can be easily scaled to provide statistically significant results.

As future work, the user interface can be refined. The mandatory push-to-talk system can be replaced with an automatic voice chat option. We could also provide a more intuitive key-mappings for easier use of the software. We could also conduct our experiments with a much larger sample space to get statistically significant results. We hope that this thesis can be used as a prototype for future research in online collaborative environments.

References

[1] R. L. Daft and R. H. Lengel. *Organizational information requirements, media richness and structural design*. Manage. Sci., 32(5):554–571, May 1986.

[2] Alan R. Dennis and Joseph S. Valacich. *Rethinking media richness: Towards a theory of media synchronicity.* IEEE Comp Soc, 1999.

[3] Andreas Lober, Sibylle Grimm and Gerhard Schwabe. *Audio vs chat: Can media speed explain the differences in productivity?* In ECIS, 2006, pages 2172–2183.

[4] Andreas Lober, Gerhard Schwabe and Sibylle Grimm. *Audio vs. chat: The effects of group size on media choice.* In HICSS 2007. 40th Annual Hawaii International Conference on System Sciences, 2007, pages 41–41.

[5] Skype support: https://support.skype.com/en/faq/FA6/what-is-skype

[6] Skype Wikipedia: https://en.wikipedia.org/wiki/Skype

[7] Zoom: https://zoom.us/about

[8] Discord: https://discordapp.com/

[9] Erdfelder F. Faul, A. E. Lang and A. Buchner. *G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences.* Behavior Research Methods, 2007, 39(2), 175-191.

[10] F-test: https://en.wikipedia.org/wiki/F-test

Vita

Broderick Michael DeSantis was born in Glens Falls, New York, to Carl and Debra DeSantis. He graduated from Appalachian State University with a Bachelor of Arts in Interdisciplinary Studies in May 2013. Having acquired a degree in the arts, he wanted to pursue a degree in the sciences, and so returned to Appalachian State University in January 2015 to pursue a Master of Science in computer science. For his thesis, he worked with the Department of Education to develop a virtual collaborative environment for academic purposes. In August 2019, he was awarded his Master of Science degree.

He is currently working for Allscripts as an associate software engineer in Raleigh, NC. He is an ardent practitioner of the Afro-Brazilian martial art Capoeira. He enjoys outdoor activities and meeting new people.