Sound Design for Stereoscopic 3D Cinema:

Exploring Current Practice and the Enhancement of Depth Perception Through the Use of Auditory Depth Cues

Christos Manolas

PhD

University of York

Theatre, Film and Television

September 2014

Abstract

Stereoscopic 3D (S3D) cinema has been the subject of experimentation for more than a century. Many attempts have been made towards its establishment both at a commercial and an artistic level. However, due to a number of technical and financial constraints most of these attempts have been, by and large, unsuccessful. With the rapid technological advance of digital technologies over the past few decades, most of these constraints have been addressed. It is, thus, unsurprising that the biggest effort in establishing S3D cinema commercially is taking place at the present time.

The introduction of S3D depth aims at changing the cinematic environment from a 2D monoscopic screen to a viewing zone with a more pronounced representation of depth. In this context it can be argued that in addition to the increased sense of depth in terms of visual cues one could also investigate the possibility of using auditory cues. This could be either in order to enhance the overall sense of depth or to support more efficiently the additional depth of the visuals.

The first part of this thesis involves the introduction of basic concepts related to the topic of the study. Firstly, the general background context is established covering relevant historical topics related to S3D cinema and multichannel cinematic sound. The next part is an introduction to basic audiovisual human perception concepts followed by the presentation of basic cinematic sound mixing topics and their relevance in S3D cinema.

The second part of the thesis involves a series of experiments that attempt to study the effectiveness of two auditory cues (high frequency attenuation and volume alteration) as a means to influence the sense of depth of S3D animation clips. The overall findings are complex and reflect the multi-faceted nature of the topic investigated. Overall they indicate that sound design techniques can enhance the S3D experience and, in particular, that the auditory depth cues studied can, in some cases, have an effect on how we perceive depth in S3D movies.

Table of Contents

Table of Figures

Chapter 1

Chapter 2

Chapter 3

Chapter 4

Chapter 5

Chapter 6

Acknowledgments

Firstly, I would like to thank my supervisor Dr. Sandra Pauletto for her consistent support and assistance throughout the course of this study in spite of the great changes in her personal circumstances and responsibilities brought about by the introduction of the new members in her family. Thanks to Nick for the time he spend proofreading my long and frequently repetitive drafts.

Special thanks to Dr. Gavin Kearney for providing the assistance, tools and technical knowledge without which this study would be much more complicated, time-consuming and costly.

I would also like to thank Professor Nicolas Holliman for allowing me to occupy the facilities and use the equipment at the newly created 3D lab in the Department of Theatre, Film and Television at the University of York, as well as Dr. David Hickman for his assistance with copyrights, paperwork and legal matters related to the study.

Special thanks to the technical and administrative personnel in the Department of Theatre, Film and Television at the University of York for the patience they demonstrated to my constant requests for resources and help and for being always available to help with the multi-natured problems I presented them with.

I would like to thank Mariana Lopez for always being there to patiently listen to my complaints and problems without ever complaining about me occupying most of the space we shared for a very long period of practical experimentation.

To my family, thank you for your understanding and the psychological and financial support throughout all the years of my studies. I would not be able to achieve my academic goals without your constant and generous help.

Finally, thanks to all the people who helped me in any way with this study throughout the past four years and all the people who kindly agreed to wear these heavy and uncomfortable 3D glasses and participate in the experiments.

Last but not least, Annie thank you for your patience with my constant questions and failed attempts to understand statistics. Your insight was key to me being in a position to understand a tiny bit of what at first seemed like an incoherent and random pile of numbers. Also, thank you for being such a good company all these years and for the legendary hotpot meals during our time in York.

Declaration

I declare that this thesis is my own work and that all source material has been referenced. I also declare that parts of this research have formed the basis for the following journal papers and conference presentations:

- Manolas, C., & Pauletto, S. (2009). *Enlarging the Diegetic Space: Uses of the Multi-channel Soundtrack in Cinematic Narrative.* The Soundtrack, 2(1), 39-55.

- Manolas, C. & Pauletto, S. (2012). *Exploring Sound Design in Stereoscopic 3D Cinema: An Investigation of Potential Uses of Auditory Cues as a Means to Enhance the Sense of Depth in S3D Cinematic Scenes*. 3D Storytelling Conference: Moving your story into the third dimension. Ravensbourne College, London, March 22-23, 2012.

- Manolas, C. & Pauletto, S. (2014). *Volume Attenuation and High Frequency Loss as Auditory Depth Cues in Stereoscopic 3D Cinema.* 3D Research. Springer. Volume 5, Issue 3, September 2014.

- Manolas, C. & Pauletto, S. (2014). *Potential Roles of the Soundtrack in Stereoscopic 3D Cinema*. Submitted to the Music, Sound and the Moving Image Journal (Liverpool University Press).

1 Introduction

1.1 Introduction

This chapter sets the basic theoretical and historical premises of this thesis introducing a number of relevant concepts from the past and present of cinema. After stating the main research questions it introduces the aims and objectives and outlines the research scope of the study. Next, a brief background overview is presented in which historical concepts of relevance to S3D cinema and cinematic sound are discussed. Additionally, concepts related to the use of audio cues in the context of S3D cinema are briefly discussed. Finally, the thesis structure is outlined.

1.2 Research Questions

The main research questions investigated in this thesis were the following:

> *Can S3D cinema be enhanced through the use of specific sound design techniques?*

And more specifically:

> *Can the perceived sense of depth of stereoscopic 3D cinematic scenes be increased by the introduction of auditory depth cues in the soundtrack?*

These research questions are based on the idea that although the novelty of S3D cinema is related to vision, both human perception and cinema as a medium rely on cross-modal perceptual processes and approaches (Shams and Kim 2010; Shelton and Searle 1980; Vroomen and de Gelder 2000) that include auditory perception. Thus, it could be possible to influence the overall sense of perceived S3D depth not only by means of adjusting the S3D visual content, but also by providing additional auditory depth cues. If such method for increasing the perception of depth is effective it could be an additional tool in the hands of S3D filmmakers and sound designers in their quest to improve the sense of depth within S3D cinematic scenes.

1.3 Aims and Objectives

This study explores the use of auditory depth cues in the context of the S3D cinema format. The topic was approached with the following aims in mind:

- To investigate possible ways of constructing soundscapes that support and enhance the unique spatial characteristics (i.e. depth) of the S3D cinematic environment.
- To propose sound design techniques and/or strategies that could support or enhance the perception of depth in S3D cinematic productions.

In order to achieve the above aims, a number of objectives were set. These are summarised below:

- To identify differences between the production cycle of 2D and S3D cinema.
- To identify the unique visual spatial characteristics of S3D cinema.
- To understand the role of sound in the perception of S3D cinematic depth.
- To propose a number of sound design techniques that could use the knowledge gained during the study to enhance the perception of depth in S3D cinema.
- To create a number of short S3D cinematic clips that use the proposed ideas.
- To practically evaluate the proposed techniques through a number of experimental perceptual tests.
- To analyze and discuss the results of the tests and derive sound design techniques that could be used to affect the sense of depth in S3D cinema.

1.4 Research Scope Outline

The introduction of S3D visual depth could potentially affect the way filmmakers approach and use the cinematic space and depth. Considering that cinema is an audiovisual medium, it is possible that auditory depth cues may be also used as cinematic tools for the enhancement and support of the sense of depth in S3D cinema. S3D filmmakers may wish to consider using auditory cues and effects that contribute to the increased sense of depth if their artistic vision or the story requires so. In this context, the study of the role of the soundtrack may become an area of interest for academics, sound design artists and film practitioners.

Admittedly, it is difficult to objectively explore and evaluate sound design techniques applied to complex audiovisual environments such as S3D cinematic scenes because the variables are many and the results can often only extend to a limited set of sounds or scenes. However, by proposing and examining ways in which auditory cues could contribute to the S3D experience one can provide a number of tested techniques and ideas that 3D filmmakers would be free to utilize, if and when required. Thus, in addition to possible academic interest the findings of such a study could provide practical tools for S3D filmmaking.

At present, an increased interest in audio technologies and systems suitable for use in the S3D cinema is observed. However, there seems to be limited interest towards understanding the ways in which audio could be used to support or enhance one of the fundamental characteristics of S3D cinema: the perception of depth. The originality of this work and the motivation behind it is to attempt an investigation of this largely unexplored area by means of a series of experiments that evaluate the effectiveness of auditory depth cues within S3D animation clips.

1.5 Background Concepts

1.5.1 Historical Attempts to Create Visually Immersive and Enveloping Cinematic Environments

Although cinema has always revolved around the existence of a 2D cinematic screen, it may be relevant to note that multiple attempts at alternative projection systems took place throughout the years. Most of these involve the enhancement of visual physical envelopment and immersion by extending the cinematic space beyond the borders of the common screen. Examples of this include the experimentation with multiple screen projection, the use of wider screens, as well as holographic and stereoscopic projection. Such historical attempts are relevant as an indication that the cinematic language has always been in a constant reassessment of itself. This reassessment may be still taking place at the present time. The cultural and technological landscape changes of the digital era may have an impact on how we approach our cinematic creations. In this context, it is interesting to try to rediscover and reassess the cinematic production and viewing process as a whole, in light of the advanced capabilities of the audiovisual systems of the present day. This, in turn, relates to the argument that forms the basis of this study: the soundtrack may be one of the tools that could be used by S3D filmmakers to explore new creative ways of using the emerging cinematic environments (in addition to the visuals).

1.5.2 The Invention of Cinema: A Cross-Modal Medium

Cinema can be regarded as the natural evolution of purely visual technologies and techniques, such as photography and painting (Braudy and Cohen 2009: 244; Chion 1994: 143). However, a more thorough examination unveils that its invention can be seen as a collaborative process involving the work carried out in a number of other fields. Motion-pictures *'appeared in the wake of the industrial revolution'* along with a number of other inventions, such as the telephone, the phonograph and the automobile (Cousins 2004: 21; Thompson and Bordwell 2003: 13). It is perhaps this timing that encouraged the use of existing, and often unrelated, technologies as building blocks for the creation of motion pictures. This concept is relevant in the context of this study as it points out the notion of cross-modality in cinema. Acknowledging this can be useful when considering the relationship of visuals and audio and the potential uses of audio cues in S3D cinema.

1.5.3 Cinematic Audio Evolution Since the Advent of Sound Film

From a strictly commercial viewpoint, it is commonly accepted that sound in cinema started in 1927 with the *'The Jazz Singer'* (Thompson and Bordwell 2003: 193). '*The Jazz Singer'* was actually the result of a long and intense battle between inventors and companies interested in incorporating their newly developed audio systems in cinematic production (Cousins 2004: 118; Mitchell 2010; Thompson and Bordwell 2003: 193-211). More specifically, it seems that it was the boom in audio reproduction technologies between the 1900s and the 1920s (Audio Engineering Society 2010; Cousins 2004: 118) that finally made the idea of audiovisual synchronisation in cinema viable, something that caused the interest of inventors, engineers and film executives around the world (Cousins 2004: 118; Thompson 2003: 194-195).

Figure 1-1: In a Warner Bros. sound studio, an exterior dialogue scene for an early talkie is shot. The orchestra plays at the side, and three cameras in booths face the couple on the bench (Source: Thompson and Bordwell 2003: 196)

Since the establishment of sound films, cinematic sound has been subject to a transformation that ultimately led to today's advanced multichannel audio systems. Today, most of the mainstream cinematic productions are released with high definition multichannel audio soundtracks and most theatres are equipped with the required audio systems capable to reproduce these soundtracks.

This evolution underlines the rather important role of cinematic sound. Although the introduction of a unique new cinematic style that employs S3D visuals might not necessarily require a soundtrack that is different to the ones used in 2D cinema, based on this observation one can propose that sound could function as an additional tool in the S3D filmmakers' toolbox.

1.5.4 Stereoscopic 3D Cinema Soundtrack Mixing Concepts

The interest in cinematic immersion has not disappeared for over a century and new digital technologies are rapidly transforming cinema. Such developments could introduce further changes not only to the technological side of audiovisual integration, but also to the creative approaches employed during the creation of the cinematic audiovisual material. In this context, the way filmmakers and sound designers use sound could be also affected, as particular conventions implied by the use of technologies and cinematic environments with limited spatial capabilities may be reconsidered (**Figure 1-2**).

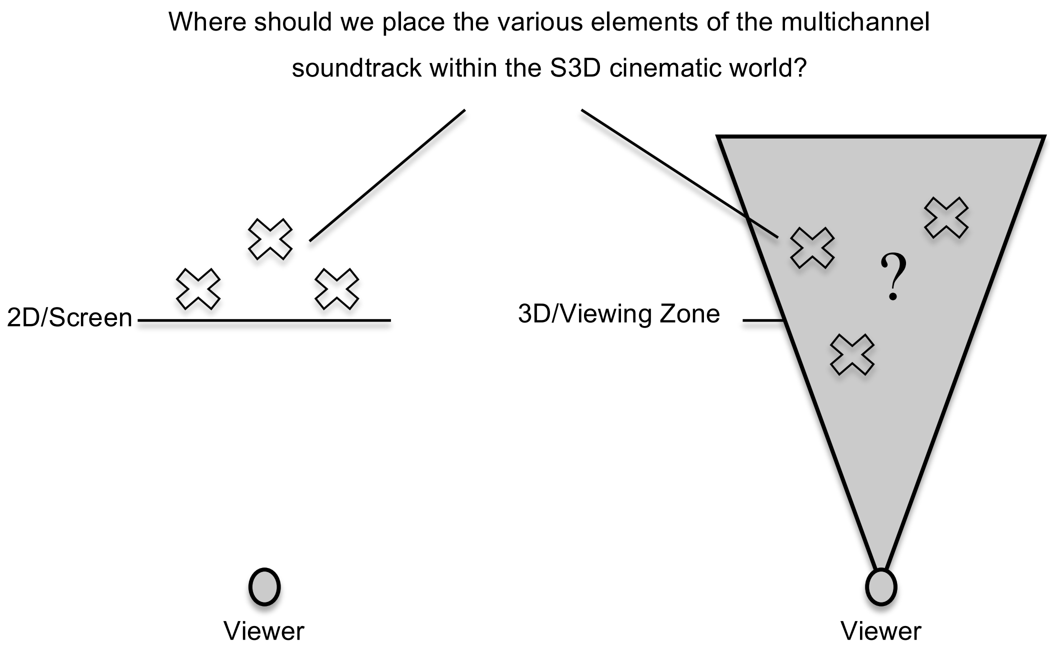


Figure 1-2: Could the introduction of S3D visuals affect the spatialisation of certain sound elements within the soundtrack? If this is the case, how we should place our sounds in this different and unique cinematic environment?

Considering this, it is not surprising that a great effort is currently being made towards the establishment of standards and technological solutions related to immersive audiovisual systems (2020 3D Media 2008; Vanhoutte *et al*. 2010, 3DIIM 2008). However, a relatively small amount of research effort focuses on the creative implications and possibilities the current technological changes may introduce. Based on this, observations regarding the creative potential of the soundtrack in the relatively unexplored area of S3D cinema could be timely and relevant. The proposal and evaluation of possible uses of sound in this emerging cinematic form could be helpful as it could provide additional tools for S3D filmmakers.

1.5.5 Identifying Potential Uses of Audio in S3D cinema

In order to identify any possible impact of S3D visuals to the creation of the cinematic soundtrack the difference between classic 2D and S3D cinema should be questioned. Arguably, the most important and obvious difference between the two types of cinema is the reproduction of visual depth. While in 2D cinema depth perception is based only on monoscopic cues, in S3D cinema stereoscopic cues are also used (Mendiburu 2009: 3; Clark 2010). This effectively transforms the cinematic screen from a flat 2D projection area to a 3D viewing space that expands beyond the actual screen (Autodesk 2008: 1; Mendiburu 2009: 27).

The increased capability of depth presentation of S3D cinema may seem as a simple addition to the existing arsenal of tools of the filmmaker that does not necessarily affect the existing 2D filmmaking workflow. However, this may not be always true, as the addition of stereoscopic depth cues comes with a different set of filmmaking constraints and rules. The S3D filmmaker is likely to devise and employ different approaches towards many aspects of the filmmaking process, such as shot composition and lighting, length of scenes, editing transitions and cuts, camera angle selection and camera movement (Autodesk 2008: 7; Scorsese on Kermode 2010).

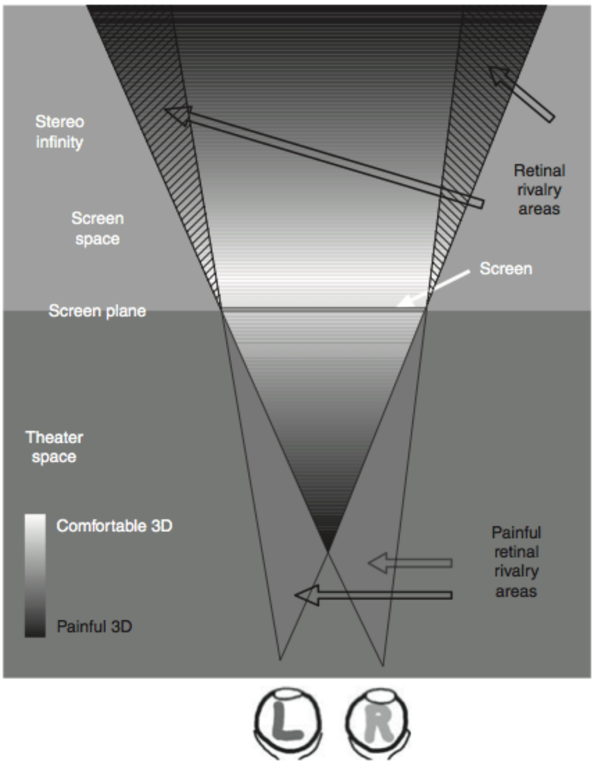


Figure 1-3: Outline of the stereoscopic viewing space (Source: Mendiburu 2009: 82)

In light of this, one may also wish to reconsider and question the possible uses of sound in this context. Although the introduction of S3D visuals in cinema does not necessarily mean that a fundamentally different soundtrack to that of 2D cinema is required, it is possible that alternative uses of sound may be utilized by S3D filmmakers in order to support or enhance the unique features of the medium. The reasons for such uses may vary. For instance, filmmakers may wish to increase the sense of immersion and envelopment of the cinematic space without altering the visual composition of a given scene. In this case, the soundtrack mix may adopt a pronounced front-to-back orientation in an attempt to enhance the sense of depth of the visual action. In addition, it is possible that audio could be used when an increased sense of movement within the S3D space is required. This may involve movements of S3D objects, characters or the camera that could be pronounced by means of appropriate audio effects. Audio could also be used as an additional tool in S3D editing. For example, auditory transitions may be used in order to make visual transitions between scenes with different S3D depth smoother. Finally, audio depth cues could be used in cases where the use of rich S3D visuals is either undesirable or problematic, such as in areas where the story or the artistic approach of the filmmaker indicates so or when the action takes place in uncomfortable S3D areas (**Figure 1-3)**.

1.6 Thesis Structure

An outline of the scope of the chapters of the study is presented below.

Chapter 1 outlines and briefly discusses the overall scope of this study. The motivation behind the research, as well as the research questions, aims and objectives are presented. A brief overview of following chapters and the topics they cover is included.

Chapter 2 firstly covers historical cinematic developments relevant to the study. A historical overview of significant advances throughout the evolution of immersive and S3D cinema is presented covering various different areas of interest, such as notable attempts to create visually immersive cinematic environments, the audiovisual relationship in the early years of cinema and the evolution of multi-channel cinematic audio systems. Next, the chapter covers a number of topics related to visual and auditory perception. An overview of fundamental audiovisual perception and cross-modal integration concepts is presented. The aim of this is to establish a basic understanding of fundamental perceptual principles, as the latter is the basis for the sound design strategies and direction of the later stages of the project. Finally, this chapter includes the introduction of a number of fundamental cinematic soundtrack spatialisation concepts and the discussion of a number of ideas related to the challenges of designing multichannel soundtracks for S3D productions. This includes a brief presentation of common sound spatialisation strategies used in 2D cinema as well as a brief discussion of the reasons and constraints that have led to their broad use. The last section includes the presentation and brief discussion of a number of ideas related to the potential role of the multichannel soundtrack in the context of S3D cinema in particular.

Chapter 3 introduces the approach and methods employed for creating and testing a number of proposed sound design techniques/strategies for use in S3D cinematic productions. The rationale behind the selection of the proposed techniques, as well as an explanation of how these techniques were practically evaluated, are presented. The apparatus and systems used for the creation of the required audiovisual material are presented and the ways this material was processed in order to express the proposed ideas are discussed. Additionally, an overview of the methodology used for the practical evaluation of the above ideas is given.

In Chapters 4 – 6, the rationale and strategies employed for each experiment is explained in detail, the testing procedures are described and the results of each experiment are presented. Chapter 4 describes experiments exploring the effectiveness of auditory depth cues as a means to increase the overall sense of depth of a given S3D animation clip. Chapter 5 describes experiments exploring the effectiveness of auditory depth cues as a means to enhance the perception of depth of S3D objects moving from inside the cinematic screen towards the viewer/listener. Chapter 6 describes one experiment exploring the effectiveness of auditory depth cues as a means to enhance the perception of depth of important static S3D objects within a S3D animation clip.

Chapter 7 summarises the results and findings of the study in relation to the original research questions. Next, the chapter includes a general discussion of based on the main findings and observations of the study and their possible connections and relevance to S3D filmmaking. Finally, an overall conclusion is presented.

1.7 Conclusion

This chapter set the basic context of this thesis by introducing a number of relevant concepts from the past and present of cinema, presenting the main research questions and aims and objectives, and outlining the research scope of the study. A brief background overview was presented in which historical concepts of relevance to S3D cinema and cinematic sound. Finally, concepts related to the use of audio in the context of S3D cinema were briefly discussed and the thesis structure was outlined.

2 Background Overview

2.1 Introduction

This study required basic knowledge of different topics, such as cinematic history and production, technological advances in the audiovisual domain, basic human perception, sound design and mixing and the implications of using stereoscopic 3D visuals in the cinematic context. It was, therefore, relevant to establish a basic understanding of key concepts from such areas of interest in order to be in a position to make informed observations during the study.

In this chapter, firstly the concept of physical immersion and envelopment in cinema is presented in a historical context. This includes the discussion of the various attempts to create physically immersive cinema in the past century, the experimentation with different projection systems and the introduction and evolution of cinematic multichannel sound systems.

Next, basic audiovisual human perception concepts are presented. These include basic visual and auditory perception, as well as their cross-modal integration and its relevance in the cinematic context.

Finally, this chapter includes the introduction and brief discussion of basic sound design and mixing concepts and ideas, as well as a brief presentation of possible ways audio could be used in the context of S3D cinema.

2.2 Cinematic Immersion and Envelopment in a Historical Context

2.2.1 Notable Attempts to Create Visually Immersive and Enveloping Cinematic Environments

Since the establishment of cinema one thing seems to have remained constant: in most cases movies are projected on a 2-dimensional (2D) screen. On a commercial level, cinematic formats and standards have changed surprisingly little over the years considering the large number of attempts that have been made towards alternative projection formats and cinematic forms throughout history (Zone 2007; National Media Museum 2009; Thompson and Bordwell 2003). This could be partly because filmmakers and exhibitors always had to conform to a large number of financial, artistic and practical conventions in order for their products to reach a wide audience. However, the constant quest for expanding the cinematic visuals into something more enveloping and physically immersive than the typical cinematic screen is evident in many stages of cinematic history. It must be noted that, in the context of this thesis, the terms *immersion*, *immersive* and *envelopment* are used to describe a sense of the spectator being physically or spatially immersed in, and enveloped by, the cinematic audiovisual environment, rather than being perceptually immersed and drawn into the story and the narrative. In other words, the term immersion is being used to describe an enhanced sense of being part of the cinematic world physically and not contextually. In the following sections some of the most significant attempts to create cinematic environments with increased visual spatial characteristics are presented.

Cineorama

Cineorama (Belton 2004; Cousins 2004: 31-32, **Figure 2-1**) is one of the earliest cases of experimentation with visually surrounding cinematic environments. It was introduced by Raoul Grimoin-Sanson in the Paris Exposition of 1900 (Belton 2004; Cousins 2004: 31-32).

A very short time after the establishment of cinema as a public form of entertainment Grimoin-Sanson came up with the idea of creating an enveloping, 360^o projection zone consisting of 10 screens placed around the audience. The movie was shot accordingly using 10 cameras placed around a flying balloon in such a way as the viewing angle of each of the cameras to correspond to the position of one of the projection screens (Belton 2004, **Figure 2-1**). During the projection, the audience stood inside a static balloon basket watching the recorded flight sequence unfolding around them in a 360^o plane.

This ambitious idea did not go a long way, mainly due to technical and practical

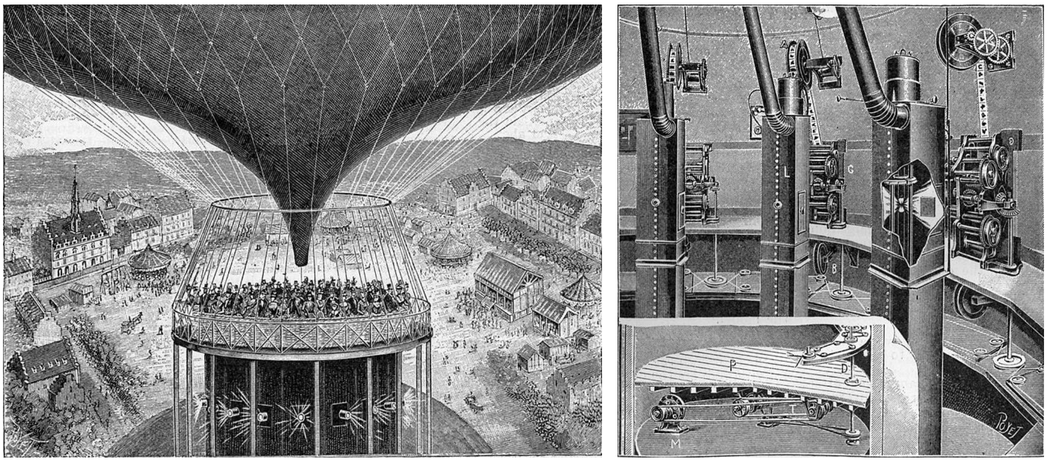


Figure 2-1: Grimoin-Sanson's Cineorama (source: Scientific American Supplement 1900)

constraints and limitations. As it was not possible for the exhibitors to deal with the extreme heat levels produced by the concentration of the required cameras and projectors, Cineorama had a short lifecycle of only 2 performances before being stopped for safety and comfort reasons (Cousins 2004: 32).

Circle-Vision 360°

Although Grimoin-Sanson's Cineorama abruptly stopped in its infancy, the idea of using multiple screens for expanding the cinematic projection zone survived and re-appeared in various different forms at later stages. Possibly one of the most widely known of these visually surrounding systems is Disney's Circlevision 360° (CircleVision 2008; Smith 1996), a system consisting of either a series of screens (usually 9) placed around the audience or of one surrounding curved screen (CircleVision 2008). The system was firstly introduced in 1955 and has been used in several occasions ever since, either as an attraction show or for long-term screenings of specially designed 360° films (e.g. 'O Canada!', 'Reflections of China') in the theme parks of the Disneyland franchise (Shatnoff 1967; Smith 1996, Disneyworld 2010).

Cinerama and Cinematic Domes

A more modest implementation of the idea of multiple screens was Cinerama, a system consisting of a curved screen forming a U-shaped panoramic projection zone in front of the audience (Thompson and Bordwell 2003, **Figure 2-2**). Although the visuals in this system did not completely envelope the audience they *'went beyond the field of vision giving the illusion that the audience was actually "in" the film'* (Abrams *et al.* 2001: 83). An enhanced implementation of the same idea appeared later in the form of the Dome systems, a bell-shaped projection environment supporting a 360 degree projection of visual content around the audience. Commercially available examples of this are the OMNI IMAX / IMAX Dome (Lantz 2003, 2006; Shaw and Lantz 1998). In these systems, the array of screens of Cinerama was replaced by a large spherical screen/dome covering the field of view of the audience, while in many implementations the separate projectors were also replaced by specialized *fish-eye* lenses providing seamless panoramic images suitable for curved screen projection (Shaw and Lantz 1998; Lantz 2006, **Figure 2-2**).

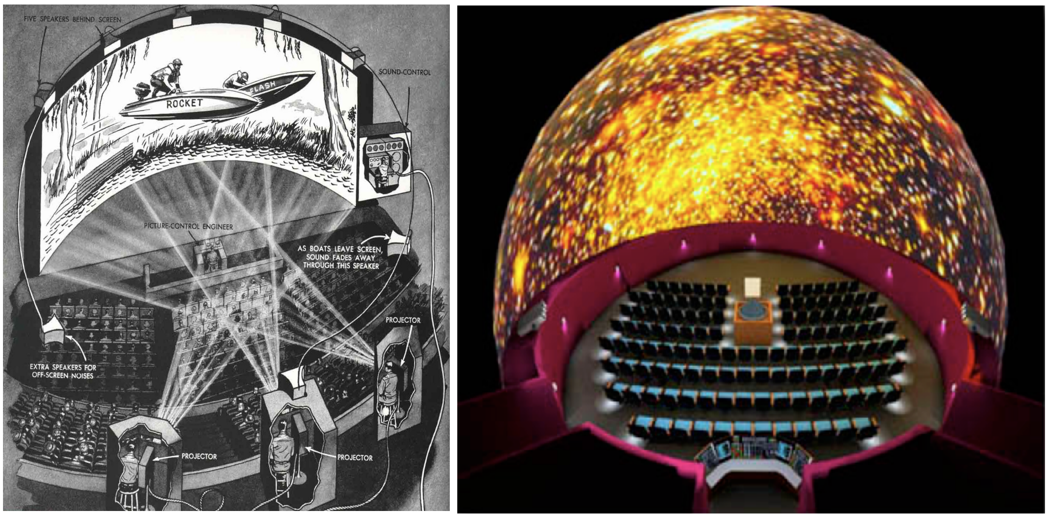


Figure 2-2: Cinerama (source: Zyber 2013) and Cinematic Dome (source: Shaw and Lantz 1998)

Cinematic domes drew from the idea of the Planetarium domes firstly introduced in the 1920s as a means *'to recreate the nightime sky'* (Shaw and Lantz 1998; Lantz 2006). The two systems are similar in the sense that they both require a spherical projection zone and panoramic projection equipment. Over the years, this similarity caused the boundaries between them to be blurred. Exhibitors realised that they could host both Planetarium shows and cinematic screenings using the same facilities and equipment. This led to several multipurpose domes installed in exhibition centres, technology museums and theme parks around the world, exploited both for educational and entertainment purposes (Shaw and Lantz 1998; Lantz 2006).

Widescreen Projection

The idea of Cinerama, and subsequently of the cinematic domes, could be also connected to the concept of widescreen projection (Thompson and Bordwell 2003: 330-332), as the latter could be viewed as the first step towards the implementation of the former. Widescreen projection, a format that meant to be the mainstream cinematic standard at later stages, was firstly introduced in 1927 in Abel Gance's *Napoleon*. Gance placed three screens/projectors side by side in order to create a flat, but extra wide, projection zone, called *the triptych* (Thompson and Bordwell 2003: 93-94, **Figure 2-3**).

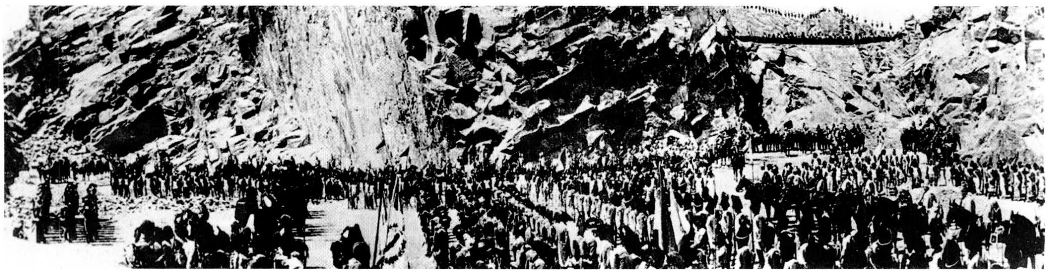


Figure 2-3: Abel Gance's Napoleon used three screens (i.e. 'The Triptych') to create an ultra-wide projection zone (source: Thompson and Bordwell 2003)

Along other early experimental attempts, Gance's triptych started a slow but steady revolution of widescreen cinema that eventually led to today's large format widescreen projections, such as the ones in modern IMAX theatres (Lantz 2003). It can be claimed that this gradual transition from the early (1.33:1 and 1.37:1 ratio) film formats to the widescreen projection of the 1950s, and later to the large format screens of modern digital cinemas, also underlines an interest in the creation of physically immersive cinematic environments as one of its aims is to spatially expand the screen in order to increase envelopment (National Media Museum 2009; Lantz 2003; Thompson and Bordwell 2003: 330-331).

Holographic Cinema

Significant attempts towards creating a 3-dimensional cinematic environment have been made by several individuals and organizations around the world using the holographic method (Holocinema 2005). Holography, a technique invented in the late 1940s by Nobel Prize winner Dennis Gabor (Nobelprize 2010), had the advantage of providing images that were truly 3D, in the sense that they could be *'viewed from different angles and without glasses'* (Holography 2008; Winslow 2007; Holocinema 2005).

The first holographic filming is attributed to M. Lehmann in 1966, while many others also experimented with holographic filming and projection in the 1960s and 1970s (Holocinema 2005). However, *'the scientists and artists soon realized that holographic films showed more the limitations than its potential, which was attributable to the immense technical difficulties involved in this technology'* (Holocinema 2005). Despite the fact that these difficulties made holography impossible to be used commercially for cinematic purposes at that point, a number of researchers kept on working on the field over the years.

Figure 2-4: The holographic image of a presenter from London was broadcast to live audience in Berlin in 2008 during a telecommunications conference (source: Payatagool 2008)

One of the most significant attempts towards developing a fully functional holographic cinema system took place at the National Cinema and Photo

Research Institute (NIKFI, Holography 2008) in the Soviet Union, under the supervision of professor Victor Komar. Over the years, professor Komar's team developed a system capable of capturing and projecting holographic colour 3D models in space for multiple viewers. The team's work was recognised and respected around the world and led to a number of technological awards including an Oscar Award for Technical Achievement in 1991 (Holography 2008; Holocinema 2005).

By the early 2000s Komar's team announced that they were able to provide Holographic cinema systems and services for commercial use (Holocinema 2005). However, the system has not been commercially exploited on a large scale as of today. Nevertheless, at present there is a renewed commercial interest on the technology with many companies offering their implementations of the idea (Payatagool 2008).

Stereoscopic 3D (S3D) Cinema

A prominent and long-lasting obsession of cinema inventors was this of adding the illusion of stereoscopic depth to their visuals. Stereoscopic depth can be achieved by introducing an offset to the images viewed by the left and right eye of the viewer. This offset corresponds to the natural distance between the human eyes and it provides additional visual depth cues when the images are viewed separately by the corresponding eye (Left or Right, **Figure 2-5**).

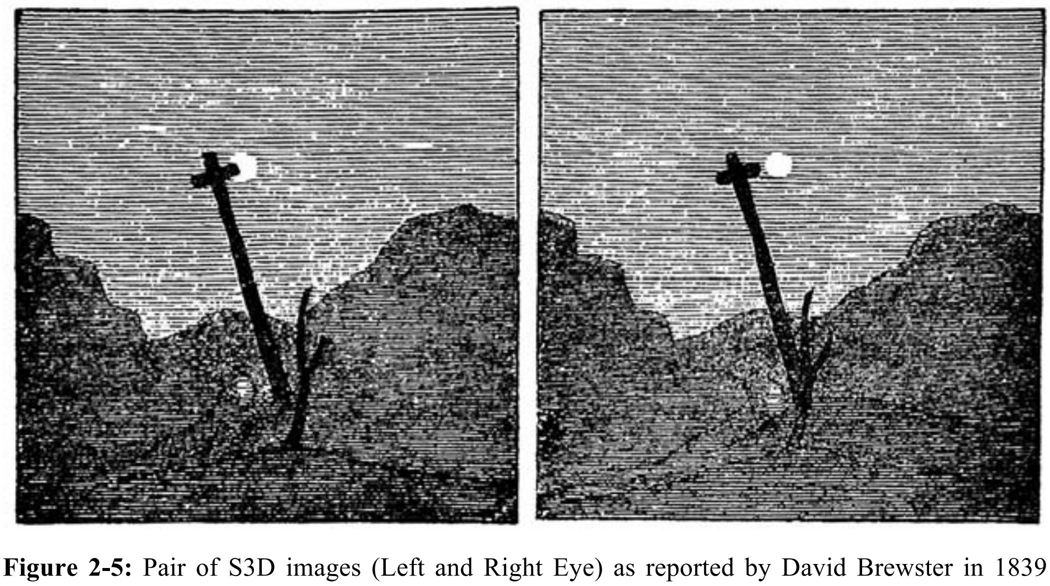


Figure 2-5: Pair of S3D images (Left and Right Eye) as reported by David Brewster in 1839 (source: Zone 2007: 10)

Stereoscopic viewing had been already achieved for still photographs long before the invention of cinema. Thus, it did not take long for early cinema inventors to start thinking about introducing this technique to moving images and films (Zone 2007: 1; Thompson and Bordwell 2003: 13, 21, **Figure 2-6**). Leading early cinema figures, such as Thomas Edison and the Lumiere brothers, appear to have been interested in giving *'a stereoscopic effect to the pictures'* they were making in the 1890s (Zone 2007; Holocinema 2005). From this early moment and until the early 1950s a number of attempts were made towards the production and exhibition of stereoscopic 3D (S3D) movies, using various techniques (Zone 2007: 1-2). However, as *'the short stereoscopic films of the novelty period* [1838 to 1952 according to Ray Zone] *were characterized by an emphasis on the technology of 3-D or the "gimmick" of the off-the-screen imagery'* they failed to be established as a long-term commercially viable form of entertainment after their novelty value faded (Zone 2007: 1-2).

Figure 2-6: Replica of a S3D camera constructed in 1890 (source: Zone 2007: 62)

The next significant collective attempt to introduce S3D cinema to the masses can be traced in the early 1950s (Cousins 2004: 223-224; Zone 2007). Between 1952 and 1955, a *'protean 3-D movie boom'* took place in Hollywood, with more than fifty stereoscopic films being released (Zone 2007: 2). However, S3D cinema failed again to be permanently established, for various reasons, such as the *'cumbersome, extremely hot filming conditions'* caused by the large amount of lights needed on the sets for such productions and the '*the awkwardness of the glasses'* (Cousins 2004: 223-224).

Despite the fact that stereoscopy failed to be established as a cinematic component on several occasions, the idea of S3D cinema survived. After being implemented for decades mainly as a theme park attraction with limited cinematic interest, it had another period of resurgence that is currently underway.

Conclusion

This brief exploration of historical attempts to create psychically immersive and enveloping cinematic formats could be taken as an indication that although the typical (2D) cinematic screen might have been the medium of choice for decades it may not have to be the only one. This is underlined by the observation that most of the attempts at alternative projection systems involve the enhancement of visual physical envelopment and immersion by extending the cinematic space beyond the borders of the common screen. This is relevant as an indication that the cinematic language has always been in constant reassessment of itself. As the cultural and technological landscape changes, so may do the way humans try to express themselves by means of their cinematic creations. It is this constant quest that makes it interesting to try to rediscover and reassess the cinematic production and viewing process as a whole, in light of the advanced capabilities of the visual systems of the present day. This, in turn, is related to the argument that forms the basis of this study: the soundtrack may be one of the tools that could be used by S3D filmmakers to explore new creative ways of using the emerging cinematic environments (in addition to the visuals).

2.2.2 Audiovisual Integration in the Silent Film Years

The main focus of the early cinema years was arguably on implementing efficient moving image capturing and screening devices. However, it seems that cinema inventors were also aware of the importance of sound as a means to achieve a fuller cinematic immersion. This is evident from the early days of cinema, where in most cases some form of audio accompanied the movie screenings. Early attempts of cinematic sound involved experimentation with live or recorded background music and synchronised sound effects or even dialogue (National Media Museum 2009; Thompson and Bordwell 2003: 37-38). Notable examples of the use of sound as an immersive cinematic tool in early cinema are presented in the following sections.

Background Music in Silent Cinema

It is broadly known that silent film screenings were usually accompanied by background music (Chion 1999: 8; National Media Museum 2009) despite the term silent being used to describe them. The music was either performed by musicians and orchestras or played back using mechanical devices, such as phonographs (Wierzbicki 2009: 14-20; Chion 1999, 8; Thompson and Bordwell 2003: 37-38). Various reasons might have led to the decision to add music in those early screenings, such as *'to identify locales and time periods, to illustrate on-screen action, to limn the basic personality traits of the characters, and to explore their thoughts and feelings'* (Wierzbicki 2009: 6). Nevertheless, even if the cinema projectionists of this early period did not treat film music in such an elaborate way, they were probably aware of the fact that a completely silent screening would not be enough to impress an audience already used to showings with rich audio content, like vaudeville theatre, circuses and live concerts (Dix 2008: 77; Wierzbicki 2009: 20; Thompson and Bordwell 2003: 22). In Wierzbicki's words, *'then as now, film music by and large has offered a gloss on whatever might be presented overtly to the audience'* (Wierzbicki 2009: 6).

Early Attempts for Audiovisual Synchronisation

Experimentation with sound in the early film days did not involve only background music, but also attempts for synchronised sound effects and dialogue. According to Thompson and Bordwell, *'in some cases, actors stood behind the screen and spoke dialogue in synchronisation with the action on the screen'* (Thompson and Bordwell 2003: 37).

Additionally, early film exhibitors also attempted to add synchronised sound effects to the visual action manually. Thompson and Bordwell report that *'people* *used noisemakers to create the appropriate sound effects'* in sync with the image (Thompson and Bordwell 2003: 37, **Figure 2-7**).

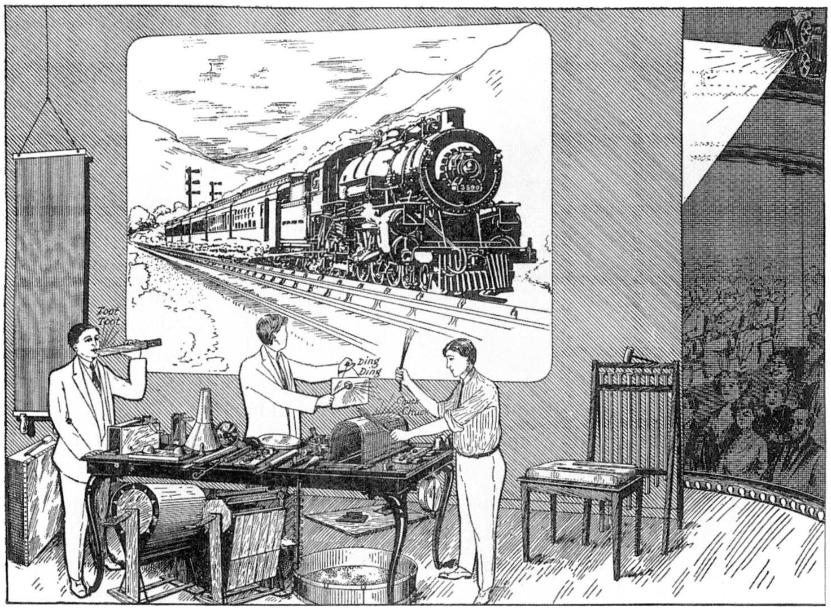


Figure 2-7: *'A drawing from a 1909 Gaumont catalogue showing how sound effects could be produced behind the screen'* (source: Thompson and Bordwell 2003: 38)

The possibility of introducing synchronised sound to cinema was also tempting to two of the leading figures of the early cinema days, Thomas Edison and W.K.L. Dickson (Dickson and Dickson 1970). In 1894, the pair attempted to achieve what is probably the first film with mechanically reproduced synchronised sound in history (Loughney 1999; The Library of Congress 2009, Murch 2000).

Figure 2-8: Frames from W.K.L Dickson's experimental sound film. This is probably the first attempt at mechanical audiovisual synchronisation (source: The Library of Congress 2009)

This was done by means of a phonograph playing pre-recorded audio material in sync with the corresponding recorded video (**Figure 2-8**). Their interest in integrating sound to their movies probably had to do with the fact that they had previously worked both on visual and on audio technologies and systems (i.e. Edison had invented the Phonograph in 1877), something that enabled them not only to envision the possible immersion created by the audiovisual integration, but also to have the technical knowledge to achieve it (Audio Engineering Society 2010; Dickson and Dickson 1970).

According to the Audio Engineering Society, by 1913 Edison had perfected his audiovisual synchronisation methods and presented *'the first talking movie using his Kinetophone process'* (Audio Engineering Society 2010). Indications that W.K.L. Dickson had achieved synchronous film sound even earlier (i.e. between 1889 and 1893) could not be confirmed, as no copies of this speculated attempt are believed to have survived (Dickson and Dickson 1970; Loughney 1999). Nevertheless, it is remarkable that by the mid-1890s not only the idea of synchronous film sound had been conceived, but also functional (albeit primitive) systems had been built in order to achieve it.

These early attempts at adding sound to motion pictures underline another key concept: the creators of cinema appear to have realised at a very early stage that sound may be helpful in order to achieve better immersion and envelopment to their performances.

2.2.3 Cinematic Audio Evolution since the Advent of Sound Film

From a commercial viewpoint, it is broadly accepted that cinematic sound was officially introduced in the *The Jazz Singer* in 1927 (Thompson and Bordwell 2003: 193). However, it appears that *The Jazz Singer* was actually the final act of a long and intense battle between inventors and companies interested in incorporating their newly developed audio systems in cinematic production (Cousins 2004: 118; Audio Engineering Society 2010; Mitchell 2010; Thompson and Bordwell 2003: 193-211). More specifically, it seems that it was the apparent boom in audio reproduction technologies between the 1900s and the 1920s (Audio Engineering Society 2010; Cousins 2004: 118) that finally made the idea of audiovisual collaboration in cinema viable, something that attracted the interest of filmmakers, inventors and film executives around the world

(Cousins 2004: 118; Thompson and Bordwell 2003: 194-195). The evolution of audio recording and reproduction technologies of this era outside the cinematic context exceeds the scope of the current work and, thus, it is not discussed further here. The reader can find a comprehensive timeline of the evolution of audio technologies in the Audio Engineering Society website (Audio Engineering Society 2010).

Sound Film as the New Industry Standard

After the commercial success of the first *talking movies* in the late 1920s, several competing sound companies that were founded over the previous years in the USA, tried to promote their proprietary sound systems to the major Hollywood firms (Thompson and Bordwell 2003: 194-211). At the same time, the studios realized that if they *'acted individually, they might choose incompatible equipment'* and, thus, decided to *'act together in adopting whichever sound system proved most advantageous'* (Thompson and Bordwell 2003: 194). Similar processes were also taking place in other technologically advanced countries, such as Germany and the USSR, while less advanced countries started importing sound technologies and incorporating them in their productions (Cousins 2004: 118-134, Thompson and Bordwell 2003: 194-211). This eventually led the major film studios around the world to negotiate, and finally agree, on universal sound film standards in order to control film distribution on a worldwide scale (Thompson and Bordwell 2003: 194-211; Sergi 2004). This decision was key in establishing sound movies as the new cinematic standard and, effectively, in forcing the filmmakers to start thinking in a largely different way.

Transformations Caused by the Advent of Sound

It can be claimed that the introduction of sound brought about changes in the cinematic production process, not only because *'filmmakers and technology workers struggled to cope with the unfamiliar, often clumsy, new* [sound recording] *technology',* but also because the sound elements had to be incorporated in the film in real-time during the shooting (Thompson and Bordwell 2003: 195, **Figure 2-9**).

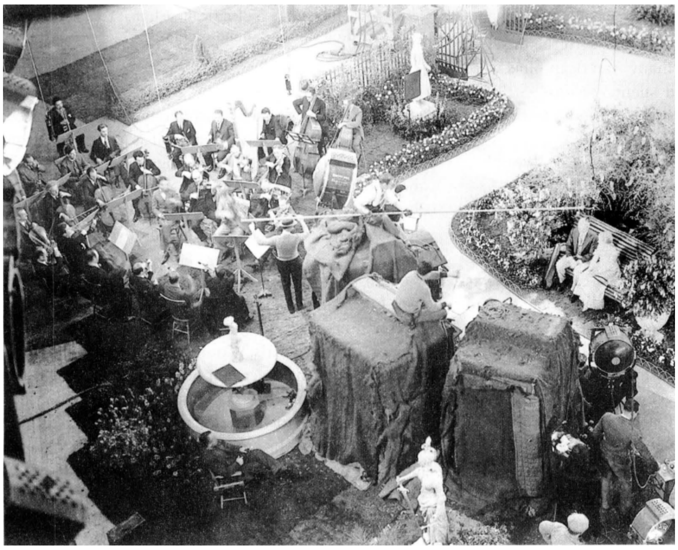


Figure 2-9: Shooting a sound scene at a Warner Bros studio. Dialogue and music are recorded live on the set (source: Thompson and Bordwell 2003: 196)

Trying to find a way to overcome the difficulties associated with *'shooting sound'*, they invented new cinematic techniques and approaches, such as having the output of multiple microphones *'combined in the printing process'* (Cousins 2004: 121), planning their editing during shooting by timing the cut of their scenes between dialogue gaps (Cousins 2004: 118-120), or even establishing whole cinema genres which would exploit the new audio features, such as the musical and the horror film (Thompson and Bordwell 2003: 195). By the late 1930s the cinematic language had been expanded by the incorporation of sound as a fundamental cinematic component (Cousins 2004: 118-150; Thompson and Bordwell 2003: 193-210).

The 1930s were not only characterized by the extensive experimentation with sound filmmaking techniques such as the above, but also by the use of new sound technologies in cinema. Advances in what is frequently refer to as *stereo sound* (Glasgal 2010; Audio Engineering Society 2010; Rumsey 2001: 10-12; Dolby Laboratories 1999), effectively laid the foundation for another key aspect of cinematic sound: multichannel audio. The various attempts to use multichannel sound in movies in the mid 1930s and relevant documented suggestions, such as how *3-dimensional* sound systems could be exploited to provide realistic cinematic audio (Gernsback 1934), indicate that the idea of a more enveloping and physically immersive cinematic sound was a point of interest for a number of individuals in the 1930s.

Fantasia and the Introduction of Multichannel Audio

This interest eventually led to the introduction of multichannel audio to a wide audience in 1941, in Walt Disney's *Fantasia* (Dolby Laboratories 1999; Miller 2004). *Fantasia*, a collection of Disney's cartoons accompanied by recorded orchestral music, made use of a multichannel audio system called *Fantasound* (Wyatt and Amyes 2005: 7; Audio Engineering Society 2010; Miller 2004). The latter was the first commercially used cinematic multichannel audio system, as cinematic sound was largely monophonic prior to that moment (Rumsey 2001: 10).

Figure 2-10: (From left to right) 8-channel recording, mixing and dubbing for Fantasound (source: Garity and Hawkins 1941)

According to Wyatt and Amyes, *Fantasound 'echoed the use of multichannel formats in use today, such as Dolby Surround, Dolby Digital and DTS'* (Wyatt and Amyes 2005: 7). Indeed, *Fantasound* can be regarded as the starting point of surround sound as we know it, although it took another decade until surround sound systems started being widely used (Dolby Laboratories 1999; Miller 2004).

The Evolution of Multichannel Cinematic Audio Systems

Between 1950 and the present day, a large number of magnetic, optical and digital multichannel audio techniques were developed, making the use of multiple audio channels of an acceptable quality possible (Dolby Laboratories

1999, DTS Digital Entertainment 2010; Miller 2004). Although a detailed technical analysis of these systems exceeds the scope of this study, the reader can find a list of notable advances in cinematic multichannel audio below:

- 1950: From the early 1950s on, filmmakers began replacing the optical sound-on-film systems with magnetic ones capable of carrying 4/6 channels of sound (Dolby Laboratories 1999; Thompson and Bordwell 2003: 329).
- 1952: Cinerama, a three-projector widescreen curved screen format was introduced, employing a magnetic six-channel sound system (Thompson and Bordwell 2003: 329; Miller 2004).
- 1953: Cinemascope widescreen anamorphic format was introduced and adopted by most major studios. Cinemascope used magnetic four-channel sound (Thompson and Bordwell 2003: 329; Miller 2004).
- 1976: Dolby Laboratories introduced Dolby Stereo, an optical four-channel audio technology (Dolby Laboratories 1999).
- 1977: Large Hollywood productions such as George Lucas' *Star Wars* and Steven Spielberg's *Close Encounters of the Third Kind* employed 6-channel Dolby stereo sound (Sergi 2004: 28), while many theatres started supporting the stereo/surround sound format in order *'to take advantage of the more powerful soundtracks'* (Thompson and Bordwell 2003: 517).
- 1978: Magnetic Stereo Surround was introduced with 6 channels of magnetic audio tracks (Dolby Laboratories 1999).
- 1983: Lucasfilm introduced the THX certification, *'a set of technical standards designed to ensure a high-quality playback of film soundtracks in specially equipped movie theaters'* (Miller 2004).
- 1987: Dolby introduced Dolby SR, an optical four-channel system that offered wider dynamic range and a broader frequency response than the previous Dolby Stereo system (Dolby Laboratories 1999; Miller 2004).
- 1992: Dolby Digital Surround (originally known as Dolby AC-3, short for audio coding 3) was introduced. This system still encoded the audio information on a film's optical tracks, but in digital format. This system

introduced the 5.1 channel configuration, as we know it today (Miller 2004; Dolby Laboratories 1999).

- 1993: Digital Theater Systems introduced the DTS surround sound technology. The system utilized a *'proprietary timecode on the film print to synchronise playback from CD-ROMs encoded with 5.1-channel digital surround sound. The soundtrack was played back through five full-range discrete speakers and a sub-woofer'* (DTS Digital Entertainment 2010).
- 1993: Sony introduced the Sony Dynamic Digital Sound (SDDS) technology, an optical system like Dolby Digital, but with 7.1 channels arranged in a unique pattern (Miller 2004).
- 1999: *'Dolby introduced yet another new surround sound technology, Dolby Digital Surround EX'* (Miller 2004). This 6.1-channel optical system used an added surround channel to rear plane (Miller 2004; Dolby Laboratories 1999).
- 2000: *'DTS introduced DTS-ES, the only digital audio format capable of delivering 6.1 channels of discrete audio in the consumer electronics market'* (DTS Digital Entertainment 2010).
- 2010: 'Toy Story 3' (Disney/Pixar 2010) was released with a Dolby Surround 7.1 soundtrack to a number of suitably equipped theatres across the world. Dolby Surround 7.1 *'used eight discrete audio channels to establish four surround zones in a theatre. This improved panning and sound localization, enhanced definition, and widened the listening sweet spot.'* (Dolby Laboratories 2010)

With a brief look into the evolution of cinematic audio systems, one can observe that the sound technology revolution that took place over the second half of the 20th century led to a strong competition between different audio companies to establish their surround systems as the new cinematic sound standard. Many of their attempts were successful, and theatres and production studios eventually added surround sound support to their products and facilities. *'By 2001, a large worldwide circuit of auditoriums, capable of playing films with multiple audio tracks, was in place'* (Thompson and Bordwell 2003: 587), while by 2015 this circuit has grown to a point where theatres without some type of surround sound compatibility are rather uncommon.

This interest in adding sound with more advanced spatial characteristics to the moving image is important in the context of this study for two reasons. Firstly, it highlights a belief in sound as a cinematic tool. Secondly, as the evolution of audio systems for cinema involves the use of increasing the number of audio channels, it seems to underline a belief that cinematic sound may benefit from the addition of advanced spatial properties. Based on these observations, it could be claimed that many filmmakers may have realised that the enhancement of the visual envelopment and immersion of cinema could be also supported by the extension of the cinematic space by means of audio.

2.2.4 General Discussion and Conclusions

It appears that the enhancement of physical envelopment and immersion has been the subject of experimentation since the advent of cinema. In an attempt to expand the cinematic world beyond the space of the typical 2D screen, a number of inventors created projection systems that surrounded the audience and produced visually immersive worlds that aimed at enveloping the viewer. Notable examples of the above include the Cineorama and the Cinerama (Cousins 2004: 29-30). Although such systems did not manage to replace the traditional cinematic screen at a commercial level, the idea of visually surrounding cinematic environments refused to disappear over the years. The concept of surrounding cinematic projection zones formed the basis for advanced cinematic systems, such as the IMAX (IMAX Corporation 2014), which make extensive use of large format curved screens or projection domes that cover the entire field of view of the viewer.

Visual immersion was also one of the reasons behind the long-lasting interest in the incorporation of S3D visuals in cinema (Zone 2007). Although S3D does not involve the extension of the field of view around the viewer, it enhances the perception of depth of the image, while it also partly bridges the gap between the viewer and the screen by creating the illusion of objects floating in front of the actual surface of the screen. This creates a visual environment that basically aims at immersing the viewer deeper into the visual action, even if the term *immersion* here does not involve the expansion of the actual cinematic projection space (screen) rather than enhanced depth of the cinematic image.

Apart from the obvious interest in creating cinematic immersion through visual means, audio immersion has been also a central point in cinematic evolution. As mentioned, filmmakers and exhibitors were interested in incorporating sound effects, music and dialogue into their film screenings even before the advent of mechanically reproduced cinematic sound. Considering that early films were *silent*, and thus in many cases not relying on audio for narrative purposes, it can be claimed that sound was used mainly as a means of increasing audience involvement in the screening. Background music enhanced the emotional involvement, while sound effects and dialogue made the visual action and characters more believable and imposing. From a perceptual perspective, this can be seen as one of the first steps towards audiovisual immersion, as sound helped in drawing the audience deeper into the visual action and, potentially, the story.

Another notable example of the use of audio for cinematic immersion was the introduction of multichannel sound systems. The advent of multichannel audio introduced the possibility of spatially expanding the cinematic soundscape. It did not take long for this possibility to become reality and for surround sound systems to become the new cinematic sound standard. One of the reasons behind this was to immerse the audience deeper into the cinematic action.

Based on these observations, it can be claimed that although extensive audiovisual immersion did not find its way to mainstream cinema in a historical context, the potential of creating physically immersive and enveloping cinematic experiences has refused to leave filmmakers' imagination. As modern digital technology breaks the boundaries and constraints of the past, the establishment of new and extensively immersive forms of cinema seems more possible than ever. If this occurs, it is possible that filmmakers may be interested in employing or devising new cinematic tools that serve the emerging medium best. In this context, although it may be true that the majority of such tools would revolve around cinematic visuals, it is also interesting to try and imagine possible ways in which sound can be used to enhance and support the developing language of modern S3D cinema. Ultimately, observations made during such an effort may provide S3D filmmakers with additional creative tools that can be used in addition to the visual ones whenever required.

2.3 Audiovisual Perception Concepts

2.3.1 Introduction

A great amount of psychological research aims at understanding how humans interpret and combine sensory information in order to understand their surroundings. Although a detailed exploration of these areas of research is beyond the scope of this study, the introduction of a number of related topics is relevant. This is because a basic understanding of the perceptual mechanisms of both vision and audition is necessary to carry out and interpret the results of the practical experimentation that follows. In this chapter, fundamental perceptual concepts that are relevant to cinema, such as basic visual and auditory perception and their cross-modal integration are presented and briefly discussed.

2.3.2 Perception in the Cinematic Context

'Since Aristotle, many philosophers and psychologists have believed that perception is the process of using the information provided by our senses to form mental representations of the world around us.' (Bregman, 1994: 3). On a broader perspective this definition could also be used for describing perception in the cinematic context. However, there are significant differences between the latter and perception in real-life conditions (Clark, 2010: 1). Firstly, perception in the cinematic context is constrained by the fact that limited sensory information is used (i.e. visual and auditory information) compared to real life (i.e. information from all five senses). Secondly, in cinema perception is not necessarily used to build representations of a realistic environment that surrounds us, but of a symbolic and imaginary one that can be largely unrealistic in many ways (e.g. the cinematic environment is spatially restricted within the screen area). While in the real world we use our perception to understand actual events taking place around us, in cinema we *suspend our disbelief* (Abrams *et al.* 2001) in order to observe and make sense of a knowingly unrealistic world. Thus, paraphrasing the previously mentioned general definition of perception, one can claim that in the cinematic context *perception is the process of using the*

information provided by some of our senses to form mental representations of knowingly unrealistic fictional worlds created by someone else. In other words, perception in cinema can be regarded as the process of intentionally and willingly allowing our own brains to create artificial worlds by means of clever manipulation of the audiovisual sensory input we provide it with. An in-depth analysis of the above concept from a physiological, psychological or philosophical perspective exceeds the scope of this study. However, it is useful to highlight the difference between perception in real-life and in cinema, as although the latter is actually a particular case of the former (Langkjær 1997: 93) the two might not necessarily coincide in all respects.

Considering this, this chapter is restricted within the scope of perception of space using visual and auditory cues, as well as the relationship between them. This is because these two senses are the dominant ones in cinema, at least in its most common formats.

2.3.3 Visual Perception

Vision *'is always trying to make sense of its input and telling us about the things we need to know about. Vision is active, not passive.'* (Snowden and Troscianko 2006: 3). The importance of this statement lies in the fact that the brain translates and decodes the visual sensory information that is available to it, and not merely receiving it (Snowden and Troscianko 2006: 5).

From a physiological perspective, vision starts when light reflected from the objects surrounding us enter the eye (Murch 1973: 10). The anatomy and functions of the eye is a complicated and highly specialized topic. A simplified description of vision can be the following: Photons (light particles) reflected by surrounding objects enter the eye through the pupil and reach the retina, an area equipped with a number of *light-sensitive cells* (Murch 1973: 11). The latter convert the incoming light stream to *nerve impulses* that are then being sent to the brain through the optic nerve (Murch 1973: 11; Snowden and Troscianko 2006: 44, **Figure 2-11**).

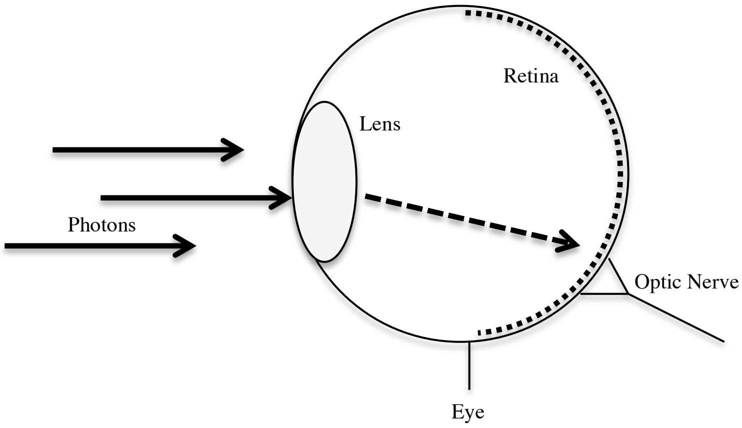


Figure 2-11: Photons pass through the eye lens and generate nerve impulses in the retina. These are transferred to the brain through the optic nerve

Borrowing a commonly and broadly used parallel, the function of each of our eyes is very similar to this of a photographic camera (Snowden and Troscianko 2006: 3). From this, more practical, perspective, the information reaching the retina can be thought of as images of our surroundings. As the brain is not able to process these images directly they have to be converted to electrical/neural signals, a form that is understood by the brain cells. These electrical signals are then analysed and categorized by the relevant brain centres in order the meaningful, 3-dimensional mental representation of our surroundings to be created. This *3D percept* created inside the brain is what we actually regard as the external world (Snowden and Troscianko 2006: 226, Murch 1973: 5).

Based on the above, the term visual perception can be attributed to the process of grouping and processing visual sensory information inside the brain. It is a multifaceted process that relies on a number of different types of information that enter the visual sensory stream. These types of information are broadly known as visual cues and they inform the brain about the various distinct qualities of light emitting objects in the surrounding world, such as color, motion, brightness or distance (Murch 1973).

The cinematic screen is by nature 2-dimensional and, thus, the actual positioning and distance of objects from the viewer is impossible to be presented in exactly the same way as in the 3-dimensional real world (Clark, 2010: 1). Nevertheless, most of the aforementioned visual cues are used in cinema in order to create the illusion of an existing environment in the brain of the viewers. However, the unique visual requirements and orientation of space of cinema (and in particular S3D cinema) makes some of these cues more important than others, in contrast to real-life conditions. The following sections revolve around a particular category of visual cues that are relevant to stereoscopic vision and, therefore, to the current study: the visual depth cues.

Monocular/Monoscopic Visual Depth Cues

One of the most striking aspects of visual perception is the ability of the brain to convert 2-dimensional (2D) images to mental representations of 3-dimensional objects and environments (Snowden and Troscianko 2006: 8). To a large extent, perceived 3-dimensionatlity relies on the different perspectives between the images received by each of our eyes, something commonly referred to as stereoscopic or binocular vision. However, the brain has also the ability to generate 3D mental representations based solely on 2D images. This is possible by the analysis and comparison of a number of cues that are available on a 2D image (de Bruijn and Boone 2002: 2), namely the monocular/monoscopic or pictorial visual cues (Mendiburu 2009: 11; Snowden and Troscianko 2006: 214). Fundamental concepts related to monocular visual cues are briefly presented below.

Size cues

'If we know the real or relative size of something then we can get an impression of its distance from us by how large its image is' (Snowden and Troscianko 2006: 215, **Figure 2-12**). In other words, the brain uses the apparent size of visual objects to calculate the distance of these objects from the viewer. This function of the brain is based on size constancy, *the assumption that objects stay the same size* irrespectively of how big or small they actually appear in the images created inside our eyes (Murch 1973: 186, 205; Mendiburu 2009: 11; Snowden and Troscianko 2006: 121). For instance, if a person appears extremely big this is most likely to be happening because it is very close to our eyes and not because we are looking at a human being of gigantic size.

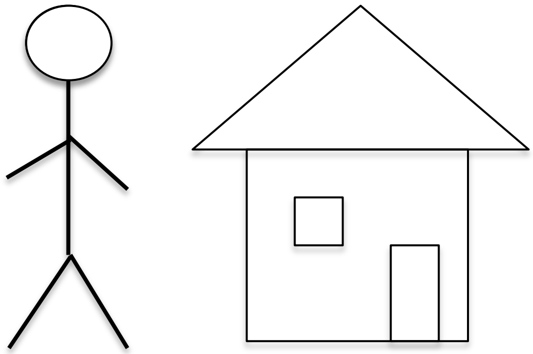


Figure 2-12: Based on previous relative size knowledge the brain is likely to assume that the human is closer to the viewer than the house although they appear to have the same absolute height

To a large extent, the calculation of a visual object's distance based on size cues relies either on previously acquired knowledge of the actual size of this particular object (e.g. the approximate size of a human being is known to most of us) or on comparisons between relative sizes (e.g. if we see a person holding a small green bag and a big red one on each of their hands we will know the relative sizes of these two bags in a possible future depth calculation).

Occlusion/Interposition

In vision, occlusion or interposition refers to situations in which a visual object is partly blocked (occluded) by another (Murch 1973: 189; Mendiburu 2009: 13, **Figure 2-13**).

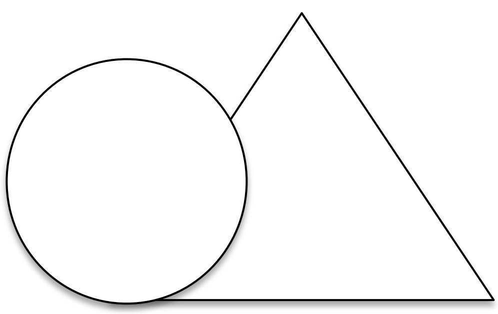


Figure 2-13: Because of occlusion the brain is likely to assume that an object (circle) is closer to another one (triangle)

For example, one can imagine a person in front of a distant mountain with the body of the human figure occluding part of the mountain. The fact that one object (the mountain) is occluded by another (the human body) is used by the brain as a distance cue, as the occluded object is logically assumed to be located behind the occluding one. The combination of occlusion and size cues can further enhance our perception of the actual distance and size of visual objects. For example, an object that has an increased size (a person) but is occluded by others (a house) is likely to be a gigantic object (a gigantic human) rather than to be located closer to the observer.

Motion Parallax Cues

From the perspective of an observer who is moving in space, visual objects located further away appear to move slower than ones located closer (Murch 1973: 193). This phenomenon is known as motion parallax, *'and* [it] *can be a very powerful depth cue.'* (Snowden and Troscianko 2006: 213, **Figure 2-14**).

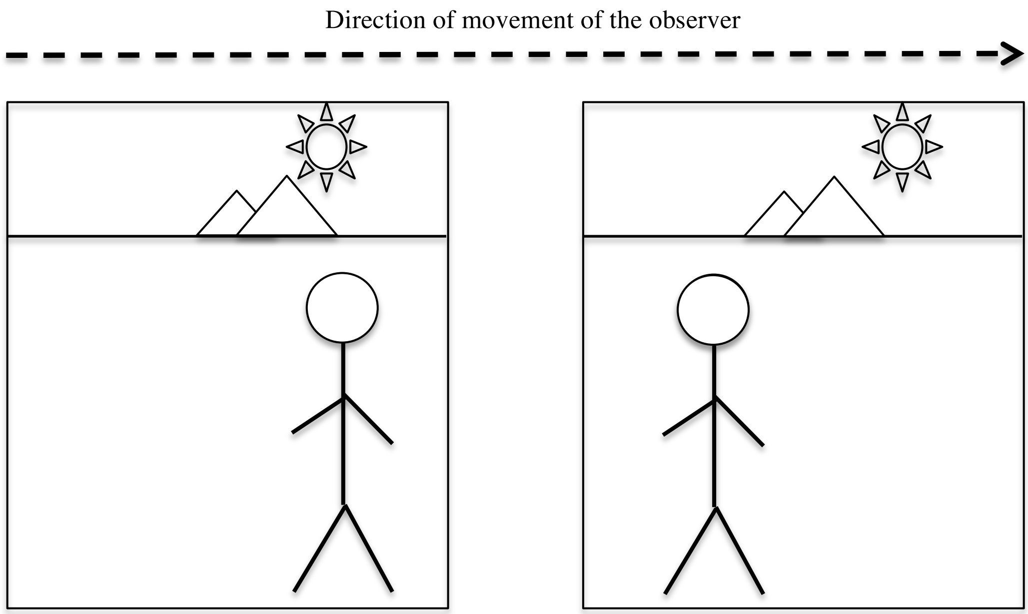


Figure 2-14: Motion Parallax. The apparent position of objects will change at different rates for a moving observer, depending on the distance of the object from the observer

One of the most obvious examples of motion parallax is the observation that for the passenger of a car, the moon seems to stay still while the houses and trees by the side of the road seem to move at high speeds. Whenever possible, the brain will add motion parallax cues to the previously mentioned size and occlusion ones in order to generate an even more accurate mental representation of the location and distance of objects from the observer.

Position Relative to the Horizon, Aerial Perspective and Other Depth Cues

Apart from the, perhaps more obvious, depth cues described above, the brain can also take into account a number of other monocular cues while calculating an object's position and distance. Firstly, in many cases an object's distance can be approximated based on its position relative to the horizon (Mendiburu 2009: 14, **Figure 2-15**).

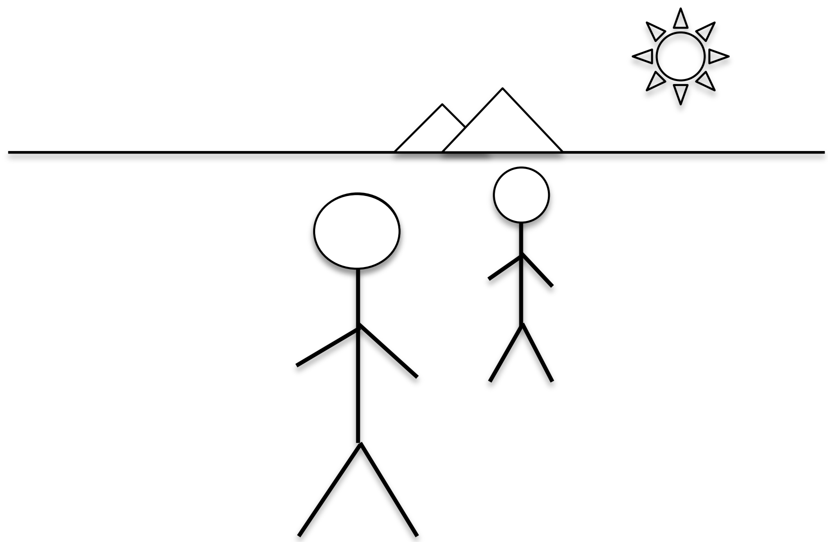


Figure 2-15: Distant objects appear higher relative to the horizon to close ones

For example, a boat sailing on the open sea at a certain distance from the observer will appear higher relative to the horizon than one that is located closer. Additionally, the brain also assumes that objects that are located further away are likely to suffer from atmosphere blur *(*Mendiburu 2009: 14). This term refers to the observation that *'light gets scattered as it travels through the air'*, so *'the further something is away from us, the more hazy it will appear.'* (Snowden and Troscianko 2006: 218, **Figure 2-16**). Other monoscopic cues, such as shading, saturation and color shift (Mendiburu 2009: 13; Snowden and Troscianko 2006:

217) can also assist the brain in calculating visual depth from 2D (monoscopic) images.

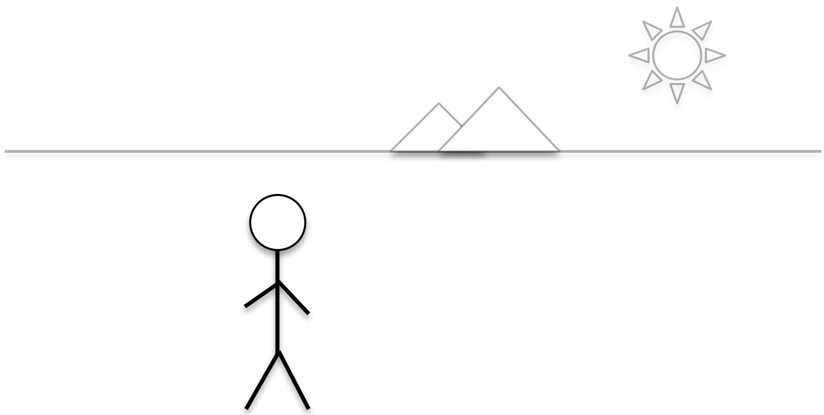


Figure 2-16: Distant objects appear more blurred than close ones

A detailed analysis of all the monoscopic visual depth and distance cues exceeds the scope of this study. However, it is crucial to stress that, in most occasions, a number of such cues are used collectively for the calculation of visual depth. In Snowden's and Troscianko's words, *'we use cues to form something in between what the two (or more) depth cues are saying. Which one carries more weight depends on which one is the most reliable signal.'* (Snowden and Troscianko 2006: 225).

Binocular/Stereoscopic Visual Depth Cues

Apart from the ability to generate 3D depth from the monoscopic visual cues within the 2D images available to each of our eyes, the brain is also capable of calculating visual depth and distance based on a comparison of the images of each eye. This ability of the brain is commonly known as stereoscopic or binocular visual depth perception (Mendiburu 2009: 17). Basic concepts related to stereoscopic depth perception are presented in the following sections.

Retinal Disparity and Stereoscopic Cues

The term retinal disparity refers to the fact that the images created in the retina of each eye are not identical (Murch 1973: 196; Snowden and Troscianko 2006: 200, **Figure 2-17**).

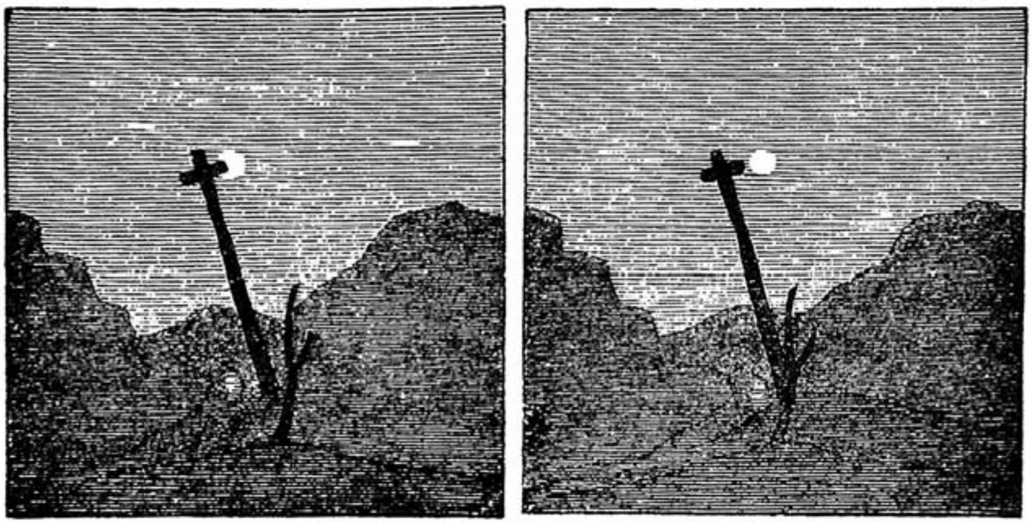


Figure 2-17: Pair of Stereoscopic 3D images (Left and Right Eye) as reported by David Brewster in 1839 (Zone 2007: 10)

This is because of the slightly different perspective of each eye. When the two images are processed inside the brain, they are '*fused together to form one, using the differences* [retinal disparity] *to calculate the depth of each part of the image'* (Snowden and Troscianko 2006: 204). The differences used for this calculation have to do *'mostly with horizontal parallaxes, occlusions revelations, some shape changes, and convergence cues'* (Mendiburu 2009: 17). These are generally known as *'stereoscopic depth cues and they are just a specific kind of* [the previously mentioned] *motion parallax cues'* (Mendiburu 2009: 17). Stereoscopic depth cues are briefly presented below.

Horizontal Parallax

Horizontal parallax refers to the magnitude of the retinal disparity (Mendiburu 2009: 17). The brain is capable of extracting information about a visual object's distance and position from the viewer based on the difference between this object's position in the left and right eye images. This is similar to how monoscopic motion parallax cues are used by the brain with the difference that in the case of stereoscopic horizontal parallax the observer or the objects do not need necessarily to move. This is because even static objects will appear in different locations on the two retinas according to their distance from the observer.

Occlusion Revelations

As described earlier, occlusion is one of the fundamental monoscopic depth cues. However, as our eyes are located at a set distance between each other, occlusion for a given object is different for each eye. Therefore, the brain uses occlusion also as a stereoscopic cue (**Figure 2-18**).

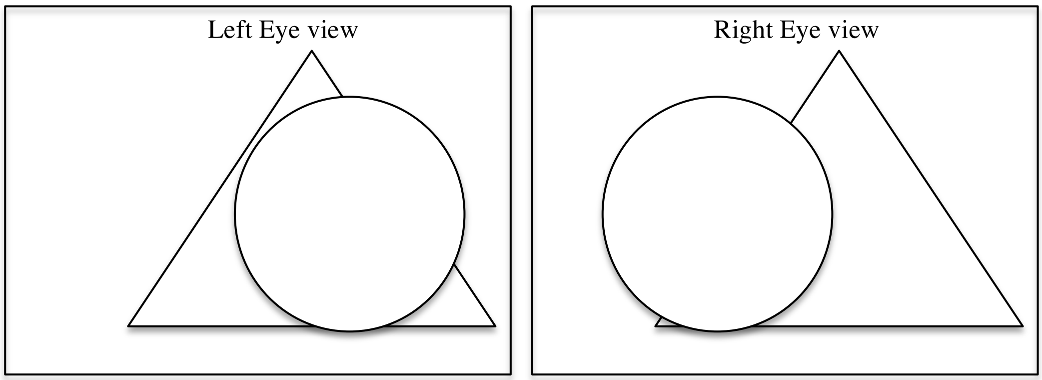


Figure 2-18: Different occlusion cues occur for the left and right eye

The basic principle behind this use of occlusion is that a part of a visual object that is occluded for one eye is visible for the other. This gives the brain strong indications regarding the distance of both the occluded and of the occluding object from the observer. For example, if the occluding object is located closer to the observer then a larger part of the occluded one will be visible to one eye and invisible to the other. This creates *'an additional texture* [the part of the object that is visible only to one eye] *that is a major cue for the brain to reconstruct a scene, to the point that occlusion will supersede any other cue'* (Mendiburu 2009: 18).

Shape Differences

The shape of visual objects may also appear different to each of our eyes, as the latter observe the former from different angles. For instance, if one places their finger in front of their eyes touching their nose it will appear to have a completely different shape for each eye. However, if the finger is viewed at an arm's length its shape will be effectively the same to each eye. Based on this difference (or the lack of it) between the apparent shapes of objects on each eye, the brain can also calculate the distance and/or size of the actual objects from the observer (Mendiburu 2009: 18).

Convergence and Accommodation

In broad terms, convergence and accommodation refer to the process of focusing our eyes to objects or areas located at a given depth and distance (Lelyveld 2009). *'Vergence is the angle one of your eyes turns relative to the other eye so that they both look at the object that you want to see'* (Lelyveld 2009). Thus, convergence refers to the angle created between the viewpoints of two eyes as they converge at a certain object or position. Accommodation refers to the adjustment of each of the eyes' lenses as they converge on an object in order to put this object in focus (Mendiburu 2009: 20; Lelyveld 2009). As the observer selectively focus on objects or positions located in variable distances and depth in space convergence and accommodation change dynamically. Based on such changes, the brain gathers additional depth information *'from your visual motor system, the muscles that control your eyes' movements'* (Mendiburu 2009: 20).

2.3.4 Auditory Perception

'Sound originates from the motion or vibration of an object. This motion is impressed upon the surrounding medium (usually air) as a pattern of changes in pressure' (Moore 2003: 2). This is commonly known as a sound wave. Sound waves are captured by the ears, where they are transduced to electrical signals (Moore 2003: 21, **Figure 2-19**).

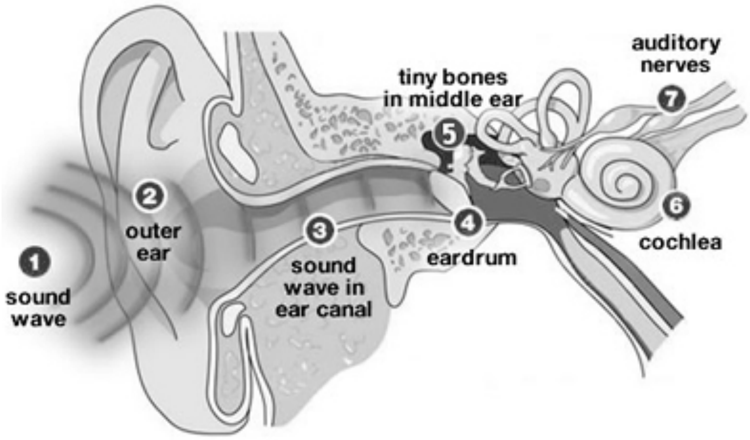


Figure 2-19: Human hearing system (source: The Hearing Care Centre 2014)

These signals are then transmitted to the relevant centres of the brain through the auditory nerves in order to be organized and processed. Based on this, the brain can then transform the incoming signals into meaningful mental representations of the nature and positions of the sound-emitting objects allowing the listener to interact efficiently with the environment. This perceptual organization and processing of incoming audio signals involves the comparison and evaluation of a number of different auditory cues, which are constantly monitored and assessed within the auditory system.

From a physiological perspective, sound waves enter the ear canal after their frequency spectrum is modified by the outer ear (i.e. pinna) (Moore 2003: 21, **Figure 2-19**). This frequency modification is crucial for sound source localization. The modified sound signal passes through the ear canal and reaches the surface of the eardrum, which in turn transmits the vibrations to the cochlea (a *spiral-shaped* fluid-containing cavity, Moore 2003: 22) through the ossicles, three interconnected small bones (Murch 1973: 15; Moore 2003: 21). As vibrations are transmitted to the cochlea, they cause the fluid it contains to vibrate according to their intensity and frequency. These vibrations, in turn, cause relevant hair cells *'to transduce mechanical movements into neural activity'* (Moore 2003: 32). A detailed examination of the physiology of the auditory system exceeds the scope of this study. The reader can find a very thorough coverage of this topic in Brian Moore's *'An Introduction to the Psychology of Hearing'* (Moore 2003).

Fundamental Auditory Perception Concepts

After sound waves have been transduced to neural activity they are transmitted to the brain for processing. From the incoming sound mixture, the brain attempts to generate mental representations of the objects that have created this mixture based on a number of assumptions and calculations. Basic concepts related to this are presented to the following sections.

Frequency, Pitch and Loudness

As mentioned earlier, *'sound originates from the motion or vibration of an object'* (Moore 2003: 2). This motion creates sound waves that vibrate at particular frequencies and amplitudes (Murch 1973: 12). In its simplest form, a sound wave is a vibration of the air molecules at a particular and constant frequency and amplitude. This simple type of sound wave is known as a sinusoidal or sine wave (Moore 2003: 2). If we take a sine wave as an example, after it enters the auditory system the brain could detect both its frequency and its amplitude. Thus, if another sine wave with different frequency and amplitude was presented after the first one, the brain could detect and register any differences between them. In Murch's words, *'when the frequency of a sound wave is varied, the listener experiences changes in pitch, whereas changes in the amplitude of a sound wave produce perceived variations in loudness'* (Murch 1973: 154).

In nature, sound waves are far more complex than simple sine waves as the various objects that create them tend to create a mixture of sounds of different frequencies and amplitudes that also vary over time. This makes the analysis of pitch and loudness of real-life sound waves a more complicated process than that of simple sine waves. Nevertheless, what is relevant to this study is that the brain has the ability to determine the pitch and loudness of such complex sound waves efficiently, albeit subjectively (Moore 2003: 2). This is crucial because such differences are used by the auditory system as cues for identifying the nature and position of the sound-emitting objects.

Timbre

'Two [complex] *sound sources may be perceived as dissimilar even though they are equal in loudness and pitch; each is said to have a different timbre*' (Murch 1973: 160). Although the validity of such definitions of timbre may be debatable, they outline that complex sound waves can vary not only in terms of pitch and loudness, but also in terms of their perceived tonal quality or character (Murch 1973: 160). Generally, this character (the timbre) of a complex sound wave is determined by the relationship between the various different frequency components it consists of. One could think of the complex sound wave as a group of different sine waves vibrating at various frequencies. If we added or removed sine waves from this group the overall perceived character of the complex sound wave would be affected. The same would also occur if we altered the volume of some (or all) of the sine waves. This is because the brain registers the changes in the sound mixture and the ratios between the different frequency components of complex sounds (Moore 2003: 206). From a practical perspective, a convenient way to understand this is given by Gerald Murch when he mentions that *'if a piano and a clarinet both play a middle C with the same loudness, the resulting tone sounds different for each instrument'* (Murch 1973: 160).

Although in real life changes in the timbre of a given sound are likely to affect other characteristics of this sound (such as pitch or loudness), in the current context it should be sufficient to claim that the perceived pitch and volume of a given complex sound could be kept steady while its timbre is modified or vice versa.

Auditory Localisation

'The term localization refers to judgments of the direction and distance of a sound source.' (Moore 2003: 233). These judgments are based on a number of auditory cues, such as time and intensity differences between the sounds reaching each of our ears, amplitude and frequency content of a sound and the amount and nature of reflections of a given sound on the surrounding objects (i.e. reverberation, Blauert 1997: 37; Moore 2003: 235; Murch 1973: 208). Based on such cues, the brain estimates the position of a sound source within a 3-

dimensional spherical space around the listener. Fundamental concepts related to this are presented below.

Direction of Sound Sources

In contrast to vision, which is restricted to a limited space (i.e. field of vision), the auditory system is capable of monitoring and registering events occurring on a 360° degree sphere around the listener. For making the description of the position of a sound within this sphere easier, a coordinate system is commonly used (Blauert 1997: 14, **Figure 2-20**). On the coordinate system, the direction of a sound is determined by its position in the horizontal (i.e. azimuth, X, Y Axis) and median planes (i.e. elevation, Z Axis; Moore 2003: 234).

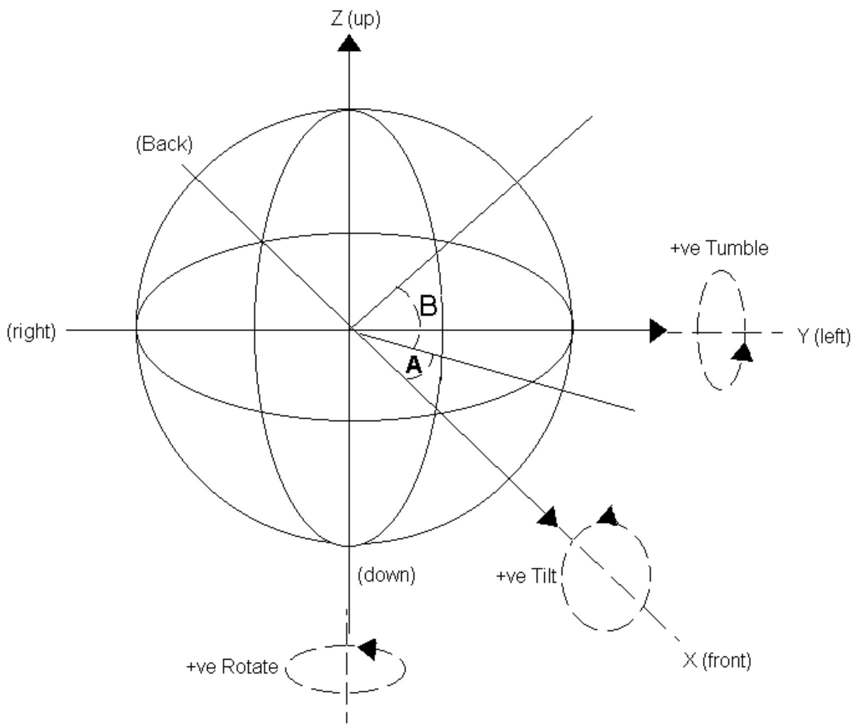


Figure 2-20: The position of a sound in relation to the listener can be described by its coordinates within a 360 degree sphere (Malham 1998)

In order to briefly examine how the brain uses auditory cues to estimate this direction, let us consider a sound source with an azimuth (X, Y Axis) of 270 degrees and elevation (Z Axis) of 0 degrees. Because of the angle of the sound source in relation to the listener's head, the sound wave would reach the left ear earlier than the right ear. This could have two effects on how the listener perceives the sound. Firstly, the sound would reach the left ear quicker. Secondly, the intensity of the sound on the left ear would be greater (Moore 2003: 235). These differences, known as *interaural time differences* (ITD) and *interaural intensity* or *level differences* (IID or ILD), are used by the brain as cues for the estimation of the position sound source (Blauert 1997: 140; Turner *et al.* 2011; Moore 2003: 235). As another example, let us consider a sound source with an azimuth of 0 degrees and an elevation of 45 degrees. As the sound source is located at equal distances from both ears, ITD and IID cues would not provide strong positional cues to the brain. In such cases, the listener could adjust the angle of their head in relation to the sound source in order to be able to use ITD and IID cues, something that apparently occurs routinely in real life (Blauert 1997: 178; Moore 2003: 249). Additionally, the brain could also judge the estimation of the position of the sound source based on the changes introduced to the frequency content of the sound wave as it hits parts of the human body (e.g. head and torso) and as it passes through the listeners' outer ear (Blauert 1997: 63, 192). Such frequency content transformations are broadly known as HRTF (Head Related Transfer Function, Moore 2003: 249). Although the frequency and the character of a sound source may influence the effectiveness of such cues, for the purposes of this study it should be sufficient to claim that they are among the most important cues for the estimation of the position of sounds, especially on the horizontal plane.

Perception of Distance

In order the apparent position of a sound source to be estimated within the sphere of audition, determining its direction alone is not sufficient. As the auditory sphere is 3-dimensional the brain has also to calculate the distance (or depth) of sound sources from the listener. For this calculation, a number of auditory cues are used. These cues include the intensity, the spectral content, and the reflections (i.e. reverberation) of a sound signal, as well as calculations based on binaural cues (Moore 2003: 265). The distance between a given sound source and the listener determines the perceived characteristics of the sound that arrives at the listener's ears (Coleman 1963).

The most obvious demonstration of this is the observation that sounds generated by distant sound sources appear quieter than ones coming from sources located closely to the listener. For instance, the intensity of a dog's barking can reveal how close the dog is to the listener. This is because the brain assumes that the actual level of the barking is, by and large, constant. Therefore, if it is perceived as quieter this is possibly because the dog is further away and not because it adjusted the volume of the barking. This '*loss of amplitude with distance'* is generally referred to as *the inverse square law* (*inverse first power loss*, Coleman 1963). Although this loss '*may not always be detectable'*, while also reflections on surrounding surfaces may cause '*major departures'* from it, *'it is the most general of the cues to distance obtaining physically for all types of sounds and at all distances'* (Coleman 1963: 303). Considering that the effect of the inverse square law in the perception of auditory distance works in a predictable manner in most cases, and also that it is an effect well-known to most humans from real life experience, for the purposes of this study it may be possible to make the following generalization: the amplitude attenuates by *'6db for each doubling of distance'* (Coleman 1963: 302, **Figure 2-21**).

In addition to the overall intensity level, the distance between the listener and the sound source might also affect the spectral content (i.e. the relationship between the different frequency components of a complex sound) of the sound arriving at the listener's ears (Moore 2003: 265; Coleman 1963).

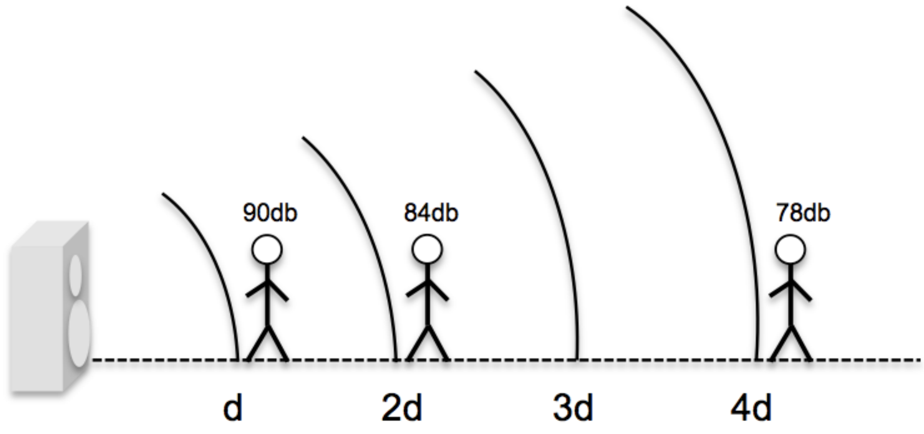


Figure 2-21: For every doubling of the distance from a sound source a 6db volume attenuation occurs

According to Moore, *'over moderate distances, the spectrum of a complex sound may be changed, owing to the absorbing properties of the air; high frequencies are attenuated more than low'* (Moore 2003: 265, **Figure 2-22**).

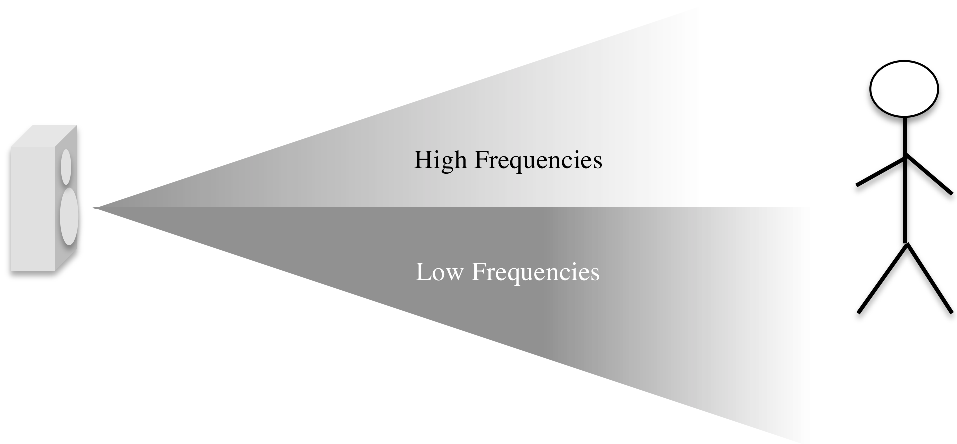


Figure 2-22: High frequencies attenuate quicker than low over moderate distances

It is important to note that this may take place *over moderate distances* and, thus, the parameter of distance is key for its effectiveness. For instance, *'at distances less than 4 feet the sound source appears to approach the observer as the lower frequencies in a sound stimulus become more prominent'* (Coleman 1963: 305-306; Bekesy 1961). Also, this cue has been proved to be effective mostly for familiar sounds, as relationships between spectral components of unfamiliar sounds cannot be evaluated (Moore 2003: 265). Additionally, it *'appears to be effective for judging the relative distances of sound sources, but not for judging their absolute distances'* (Little *et al.* on Moore 2003: 265). However, assuming that many of the cinematic scenes and objects would normally fall within the range of *moderate distances* and that many of the sounds accompanying the cinematic object would be familiar to the viewer/listener from previous experiences, the following generalization could be made in the context of this study: the brain may frequently use frequency loss as a cue by assuming that distant sound sources are expected to produce sounds with lower frequency content.

For sound sources that are close to the listener's head, the brain can also use binaural (ITD or IID) cues to calculate the distance of the sources. According to Turner *et al.*, *'the greater the ITD or IID the shorter the perceived distance'* (Turner *et al.* 2011). Similarly, Moore mentions that *'larger than normal IIDs (especially at low frequencies) provide a cue for distance'* (Brungart on Moore 2003: 265).

In addition to the above, it must be stressed that distance can be also calculated based on the examination of the reflections of a sound on the surrounding objects, something known as reverberation. The detailed analysis of a multifaceted and complex topic, such as the study of the perceptual and physical characteristics of reverberation, exceeds the scope of this work. However, the study of the role of reverberation as an auditory depth cue in the context of S3D cinema may be an interesting topic for future work.

2.3.5 Cross-modal Integration

All our senses essentially have the same purpose: to inform us about objects and events in our surroundings. This is a relevant observation, as it underlines that in most cases our brain expects that information about a given event or object could be available in more than one of the sensory streams reaching our brain (Bolognini *et al.* 2005: 273). For example, humans can see and hear the raindrops on a rainy day, while their sense of touch also validates this event when the water gets in touch with their skin. Additionally, it is possible that the smell of the wet grass and dirt could be also detectable. Our brain is good at combining and analyzing the cues coming from different sensory streams and identifying whether they belong to the same external event. This ability *'can enhance perceptual clarity and reduce ambiguity about the sensory environment'* (Lippert *et al.* 2007: 102). The above is commonly known as cross-modal integration and is a heavily studied topic of modern psychology (Holt 1997; Bolognini *et al.* 2005). Arguably, in the context of cinema the examination of cross-modal integration is largely related with the consideration of two sensory streams: the visual and the auditory. Relevant concepts are briefly presented in the following sections.

Audiovisual Integration as a Collaborative Process

In the context of cross-modal integration, the perceptual organization of information obtained from the auditory and the visual sensory streams can be regarded as a unified (albeit multifaceted and complex) process. The brain can use cues obtained from each of the two sensory streams in order either to reinforce and verify a perceptual event or to dismiss it. For example, let us assume that one hears a strange noise behind them while walking outdoors. Their auditory sensory stream informs them about an event that may need attention (i.e. something potentially dangerous) in their surroundings. When they turn their head to verify the event with their vision they do not see anything strange. Thus, they might assume that either the sound was nothing worth further attention or even that they were misled by their own auditory system. In this example, one could make two observations. Firstly, the information obtained by one of the sensory streams (audition) could cause the brain to search for information in another (vision) in order to verify the source and nature of the object or event that generated the sensory information (sound wave). Secondly, if the object or event that presumably generated the initial sensory information (sound) could not be verified by cues obtained from another sensory stream (vision) less attention may be given to it. Therefore, in addition to comparing and organizing auditory and visual cues separately the brain also monitors and compares cues across the two different modalities. If cross-modal cues confirm each other then more accurate and solid conclusions could be drawn about a given event or object

(Lippert 2007: 102). Accordingly, contradicting cross-modal cues could force the brain to dismiss the ones that seem less reliable. Admittedly, this is a rather simplistic introduction to the topic. However, in the current context it should be sufficient as it highlights a relevant idea: it is likely that the brain may try to verify auditory events through vision and vice versa.

Visual Dominance

When an event is detectable by both vision and audition, in most cases the visual cues have a stronger influence than the auditory ones to the mental representation of this event (Shams and Kim 2010: 2; Meares 1993: 172). This well documented phenomenon is commonly known as *visual dominance* or *visual prepotency* (Holt 1997: 115; Woszczyk *et al.* 1995: 4). Visual dominance is particularly prominent in the spatial domain (Kitagawa and Ichihara 2002). According to Woszczyk *et al.*, *'a sound is perceived as if it were originating from a point in the visual environment where something is moving or changing, even when the physical source of the sound is elsewhere'* (Woszczyk *et al.* 1995: 4). This *'is due to a general attention bias towards visual modality'* (Posner and Nissen on Woszczyk *et al.* 1995: 4) and is commonly referred to as the *ventriloquism effect* (Kitagawa and Ichihara 2002; Bregman 1994: 183, 290). The ventriloquism effect is largely exploited in the cinematic context. The perceived spatial position of sounds that correspond to onscreen visual events or objects is largely dictated by the apparent position of the latter (Meares 1993: 172). Michel Chion uses the term *sound magnetization* to describe this cinematic audiovisual relationship (Chion 1994: 69), as the sound appears to be *magnetized* by the corresponding visual event or object. As an example, one could think of a scene where the voice appears to be emanating from the lips of a character located on the right side of the screen while in reality it is played through the centre speaker in the middle of the screen (**Figure 2-23**).

The power of this phenomenon may be also highlighted by the observation that, to the present day, an arguably limited number of audio channels (e.g. usually 3 to 6) are used to carry the audio information that accompanies the onscreen visual action in cinema. Although the positions of the various sound elements within the soundtrack do not match spatially these of their corresponding visuals, the audience perceives the two as a unified audiovisual event.

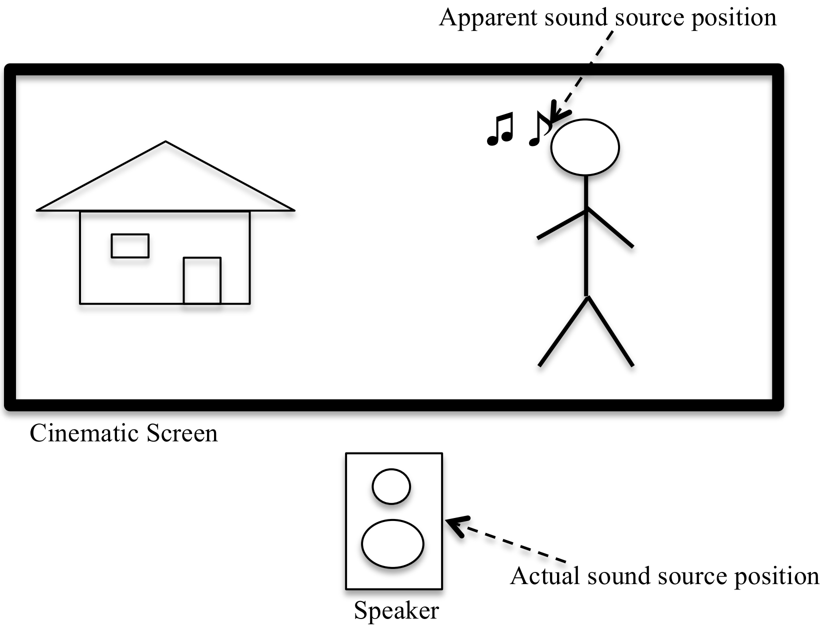


Figure 2-23: Sound magnetisation from visual events occurs regularly in cinema

Effects of Audition on Visual Perception

Visual dominance over audition is one of the most broadly known and used examples of audiovisual integration. However, audition has also some power to effect visual perception in various ways. Perhaps the most well documented effect of audition over vision is *'the Auditory Driving Effect, […] the auditory temporal capture of vision'* (Mastoropoulou 2006: 25), which refers to the observation that the rate of auditory information can affect the perceived rate of corresponding visual information.

Audition can also affect other perceptual parameters of audiovisual events (Ecker and Heller 2004; Vroomen and Gelder 2000; Lippert *et al.* 2007). For instance, Lippert *et al.* (2007: 103) have reported that *'a simultaneous tone improved detection of a dimly flashed light, enhanced the discriminability of briefly flashed visual patterns or increased the perceived luminance of light'*. Similarly, Freeman and Driver demonstrated that the apparent motion of a sequence of visual stimuli could be altered *'solely by timing of a static sound'* (Freeman and

Driver 2008), while Mastoropoulou proposed that visual attention could be guided to certain parts of a CG animation sequence solely by means of sound (Mastoropoulou 2006). Ecker and Heller also have shown that the *'path of a ball moving in 3D-space can be altered solely by the addition of an auditory cue'* (Ecker and Heller 2004).

The effect of audio on the way an audiovisual stimulus is perceived is also widely used and evident in the cinematic context. For instance, a loud sound of a monster coming from the left side of the screen is likely to attract visual attention to this part of the screen. The bigger the cinematic screen the most significant this effect may be. This is because in large screens that exceed the field of vision, visual attention may have to be given to the parts of the screen that carry significant information. The use of auditory cues in order to influence visual attention could be especially important in S3D cinema. This is because the visual complexity of the S3D environment could result in an increased need for guiding the viewer's visual attention to those parts of the scene that are more relevant to the story.

Other Perceptual Effects of Audio in Cinema

In the cinematic context, audio sensory information is also routinely used to alter the emotional impact of a scene, to strengthen *suspension of disbelief* (Abrams *et al.* 2001: 211), or to set the context for the visual information. For example, let us assume that an explosion is displayed onscreen. A corresponding explosion sound effect could make the scene more believable and impressive (Dykhoff 2008), while the sound of a heartbeat might underline the emotional state of the protagonist while watching the explosion. Another example of the potential emotional effect of audio on cinematic visuals is the way a musical piece could transform the way a scene is perceived.

As a whole, the impact of audio in cinema is a broad topic that exceeds the scope of this study. For our purpose it should be sufficient to claim that auditory sensory cues can affect the way cinematic visuals are perceived in various ways. This is crucial when considering the creation of soundtracks that support and accompany S3D visuals, as the effectiveness and artistic appeal of the latter could be also affected by the use of appropriate sound design.

2.3.6 Summary

A great amount of psychological research aims at understanding how humans interpret and combine sensory information in order to understand their surroundings. This section introduced basic audiovisual perception concepts that are relevant to cinema and to the topics studied. A more detailed exploration of these areas of research is beyond the scope of this study.

2.4 S3D Cinematic Soundtrack Mixing Concepts

2.4.1 Introduction

It is possible that cinematic production may be affected by the addition of S3D visuals in various ways. Decisions related to camera angles, scene length and composition, visual focus, lighting, coloring and post-production may be affected by the unique nature and spatial characteristics of the S3D cinematic space. Should this assumption prove to be correct, it may have an impact on the decisions related to the construction of the soundtrack in addition to the mentioned implications on the visual side. Certain sound design spatialisation/mixing techniques and approaches that are commonly employed in 2D cinema may not be the only alternatives for use in S3D cinema. In this scope, it may be interesting firstly to establish a basic understanding of typical uses of the soundtrack in 2D cinema and, secondly, to try to question how the possible changes brought by the use of S3D visuals could affect it. Ultimately, a number of possible uses of the soundtrack in the context of S3D cinema could be proposed in light of the nature and the spatial or technical characteristics of this unique medium. Although such uses may not be necessarily the only alternatives for S3D filmmakers, they may provide an additional creative tool that can be used at will if and when the narrative requires so.

This chapter includes a brief presentation of some typical uses of the soundtrack in 2D cinematic productions. Next, concepts related to the differences between the S3D and 2D cinematic production, as well as the possible impact of these differences on the creation of the soundtrack, are briefly presented. Finally, a number of possible ways the soundtrack could be used in S3D cinematic productions are proposed.

2.4.2 Classic (2D) Cinema Soundtrack Mixing Concepts

Before examining the role of the soundtrack in S3D cinema, it is useful to briefly look at the way the spatial features of the multichannel soundtrack are typically used in classic 2D cinema. This could potentially assist in understanding possible constraints and limitations imposed to the use of multichannel audio by the 2D nature of the visuals, something that in turn could assist in identifying what areas of the multichannel soundtrack mixing are likely to be affected by the introduction of S3D visuals.

Typical Sound Spatialisation Approaches Used in the Multichannel Soundtrack

A strict categorization of cinematic sound spatialisation strategies is perhaps impossible, as with many other processes involving creative and artistic input. However, it can be claimed that filmmakers and sound designers commonly employ certain approaches towards spatialising the elements of the soundtrack within the surround soundscape. Examples of such approaches are the placement of dialogue in the front centre channel and the common use of the surround channels for transient sound effects or ambiences (Kerins 2011; Beauchamp 2005). Typical spatialisation conventions such as the above are briefly presented below. It must be noted here that the concepts presented in the following sections are related mainly to the creation of soundtracks for 5.1 and 7.1 channel surround sound systems (e.g. Dolby Surround and DTS), as these are the dominating formats within the cinema industry at the present time.

Dialogue

For the vast majority of filmmakers, dialogue clarity is essential. This is not only because dialogue usually carries information that is crucial for the narrative, but also because *'additional cognitive processing is needed in order to derive both literal and implied meaning'* from it (Beauchamp 2005: 146). It is, thus, reasonable that usually the dialogue is kept as clear and distinctive as possible in order to cut through the overall soundtrack and become intelligible. This requirement for clarity inevitably has an effect on the dialogue spatialisation approaches taken during the soundtrack production (Kerins 2011: 71). By placing the dialogue on certain channels (e.g. centre front channel, **Figure 2-24**) and keeping it separated from the rest of the audio elements of the soundtrack (e.g. background sounds/ambiences or sound effects) filmmakers avoid a number of intelligibility and perception issues (Dolby Laboratories 2005: 5-1).

For example, if dialogue were to be placed on the rear left channel where large amounts of the surround ambience and background music are also routinely sent, the clarity of the dialogue could be compromised. Additionally, dialogue clarity could be inconsistent among members of the audience seated in different locations of the theatre, while many filmmakers also believe that extensive off-screen spatialisation could possibly distract the viewer from the visual action (Hirst 2006: 20; Rydstrom 2010; Boyes 2010a).

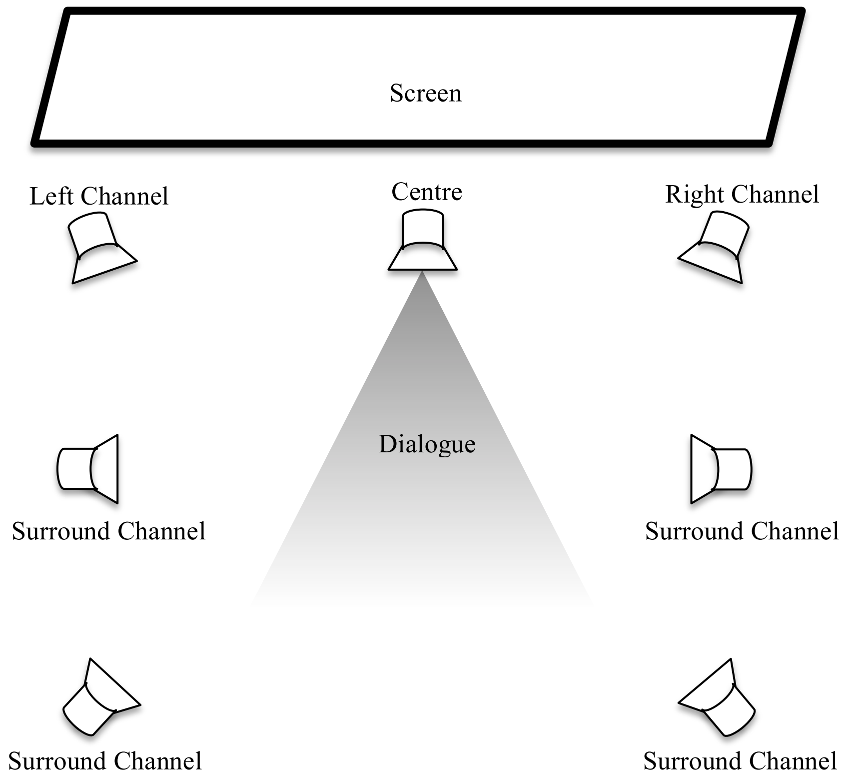


Figure 2-24: Dialogue is commonly sent to the centre front channel

Most importantly, sound designers have to ensure that the dialogue will be clearly audible in a variety of different sound systems employing different channel configurations (Kerins 2011: 14, 44). Placing the dialogue in channels that are guaranteed to exist in virtually all available cinematic sound systems, such as the front Left and Right channels and/or the centre front channel, could be clearly audible in virtually all channel configurations in theatres and home cinema systems.

Perhaps there are also other factors that lead to the generally conservative use of dialogue spatialisation, such as the observation that *'audiences focus on the* *content rather than the placement of dialog in the mix'* (Beauchamp 2005: 146) or the observation that sound is *magnetised* by the image (Chion 1994: 69). However, one can assume that if sound system compatibility was not an issue (for instance, if a universal surround sound format and speaker configuration was followed closely and strictly by all theatres and home entertainment systems), filmmakers could be more adventurous with their dialogue spatialisation approaches. Mark Kerins' observation that the introduction of digital surround sound resulted in an increased use of spatialised dialogue (Kerins 2011) highlights the premise of this assumption. However, for various reasons, the spatialisation approach towards dialogue at the present time in the majority of cases follows the conventional safety of the past sixty years: the dialogue more often comes from the front speakers both for onscreen and off-screen characters.

Sound Effects

In contrast to dialogue, pronounced spatialisation of sound effects has been a common cinematic sound design practice for decades. This is unsurprising considering that the surround channels were initially referred to as the *Effects Channels*, something that underlines their intended use. Since the appearance of surround sound systems, spatialised sound effects have been routinely used to enhance immersion, increase the impact of visual events or even expand the cinematic world beyond the limits of the screen. Many examples of diegetic sounds corresponding to moving cinematic objects come to mind, such as fly-over helicopters, bullets and arrows traveling through the theatre or static foley sounds corresponding to off-screen objects or characters like *'the roar of a monster in the back'* (Vanhoutte *et al.* 2010: 11). The fact that the spatialisation capabilities of surround systems have been employed far more frequently for sound effects rather than for dialogue may be partly because in most cases the former is not as crucial for the narrative as the latter. Filmmakers could be more experimental with spatialised sound effects without running the risk of the more important parts of the narrative going unnoticed or masked.

Considering the above, it should not come as a surprise that one of the most commonly employed uses of the multichannel soundtrack has been the occasional spatialisation of off-screen diegetic and foley sounds (**Figure 2-25**).

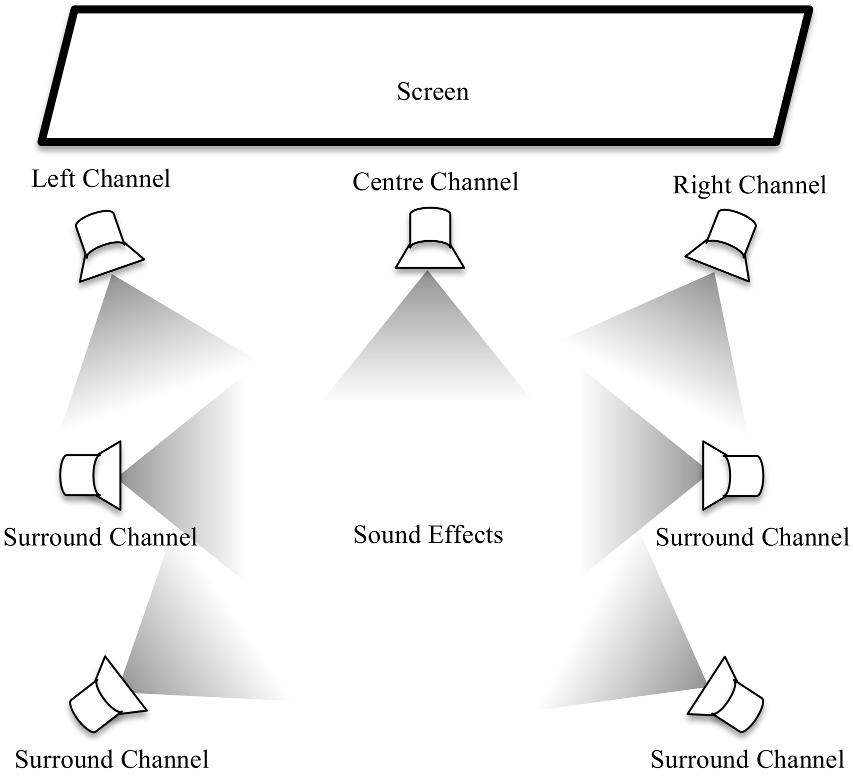


Figure 2-25: Sound effects are frequently sent to the surround channels

However, it must be stressed that even though spatialisation of sound effects is employed more frequently than that of other elements of the soundtrack it is still generally used selectively and in moderation, possibly because filmmakers may not intend to drive the viewers' attention away from the 2D space of the visual action.

Ambiences, Background Sound and Reverberation Effects

Although the surround ambience (including surround reverberation effects) panning approaches used by different sound designers might vary (Rydstrom 2010; Boyes 2010; Beauchamp 2005: 149), the use of the surround channels for the creation (or simulation) of acoustically immersive spaces by means of both diegetic and non-diegetic ambiences and background music is also a common practice. However, in most cases the created soundscapes aim mainly at immersing the viewer into the visual action that takes place in front of them, rather than at acoustically expanding the cinematic space to a realistic 3D sphere.

This is justified, as it is possible that the creation of a realistic 3D soundscape that includes distinctive spatialised off-screen audio or prominent diegetic off-screen sound objects may distract the viewer from the visual action (Rydstrom 2010; Boyes 2010).

Large amounts of the soundtrack's ambience sound textures, reverberation and background music are routinely mixed to the surround channels. This approach is most likely employed in order to expand the soundscape toward the sides of the audience rather than as an attempt to create a realistic aural environment. By expanding the audio scene to the surround channels filmmakers may achieve a better envelopment, but also make space for more crucial elements of the soundtrack (e.g. dialogue, onscreen sound effects) in the centre front of the soundscape.

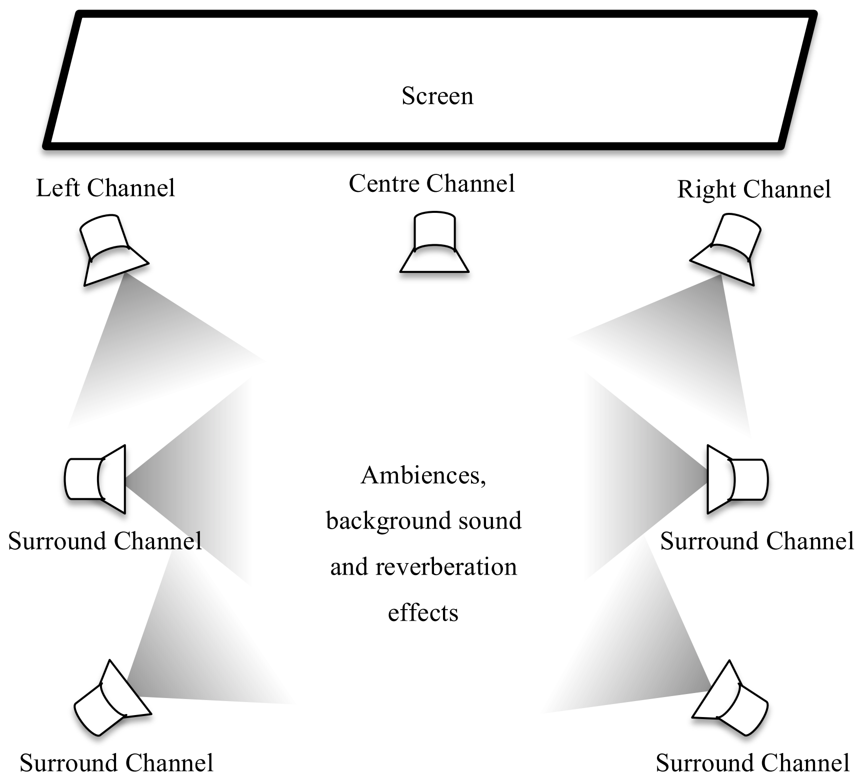


Figure 2-26: Ambiences, background sound and reverberation effects are commonly sent to the Left / Right or surround channels

It can be observed that a screen-oriented approach is taken towards cinematic sound spatialisation, despite the fact that modern sound technologies provide the means for the creation of realistic spatial audio representation on a 360 degrees sphere (Boyes 2010). Nevertheless, the broad use of the surround soundscape for the expansion of the cinematic audio environment (ambience, reverberation, background music) suggests that audience immersion may be improved by spatially extending the soundscape, despite the fact that the visuals are limited to the 2D screen in front of the viewer. At the same time such audio spatialisation approaches seem to be used with caution and on a selective basis. Among other reasons, this may be again because filmmakers may not intend to distract from the 2D visuals displayed on the cinematic screen.

Considerations Regarding the Adoption of Certain Sound Spatialisation Strategies

Although cinematic multichannel sound systems were admittedly not designed for realistic 3D spatialisation, they were more than capable of fulfilling the spatialisation demands of cinema. However, for a number of reasons, even the most adventurous of sound designers and filmmakers do not appear to explore the increased spatial capabilities of modern multichannel sound systems extensively. Admittedly, surround spatialisation is employed more frequently compared to the past, especially after the introduction of surround systems with discrete channels (Kerins 2011; Dolby Laboratories 1999, 2005) something that has also affected pre and post-production decisions and strategies. In spite of this, it appears that many filmmakers are still using the multichannel soundtrack on an occasional and selective basis, rather than as a fundamental cinematic component. To some extent, this use of multichannel sound could be justified by the suggestions of numerous sound designers and filmmakers that just because multichannel sound is available it does not mean that it should be employed constantly, but its use should be dictated by the most important element of the movie: the story (Kerins 2011: 148). However, on Kerins' words:

> *'despite the importance of story and narrative as a determining factor in deciding whether the digital surround style might fit a given movie, story is not the ultimate determinant of whether that style is actually used. Often aesthetic choices are limited by logistical*

> *and practical concerns.' (Kerins 2011: 148)*

Such concerns have to do with various issues that may be beyond the interests or the authority of the filmmaker. For example, production companies have to take into consideration that their soundtracks must be both compatible with, and intelligible in, sound systems with no (or with limited) surround sound capabilities (e.g. TV speakers or stereo systems). Again, Kerins accurately outlines this constraint:

> *'the makers of Jurassic park knew that they could use the DTS soundtrack to place sounds very specifically [e.g. to the right surround speaker]. What they could not do, though, was make it crucial to know that particular sound came from the right surround speaker, as viewers watching the film on video or TV would not receive this piece of information.' (Kerins 2011: 44).*

In addition, even theatres supporting the common multichannel sound formats may have significant variations in terms of their acoustics. Reasons for these variations could be differences on speaker/sound quality, room/space acoustics and reverberation, cabling, alignment of speakers and others. These expected variations between reproduction systems effectively mean that a soundtrack could be experienced differently on different environments, something that could make the heavy use of the surround channels for soundtrack spatialisation problematic. This is highlighted in Ben Burtt's comment that *'sound events that had been carefully placed on the surround channels were simply not being properly played in theatres'*, something that led him (and other sound designers) to place *'all narrative information on the front speakers, with the surrounds used for spectacular (but nonessential) enhancements.'* (Burtt on Kerins 2011: 162).
Similarly, the soundtrack should be experienced in a consistent manner from all members of audience sitting in different areas of a theatre, something that is not possible without either using extremely elaborate and/or personalised speaker arrangements or headphones. This is also something that could add another constraint to the creative freedom of the sound designers and mixers. Therefore,

in most cases even those interested in using extensive sound spatialisation in their soundtracks may be unable to experiment heavily with such practices, perhaps not because of lack of skill or imagination but because of certain structural and technological conventions within the cinema industry (**Figure 2-27**).

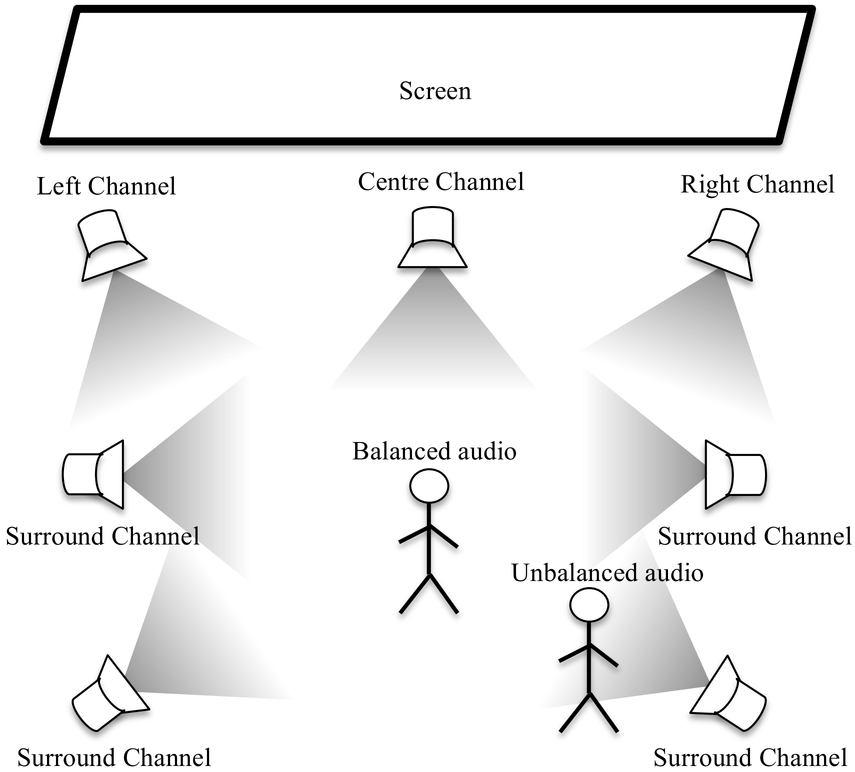


Figure 2-27: For members of the audience seating outside the optimal listening position audio may seem unbalanced

Another constraint on the work of sound designers could be that filmmakers are often *'perceived by film sound professionals'* (Kerins 2011: 159) as having very cautious and conservative approaches regarding the use of surround spatialisation in their movies. In many cases, this filmmakers' seeming *'aversion to active surround channels'* could lead sound teams to also employ conventional approach to surround panning, possibly against their intention (Kerins 2011: 159). Possible reasons for filmmakers employing such conservative and cautious approaches towards the extensive use of multichannel sound spatialisation may include personal preferences, professional background, lack of *sound education* or fear of distracting the audience's attention away from the screen (i.e.

commonly described as *the exit door effect*, Kerins 2011: 159). To the above, one could also add that in a cinematic environment that remains by and large 2-dimensional, a soundtrack that makes extensive and consistently heavy use of channels located outside the field of view (i.e. rear/surround channels) may be perceived as unnecessary (Kerins 2011: 159).

Finally, it must be stressed that modern cinematic productions are usually subject to quite strict time and budget constraints. This is something that could lead to the adoption of more conventional approaches at various production levels. As the extensive use of sound spatialisation can be a rather unconventional and debatable cinematic strategy, the odds are that in most productions the, frequently limited, available time and money would be spent on different production areas that are traditionally considered more important than the multichannel soundtrack (Kerins 2011: 148).

This list of reasons for the current limitations on the use of multichannel sound spatialisation in cinema is not exhaustive. However, it highlights the fact that the use of sound spatialisation in major studio productions may be dictated not only by creative decisions, but also by commercial, financial and technological constraints as well as by the preconceptions and preferences of many of those involved in the production (Kerins 2011: 159). To some extent, the aforementioned issues may provide an explanation for the, rather conservative, sound spatialisation approaches employed in the vast majority of modern productions.

2.4.3 Stereoscopic 3D Cinema Production Concepts

It is possible that the introduction of S3D visuals could be viewed not merely as the addition of stereoscopic depth to an otherwise unchanged type of cinema, but something that may bring new challenges and implications both to the cinematic production and experience as a whole. Considering the unique cinematic environment that may result from taking this viewpoint, it is interesting to question the ways in which this may be reflected in the multichannel soundtracks that accompany such S3D productions. In the following sections relevant concepts and ideas are presented and briefly discussed.

A large amount of work on the general field of sound spatialisation revolves around the concept of accurate and realistic reproduction of 3D soundscapes. However, although systems designed with such orientation may provide superior spatialisation to those of the multichannel systems used for cinematic purposes (e.g. 5.1 or 7.1 speaker surround sound systems) they either have failed to replace the latter as widely used cinematic sound standards or are used on a rather limited scale. One of the reasons for this may be that cinema does not necessarily require a realistic spatial representation of sound elements. This may apply both to on-screen and off-screen cinematic elements and objects (Chion 1994). To some extent, the reason for this could be that cinema is not merely an attempt to mechanically reproduce reality (Abrams *et al.* 2001: 113; Braudy and Cohen 2009: 285), but a form of artistic expression where symbols, associations and metaphors are equally important to (or perhaps even more than) realism (Chion 1994: 107).

Over the years, a cinematic language has been established that is capable of expressing the artistic and subjective view of the filmmaker, while also being acceptable and understandable by the viewer. The effectiveness and believability of this language is based on conventions and the assumptions these conventions lead to (Dix 2008: 71; Chion 1994: 108), rather than on how real the cinematic world looks (or sounds). Therefore, although a certain level of *believability* is arguably needed, realistic representation of events (in the strict notion of the term) may not be the main goal or concern of cinema (Kerins 2011).

Examples of this can be found on many aspects of cinematic production. For instance, 2D cinema projection does not display 3D depth in a realistic way (de Bruijn and Boone 2002: 2). Thus, we rely on perspective (e.g. the relative sizes of the depicted objects) in order to mentally create the impression of depth (de Bruijn and Boone 2002: 2). However, this unrealistic depiction of 3D worlds in a 2D screen does not prevent viewers from being immersed in the cinematic world and the story. In fact, the viewers consciously allow themselves to be immersed in the cinematic world through the psychological mechanism of suspension of disbelief (Abrams *et al.* 2001) despite the fact that the environment in front and around them is not an accurate representation of the real world. Similarly, it can be pointed out that the borders of the screen and the theatrical space around the viewer are visible, and the camera movements usually do not resemble a realistic point of view of a human being (de Bruijn and Boone 2002: 2).

From an audio perspective, the creation of realistic soundscapes and sound effects also may not always provide the optimal choice for production (Back and Des 1996: 1; Sergi 1999). On many occasions, the use of unrealistic sound effects could be more effective in cinema than the recordings of the real sound events (Sergi 1999, Bordwell and Thompson 2008: 287). For instance, it is widely known that largely unnatural and exaggerated sound effects like gunshots and punches can sound more convincing to the viewer than the respective realistic sounds. In addition, sound localisation can be significantly altered by phenomena such as sound/screen magnetisation (Chion 1994: 69) or the *exit door* effect (Chion 1994: 83; Kerins 2007: 3). Nevertheless, the fact that cinematic audio is not a strict representation of auditory reality does not seem to prevent audiences from being immersed in the cinematic. Based on this, it can be claimed that spatialising audio in a realistic way around the audience may not be an essential concern for cinema (including S3D cinema), in contrast to other forms of immersive media (e.g. video games, Virtual Reality systems etc., Hirst 2006: 20). This view may be also supported by the spatial layout of the cinematic environment where the visual cinematic action takes place mainly in front of the viewers without expanding around them in a realistic manner (**Figure 2-28**).

Thus, from a cinematic perspective, one of the main concerns of the sound designer may be how to use the multichannel soundtrack to enhance the cinematic experience and support the story (even if this means that largely unrealistic sound effects have to be used) rather than how to realistically and accurately reproduce 3D soundscapes around the audience (Field 2002: 2). This could be also supported by the fact that cinematic sound has followed a separate evolution from that of 3D audio systems capable of accurate spatialisation, despite the fact that the latter have been available for decades. Although the integration of 3D audio systems in cinema could admittedly increase the creative potential for sound designers and filmmakers, it is likely that in the cinematic context the creation of inventive, strategically planned and carefully designed multichannel soundtracks that serve best the cinematic story should be a priority over the creation of spatially realistic 3D soundscapes.

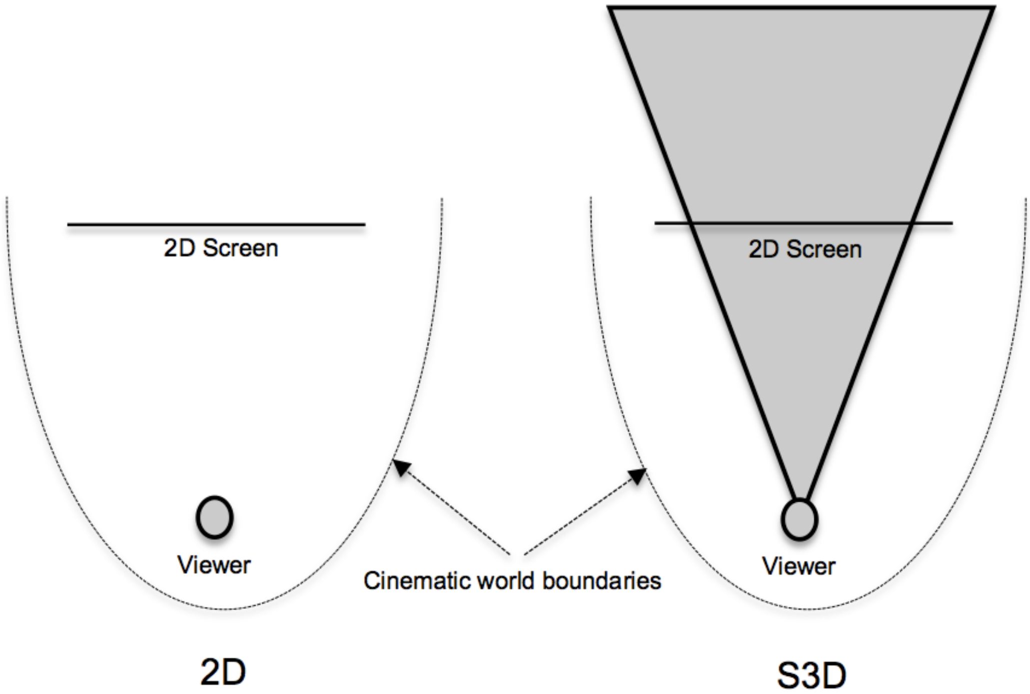


Figure 2-28: In both 2D and S3D cinema the boundaries of the cinematic world are determined by the dimensions and location of the screen. The created space is not a realistic 360 degree 3D environment but a symbolic, front-oriented one

Spatial Characteristics of the S3D Cinematic Environment

S3D cinema is based on stereopsis, the use of differences between the visual information on the left and right eye *'to get a strong impression of depth'* (Snowden and Troscianko 2006: 200-204). In the cinematic and photographic context, the term stereoscopic refers the projection of different images to each of the viewer's eyes, with each of these images showing a slightly different perspective of a visual scene. This causes the brain to create the illusion of objects appearing in front or behind the screen, as it resembles the natural angle difference from which our eyes observe the environment (Morton and Edwardz 2010; Mendiburu 2009; Clark 2010). A detailed analysis of stereopsis and S3D visuals for cinematic use is beyond the scope of this study. However, what is relevant for our purposes is the spatial characteristics of the created cinematic environment. As the two different stereoscopic image streams (Left and Right eye) are projected onto the screen, the stereoscopic cinematic space is inevitably shaped by its physical borders. The result is a viewing space '*more like a pyramidal box'* (Mendiburu 2009: 27) in front of the viewer commonly referred to as the *stereoscopic window* (Clark 2010: 9; Autodesk 2008: 3, **Figure 2-29**).

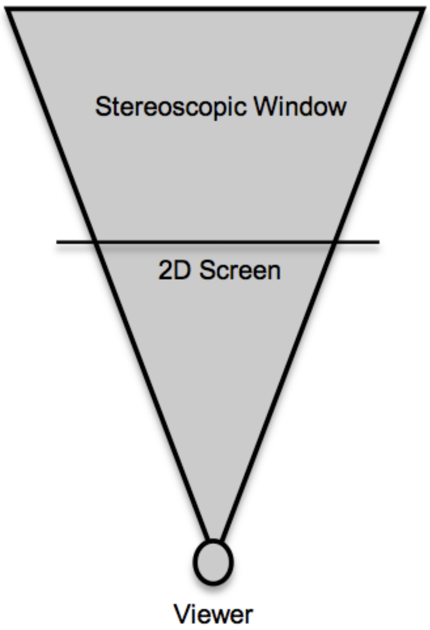


Figure 2-29: The Stereoscopic Window

One of the reasons that make the spatial layout of the S3D space relevant is that it can affect decisions regarding shot composition, camera placement and camera movement. For instance, Mendiburu (2009: 92) claims that in 3D *'you have to think in terms of volume composition, instead of picture composition'*, as '*you will box the action more than you will frame it'*. Similarly, Clark's (Clark 2010: 8) view towards S3D scene composition is that *'to take advantage of the unique capabilities of 3D, scenes may be composed in a way that emphasizes the depth of the set'*. This may include shots consisting of more depth layers or particular camera placements in order to stress the enhanced depth of the cinematic world (Clark 2010: 8-9). For example:

> *'while a 2D cinematographer might stage a conversation between two actors as they stand against a wall, a 3D filmmaker, to take full advantage of the medium, might prefer to stage the same scene with the actors conversing as they stroll down a street, passing other pedestrians and moving through multiple layers of depth'* (Clark 2010: 8).

Additionally, more frequent or pronounced camera movements might be used in S3D:

> *'filmmakers working in 3D generally opt to keep their cameras in motion, relying less upon static shots and more upon carefully-planned camera choreography to engage the audience in the story'* (Clark 2010: 9).

Another reason that makes the spatial layout of the stereoscopic window important is that it is subject to *stereoscopic window violations* (Mendiburu 2009: 97). Such violations may occur when visual objects that appear in front of the screen (i.e. negative parallax space) partly cross the edges of the stereoscopic viewing space. From the perspective of the audience this can be confusing, as the visible parts of the object could appear to be floating in mid-air with its other parts unexpectedly missing (Gardner 2009). S3D filmmakers use a number of strategies to address this issue, such as avoiding to place objects close to the edges or masking the edges of the screen with *movable mattes* (or *floating windows*) (Clark 2010: 9). This is a term used for the application of black borders to the edges of the stereoscopic images in order to make the parts that violate the stereoscopic window invisible. Irrespectively of what strategies may be available for avoiding such violations, it appears that in S3D it is perhaps even more important to keep the viewers' visual attention away from the edges of the screen than it is in 2D cinema (Mendiburu 2009: 27, 92). This may impose a unique set of constraints to the stereoscopic production process, something that could also affect aesthetic and creative filmmaking decisions.

Another relevant concept related to S3D cinema production is that the use of selective focus to *'direct the viewer's eyes to the key point'* does not seem to work in S3D in the same manner as in 2D (Bayon 2010; Lelyveld 2009). Although subtle use of selective focus has been already successfully utilized in S3D films (Gardner 2009), it is probable that out-of-focus portions of a S3D scene could create *visual confusion* and *discomfort* (Bayon 2010; Lelyveld 2009). This can cause filmmakers to avoid the use of selective focus as their main attention-guiding device. In such cases, other visual means can be used, such as stronger lighting and color cues or drastic changes in the structure of the shot composition (Bayon 2010; Clark 2010: 8). It is possible that sound could also be added to the list of cinematic devices that can be used for guiding the attention of the viewer within the stereoscopic window.

Finally, postproduction, and in particular editing, could be also affected by the introduction of stereoscopic cues, although this is still a largely debatable statement (Mendiburu 2009: 92, 151; Autodesk 2008: 2). Traditional editing techniques that work well in 2D cinema are not guaranteed to be effective in 3D (Autodesk 2008: 7, Mendiburu 2009: 96). For instance, the presence of stereoscopic depth cues may create a need for longer scenes (Clark 2010: 10; Bayon 2009), which may be also *'blended together at a slower pace'* (Mendiburu 2009: 26, 151), in contrast to the rapid scene transitions commonly used in many contemporary 2D productions (Kerins 2006: 44). This is not only because the viewer may need more time to explore a S3D image than a 2D one (Clark 2010: 9; Mendiburu 2009: 26), but also *because 'the audience tends to scan the whole scene before going back to the subject'* (Mendiburu 2009: 151). Additionally, scene cuts or transitions may have to conform to certain rules related to stereoscopic depth, a concept commonly known as *depth continuity* (Mendiburu 2009: 88, 92). Viewers have to readjust their point of focus (or convergence) between scenes that employ different stereoscopic depth, so this depth may have to be matched during editing (Clark 2010: 10; Lelyveld 2009). This is something that could make S3D editing different than that of 2D cinema.

Based on these ideas, one could argue that stereoscopy may be regarded by those more keenly interested in the S3D cinematic form as an opening of an '*uncharted land'* with its own rules and challenges (Mendiburu 2009: 2). As Mendiburu points out:

> *'while you start envisioning your movie, you start making choices that will most often stick until the end of the project and eventually show on the screen. If you make artistic choices that run against 3D effects, mood, or quality, you'll have to work around or even against them. (Mendiburu 2009: 96).*

In this case, certain artistic and technical decisions may need to be made during pre-production and production, and then reviewed throughout post-production (Mendiburu 2009: 38, 43). Such considerations related to the S3D visuals may, in turn, have an impact on the way we conceive and create the multichannel soundtracks that will accompany these visuals. In the following sections, a number of possible approaches to the design of soundtracks specific to the S3D form are briefly discussed.

Using Audio to Draw Visual Attention

Guiding the viewers' attention to particular visual objects or areas within a scene may be different in S3D than in traditional 2D cinema (Bayon 2010; Mendiburu 2009: 87, 99). To a large extent, this is because *selective focus* and *depth of field* do not work in S3D as effectively as in 2D (Clark 2010: 8; Lelyveld 2009; Hayes on Bayon 2010), although such cinematic techniques are sometimes used in also S3D productions. From a practical perspective, a reduced control over selective focus may be an undesired feature. This is because filmmakers more often than not may wish to control and guide the viewers' visual attention to particular areas or subjects of a given scene. Considering that human visual attention and localisation are multisensory perceptual processes that rely heavily on audition (Mastoropoulou *et al.* 2005: 2), it is proposed that strong direction and depth/distance auditory cues could be also used as visual attention-guiding devices.

Taking into account that modern cinematic audio systems possess increased spatialisation capabilities, it is expected that the required level of control over the positioning of sounds across the stereoscopic window should be attainable.

Additionally, modern audio technologies allow for detailed manipulation of the timbral and dynamic characteristics of the sound elements. This could result in filmmakers being able to adjust not only the perceived position of the sound across the length of the viewing space, but also the perceived depth of these sources within the S3D environment. Therefore, the viewers' visual attention could be guided even more accurately towards the positions of visual objects within the S3D scene by means of appropriately processed auditory cues. In addition to position, attention may be also attracted by making the sound of a certain object prominent in the mix: this can be done by choosing the sound carefully in relation to what else is in the mix at the same time and balancing it appropriately.

One of the main advantages of such approaches is that they do not require changes on the overall visual character and layout of the scene.

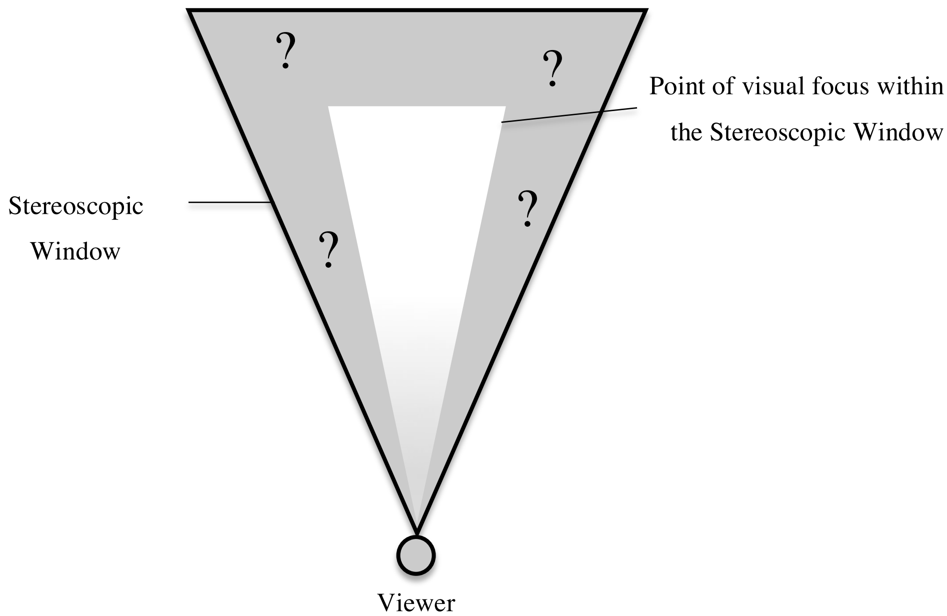


Figure 2-30: Focusing within the Stereoscopic Window is more complex than in a 2D screen. The viewer needs to adjust their visual attention within a 3D space rather than just a part of the 2D screen

For example, let us consider a scene that takes place inside a jungle. Assuming that a character appears in the centre of the stereoscopic window (for instance, among tree branches) an enemy partly hidden on the back of the viewing space aims at him with his arrow. As the audience is likely to be looking at the protagonist in the centre of the scene they might not notice the hidden enemy. In 2D cinema, the visual attention of the viewer could be led to the enemy by means of selective focus (by putting the protagonist out of focus while sharpening the focus of the enemy). However, this approach might not be as effective in S3D as the viewer is likely to attempt to readjust their focus rather than turn their attention to another part of the S3D scene (Lelyveld 2009; Clark 2010: 8). This could result in the viewers not focusing on the desired area. In addition, it could result in discomfort and confusion (Lelyveld 2009; Bayon 2010). By positioning appropriately processed auditory cues in the direction of the enemy, the viewers' visual attention could be attracted to the desired part of the viewing space, where they would visually identify the hidden character (Mastoropoulou 2005: 2).

Using Audio to Support Camera and Visual Object Movements

One of the most commonly exploited features of S3D cinema is the *viewer space* effect (Autodesk 2008: 4). This term refers to the creation of the illusion of Front-to-Back/Back-to-Front visual object movements between the viewer and the screen. In many cases, this may create a requirement for distinct corresponding sound movements within the soundscape, in order for these visual movements to become more imposing. Filmmakers have been using appropriate sound spatialisation to create or support the broadly known *fly-over effects* for decades (Vanhoutte *et al.* 2010: 11; Nudds 2007: 37). However, such uses of sound are rather limited and used sparingly in most S3D cinematic productions. In many instances the visual S3D *viewer space* objects are silent. One possible reason for this may be that fly-over effects can be distracting when viewing the 2D version of the film that may use the same soundtrack as the S3D version. Admittedly, not every visual cinematic object or action within a movie needs to be consistently supported by a corresponding sound event. However, in the case of viewer space effects the aesthetic intention is to create the illusion of objects moving inside the theatre space. It is possible that this illusion could be enhanced and/or extended beyond the viewing space if the movement was supported by the corresponding fly-over sound effects, as *'our auditory experience of sounds commits us to the existence of objects'* (Nudds 2007: 37). It is proposed that viewer space visual object movements in S3D may benefit by the accompaniment of appropriately designed corresponding sound events (Woszczyk *et al.* 1995: 5).

The above concept may apply not only to the typical *viewer space* effects intended to surprise and impress the viewer, but also to various other S3D objects that have no particular narrative or emotional significance, and that appear to move within the theatre space as the camera moves through the stereoscopic scene.

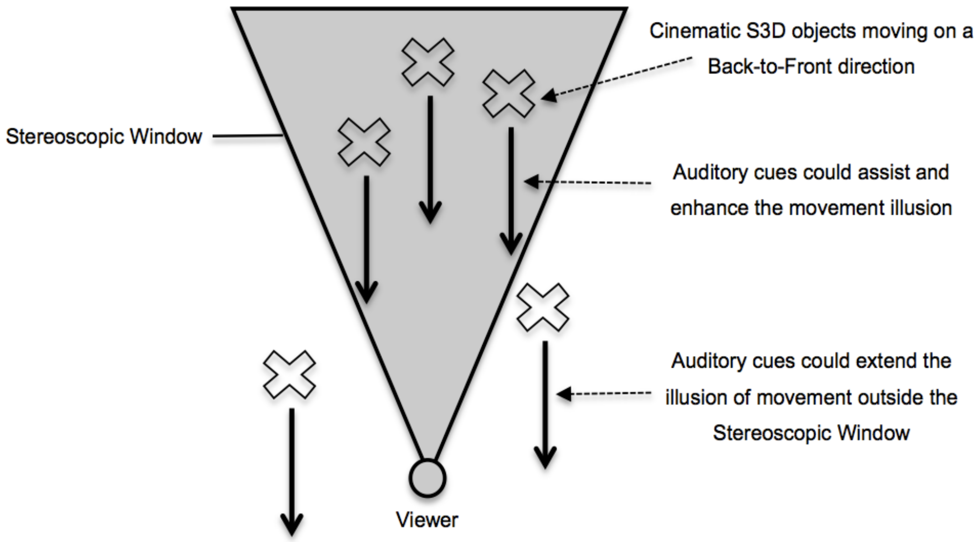


Figure 2-31: Auditory cues could enhance and/or extend the illusion of movement within the Front-to-Back/Back-to-Front axis of the Stereoscopic Window

For example, let us imagine a scene inside a forest. As the camera moves through the forest with a forward direction, subtle symbolic or realistic sounds could accompany the trees movements across the theatre space. As the trees exit the stereoscopic window the sounds could be panned to the rear, expanding the cinematic world backwards. This could be used to create a sense of the viewer actually traveling through the forest while leaving the trees behind. Such use of sound could accentuate the apparent movement of the camera and enhance the overall sense of immersion of the scene (Woszczyk *et al.* 1995: 5: Nudds 2007: 37, **Figure 2-31**).

It is possible that prominent sound movements outside the stereoscopic window could distract the viewer from the visual action. However, this may be avoided by arranging the sound spatialisation on a front-to-back orientation without drawing attention to the left and right edges of the stereoscopic window, as the viewer is less likely to search for visual cues in the rear of the cinematic space.

It is proposed that the use of sound events that correspond to visual object movements within the viewer space could be used to increase the presence and impact of these movements (Nudds 2007: 37). Ultimately, it may be possible that this could be used to enhance the sense of immersion and the envelopment of the cinematic experience as a whole without altering the visual composition.

Using Audio to Support Stereoscopic Editing

The introduction of stereoscopic cues may affect decisions related not only to the spatial character of each scene, but also to how different scenes are combined together during editing (Autodesk 2008: 7; Mendiburu 2009: 4, 92). For instance, a cut between a scene where the main visual object appears in front of the screen to one where it is far beyond it may be avoided in many occasions. This is because it could cause visual discomfort and confusion to the viewer (Mendiburu 2009: 88; Lelyveld 2009). S3D filmmakers usually take this into account during the preproduction and production phases and adjust the depth of succeeding scenes according to a relevant *depth script* (Autodesk 2008: 7; Mendiburu 2009: 93), a specially designed plan for the alterations of S3D depth throughout the movie. However, this may compromise the options and the creative input of the editor, as direct cuts between existing scenes with great stereoscopic depth differences (commonly known as *jump cuts*) would be generally avoided (Mendiburu 2009: 88, 153). It is proposed that a possible way to achieve transitions between such scenes is a combination of visual fades with appropriate audio cross-fades. In particular, a gradual cross fade between the soundscapes of the two succeeding scenes can start before the first scene ends. The visuals of the first scene can then fade to black, while the soundscape of the second scene becomes prominent. The fact that the visuals fade to black is crucial here, as *'once we remove visual stimulus and surround ourselves with appropriate auditory cues immersion into a virtual environment is almost guaranteed'* (Haines and Hooker 1997: 3). It is possible that by creating an appropriate soundscape that includes strong distance and directional cues the viewers' spatial perception of the cinematic world can be manipulated (Chueng and Marsden 2002). This could be regarded not merely as a transition from one soundscape to another, as is the case in many audio transitions in classic 2D cinema, rather than as one from an acoustic space with its unique shape and dimensions to another. After the desired spatial expectation is established by means of sound, the visuals of the second scene could fade in.

Cross-fades between scenes are also frequently employed by editors of S3D movies (Mendiburu 2009: 154). Such transitions could be also supported by the use of appropriate sound spatialisation effects. As an example, let us assume a transition from a scene with few or no stereoscopic depth cues to one with rich S3D content. One of the ways to achieve a smooth transition between these two scenes would be a visual cross-fade (Mendiburu 2009: 154). The latter could be supported by the creation of a relevant audio cross-fade, with the soundscape also gradually expanding from the front plane towards the back of the theatre. Such an approach could help in achieving two goals. Firstly it could create a more immersive and imposing transition effect. Secondly, it could prepare the audience for the change of S3D depth between the scenes, especially if the sound transition occurs slightly in advance of the visual one.

Editing might be also affected by the fact that S3D scenes could take longer to read (Clark 2010: 9; Mendiburu 2009: 151), something that could cause editors to allow more playing time for scenes with greater stereoscopic depth. This, in turn, may affect the overall pace and rhythm of the movie. According to relevant research, sound has a strong temporal effect on audiovisual perception (Holt 1997: 49; Mastoropoulou 2006: 6). For instance, while studying the effect of music on the perception of display rate of animations, Mastoropoulou (2006) reported that *'slow tempo/relaxing music resulted in longer duration estimations and slower perceived temporal rates'*. It is, thus, possible, that the soundtrack could be used for adjusting the perceived pace of particular scenes or sequences if and when such a perceptual effect is required by the filmmaker.

Finally, the multichannel soundtrack mix could be used in order to increase the sense of immersion and envelopment in scenes with little or no stereoscopic depth. As the viewing of S3D material for extended periods of time can be tiring, scenes with shallow stereoscopic depth are routinely placed between ones with extensive depth cues during editing. These shallow depth scenes are commonly known as *'rest areas'* and their inclusion at regular intervals throughout the movie can be a major concern for many S3D filmmakers (Mendiburu 2009: 88). However, the decreased stereoscopic visual depth of such scenes could possibly compromise the sense of immersion and envelopment. It is possible that by using appropriate enveloping soundscapes that include strong distance and spatial audio cues filmmakers could compensate for the lack of visual depth and keep the viewers immersed in the cinematic world by means of sound. Admittedly, there may be concerns that the introduction of such cues may create intelligibility issues by masking or interfering with the dialogue. However, if such cues are deemed by the filmmaker to be required for the enhancement of depth perception, audio manipulation techniques such as temporal or spectral separation of the audio cues and the dialogue could be utilized. Additionally, spatial separation may be used, for instance the dialogue and audio cues may be carried by different channels (e.g. the sound containing the depth cue could be sent to the Left or Right front channels while the dialogue could be carried by the centre channel).

Considering the General Spatial Orientation of the S3D Soundtrack Mix

Based on the increased sense of depth in front of the viewer and the spatial layout of the stereoscopic window it is possible that soundtrack mixes for stereoscopic scenes may adopt a more pronounced front-to-back orientation. This view is partly underlined by the increasing use of the Dolby Surround 7.1 sound system in recent S3D productions (Dolby 2010). Dolby 7.1 utilizes two extra audio channels compared to the widely established 5.1 audio formats, and these extra channels are placed to the rear of the viewer. They are, thus, used to enhance the front-to-back sound spatialisation capabilities of the system. In this case, the underlined interest towards a front-to-back soundtrack mix character is in line with front-to-back orientation of the stereoscopic window (**Figure 2-32**).
In addition to matching the visual orientation, such an audio spatialisation approach may provide another advantage: it is less likely to draw the viewers' attention towards the edges (and, thus, outside) the stereoscopic window. Distracting the viewers' attention from the screen has been a main concern in 2D filmmaking for years (Chion 1994: 83; DeLancie 2000; Kerins 2011: 72). In S3D

cinema these effects could be even greater, as the edges of the stereoscopic window *'are dangerous places where images can be painful'* (Mendiburu 2009: 27, 87, 92).

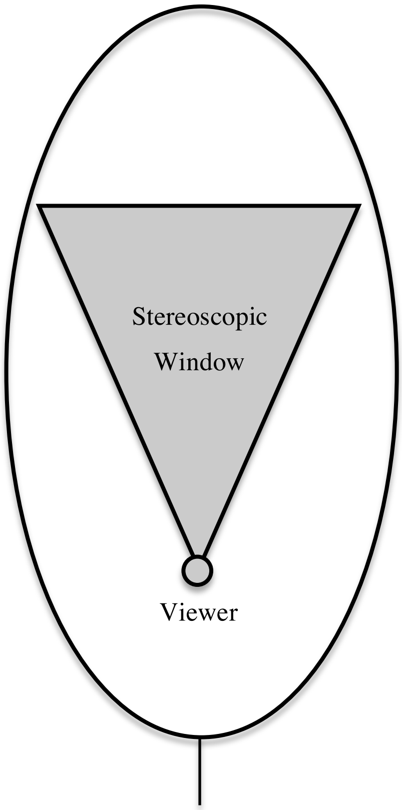


Figure 2-32: The spatial characteristics of the Stereoscopic Window make the use of Front-to-Back soundscape more appropriate than a Left-to-Right one

In the case of sound spatialisation within the stereoscopic window (i.e. in front of the viewer and extending towards the screen), a front-to-back sound spatialisation approach is likely to keep the viewers/listeners' attention to the safer parts of the S3D space: the front and centre of the stereoscopic window. In the case of sounds emanating from the rear of the theatrical space, it is also possible that the viewer/listener is unlikely to turn and look at the apparent source of the sound, as they would be aware that all the visual action takes place in the front. It is proposed that a possible way to expand the sense of depth (front-to-back orientation) of the soundtrack mix is to introduce sound objects with strong or exaggerated distance cues (Turner *et al.* 2011; DeLancie 2000;

Mastoropoulou 2006: 20, **Figure 2-33**). Although such distance cues could be used to reflect distance both on the front and the rear of the stereoscopic space, the current study focuses on their effect in the front plane of the viewing space and within the stereoscopic window. This is because the effect of auditory depth cues in the front plane of the S3D viewing space appeared to be an appropriate starting point for the exploration of such effects in S3D cinema, as it matched the main goal of the S3D visuals: to enhance the sense of depth within the stereoscopic window. The exploration of such auditory cues on the rear plane of the viewing space is also an intriguing idea, but it may come with a different set of considerations and constraints that exceed the scope of this study. However, it may be a topic that can form the basis of further work in the area of sound spatialisation for S3D cinema.

As distance perception is a multimodal process greatly affected by auditory cues (Mastoropoulou 2006: 2), it may be possible that by altering the timbral and dynamic characteristics of cinematic audio objects their perceived distance from the viewer could be increased or decreased as necessary (Kaye and Lebrecht 2009: 182; Mastoropoulou 2006: 15, 20). Such audio manipulation could, in turn, affect the impression of depth of the stereoscopic scene as a whole (Turner *et al.* 2011).

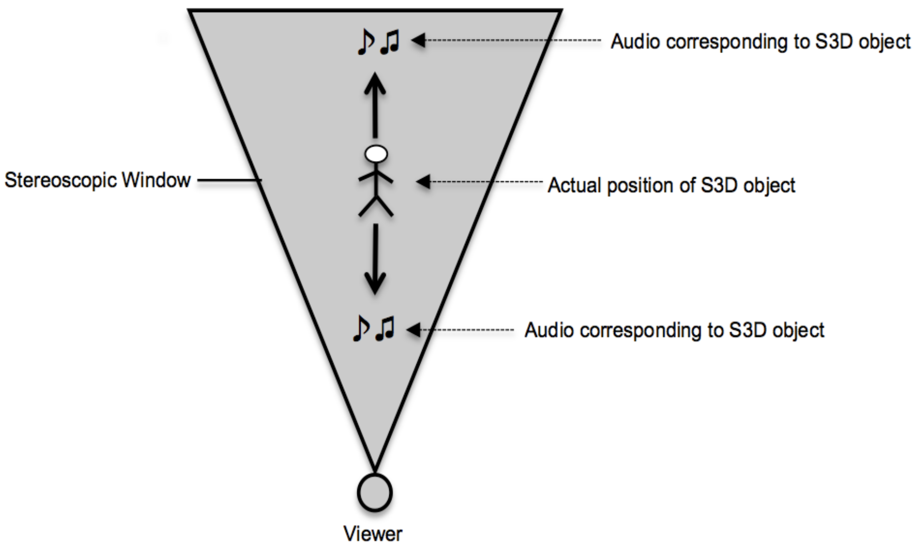


Figure 2-33: Strong and/or exaggerated auditory distance cues could be used in order to increase or decrease the perceived distance between S3D objects and the viewer/listener

For example, let us consider a scene where a number of characters are located around the viewer and at various distances. Characters appearing at a given stereoscopic visual depth within the stereoscopic window could seem more intimate to the audience if the sounds they produce are dry, loud and at their full frequency range. Similarly, sounds produced by characters further away from the viewer could be filtered (e.g. attenuate high frequencies), more quiet and with a certain amount of reverberation added (Kaye and Lebrecht 2009: 182). This is in line with recent research exploring whether the introduction of distinctive audio samples (i.e. samples that could function as distance and spatial cues) could significantly affect the viewers' perception of stereoscopic visual depth of a given S3D object (Turner *et al.* 2011). Turner *et al.* (2011) also reported results indicating that audio depth cues have the ability to influence the depth perception of S3D visuals.

In addition, it is possible that a mixture of sound objects representing characters located at different distances from the viewer could be positioned to the rear. Perceptually, this could potentially expand the cinematic world backwards, therefore enhancing the overall perceived depth (and consequently the immersion and envelopment) of the scene. This may apply to characters, objects and events within the S3D scene or even its overall sound-emitting background.

Admittedly, the use of audio distance cues in cinema is not a new concept as filtering, reverberation and volume level control have been consistently used in cinema for the enhancement of depth perception. However, such audio processes could be more relevant in a cinematic environment heavily based on visual depth cues, such as S3D cinema, than they are in 2D cinema. Based on this, the extensive use of audio depth cues within the soundtrack mix of S3D movies may be an idea that is worth exploring further, as it could potentially provide filmmakers with an additional creative tool. The main area of interest of the current study is to explore whether the introduction of distinctive auditory depth cues within the stereoscopic window (i.e. samples that could function as distance and spatial cues) could affect the viewers' perception of the depth of a given S3D scene (Turner *et al.* 2011).

2.4.4 Summary

In many cases, cinematic production may be affected by the addition of S3D visuals in various ways. Decisions related to camera angles, scene length and composition, visual focus, lighting, coloring and postproduction may be based on the unique nature and spatial characteristics of the S3D cinematic space. It is possible that the implications of using S3D could also have an impact on the decisions related to the construction of the soundtrack. Certain sound design spatialisation and mixing techniques and approaches that are commonly employed in 2D cinema may not always be the optimal ones for use in S3D cinema or they might assume a different role in the different context of S3D. A number of possible uses of the soundtrack in the context of S3D cinema in light of the nature and spatial or technical characteristics of this unique medium have been proposed.

2.5 Conclusion

This is a multidisciplinary study involving information from various different topics and fields of expertise. In this context it was relevant to establish a basic understanding of key concepts from different areas of interest, such as basic cinematic history and production, technological advances in the audiovisual domain, basic human perception, sound design and mixing, as well as possible implications of using S3D visuals in the cinematic context.

In this chapter, firstly the concept of cinematic immersion and envelopment was presented in a historical context. This included topics such as the numerous attempts to create increased visual depth in cinema in the past century and the introduction and evolution of multichannel sound systems.

Next, basic audiovisual human perception concepts were presented. These included basic visual and auditory perception, as well as their cross-modal integration and its role in the cinematic context.

Additionally, this chapter included the introduction and brief discussion of basic sound design, sound spatialisation and mixing concepts. Initially, typical sound spatialisation techniques and approaches in classic (2D) cinema were presented.

Then, possible differences between 2D and S3D cinema production were highlighted followed by a discussion on the impact they may have on the creation of the S3D multichannel soundtrack.

Based on this, a number of possible ways in which the multichannel soundtrack could be used in the context of S3D cinema were proposed.

3 Experimental Methodology

3.1 Introduction

It may be possible that audio could be used to support and enhance the overall sense of depth of a S3D scene, draw visual attention to certain parts of the stereoscopic window, accompany camera or visual object movements or extend the S3D cinematic world beyond the limits of the S3D visual cinematic environment (see: **Chapter 2.4 - S3D Cinematic Soundtrack Mixing Concepts**). The experiments presented in this thesis focus on the possible impact of on-screen sounds accompanying S3D clips and corresponding to the displayed cinematic setting (background sounds) and the on-screen moving or static S3D objects.

In classic 2D cinema, the illusion of visual depth is created by means of monocular depth cues (Mendiburu 2009: 26). Accompanying auditory depth cues, such as volume level attenuation, high frequency loss and reverberation, are also commonly used to support the visuals and enhance the sense of depth of the cinematic scene (Beauchamp 2005: 145-149; Orpen 2003: ix-34). This should not come as a surprise, as it is broadly accepted that the perception of depth in humans is a cross-modal process that relies on both visual and auditory sensory cues (Vroomen and de Gelder 2000; Turner *et al.* 2011; Ecker and Laurie 2004).

In S3D cinema, the introduction of binocular (stereoscopic) visual cues aims at the creation of an increased sense of visual depth within the cinematic world (Mendiburu 2009: 26). Admittedly, this may not have a direct impact on the soundtrack and its creation. In the majority of commercial cinematic S3D productions, the 3D enhancement of visual depth may not be necessarily reflected in the audio domain. However, the possibility of this increased depth being also reflected or transferred in the audio that accompanies these visuals should not be ignored, as the audiovisual interaction may function as an additional creative tool in S3D filmmaking.

Although a great amount of research has been conducted on the field of auditory depth perception (Coleman 1963; Ecker and Laurie 2004; Turner *et al.* 2011), the possibility of using pronounced auditory depth cues with the aim of reflecting, supporting or enhancing the sense of depth in S3D cinema has not been extensively studied yet. The aim of the experiments described in chapters 4 to 7, is to make a contribution to the research of this largely unexplored area by investigating a small number of ways in which auditory depth cues could be used in the S3D cinematic context. In this chapter, an overview of background concepts used in the experiments is provided, followed by a presentation of the experimental methodology used.

3.2 Background Concepts

Research shows that humans use cross-modal sensory cues to perceive the spatial characteristics of their surroundings (Shams and Kim 2010; Vroomen and de Gelder 2000; Lippert *et al.* 2007). In this study it is hypothesised that the introduction of prominent auditory depth cues within the soundtrack could enhance and affect the overall sense of depth in S3D animation clips. Such auditory depth cues may include, among others, the selective filtering of the high frequencies and the alteration of the volume levels of audio components of the soundtrack (Coleman 1963).

Based on practical everyday life experience we know that in many occasions the more distant a sound-emitting object is located in relation to the viewer/listener the lower its volume may appear to be (Ingard *et al.* on Coleman 1963: 306). Secondly, the high frequencies of the signal tend to attenuate at different rates than low frequencies as sound travels through air (Ingard *et al.* on Coleman 1963: 306). In other words, in many occasions the overall volume level and the high frequency content of a given sound tend to increase as a sound moves closer to the viewer/listener and decrease as it moves further away. Admittedly, this is not true in all cases and situations, as other factors also affect our perception of auditory depth. However, it may reflect the general empirical knowledge available to most people for a large amount of familiar everyday life situations. For the sake of simplicity, the above could be summarized in the following argument:

Sounds supposed to be emitted by distant cinematic sources in many cases could be expected to have decreased high frequency content and overall volume level based on a large amount of real-life experiences of the spectator.

Based on this, it is proposed that by altering the volume level and the high frequency content of a given sound, one could also alter the perception of depth of a spectator of a S3D scene and create the illusion that the source of a sound is closer or more distant to what the visuals suggest.

From an experimental perspective, one way to approach this is to investigate cases where the visual cues related to the sound-emitting cinematic object remains at the same apparent distance and only the depth cues of the accompanying sound are modified. If the assumption that such a change of the sound could affect depth perception, it may be possible to use audio depth cues to affect the perception of depth of both S3D objects within a scene and of the S3D scene as a whole without altering the actual S3D visual composition of the scene. This, in turn, may be useful for S3D filmmakers, especially in situations where the alteration of the S3D composition is either undesirable or difficult. It must be noted here that such audio manipulation techniques have been used regularly in cinema in the past. However, in an environment that is increasingly based on the illusion of extensive depth, such as S3D cinema, the role of these auditory depth cues could potentially change.

3.2.1 Depth versus Horizontal/Vertical Auditory Cues

Auditory depth is only one of the spatial elements of the cinematic environment. Using modern audio technologies filmmakers can also dictate the position of sounds on a horizontal and vertical axis (i.e. front left, right and centre channels or height channels in some systems), while sounds can also be placed outside the field of view (i.e. surround channels). The reason the experiments described in this thesis address only auditory depth and not the other two spatialisation dimensions is mainly the unique spatial characteristics of the S3D cinematic environment. The shape of the S3D viewing space (i.e. stereoscopic window) makes auditory depth an interesting starting point for studying auditory cues in

S3D cinema. It must be stated that sound spatialisation on the horizontal or vertical axis is also relevant to the concepts studied here. However, taking into account the depth-oriented spatial layout of the S3D visuals it was decided that priority should be given to auditory depth before the other auditory spatialisation directions are studied. Nevertheless, this should not be regarded as a preference or prioritization of one auditory cue over others, but as an attempt to start the experimental work from the point that seemed most obvious and relevant.

3.2.2 Relative versus Absolute Auditory Depth Cues

The two auditory depth cues used in the experiments (volume alteration and high frequency loss) are not absolute depth perception cues but relative ones (Mershon and Bower 1979). This means that any differences between two versions of an otherwise identical S3D animation clip (one with the presence of prominent auditory depth cues and one without) would be most likely detectable if there was a direct reference available to the viewer/listener or if the two versions were presented to the viewer in succession so that judgments could be made on the basis of a clear reference. However, such an experimental design may not be directly relevant in the cinematic context, as cinema viewers are usually only presented with a particular cinematic scene once. During a cinematic viewing there is no direct comparison of scenes and, therefore, no clear reference for the perception of depth. In this context, it is likely that the cinema spectator could resort to other available sensory streams, such as the combination of visual and auditory cues or the approximate estimation of distances between different objects within the cinematic scene.

In the S3D cinema context, it may be possible that viewers/listeners could resort to the available monoscopic and stereoscopic visual cues, using auditory cues as a confirmation for their depth calculation. This idea was used in the experiments carried out in this study in the majority of which auditory depth cues are treated as cues that are relative to a reference distance point established by the available S3D visual cues (**Figure 3-1**).

Additionally, at the start of the experiments, two clips representing the opposite end of the auditory and visual depth scale found in the S3D material were presented to the subjects so that the overall depth range was indirectly established before the actual test began.

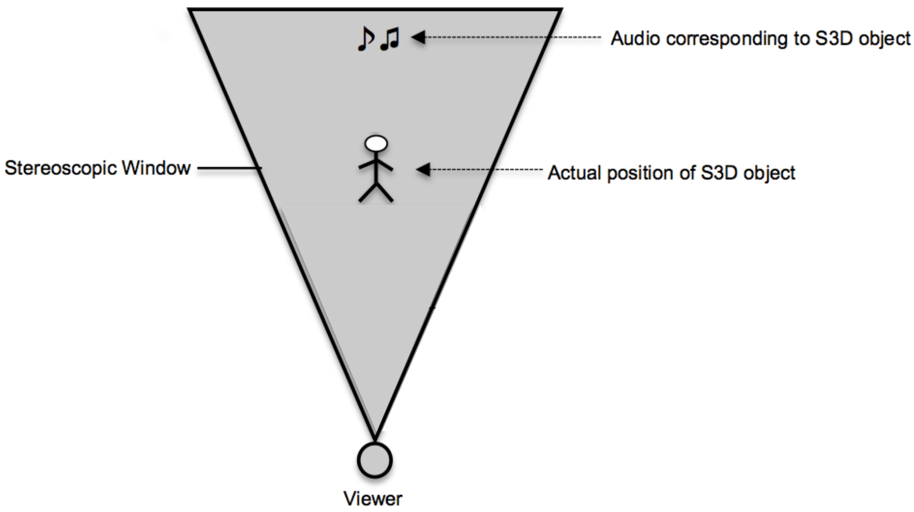


Figure 3-1: Visual cues can be used for judging the depth/distance of the cinematic object. This would establish a multisensory reference for judging the depth/distance of the corresponding sound

3.2.3 Cross-modal Integration and Auditory Scene Analysis

Perceptual confirmation of events by comparing inputs from different sensory streams is an important factor in determining depth and distance in our environment. The human brain routinely combines and analyzes cues coming from different sensory streams and is capable of identifying whether they belong to the same external event (Ernst and Bulthoff, 2004 in Lippert *et al.* 2007). This ability of the *brain 'can enhance perceptual clarity and reduce ambiguity about the sensory environment'* (Stein and Meredith, 1993 in Lippert *et al*, 2007: 102) and is frequently referred to as cross-modal integration (Holt 1997; Bolognini *et al.* 2005).

The perceptual organization of information obtained from the auditory and the visual sensory streams could be seen as a unified process. The brain uses cues obtained from each of the two sensory streams to reinforce and verify a perceptual event. The information obtained by one of the sensory streams (e.g.

audition) can cause the brain to search for confirmation in another (e.g. vision). For instance, *'when you hear a sudden sound you tend to visually orient to the location of the sound and that is an example of overt orienting of attention'* (Shams 2010). In addition to this, changes of the characteristics of one of the sensory streams may perceptually affect another. For example, *'if auditory and visual stimuli are presented synchronously but from different positions, the auditory event is mislocated towards the locus of the visual stimulus'* (Kitagawa *et al.* 2002). This indicates that in addition to comparing and organizing auditory and visual cues separately, the brain also monitors and compares cues across the two different modalities. If cross-modal cues confirm each other then more accurate conclusions can be drawn about a given event or object (Lippert *et al.,* 2007: 102).

Previous studies have shown that the presence of cross-modal cues can affect the perception of depth in 3D visual environments. Turner et al. (2011) have shown that the perception of distance of S3D visual objects can be affected by changing the distance of a loudspeaker playing a corresponding sound from the audience. Ecker and Laurie (2004) have shown that the addition of an auditory cue can change the perceived path of a ball moving in 3D-space. Kitagawa et al. (2002) have shown that a steady sound can be perceived to be changing in loudness when watching a synchronous 3D object moving in depth.

Research has also demonstrated that the brain uses the same universal principles to deal both with audio and visual stimuli. Auditory Scene Analysis is a term that describes how the brain utilizes temporal and spectral characteristics of sounds to identify and make sense of sound emitting objects in the environment (Bregman 1994). Bregman demonstrated that the brain uses temporal and spectral analysis of a given sound mixture to determine the nature, size, distance and importance of sound emitting objects around the listener. His work shows that Gestalt psychology principles apply to auditory perceptual processes in a similar way to how they apply to visual processes. For example, *'principles of perceptual organization like similarity* (in volume, timbre, spatial location), *good continuation, and common fate seem to play a similar role in the two modalities'* (Vroomen *et al.* 2000).

In addition to comparing cues across different modalities and using similar principles in audition and vision to determine distance and depth of objects, the brain also compares different cues within the same modality (vision or auditory) to make sense of an object presence. Snowden *et al.* (2006) showed that the brain combines all available visual cues to calculate depth and make decisions regarding the perceived depth and distance of an object. This heuristic approach to decide which cue is the most reliable for depth calculation has many similarities to processes described in auditory scene analysis that are used to identify depth through auditory depth cues. For example, familiarity with a sound can be an important cue when judging the depth and distance of a sound-emitting object (Mershon *et al.* 1979). When the timbre and approximate volume level of a given object is known, the brain can make informed estimations about its distance based on its level and spectral content as well as knowledge that comes from familiarity of the size and nature of the sound emitting object.

3.2.4 Simple and Rich S3D content

If the brain uses both visual and auditory cues to calculate depth, then it could be possible that the amount of visual and auditory information provided could influence our ability to make this calculation. In scenes with rich S3D content the viewer/listener may be encouraged to visually scan the image more actively (Mendiburu 2009: 26, 151), thus paying more attention to the visual environment, while in simpler S3D scenes less effort may be required from the viewer/listener to scan the visual environment. In such cases, the degree of S3D richness may influence how much attention the viewer/listener pays to the auditory cues. This, in turn, could mean that the degree of S3D richness in a scene could influence the way the overall depth is perceived in this context. In this study, clips with both rich and simple S3D content were considered. This approach allowed for the coverage of a range of possibilities that S3D cinema can present us with. The S3D visual richness of the clips was determined on the basis of:

a) how many distinct layers of S3D graphics were available in each clip;
b) how complex these S3D graphics were.

The clips in the rich category included at least three distinct layers of S3D graphics and each one of them would extend to more than half the width of the

screen. The clips in the simple category included no more than two distinct layers that would extend to less than half the width of the screen.
These criteria ensured that between simple and rich clips there was a significant increase of S3D information (extending towards both the depth and the width of the cinematic world) that may require a substantial increase in viewing effort by the viewer/listener. More information regarding the use of this categorization can be found in section **3.3.4 – Apparatus and Material**.

3.3 Method

3.3.1 Naming Conventions

For clarity reasons, the names of the experiments use the following abbreviations according to the independent variable studied:

- (HFF) for High Frequency Filtering
- (LA) for Level Alteration
- (LAHFF) for the combination of Level Alteration and High Frequency Filtering

For example, the complete name for Experiment 1 is Experiment 1 (HFF), for Experiment 2 it is Experiment 2 (LA) and for Experiment 7 it is Experiment 7 (LAHFF) reflecting the independent variable.
In the cases of Experiments 3 and 6, where different experimental methods were used the names of the experiment also includes information regarding the experimental method. When the clips where viewed in a non-successive, randomized order the letters NS (for Non-Successive) were added to the name, while when the experimental method involved successive viewings the letter S (for Successive) was used:

- Experiment 3NS (LAHFF)
- Experiment 3S (LAHFF)

In the case of Experiment 3S (LAHFF), which was carried out twice, the letters **a** or **b** were added to the name in order to distinguish between the two different instances:

- Experiment 3Sa (LAHFF)
- Experiment 3Sb (LAHFF)

3.3.2 Design

Two types of experimental design where used in this study. The first type presented the two versions of each S3D clip (*control* – clip with the original unaltered sound and *experimental* – clip with auditory depth cues integrated in the sound) not in succession and with at least 3 different clips in between the *experimental* and *control* version of the same clip. This was in order to simulate cinema viewing where the same scene is not repeated twice in succession in the majority of occasions. The experiment employed a repeated measures design and the subjects were asked to rate the overall depth of given S3D clips (Experiments 1 (HFF), 2 (LA), 3NS (LAHFF) and 7 (LAHFF)) or the depth of objects movements within S3D clips (Experiments 4 (HFF), 5 (LA), 6NS (LAHFF)).

This design employed an Absolute Rating Scale/ACR (ITU 1998) with a scale ranging from 1 to 9 (1 = no significant depth of scene or movement, 9 = very prominent depth of scene or movement). The scores obtained during the viewings were later analysed using the Wilcoxon Signed-Ranks test (Robson 1994: 118). The decision to use this particular test was made because in many cases the data were not normally distributed (**Table 3-1**; **Figure 3-2**; Field 2005). This is in line with the methods used in similar past experiments and with the relevant ITU recommendation and literature (ITU 1998: 5; Robson 1994; Harris 2008).

The second type of experiment presented the two versions of each clip (*control* and *experimental*) in succession, so that the subjects could make a direct comparison. In this design, subjects were forced to select the clip that appeared to have a more pronounced depth between the two. This design was employed in three experiments (3Sa/3Sb (LAHFF), 6S (LAHFF) and 7 (LAHFF)). Results were analysed using the Binomial Distribution test (Robson 1994).

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Clip_A_ctl	.159	35	.026	.939	35	.053
Clip_A_exp	.180	35	.006	.926	35	.021
Clip_B_ctl	.163	35	.020	.947	35	.090
Clip_B_exp	.194	35	.002	.922	35	.017
Clip_C_ctl	.243	35	.000	.771	35	.000
Clip_C_exp	.289	35	.000	.751	35	.000
Clip_D_ctl	.186	35	.004	.918	35	.013
Clip_D_exp	.207	35	.001	.925	35	.020

a. Lilliefors Significance Correction

Table 3-1: Results of normality tests for Experiment 1 (HFF)

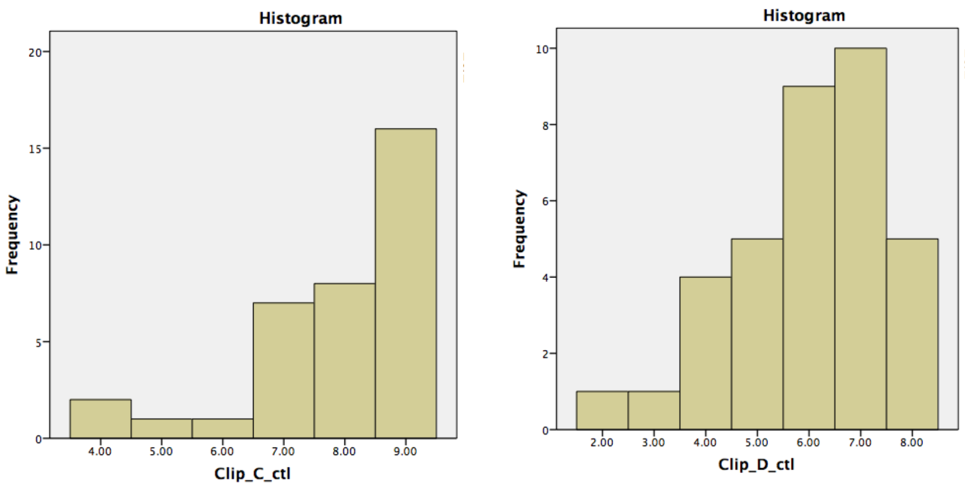


Figure 3-2: Histograms displaying results of normality tests for Clips C and D in Experiment 1

Overall Depth of a Scene: Experiments 1-3

The effect of high frequency and overall volume attenuation on the perception of S3D scenes was first studied in isolation (Experiments 1 (HFF) and 2 (LA)) and then when the two auditory cues were combined (Experiment 3NS (LAHFF)).

This was in order to be able to identify possible perceptual effects produced by the combination of the two different cues.

In Experiments 1 (HFF), 2 (LA) and 3NS (LAHFF) the independent variable was the frequency content, the overall volume level of the background soundtrack that accompanies the S3D visual clips and the combination of these two cues. The background soundtrack consisted of mechanical noises and hums that corresponded to the background of the S3D animation clip (i.e. the selected S3D clips portray a gigantic machine that envelops the characters). It must be noted that the background sound was the only sound present in the soundtrack and that there were no other sounds in the background or foreground.

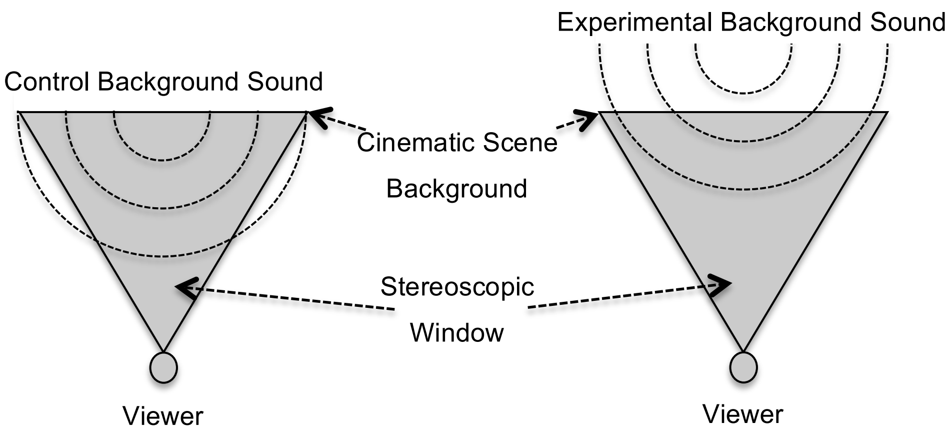


Figure 3-3: Background sound displacement in Experiments 1 (HFF), 2 (LA), 3NS/S (LAHFF)

The independent variable had two levels: *control* (the original, unfiltered background sound) and *experimental* (the same background sound but with its higher frequencies filtered or/and its overall volume level attenuated by a given amount, **Figure 3-3**).

The dependent variable was the perceived sense of depth of the overall S3D animation clip from the participants' viewpoint. The order of presentation of the control and experimental versions of the S3D clip was randomised between subjects and two versions of the same clip were not presented in succession. This was in order to mimic a real cinematic experience in which two versions of the same clip would not be directly compared.

Experiment 3Sa (LAHFF) is a follow up to Experiment 3NS (LAHFF). It used the same clips as Experiment 3NS (LAHFF) (i.e. both high frequency and overall volume attenuation applied to the experimental soundtrack), with the only difference being that this experiment involved successive viewings of the clips rather than randomized ones. The purpose of this follow up experiment was to test whether an experimental procedure that included a direct comparison of depth could provide different results to the results produced by Experiment 3NS (LAHFF).

Experiment 3Sb (LAHFF) is a repetition of Experiment 3Sa (LAHFF) in a different setting and with different participants. It was carried out in order to verify the findings of Experiment 3Sa (LAHFF).

Depth of moving objects: Experiments 4-7

In Experiments 4 (HFF), 5 (LA), 6NS (LAHFF) the independent variable was the gradual change of the frequency content and/or the overall volume level of a sound that accompanied a particular object moving from a given depth within a S3D animation clip and towards the viewer/listener.

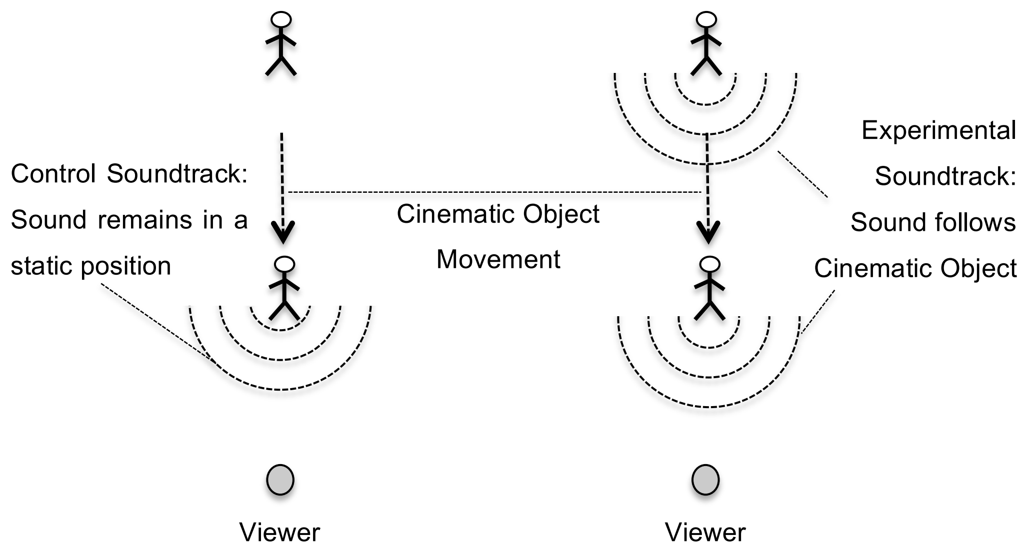


Figure 3-4: Sound corresponding to the S3D object movement was used in Experiments 4 (HFF), 5 (LA), 6NS/S (LAHFF)

The independent variable had two levels: *control* (the original, unaltered accompanying sound that did not change over time) and *experimental* (the same sound but with its higher frequencies or its overall volume level increased as it approached the viewer/listener, **Figure 3-4**).

The dependent variable was the perceived depth of movement of a particular sound-emitting object within the S3D clip. The order of presentation of the control and experimental versions of the S3D clip was randomised between subjects and the two versions of the same clip were not presented in succession.

Experiment 6Sa (LAHFF) was a follow up to Experiment 6NS (LAHFF) sharing the same clips of Experiment 6NS (LAHFF) but using the second experimental design (successive viewings). The purpose of this follow up experiment was to test whether an experimental procedure that included a direct comparison of depth could provide different results to the results produced by Experiment 6Sa (LAHFF).

In Experiment 7NS (LAHFF) the independent variable was the change of the frequency content and/or the overall volume level of a sound that accompanied a particular object within the S3D clip. The variable had two levels: *control* (the original, unaltered accompanying sound) and *experimental* (the same sound but with its higher frequencies and its overall volume level decreased). The order of presentation of the control and experimental versions of the S3D clip was randomised between subjects. The motivation for this experiment was to investigate whether the volume level alteration and high frequency filtering of a distinct and unambiguous sound (i.e. the sound emitted by a specific object that is visible on-screen) would have a different perceptual effect to the background sound mixture studied in Experiments 1-3.

Further details regarding each experiment and how the filtering and volume level attenuation were applied to the soundtracks are provided in the relevant sections of each experiment (see: **Chapters 4, 5** and **6**).

3.3.3 Participants

The sample size for the experiment was determined based on recommendations in the literature (Field 2005; Harris 2008; Robson 1994) and information derived from the GPower software (Faul and Lang 2007; Heinrich-Heine-Universität 2013), considering a statistical power of 0.8, a significance level of 0.05 and an effect size of 0.5.

The sample population for all experiments was selected in the following manner. Invitations were sent to candidates of all backgrounds, genders, nationalities and ages from the general population of the University of York using appropriate centralized email notifications and public advertisements. From the pool of candidates who responded to the invitation, the selection of the participants was based strictly on the order in which they replied to the invitation. This was in order to avoid any effects caused by the active involvement of the experimenter in the selection process.

All precautions were taken to ensure that prior to the test participants were unaware of the actual purpose of the experiment and of the actual research question. Participants had normal or corrected to normal vision and no significant hearing difficulties. This was verified by means of the following tests:

- ThinkQuest Library: The Eye Exam test (ThinkQuest 2011)
- Media College: Depth Perception Test (Media College 2011)
- British Society of Hearing Aid Audiologists: Online Hearing Test (BSHAA 2011)

Upon arrival, each participant was asked to take all three tests in order to ensure their suitability for the experiment.

In terms of previous knowledge and expertise, participants were non-experts in the field of audio production and/or sound design. This information was obtained by asking the participant about their professional and academic background, as well as their knowledge and expertise of audiovisual material production.

Further details about the participants of each experiment are presented in the respective chapters as different sample populations were used for each experiment.

3.3.4 Apparatus and Material

The following hardware and software parts were used for the experiments:

- PC Workstation
- NVIDIA Geforce 550Ti 3D Graphics Card (NVidia Corporation 2014)
- Optoma GT750 3D projector (Optoma Europe Ltd. 2012)
- Optoma DLP 3D Glasses (Optoma Europe Ltd. 2012)
- Genelec 8050A Loudspeaker (Genelec 2014)
- NVIDIA's Stereoscopic Player (NVidia Corporation 2014)

The distance between the projector and the screen was 200cm and the diagonal size of the viewing area of the screen was 90 inches (200cm) with an aspect ratio of 16:9. The single speaker used for the experiments was placed at the same distance as the screen (either below the bottom side of the screen frame or behind the fabric of the screen when acoustically transparent fabric was used). Prior to the experiments the speaker was calibrated in accordance to the relevant Dolby recommendations (Dolby Laboratories 2000).

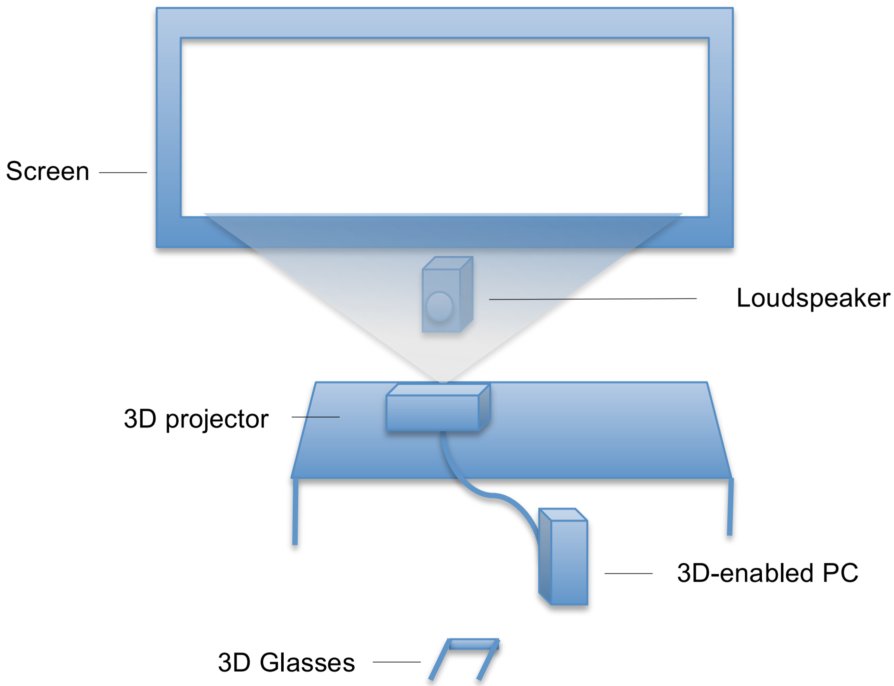


Figure 3-5: Illustration of the hardware setup

The footage was played back using the NVIDIA Stereoscopic Player through NVIDIA's 3D TV Play technology (Nvidia Corporation 2011).
The following handouts were also created and used for the experiments:

- Test instructions
- Informed consent form
- Demographic Information form

These documents can be found in **Appendix 3 – Experiment Documents and Forms**.

Experimental Space

For the purposes of Experiments 1 (HFF), 2 (LA), 3NS/3Sa (LAHFF), 4 (HFF), 5 (LA) and 6NS (LAHFF), a dedicated space was created in the department of Theatre, Film and Television at the University of York (**Figure 3-6**).

Figure 3-6: Experimental space for Experiments 1 (HFF), 2 (LA), 3NS/Sa (LAHFF), 4 (HFF), 5 (LA), 6NS (LAHFF)

The space was created using a cube-shaped metal frame, on the four sides of which curtains were attached. The curtains were made of thick fabric and were used both to visually isolate the participant from the surroundings and to provide a certain level of sound absorption in order to decrease possible sound reflections on the surrounding walls and objects. On the top side of the space pieces of plastic film were used to block light from the outside and to prevent other possible distractions. The dimensions of the space were: Height: 230cm, Width: 230cm, Length: 300cm.

Due to constraints and limitations outside the authority of this research, Experiments 3Sb (LAHFF), 6S (LAHFF) and 7 (LAHFF) were carried out in a different location in the post-production facilities of the Theatre, Film and Television department at the University of York (**Figure 3-7**). As this space was specifically designed for video processing and post-production tasks it had different audio isolation characteristics than the space created for Experiments 1 (HFF), 2 (LA), 3NS/Sa (LAHFF), 4 (HFF), 5 (LA) and 6NS (LAHFF). Additionally, the lighting could be adjusted in a more detailed manner. This resulted in the experimental space being more darkened providing, thus, improved viewing conditions.

Figure 3-7: Experimental space for Experiments 3Sb (LAHFF), 6S (LAHFF) and 7 (LAHFF)

Despite the fact that the space was different to the one used in the majority of the experiments, special attention was given to keep the equipment, software and parameters identical to the ones described earlier.

It must be stressed that the arrangement of the experimental space allowed for only one participant at any given viewing and they were also supervised. As such, the experience of the participant was different to that of a normal cinematic viewing where many participants co-exist and there is no apparent supervision of the actions of the viewers. Admittedly, the controlled viewings may not allow for conclusions relating to the possible role of cinematic spectacle on viewers' perception, as the experience is greatly different to that of a real-life cinematic viewing. However, such viewings could be helpful in evaluating whether the proposed ideas may work in a controlled experimental environment first before being tested and applied to full scale cinematic viewings, something that may form the basis for further work.

S3D Material Sourcing and Creation

For the purposes of Experiments 1 (HFF), 2 (LA) and 3NS/3Sa/3Sb (LAHFF), four S3D animation clips were created. Each clip was 15 seconds long. The original material used was sourced from *Elephants Dream* (Blender Foundation, Netherlands Media Art Institute 2006), a short science fiction animation that is available under a Creative Commons license (Creative Commons 2013). From this original animation, only the visual material was used. The subject of the animation is the adventures of two characters in an industrial setting where their thoughts seem to affect the behavior and appearance of a gigantic machine that surrounds them.

For the purposes of Experiments 4 (HFF), 5 (LA) and 6NS/6S (LAHFF), a different set of four S3D animation clips were created between 7 and 13 seconds long. The original footage used was sourced from the following animation feature films and S3D demo movies:

- Fly Me to the Moon 3D (nWave Pictures 2007)
- Dracula 4D (Red Star Studio 2010)
- Depth Q S3D Demo Movies (Lightspeed Design 2014)

Permission to use this original work for the purposes of the experiments was obtained by the respective companies and copyright holders (nWave Pictures 2007; Red Star Studio 2010; Lightspeed Design 2014).

From the original footage, only the visual material was used. The clips were selected and edited in such a way as to include a prominent movement of a S3D object within the S3D space and towards the viewer/listener.

For the purposes of Experiment 7 (LAHFF) two S3D animation clips were created. Each clip was 15 seconds long. The original material used was sourced from *Elephants Dream* (Blender Foundation, Netherlands Media Art Institute 2006, Creative Commons 2013). From this original animation, only the visual material was used. Screenshots from the visual clips used can be found in **Appendix 4 - Visual Clip Screenshots**.

It must be noted that the clips used for the experiments were of a short duration (10-15 seconds) and, therefore, they did not resemble a normal cinematic viewing where the spectator attends a large number of scenes in succession for 90 minutes or more. Additionally, only clips containing 3D animation footage were included. This was mainly due to copyright, time and budget constraints. Therefore, the clips could not be used to test whether participants may have a different response viewing S3D scenes with actual cinematic shots of real-life objects and human actors.

Richness of S3D Content Criteria (Experiments 1 (HFF), 2 (LA), 3 (LAHFF))

During the sourcing and editing of the clips special attention was given to the nature and richness of the S3D content. As mentioned in section 3.2.4, richness was determined by a) how many distinct layers of S3D graphics were available in each clip, and b) how complex these S3D graphics were. For instance, a clip with a very dark background where only the protagonist is visible would not be considered to have rich S3D content. Conversely, a clip that takes place within an industrial setting with multiple layers of complex S3D visual objects (e.g. machinery or building structures) would be considered to have rich S3D content. In scenes with rich S3D content the viewer/listener is encouraged to visually scan the image more actively (Mendiburu 2009: 26, 151), thus paying a lot of attention to the visual environment. In rich clips, it was expected that the viewer/listener attention would focus more on the visual cues rather than the auditory ones, as more realistic and numerous visual cues were provided. In simpler S3D clips it was expected that the viewer/listener would pay more attention to auditory cues because the visual cues were simpler.

According to this categorization, the clips were separated into two categories according to the richness of their S3D content. This was indicated in the clip file name with the addition of the letters R for Rich and S for Simple. The criteria for this categorization were the following:

- Clips involving distinct and prominent S3D objects occupying more than half the width of the screen and located in more than two distinct depth layers within the stereoscopic window were classified under the Rich (R) category
- Clips involving S3D objects located in two or less distinct depth layers within the stereoscopic window, extending less than half the width of the screen and which did not display prominent movement on the background layer, were classified under the Simple (S) category
- Clips with a very dark background and little visual detail were classified under the Simple (S) category

Overall, four clips were created for the main test using this method, with two being classified under the S and two under the R category:

- 1_R: Rich S3D content clip
- 2_S: Simple S3D content clip
- 3_R: Rich S3D content clip
- 4_S: Simple S3D content clip

Figure 3-8: Screenshots of Rich (left) and Simple (right) S3D content clip

In addition to the four clips created for the main test, another pair of clips was also created:

- DemoA_R: Rich S3D training and familiarization clip;
- DemoB_S: Simple S3D training and familiarization clip.

The Clip Soundtracks

In each of the experiments a different soundtrack variable was studied. Therefore, the production of the soundtracks that accompanied the S3D clips is covered in detail in the respective sections of each experiment, as they may be different and unique for each set of clips. However, the general structure followed in all experiments is summarized below:

- For each clip an original soundtrack was created. This was the *control* soundtrack.
- For each *control* soundtrack an *experimental* version was created in which the parameter relevant to the experiment was modified (e.g. high frequency were attenuated).
- Each S3D visual clip was merged once with the *control* soundtrack and once with the *experimental* soundtrack, producing two versions of each S3D clip.
- Letters **C** or **E** were added to the names of the resulting clips according to their soundtrack version (Control or Experimental).

This resulted in a number of pairs of clips each consisting of a control and an experimental version.
In Experiments 1 (HFF), 2 (LA) and 3NS/3Sa/3Sb (LAHFF) the names of the clips were updated accordingly in order to reflected their version (C or E) in addition to the richness of content (R or C):

	Pair 1	Pair 2	Pair 3	Pair 4
Control	1_R_C	2_S_C	3_R_C	4_S_C
Experimental	1_R_E	2_S_E	3_R_E	4_S_E

In Experiments 4 (HFF), 5 (LA), 6NS/6S (LAHFF) and 7 (LAHFF) the names of the clips were updated in order to reflect their version only (C or E) as the parameter of richness did not apply:

	Pair 1	Pair 2	Pair 3	Pair 4
Control	1_C	2_C	3_C	4_C
Experimental	1_E	2_E	3_E	4_E

These pairs of clips could be then used for the comparisons required in each of the experiments. For spectrum analysis graphs of the sounds used see **Appendix 2 – Soundtrack Settings**.

3.3.5 Procedure

Order Randomisation

The order in which the different clips were to be presented to the subjects was randomly determined using the online randomisation tool *Research Randomiser* (Urbaniak and Plous 2008).

For all experiments involving non-successive viewings (i.e. Experiments 1 (HFF), 2 (LA), 3NS (LAHFF), 4 (HFF), 5 (LA) and 6NS (LAHFF)), firstly the viewing order of the four clips (clip 1, 2, 3 or 4) was determined and then the background soundtrack version (*control* C or *experimental* E) for each of the clips was also randomly allocated. This was done by randomly assigning either the number 1 or number 2 (with 1=Control, 2=Experimental) to each of the clips, following their predetermined order. The randomisation process for Experiments 1 (HFF), 2 (LA), 3NS (LAHFF), 4 (HFF), 5 (LA) and 6NS (LAHFF) is illustrated in the following example:

Example of the Randomisation Procedure for Experiments 1 (HFF), 2 (LA), 3NS (LAHFF), 4 (HFF), 5 (LA) and 6NS (LAHFF)

- *Randomiser output **2 4 3 1** would correspond to the clip viewing order.*
- *A second Randomiser output was produced, which generated a sequence of random occurrence of numbers 1 and 2. The sequence was set to*

generate a total of numbers identical to the ones in the first sequence (i.e. four numbers, corresponding to the four clips). As an example, ***1 2 2 1*** *would correspond to the following control/experimental background soundtrack allocation* ***C E E C*** *(i.e. 1=Control, 2=Experimental). Therefore, the version of each of the clips (i.e. control or experimental) was randomly determined and allocated by the random allocation of numbers 1 and 2.*

- *Putting all the above together, the clips would be then presented to the participant in the following order:* ***2C 4E 3E 1C.***

After the first viewing a second one followed. The order of the clips was identical but the background soundtrack versions were inverted as described below:

First viewing:

2C 4E 3E 1C

Second viewing:

2E 4C 3C 1E

This procedure ensured that each viewer would watch both the control and the experimental version of each of the four clips, but the two versions of any given pair would not be viewed in succession. This was deemed to be important as successive viewings of clips with identical visual content could introduce practice and/or order effects (Robson 1994: 20). The order in which the trial test clips (DemoA_R, DemoB_S) was played to the participant was also randomised.
As mentioned, in Experiments 3Sa (LAHFF), 3Sb (LAHFF), 6S (LAHFF) and 7 (LAHFF) the two versions of the clips were shown to the participants in succession. This enabled the participant to do a direct comparison between the first and second clip (i.e. the first clip could be used as a reference for the second clip). The order in which the pairs of clips were presented was also randomised between subjects. Similarly, the order in which the two versions of each clip pair (control or experimental soundtrack) were presented was randomised. This was

in order to avoid potential order effects (Harris 2008: 157). The clip order was firstly determined by randomising a set of unique numbers from 1 to 4 (1 to 2 in the case of Experiment 7). Secondly, numbers 1 or 2 were randomly allocated to each of the pairs of clips. This was in order to determine the order of the versions of the clip within each pair. The randomisation process for Experiments 3Sa/3Sb (LAHFF) and 6S (LAHFF) is illustrated in the following example:

Example of the Randomisation Procedure for Experiments 3Sa/3Sb (LAHFF) and 6S (LAHFF)

- *Randomiser output **2 4 3 1** would correspond to the order of the four pairs of clips.*
- *A second Randomiser output was produced, which generated a sequence of random occurrence of numbers 1 and 2. The sequence was set to generate a total of numbers equal to four (i.e. four numbers, corresponding to the four pairs of clips). Numbers 1 and 2 were then changed to letters C and E standing for the Control and Experimental version of soundtrack (i.e. 1=Control, 2=Experimental). As an example, **1 2 2 1** would correspond to the control/experimental background soundtrack allocation **C E E C**. This would determine which version of the clip would be presented to the participant first for each pair.*
- *Putting all the above together, the clips would be then presented to the participant in the following order: **2C-2E, 4E-4C, 3E-3C, 1C-1E.***

As Experiment 7 (LAHFF) involved only two clips instead of four, the procedure was the following:

Example of the Randomisation Procedure for Experiment 7 (LAHFF)

- *Randomiser output **2 1** would correspond to order of the two pairs of clips.*
- *A second Randomiser output was produced, which generated a sequence of random occurrence of numbers 1 and 2. The sequence was set to*

generate a total of numbers equal to two (i.e. two numbers, corresponding to the two pairs of clips). Numbers 1 and 2 were then changed to letters C and E standing for the Control and Experimental version of soundtrack (i.e. 1=Control, 2=Experimental). As an example, ***1 2*** *would correspond to the control/experimental background soundtrack allocation* ***C E****. This would determine which version of the clip would be presented to the participant first for each pair.*

- *Putting all the above together, the clips would be then presented to the participant in the following order:* ***2C-2E, 1E-1C.***

Procedure

Upon arrival to the test location, each participant was asked to read and sign an informed consent form and fill in an electronic personal/demographic information form. The participant was informed that the personal details provided would be kept confidential and that their name and contact details would be deleted from the experimenter's records as soon as the test result analysis was finished. Next, the visual acuity, stereoscopic vision and hearing ability of the participant were checked using relevant eye exam, depth perception and hearing tests (ThinkQuest 2011; Media College 2011; BSHAA 2011).

Once the preparatory procedures were completed, the participant was asked to take a seat inside the experimental space and at a set distance of 300cm from the screen. They had access to a computer mouse located on a desk in front of them. This was in order to be able to provide their ratings on a relevant form appearing on-screen after each clip. Next, a description of the test procedure was given verbally by the experimenter, while a printed test instructions sheet was also handed to the participant (see: **Appendix 3 – Experiment Documents and Forms**).

In Experiments 1 (HFF), 2 (LA), 3NS (LAHFF), 4 (HFF), 5 (LA) and 6NS (LAHFF) the two training clips were played to the participant in pre-determined randomised order. This was both in order to enable the participant to get familiar with the clip nature and S3D content and to establish an indirect visual depth reference against which the participant could evaluate the clips of the main experimental sequence. After viewing each clip the participant was asked to rate the overall depth of the scene on a 1 to 9 Absolute Rating Scale/ACR (i.e. 1=no depth, 9=very prominent depth), using a specially designed electronic form that appeared on-screen (see: **Appendix 3 – Experiment Documents and Forms**). In Experiments 3Sa (LAHFF), 3Sb (LAHFF), 6S (LAHFF) and 7 (LAHFF) no training clips were played, as there was no need for establishing initial reference points (i.e. the clips would be played in a successive order so the depth of the first clip could be used as a reference for that of the second clip).

In Experiments 1 (HFF), 2 (LA), 3NS (LAHFF), 4 (HFF), 5 (LA) and 6NS (LAHFF) once the training test was completed, the main experiment was carried out. The clips were played in the predetermined order and with the control/experimental background soundtrack versions for each also randomly allocated, as described earlier. After the viewing of each clip the participant was asked to rate the depth of the viewed clip in the on-screen electronic answering form. The sequence of both the trial and main tests is demonstrated in the following example:

Example (Experiments 1 (HFF), 2 (LA), 3NS (LAHFF), 4 (HFF), 5 (LA) and 6NS (LAHFF*))*

*(**ans** = Answering Form Screen)*

Trial Test

***Demo1** -> ans -> **Demo2** -> ans*

First viewing

***2C** -> ans -> **4E** -> ans -> **3E** -> ans -> **1C** -> ans*

Second viewing

***2E** -> ans -> **4C** -> ans -> **3C** -> ans -> **1E** -> ans*

In Experiments 3Sa (LAHFF), 3Sb (LAHFF) and 6S (LAHFF), the four pairs of clips were also played in the predetermined order and with the control/experimental background soundtrack versions for each randomly allocated. After the viewing of each clip the participant was asked to force-select one of the two versions (i.e. 1=Control or 2=Experimental) using the on-screen electronic answering form. The sequence of the tests is demonstrated in the following example:

Example (Experiments 3Sa (LAHFF), 3Sb (LAHFF) and 6S (LAHFF))

*(**ans** = Answering Form Screen)*

***2C-2E** -> ans -> **4E-4C** -> ans -> **3E-3C** -> ans -> **1C-1E** -> ans*

Similarly, in Experiment 7 (LAHFF) the two pairs of clips were played in the predetermined order and the participant was asked to force-select one of the two versions (i.e. control or experimental) using the on-screen answering form. The sequence of the tests is demonstrated in the following example:

Example (Experiment 7 (LAHFF))

*(**ans** = Answering Form Screen)*

***2C-2E** -> ans -> **1E-1C** -> ans*

Specific details for each of the experiments are presented in the relevant chapters.

3.4 Conclusion

It is possible that auditory depth cues within the soundtrack may provide an additional tool to S3D filmmakers by supporting and/or enhancing the sense of depth of S3D scenes. The experiments presented in this thesis focus on the impact of on-screen sounds accompanying S3D animation clips and corresponding to the displayed cinematic setting (background sounds) and moving or static on-screen S3D objects.

In classic 2D cinema, the illusion of visual depth is created by means of monocular depth cues (Mendiburu 2009: 26). Accompanying auditory depth cues, such as volume level attenuation, high frequency loss and reverberation, are also commonly used in cinema to support the visuals and enhance the sense of depth of the cinematic scene (Beauchamp 2005: 145-149; Orpen 2003: ix-34). This is based on existing knowledge that the perception of depth and distance in humans is a cross-modal process that heavily relies on both visual and auditory sensory cues (Vroomen and de Gelder 2000; Turner *et al.* 2011; Ecker and Laurie 2004).

In S3D cinema, the introduction of stereoscopic visual cues attempts to further enhance and pronounce the illusion of visual depth of the cinematic world (Mendiburu 2009: 26). It is possible that this could be also reflected in the audio that accompanies these visuals. Although a great amount of research has been conducted on the field of auditory depth perception (Coleman 1963; Ecker and Laurie 2004; Turner *et al*. 2011), the possibility of using extensive and pronounced auditory depth cues with the aim of reflecting and/or enhancing the sense of depth in S3D cinema has not been extensively studied yet. The aim of the experiments described in the present chapter, as well as in Chapters 4, 5 and 6 is to fill this gap and start an exploration towards this direction.

In this chapter, an overview of background concepts relevant to these experiments was provided, followed by a presentation of the methodology employed for the experiments.

4 Experiments 1-3: Alteration of S3D Depth with a Broadband Background Sound

4.1 Introduction

This chapter presents a series of experiments studying the effectiveness of auditory depth cues as a means to increase the perceived sense of depth of S3D animation clips. The auditory cues used in Experiments 1-3 are: high frequency loss (Experiment 1 (HFF)), volume level alteration (Experiment 2 (LA)) and the combination of the two (Experiment 3 (LAHFF)) (Coleman 1963). The idea behind these experiments is that by altering the high frequency content and/or the volume level of a background sound (corresponding to the background of the clip) the perception of depth of the S3D animation clip as a whole could be influenced without altering the S3D visual content. Details on the methodology and the results of each experiment, as well as a brief discussion of the findings, are presented in the following sections.

4.2 Experiment 1 (HFF)

4.2.1 Introduction

The aim of Experiment 1 (HFF) was to test whether the sole use of selective high frequency filtering of broadband background sounds that accompany S3D animation clips can affect the perception of depth of these clips as a whole. This idea was tested on naive viewers (viewers with little or no knowledge of audio processing and of audiovisual perception as a whole), a population assumed to be comparable to the majority of the general cinematic audience.

The hypothesis for this experiment was:

H_1: The perception of depth of a given S3D cinematic scene can be increased by attenuating the high frequency content of an accompanying broadband sound.

Participants were presented with a number of S3D animation clips twice: once accompanied by an unfiltered broadband background sound (control soundtrack) and then accompanied by the same broadband background sound with its high frequencies appropriately filtered (experimental soundtrack, **Figure 4-1**).

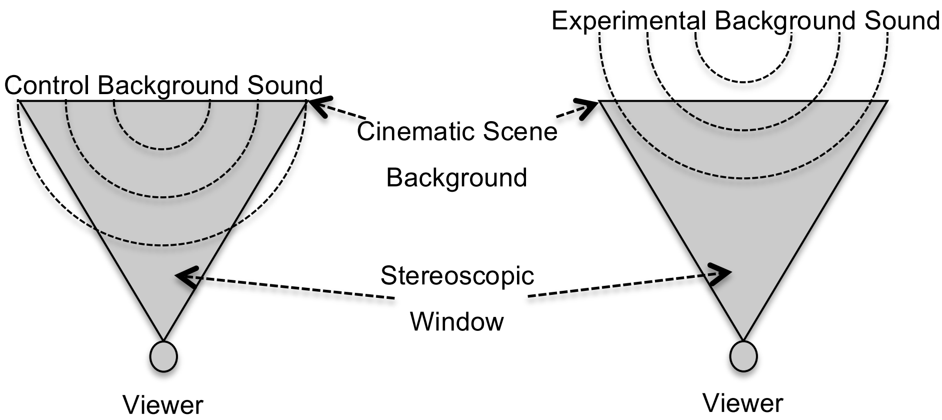


Figure 4-1: The accompanying broadband background sound was displaced from its original position in the experimental version of the clip

The order of presentation was randomised (see: **Chapter 3 - Methodology**). After viewing each clip, participants were asked to rate the perceived level of depth of the S3D clip on a 9-point scale (see: **Chapter 3.3.5 – Procedure**). If the hypothesis proved to be correct, the majority of participants would rate the level of perceived 3D depth of the clips accompanied by the experimental soundtrack higher than that of the clips with the control soundtrack. More specifically, it was expected that the high frequency loss of the experimental (filtered) soundtrack would be attributed to the sonic background being located further away from the background of the clips with the control (unfiltered) soundtrack. If the difference in ratings of depth perception between experimental and control clips was statistically significant in the expected way, it would be possible to be attributed to the only variable changing: the auditory depth cue provided by filtering the high frequencies of the background sound. The background for this hypothesis is based on relevant research and findings of auditory and multi-sensory depth perception in humans (Coleman 1963; Ecker and Laurie 2004; Turner *et al.* 2011).

4.2.2 Method

Design

The experiment employed a repeated measures design (Robson 1994: 100). The independent variable was the frequency content of the background sound that accompanied the S3D animation clips. The variable had two levels: control (the original, unfiltered background sound) and experimental (the same background sound but with its high frequencies appropriately filtered/attenuated). The dependent variable was the perceived sense of depth of the S3D animation clip as a whole from the participants' viewpoint. The test used an Absolute Rating Scale/ACR (ITU 1998) with a scale ranging from 1 to 9 (1 = no depth to 9 = very prominent depth). The scores obtained were analysed using the Wilcoxon Signed-Ranks test (Robson 1994: 118). This is in line with the relevant ITU recommendations and literature (ITU 1998: 5; Robson 1994; Harris 2008).

Participants

For the purposes of this experiment a sample of 35 participants were recruited. The sample consisted of 18 female and 17 male participants from the student and staff population of the University of York, UK. The mean age of the participants was 27.2 years (standard deviation = 9.23) with a maximum age of 59 years and a minimum age of 16 years.

Apparatus and Material

A general description of the apparatus and material used in all the experiments is given in **Chapter 3.3 - Method**. The creation and sourcing of the original S3D visual material used in the experiments is presented in **Chapter 3.3.4 - Apparatus and Material**. The source material used in this experiment was also used in Experiments 2 (LA) and 3NS (LAHFF) so that the final results are comparable.

After the sourcing, editing and categorisation of the S3D clips, new background sounds were created (see: **Chapter 3.3.4 - Apparatus and Material**). These sounds were created in such a way to reflect the apparent mechanical and electrical activity that takes place in the background of the S3D animation clips (**Figure 4-2**, **Appendix 4 - Visual Clip Screenshots**).

Figure 4-2: Mechanical and electrical activity takes place in the background of the clips

For each background sound, two different versions were created: control and experimental (see: **Chapter 3.3.4 - Apparatus and Material**). In the experimental version, the high frequencies of the signal were attenuated (**Figure 4-4**) as to roughly represent the high frequency attenuation caused by a displacement of the sound source (machinery and electrical devices that form the visual background of the S3D animation clips) by approximately 200 meters from its original position (the control soundtrack).

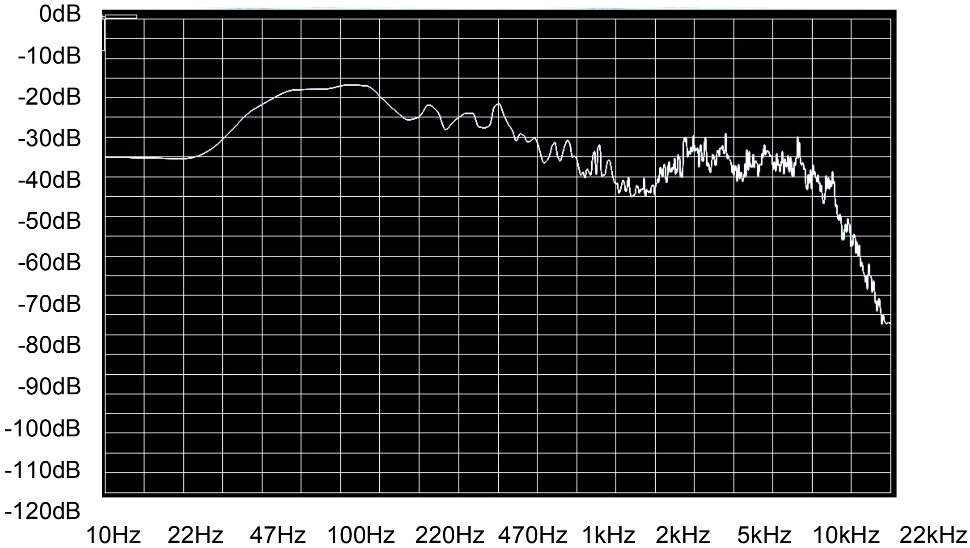


Figure 4-3: Spectrum Analysis for Control sound for (Clips 1_R and 2_S) for Experiments 1 (HFF), 2 (LA) and 3 (LAHFF)

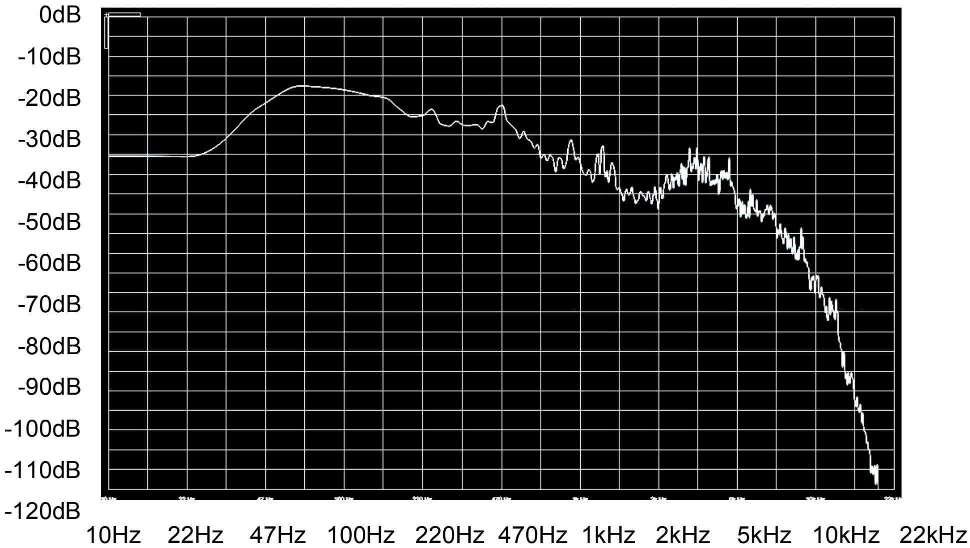


Figure 4-4: Spectrum Analysis for Experimental sound (Clips 1_R and 2_S) for Experiment 1 (HFF - High Frequency Filtering)

This resulted in certain frequencies to drop in volume by amounts ranging from 0.2dB to 12dB (**Table 1, Figures 4-3, 4-4**). For frequency analysis graphs of all the sounds used in the experiment see: **Appendix 2 – Soundtrack Settings**.

The decision to use this particular distance range for the displacement was based on the following observations. Auditory depth cues appear to be underestimated in larger distances (Cochran *et al.* 1968; Mershon and Bower 1979; Kearney *et al.* 2010). Therefore, if a relatively small auditory depth displacement was used (e.g. a displacement of 50-100 meters) it could be more unlikely to register as a significant cue. The choice of the rather exaggerated 200 meters auditory cue displacement was expected to overcome this perceptual shortfall and be strong enough a difference to affect depth perception.

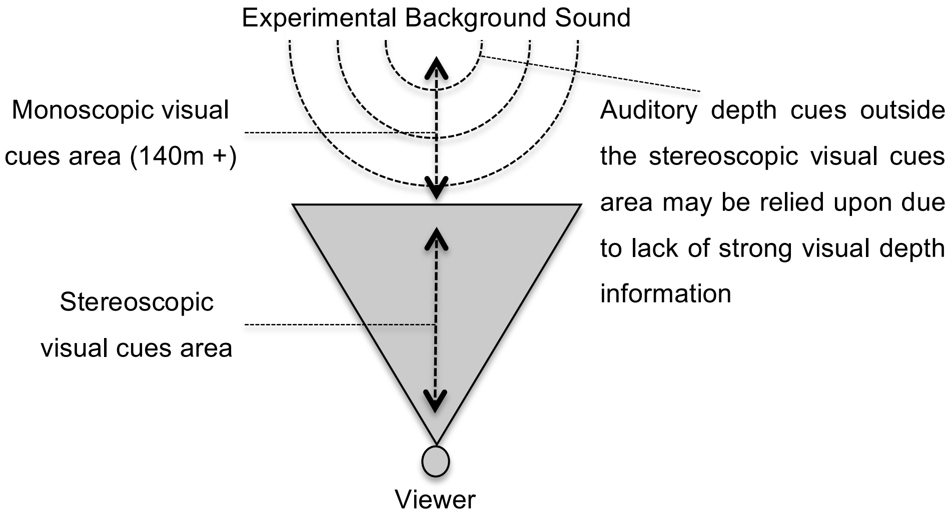


Figure 4-5: S3D visual depth cues are effective for distances between approximately 30 to 140 metres. The lack of S3D cues beyond this point may encourage the participant to search for additional depth cues (e.g. auditory cues)

This decision was also partly related to the suggested distances at which S3D visual cues start becoming ineffective (Mendiburu 2009: 21, 26). More specifically, S3D visual depth cues tend to be used by humans for distances between approximately 30 to 140 metres (Mendiburu 2009: 21), beyond that they become ineffective. Therefore, for scenes containing landscapes and objects with apparent distances that could be estimated by the viewer to be at, or beyond, the upper limit of this distance range or where the S3D depth cues are ambiguous or weak (e.g. clips 2_S_C / 2_S_E and 4_S_C / 4_S_E), it could be possible that strong auditory cues could be taken into consideration for distance calculation (Holt 1997; Bolognini *et al.* 2005: 273; Lippert *et al.* 2007: 102; Mendiburu

2009: 21, 26, **Figure 4-5**). In this case, an apparent auditory displacement that clearly exceeds the S3D visual perception range (30-140 meters) could be possibly strong enough to affect the perceptual decisions.

In order to filter the background sounds in such a way as to reflect the desired high-frequency attenuation, a table (**Table 4-1**) provided by Coleman was used as a general guide (Coleman 1963: 307).

TABLE 1

ABSORPTION COEFFICIENT IN DB/100 FT AT 50% RELATIVE HUMIDITY AND 20° C.

Frequency in cps	a_1	a_2
1000	.2	.2
2000	.3	.3
3000	.5	.6
4000	.7	1.0
5000	1.1	1.5
6000	1.5	2.1
7000	2.0	2.6
8000	2.4	3.3
9000	2.9	4.0
10000	3.5	4.7

Note.—a is absorption coefficient.

Table 4-1: Indicative absorption coefficients in DB per FT (source: Coleman 1963: 307)

The table provides absorption coefficients for different sound frequency bands in a free space environment and under standard environmental conditions (Coleman 1963: 307). Such absorption coefficients reflect the amount of attenuation of different frequency bands as a sound signal travels through air. This frequency specific attenuation is caused because different frequency components are absorbed by the air at different rates and distances and not as a single entity. The absorption coefficients were used to emulate the required high frequency attenuation in the free field were applied to the soundtrack using a digital multiband equalizer (EQ). The displacement of the sound source was calculated in relation to the original (unfiltered) background broadband sound, which functioned as a reference for the creation of the experimental sound. The source distance reference was considered to be the calculated distance of the visual background from the camera (between 50-100 meters). The calculation of this reference distance was based on measurements and conversions within the Blender 3D map of the original footage (Blender Foundation, Netherlands Media Art Institute 2006). In the case of clips with no visible background (clips 2_S_C / 2_S_E and 4_S_C / 4_S_E) the same distance was assumed. Firstly, this was because the clips took place in the same cinematic space as those with visible background (and also the two trial clips – DemoA and DemoB) and, thus, it was expected that the viewer could assume a comparable or similar distance for the invisible scene background. Secondly, this was in order to maintain a consistent point of reference among all the experimental clips (clips with and without visible background).

The filtering was applied to the signal using the built-in digital EQs on a Logic Pro 9 digital audio workstation (Apple Inc. 2014). As mentioned, the filtering values for each EQ band (each frequency band) were calculated in such a way as to simulate the attenuation of frequencies of a sound-emitting source that has been displaced away from its original position by an approximate distance of 200 meters. The detailed EQ settings used can be found in **Appendix 2 – Soundtrack Settings**. Details regarding the procedure followed in this experiment can be found in the relevant section (**Chapter 3.3.5 - Procedure**).

4.2.3 Results

The data were collected using the experimental procedure described in the relevant chapter (**Chapter 3.3.5 - Procedure**) and analysed using the Wilcoxon Signed Ranks test (Sani and Todman 2006; Harris 2008). The sample size was determined using an alpha level of 0.05 (Sani and Todman 2006), with a statistical power of 80% and effect size of 0.50. The sample size was determined using the GPower software (Faul and Lang 2007; Heinrich-Heine-Universität 2013).

The results suggested that there was no statistically significant difference between the perception of depth in S3D clips with the control soundtrack and the experimental soundtrack. This was the case for all four different clips of the experiment, irrespectively of whether they had rich (clips 1_R_C / 1_R_E and

3_R_C / 3_R_E) or simple S3D content (clips 2_S_C / 2_S_E and 4_S_C / 4_S_E; see: **Appendix 1 – Statistical Data**). For details about the definition of simple and rich S3D content see **Chapter 3**.

The two-tailed significance level values obtained were 0.76, 1.00, 0.51 and 0.43 for pairs of clips 1_R, 2_S, 3_R and 4_S respectively. The calculated one-tailed significance level values were 0.38, 0.50, 0.25 and 0.21 for pairs of clips 1_R, 2_S, 3_R, and 4_S respectively.

The values of the cumulative scores of pairs of clips 1_R and 3_R (rich S3D content), as well as the cumulative scores of clips 2_S and 4_S (simple S3D content) were as follows. The two-tailed significance level values obtained were 0.49 and 0.58 for the cumulative scores of clips 1_R / 3_R and 2_S / 4_S respectively. The calculated one-tailed significance level values were 0.24 and 0.29 for the cumulative scores of clips 1_R / 3_R and 2_S / 4_S respectively. The cumulative scores of all four clips returned a two-tailed significance level value of 0.98. The calculated one-tailed significance level value was 0.49. Detailed statistical data are provided in **Appendix 1 – Statistical Data**.

4.2.4 Discussion

Based on the results of the test, the null hypothesis:

> *H_0: The perception of depth of a given S3D cinematic scene cannot be increased by attenuating the high frequency content of an accompanying broadband sound*

cannot be rejected. This result could be due to a number of different reasons.

1) In this experiment, only one auditory cue (i.e. selective filtering of high frequencies) was used to increase the perception of depth. It is well documented that human perception of space is a cross-modal process during which the brain actively seeks confirmation for external events using more than one sensory cue (Holt 1997; Bolognini *et al.* 2005, Lippert *et al.* 2007). If only one sensory cue confirms an event while others either do not or are completely absent, it is

possible that the only cue available will be considered by the brain not to be trustworthy. It is therefore possible that the filtering of high frequencies in a background sound does not have the desired effect because it is not confirmed by other visual and/or auditory cues. This possibility forms the basis of another experiment carried out during this study (Experiment 3NS (LAHFF)) in which the cumulative effect of two auditory depth cues (high frequency filtering plus overall volume attenuation) is explored.

2) It is possible that the effect sought in this experiment is dependent on the spectral and/or cognitive content of the sounds used. In this experiment, control sounds were chosen to have a broadband frequency spectrum. This was done to ensure that when high frequencies were attenuated the result was clearly audible. However different sounds, with varied spectral and cognitive content, could theoretically produce different results. To fully disregard high frequency attenuation as a cue that can, on its own, affect our perception of depth in S3D scenes we should investigate a large number of sounds and clips. Such a proposition goes beyond the scope of this study, but could be considered as a topic for further work.

3) The fact that this experiment focused on the background of S3D animation clips while the main characters and/or the main points of interest of the clips were in the foreground (see: **Appendix 4 - Visual Clip Screenshots**) may have had affected the perceptual effect. It is possible that as the attention of the viewer/listener was naturally taken to the foreground, any differences of distance of the background were not registered.

4) It is possible that the auditory effect was not strong enough to affect the overall sense of depth of the S3D clip More extreme differences between the experimental and control versions of the soundtrack may be strong enough to overcome visual dominance and register as important audiovisual perceptual cues.

4.3 Experiment 2 (LA)

4.3.1 Introduction

The aim of Experiment 2 (LA) was to test whether the sole use of volume attenuation of broadband background sounds that accompany S3D animation clips can affect the perception of depth of the clips as a whole. This idea was tested on naive viewers, similarly to Experiment 1 (HFF).

The hypothesis for this experiment was:

> *H_1: The perception of depth of a given S3D cinematic scene can be increased by attenuating the overall volume level of an accompanying broadband sound.*

Participants were presented with a number of S3D clips once accompanied by a broadband background sound at a given volume (control soundtrack) and then accompanied by the same broadband background sound with its volume appropriately attenuated (experimental soundtrack). The order of presentation was randomised (see: **Chapter 3 - Methodology**).

Similarly to Experiment 1 (HFF), after viewing each clip, participants were asked to rate the perceived level of depth of the S3D clip on a 9-point scale (see: **Chapter 3.3.5 – Procedure**). It was expected that the majority of participants would rate the level of perceived S3D depth of the clips accompanied by the experimental soundtrack higher than that of the clips with the control soundtrack due to the volume attenuation of the experimental soundtrack being attributed to a sound-emitting background further away from the one in the control soundtrack (**Figure 4-1**).

If the obtained results were in the expected direction, it would be possible to attribute this difference to the auditory depth cue provided by attenuating the volume of the background sound, as this would be the only variable changing.

4.3.2 Method

Design

Similarly to Experiments 1 (HFF) and 3 (LAHFF), the experiment employed a repeated measures design (Robson 1994: 100) with the independent variable being the volume level of the background sound that accompanied the S3D visual clips. The variable had two levels: control (the background sound in its original volume level) and experimental (the same background sound but with its volume level appropriately attenuated). The dependent variable was the perceived sense of the overall depth of the S3D clip. The test used an Absolute Rating Scale/ACR (ITU 1998) with a scale ranging from 1 to 9 (1=no depth to 9=very prominent depth). The scores were analysed using the Wilcoxon Signed-Ranks test (ITU 1998: 5; Robson 1994; Harris 2008).

Participants

In Experiment 2 (LA), a sample of 35 participants was recruited. The sample consisted of 21 female and 14 male participants from the student and staff population of the University of York, UK. The mean age of the participants was 24.3 years (standard deviation = 10.42) with a maximum age of 62 years and a minimum age of 18 years.

Apparatus and Material

A general description of the apparatus and material used in all the experiments of this study is given in **Chapter 3.3 - Method**. The source material used in this experiment was also used in Experiments 1 (HFF) and 3NS (LAHFF) so that final results could be comparable.

For each background sound used with the clips, two different versions were created: control and experimental. In the experimental version the volume level of the signal was attenuated (**Figure 4-7**) as to roughly represent the volume attenuation caused by a displacement of the sound source by approximately 200 meters from its original position (the control soundtrack, **Figure 4-6**). For frequency analysis graphs of all the sounds used in the Experiment see: **Appendix 2 – Soundtrack Settings**. Similarly to Experiment 1 (HFF), the distance range was chosen based on the observation that auditory depth cues tend to be underestimated in larger distances (Cochran *et al.* 1968; Mershon and Bower 1979; Kearney *et al.* 2010) and to the suggested distances at which S3D visual cues start becoming ineffective (Mendiburu 2009: 21, 26, **Figure 4-5**).

The volume level of the background sounds was adjusted in such a way as to reflect the volume attenuation caused by a displacement of the sound source by a given distance based on calculations using the inverse square law. The latter states that *for every doubling of the distance from a sound source a 6db volume attenuation occurs* (Coleman 1963: 306). In plain terms, in real-life conditions, the more distant a sound emitting object is located in relation to the viewer/listener the lower its volume appears to be (Ingard *et al.* on Coleman 1963: 306). The sound source distance reference was considered to be the calculated distance of the visual background of the scene from the camera (50-100 meters based on measurements and conversions within the Blender 3D map of the original footage).

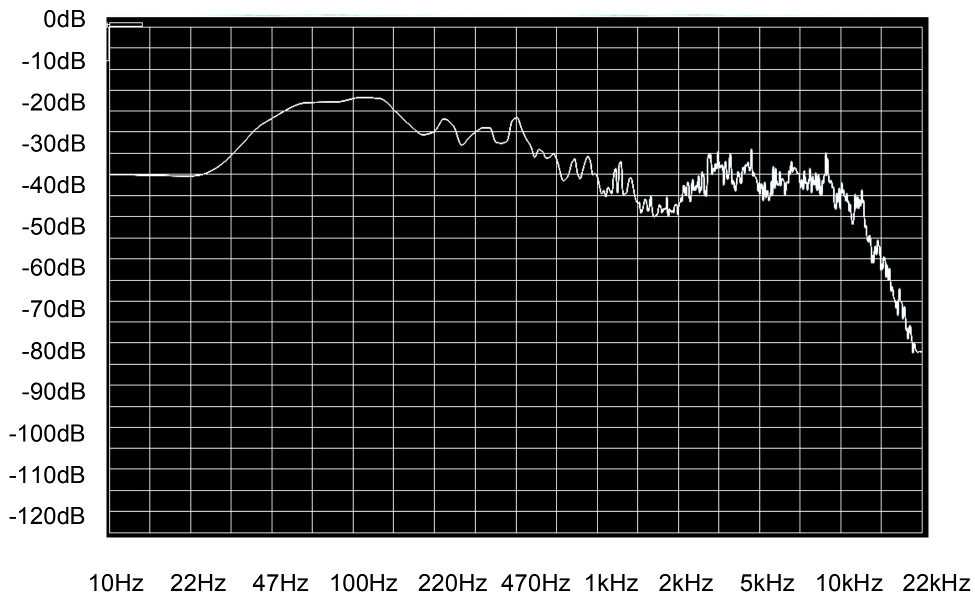


Figure 4-6: Spectrum Analysis for Control sound for (Clips 1_R and 2_S) for Experiments 1 (HFF), 2 (LA) and 3 (LAHFF)

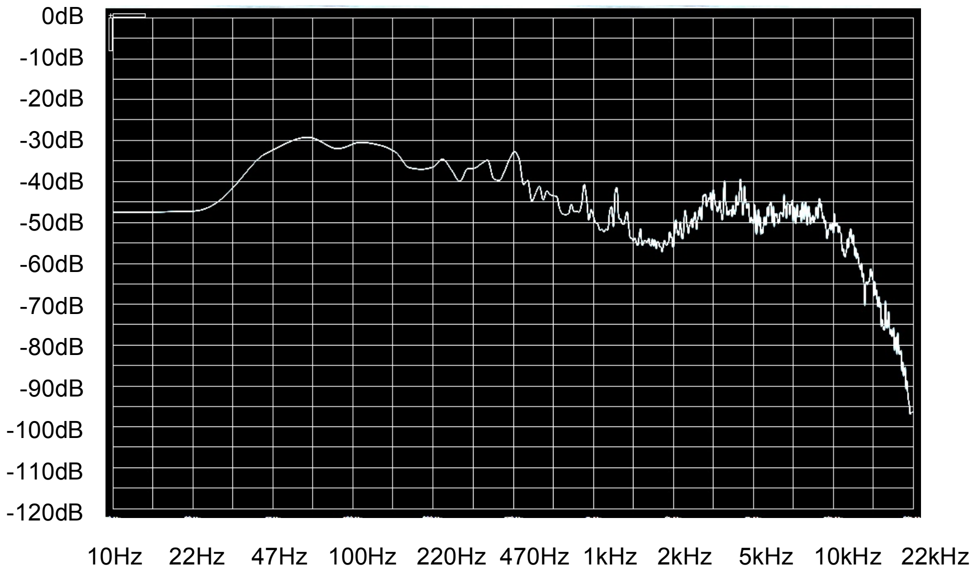


Figure 4-7: Spectrum Analysis for Experimental sound (Clips 1_R and 1_S) for Experiment 2 (LA – Volume Level Alteration)

As a result, a drop in volume of approximately 12dB was expected to roughly correspond to a supposed displacement of the sound source of about 200 meters (max) from the point of reference (**Figures 4-6, 4-7**).

The calculated volume attenuation was applied to the soundtracks using a Logic Pro 9 digital audio workstation (Apple Inc. 2014). More details regarding the settings used for the volume attenuation can be found in **Appendix 2 – Soundtrack Settings**.

4.3.3 Results

The sample size for this experiment was determined using an alpha level set to 0.05 (Sani and Todman 2006), with a statistical power of 80% and effect size of 0.50. The sample size was calculated using the GPower software (Faul and Lang 2007; Heinrich-Heine-Universität 2013).

The results suggested that in two out of four cases (clips 2_S and 4_S) statistically significant difference existed between the perception of depth in S3D clips with the control soundtrack and the clips with the experimental soundtrack (**Figures 4-8, 4-9**). It must be noted that the two cases in which significant results were observed were the ones involving simple S3D content, while no significant results were observed in the cases of clips with rich S3D content (clips 1_R and 3_R). For details about the definition of simple and rich S3D content see **Chapter 3**.

In the case of clip 2_S, the results suggested that participants rated the version with the experimental soundtrack as having a more pronounced S3D depth compared to the control soundtrack version. The one-tailed significance level value for clip 2_S was .04, while the two-tailed one was 0.09. In the case of clip 4_S, participants rated the control version as having a more pronounced S3D depth compared to the experimental version. The one-tailed significance level for this clip was 0.01, while the two-tailed one was 0.03. It must be noted that the direction of this result was against the predicted one and also against the significant result for clip 2_S.

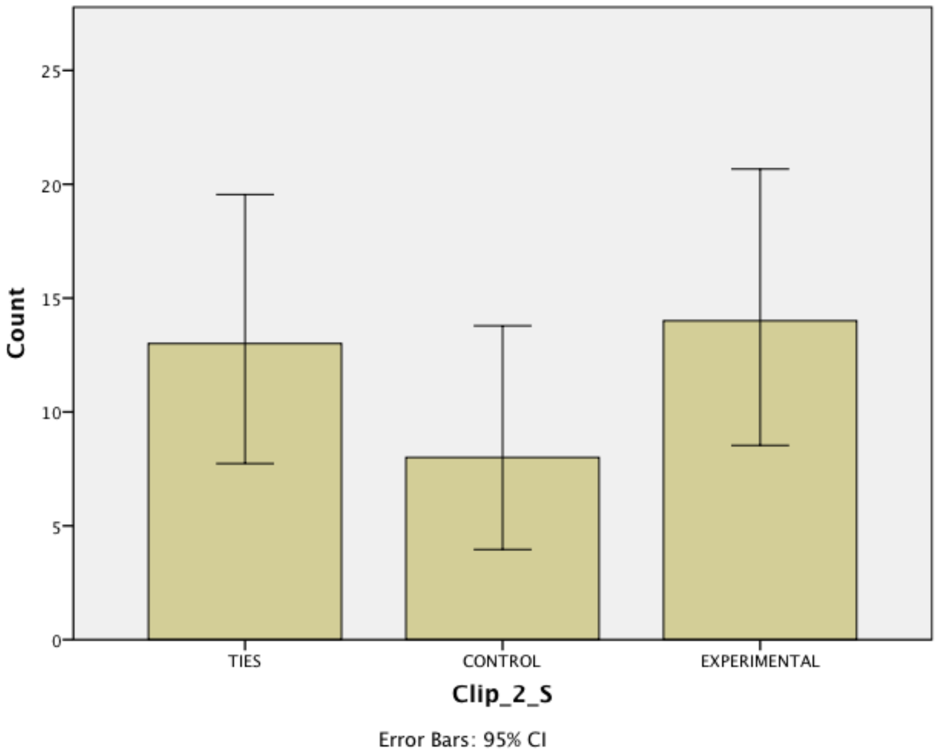


Figure 4-8: Participant response data graph for Clip 2_S

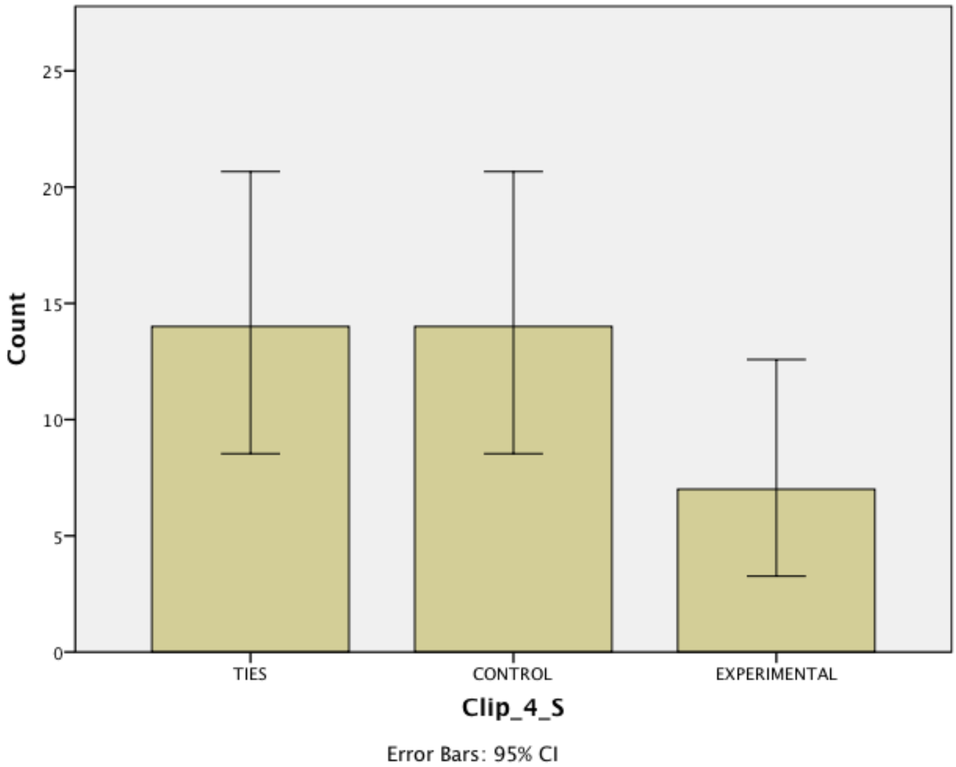


Figure 4-9: Participant response data graph for Clip 4_S

In the cases of clips 1_R and 3_R, no significant results were observed. The one-tailed significance level value for clip 1_R was .12 and the two-tailed one was 0.25. The one-tailed significance level value for clip 3_R was .31 and the two-tailed one was 0.62.

The cumulative scores of clips 1_R and 3_R (rich S3D content), as well as the cumulative scores of clips 2_S and 4_S (simple S3D content) were also analysed. The two-tailed significance level values obtained were 0.64 and 0.74 for the cumulative scores of clips 1_R / 3_R and 2_S / 4_S respectively. The calculated one-tailed significance level values were 0.32 and 0.37 for the cumulative scores of clips 1_R / 3_R and 2_S / 4_S respectively. The cumulative scores of all four clips returned a two-tailed significance level value of 0.58. The calculated one-tailed significance level value was 0.29. Detailed statistical data are provided in **Appendix 1 – Statistical Data**.

4.3.4 Discussion

Comparing the results for the four clips (1_R, 2_S, 3_R, 4_S) a useful observation can be made. In the case of clip 2_S, participants rated the experimental clip as having a more pronounced overall S3D depth (**Figure 4-8**), while in the case of clip 4_S participants rated the control clip as having the more pronounced overall S3D depth (**Figure 4-9**). The fact that the direction of the results of the two clips is contradictory does not allow for clear conclusions and further experimentation is needed in order to examine the reasons for this. However, it is interesting that these two significant results correspond to the two clips that had a *simple* S3D content, while no significant effect was observed in the two clips that had a *rich* S3D content (see: **Appendix 4 - Visual Clip Screenshots**). The significant results could be viewed as an indication that the absence of strong stereoscopic visual background influenced the subjects' ability to judge the depth of scene. If this is proved to be the case, it may offer S3D filmmakers the indication that they could use auditory depth cues in scenes where the S3D visual background is either weak or absent (e.g. dark or very dim scenes) in order to enhance the overall sense of depth of the scene. It is also possible that this may also apply to parts of a S3D movie where the depth script dictates that strong S3D cues are undesirable. However, the current results are not conclusive and further experimentation in this direction may be necessary in order to clarify the effectiveness of such approaches.

The contradictory direction of the significant values in clips 2_S and 4_S could be due to the following reasons.

1) It is possible that although volume attenuation may be an auditory cue that can affect the participants' sense of S3D depth in clips with simple S3D content this is not done in a simple and easily predictable manner. In this case, auditory cues may be taken into consideration in the absence of strong S3D depth cues but the brain may employ different strategies to calculate the depth. These could be based on real life expectations, but also previous experience, cinematic expectations, cultural expectations, etc. Thus, the difference between the control and experimental soundtracks is registered but does not trigger a consistent reaction towards the expected direction (in one case the control clip was

perceived as having a greater S3D depth although it had a higher volume level dictating a closer distance of the sound emitting object: namely the background of the scene).

2) It is possible that either or both of the significant values observed in pairs 2_S and 4_S could be the result of a Type I statistical error.

The insignificant results for clips 1_R and 3_R could be due to a number of different reasons.

3) In relation to the lack of significant results in the cases of clips 1_R and 3_R, in this experiment only one auditory cue (volume attenuation) was used to increase the perception of depth. As human perception of space is a multimodal process (Holt 1997; Bolognini *et al.* 2005, Lippert *et al.* 2007) if only one sensory cue confirms an event while others do not, it is possible that the available cue will not be considered trustworthy. This possibility is investigated in Experiment 3NS (LAHFF) in which the cumulative effect of two auditory depth cues (i.e. high frequency filtering plus overall volume attenuation) is explored.

4) Similarly to Experiment 1 (HFF), it is possible that the effect sought in this experiment is dependent on the cognitive content of the sound. A large number of sounds and clips should be studied in order to investigate this possibility. This proposition goes beyond the scope of this research, but should be considered for further work.

5) The fact that the main characters and/or points of interest of the clips were in the foreground (see: **Appendix 4 - Visual Clip Screenshots**) while experiment focused on the background of S3D clips may have had affected the perceptual effect.

6) Similarly to Experiment 1 (HFF), it is possible that the auditory effect was not strong enough to affect the overall sense of depth of the S3D clip. In order to investigate this claim more extreme differences between the experimental and control versions of the soundtrack may be used in related future work.

Based on the overall results of the test, the null hypothesis:

> *H_0: The perception of depth of a given S3D cinematic scene cannot be increased by attenuating the overall volume level of an accompanying broadband sound*

cannot be rejected.

4.4 Experiments 3NS / 3Sa / 3Sb (LAHFF)

4.4.1 Introduction

The aim of Experiments 3NS / 3Sa / 3Sb (LAHFF) was to test whether the combination of the two variables tested in Experiments 1 (HFF) and 2 (LA) (selective high frequency filtering and volume alteration of broadband background sounds) can affect the perception of depth of the S3D animation clips as a whole. The hypothesis for this experiment was:

> *H_1: The perception of depth of a given S3D cinematic scene can be increased by attenuating both the high frequency content and the overall volume level of an accompanying broadband sound.*

As in the other two experiments, participants were presented with a number of S3D clips twice: once accompanied by an unfiltered broadband background sound of a given volume level (control soundtrack) and once accompanied by the same broadband background sound with its high frequencies appropriately filtered and its overall volume level decreased accordingly (experimental soundtrack).

In Experiment 3NS (LAHFF), the order of presentation was randomised and, after viewing each clip, participants were asked to rate the perceived level of depth of the S3D clip on a 9-point scale (see: **Chapter 3.3.5 – Procedure**).

In Experiment 3Sa (LAHFF) each pair of clips was presented to the participant in succession, but the order of the clips and the order of the presentation of the different versions within each pair of clips were randomised (see: **Chapter 3 - Methodology**). As participants viewed the two versions of each clip in succession they had to force-select the clip that appeared to have the most profound S3D depth effect among the two.

Experiment 3Sb (LAHFF) was a follow up to Experiment 3Sa (LAHFF), with the same clips being presented in succession but carried out in a different experimental environment and with different participants. The aim of this experiment was to verify the findings of Experiment 3Sa (LAHFF).

In all three experiments, it was predicted that the majority of participants could find the clips accompanied by the experimental soundtrack to have a higher level of perceived 3D depth than those with the control soundtrack. The reason for this was that the combined effect of high frequency loss and volume level attenuation of the experimental (filtered / attenuated volume level) soundtrack would be potentially attributed to the sonic background being located further away from the background of the clips with the control (unfiltered / original volume level) soundtrack (**Figure 4-1**). Thus, any statistically significant results in the predicted direction could be attributed to the combined auditory depth effect of filtering the high frequencies and attenuating the overall volume level of the background sound.

4.4.2 Method

Design

Experiment 3NS (LAHFF) employed a repeated measures design (Robson 1994: 100). The independent variable was the combination of the frequency content and the overall volume of the background sound that accompanied the S3D clips. The variable had a control level (the original, unfiltered background sound with original volume level) and an experimental level (the same background sound but with its high frequencies appropriately filtered and the overall volume level attenuated). The dependent variable was the perceived sense of depth of the S3D clip as a whole from the participants' viewpoint.

Experiment 3NS (LAHFF) used an Absolute Rating Scale/ACR (ITU 1998) with a scale ranging from 1 to 9 (1 = no depth to 9 = very prominent depth). The scores obtained during the viewings were analysed using the Wilcoxon Signed-Ranks test (ITU 1998: 5; Robson 1994; Harris 2008).

Experiment 3Sa (LAHFF) used the same clips and material as Experiment 3NS (LAHFF) but the experimental design was different. In this case, the control and experimental clips were presented to the participant in succession. The order in which the two versions were presented was randomised prior to the experiment

(see: **Chapter 3.3.5 - Procedure**). This was in order to reduce or eliminate potential order effects. This was also the case in Experiment 3Sb (LAHFF).
This method allowed verifying whether successive viewings of the different versions of the same clip (control and experimental versions) would produce different results to Experiment 3NS (LAHFF), in which the clips were presented in a non-successive order. After each pair of clips was viewed participants were asked to select which of the two versions of the clip appeared to have a more pronounced overall S3D depth. Participants had to force-select one of the two versions. The data collected using this method were analysed using the Binomial Test analysis (McCluskey and Lalkhen 2007).

Participants

The participants of all three experiments were recruited from the student and staff population of the University of York.
For Experiment 3NS (LAHFF) a sample of 35 participants were recruited. The sample consisted of 18 female and 17 male participants. The mean age of the participants was 24.5 years (standard deviation = 6.07) with a maximum age of 39 years and a minimum age of 18 years.
For Experiment 3Sa (LAHFF) the sample was 35 participants. The sample consisted of 21 female and 14 male participants, while the mean age of the participants was 26.2 years (standard deviation = 7.23) with a maximum age of 44 years and a minimum age of 18 years.
For Experiment 3Sb (LAHFF) the sample was 30 participants with 15 being female and 15 male. The mean age of the participants was 23.6 years (standard deviation = 4.90) with a maximum age of 38 years and a minimum age of 18 years.

Apparatus and Material

A general description of the apparatus and material used in all the experiments of this study is given in **Chapter 3.3 - Method**.
For each background sound a control and an experimental version were created. In the experimental version the high frequencies and the overall volume level of

the signal were attenuated (**Figures 4-10, 4-11**) in order to reflect an attenuation caused by a displacement of the sound source (the S3D clip background) by approximately 200 meters.

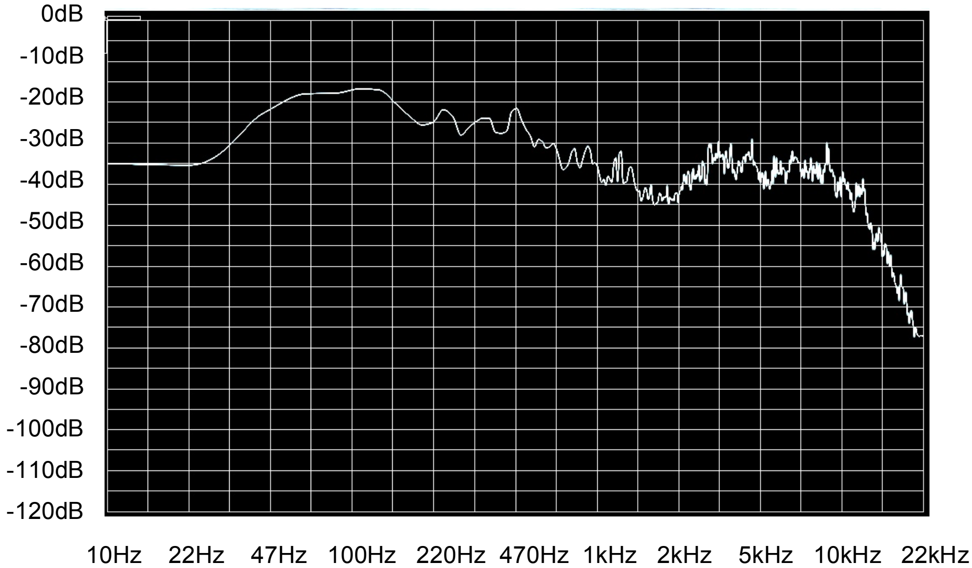


Figure 4-10: Spectrum Analysis for Control sound (Clips 1_R and 2_S) for Experiments 1 (HFF), 2 (LA) and 3 (LAHFF)

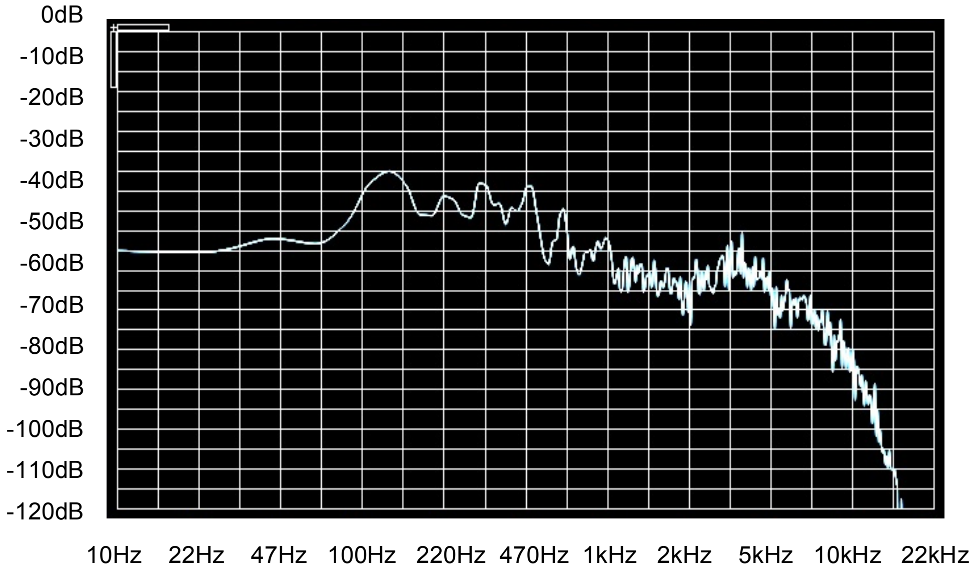


Figure 4-11: Spectrum Analysis for Experimental sound (Clips 1_R and 2_S) for Experiment 3 (LAHFF – Volume Level Alteration and High Frequency Filtering)

For frequency analysis graphs of all the sounds used in the Experiment see: **Appendix 2 – Soundtrack Settings**. The decision regarding the selected distance range was made for the same reasons as discussed in Experiments 1 (HFF) and 2 (LA). High frequency filtering and volume attenuation was achieved using the same techniques as in Experiment 1 (HFF) and Experiment 2 (LA), and used the same absorption coefficient table provided by Coleman (1963) (**Table 1, Appendix 2 - Soundtrack Settings**).

Similarly to the previous two experiments, the apparent position of the sound source (background sound) in the version of the clip with the control soundtrack was used as a reference for the calculation of the displacement. Detailed soundtrack settings can be found in **Appendix 2 – Soundtrack Settings**. Details regarding the procedure followed in this experiment can be found in the relevant section (**Chapter 3.3.5 - Procedure**).

4.4.3 Results

For all three experiments (Experiments 3NS, 3Sa and 3Sb) the alpha level was set to 0.05 (Sani and Todman 2006), with a statistical power of 80% and effect size of 0.50

In Experiment 3NS (LAHFF) the data were analysed using the Wilcoxon Signed Ranks test (Sani and Todman 2006; Harris 2008). The results suggested that there was no statistically significant difference between the perception of depth in S3D clips with the control soundtrack and the clips with the experimental soundtrack. This was the case for all four clips of the experiment, irrespectively of whether they had rich (1_R and 3_R) or simple (2_S and 4_S) S3D content. For details about the definition of simple and rich S3D content see **Chapter 3.3.4 - Apparatus and Material**.

The two-tailed significance level values obtained were 0.67, 0.54, 0.48 and 0.43 for pairs of clips 1_R, 2_S, 3_R and 4_S respectively. The calculated one-tailed significance level values were 0.33, 0.27, 0.24 and 0.21 for clips 1_R, 2_S, 3_R and 4_S respectively.

The values of the cumulative scores of clips 1_R and 3_R (rich S3D content), as well as the cumulative scores of clips 2_S and 4_S (simple S3D content) were as follows. The two-tailed significance level values obtained were 0.89 and 0.86 for the cumulative scores of clips 1_R / 3_R and 2_S / 4_S respectively. The calculated one-tailed significance level values were 0.44 and 0.43 for the cumulative scores of clips 1_R / 3_R and 2_S / 4_S respectively. The cumulative scores of all four clips returned a two-tailed significance level value of 0.81. The calculated one-tailed significance level value was 0.40. Detailed statistical data are provided in **Appendix 1 – Statistical Data**.

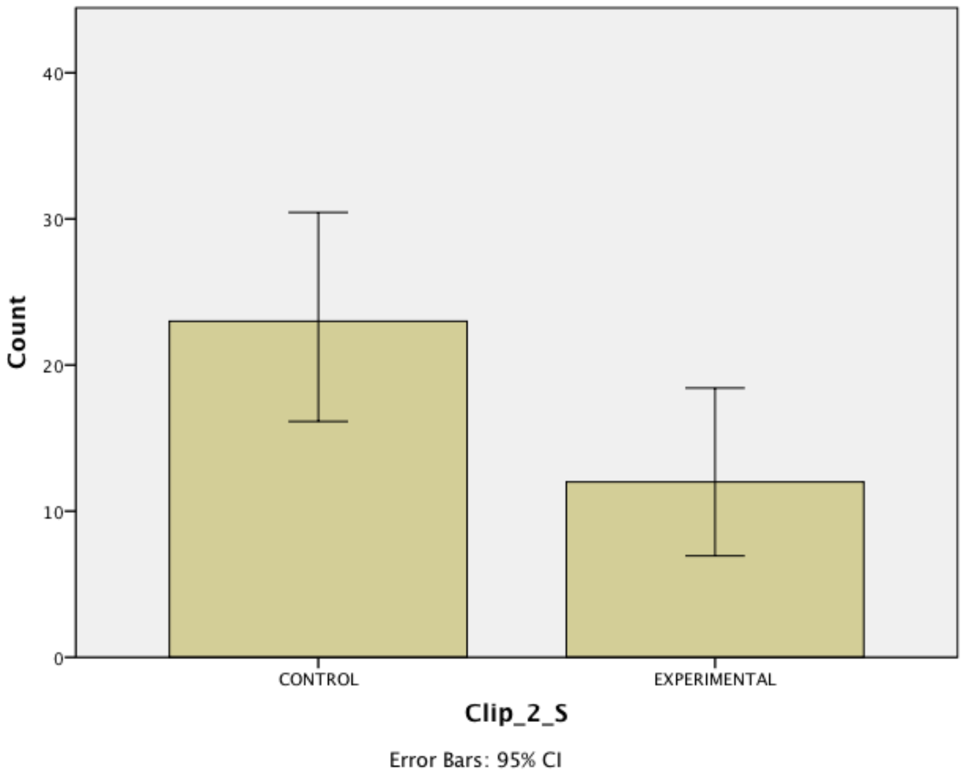


Figure 4-12: Participant response data graph for Clip 2_S (Experiment 3Sa (LAHFF))

In Experiment 3Sa (LAHFF) the data were analysed using the Binomial Test analysis (McCluskey and Lalkhen 2007). The results showed that statistically significant difference between the perception of depth in S3D clips with the control soundtrack and the clips with the experimental soundtrack existed in the cases of clips with *simple* S3D content (i.e. clips 2_S and 4_S, **Figures 4-12, 4-13**, **Appendix 4 - Visual Clip Screenshots**). The difference observed in this experiment was in the opposite direction to the one predicted (control clips were rated as having a more pronounced S3D depth than experimental ones).

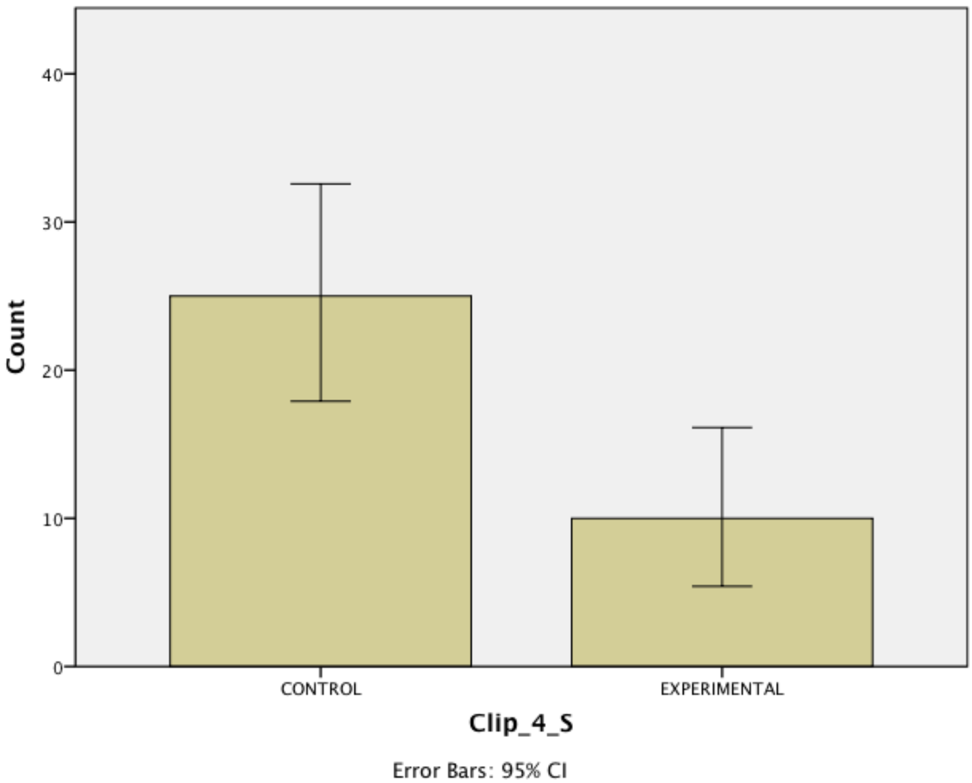


Figure 4-13: Participant response data graph for Clip 4_S (Experiment 3Sa (LAHFF))

In the two cases involving pairs with rich S3D clips (1_R and 3_R) no statistically significant differences between the perceived S3D depth of control and experimental clips were observed. For details about the definition of simple and rich S3D content see **Chapter 3.3.4 - Apparatus and Material**.

It must be stressed that the cumulative scores for the clips with simple S3D content, as well as the cumulative scores for all four clips also returned statistically significant results (**Figures 4-14, 4-15**). The two-tailed significance level values obtained for the four clips were 0.50, 0.09, 1.00 and 0.01 for clips 1_R, 2_S, 3_R and 4_S respectively. The calculated one-tailed significance level values were 0.25, 0.04, 0.50 and 0.005 for clips 1_R, 2_S, 3_R and 4_S respectively.

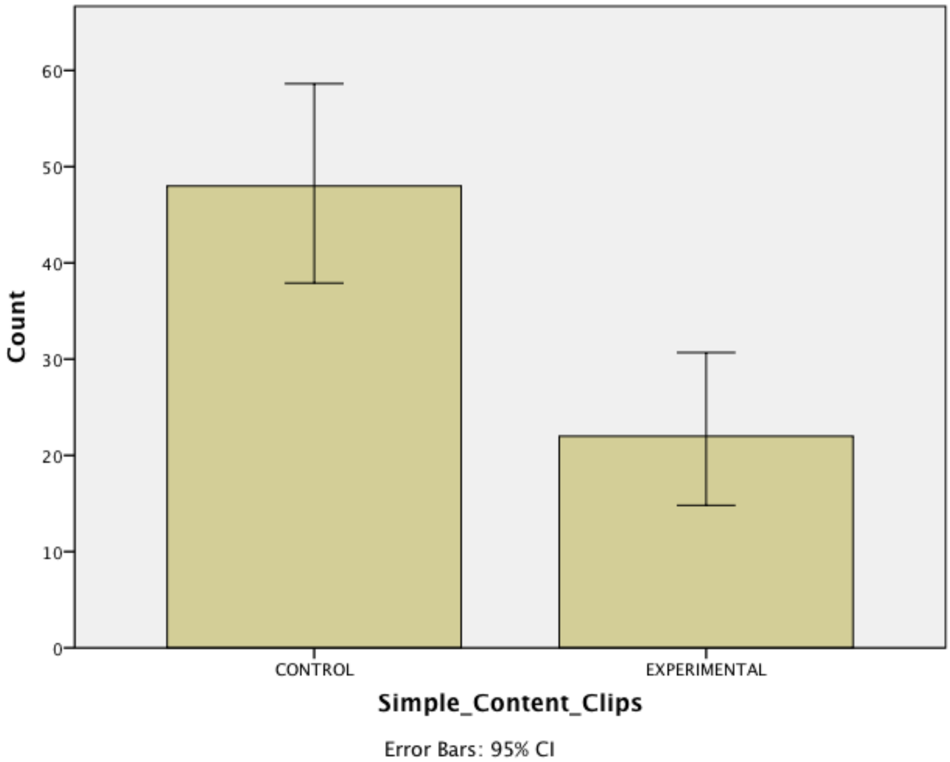


Figure 4-14: Participant response data graph for simple S3D content clips (Experiment 3Sa (LAHFF))

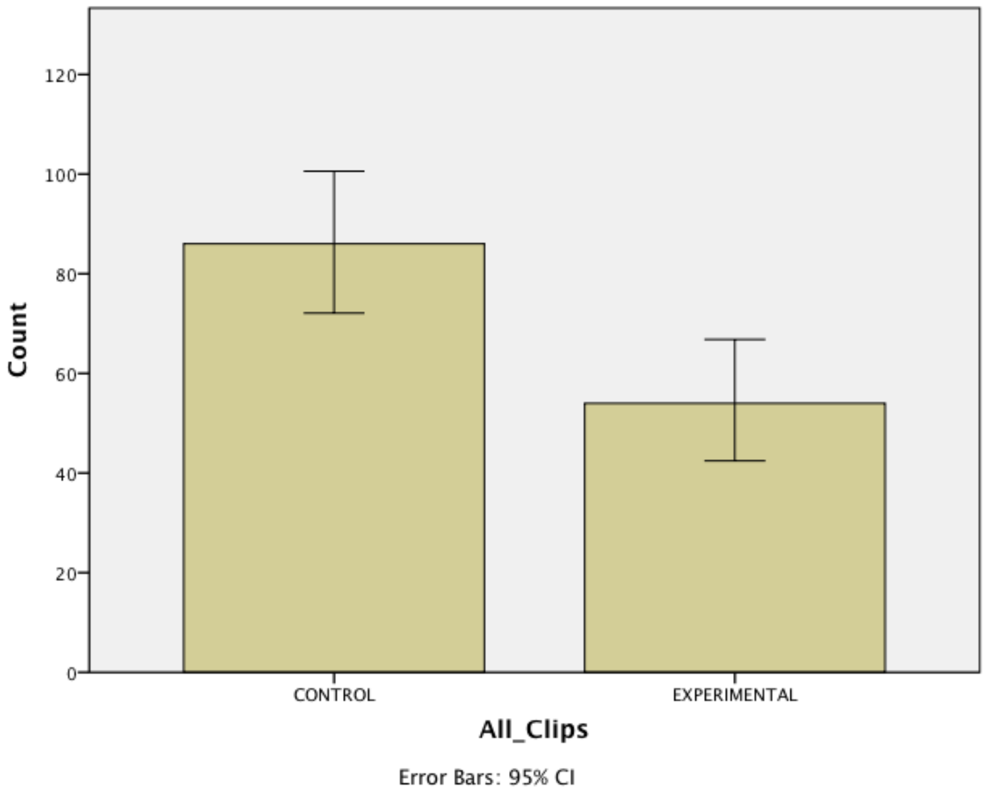


Figure 4-15: Participant response data graph for all clips (Experiment 3Sa (LAHFF))

The values for the cumulative scores of clips 1_R and 3_R (rich S3D content), as well as the cumulative scores of clips 2_S and 4_S (simple S3D content) were as follows. The two-tailed significance level values obtained were 0.55 and 0.003 for the cumulative scores of pairs 1_R / 3_R and 2_S / 4_S respectively. The calculated one-tailed significance level values were 0.27 and 0.001 for the cumulative scores of clips 1_R / 3_R and 2_S / 4_S respectively. The cumulative scores of all four clips returned a two-tailed significance level value of 0.009. The calculated one-tailed significance level value was 0.005. Detailed statistical data are provided in **Appendix 1 – Statistical Data**.

For Experiment 3Sb, (LAHFF) the data were analysed using the Binomial Test analysis (McCluskey and Lalkhen 2007). The results showed no statistically significant difference between the perception of depth in S3D clips with the control soundtrack and the clips with the experimental soundtrack for each of the four clips in isolation. However, similarly to Experiment 3Sa (LAHFF), statistically significant results were observed when the cumulative scores for clips with simple S3D content were analysed (**Figure 4-16**).

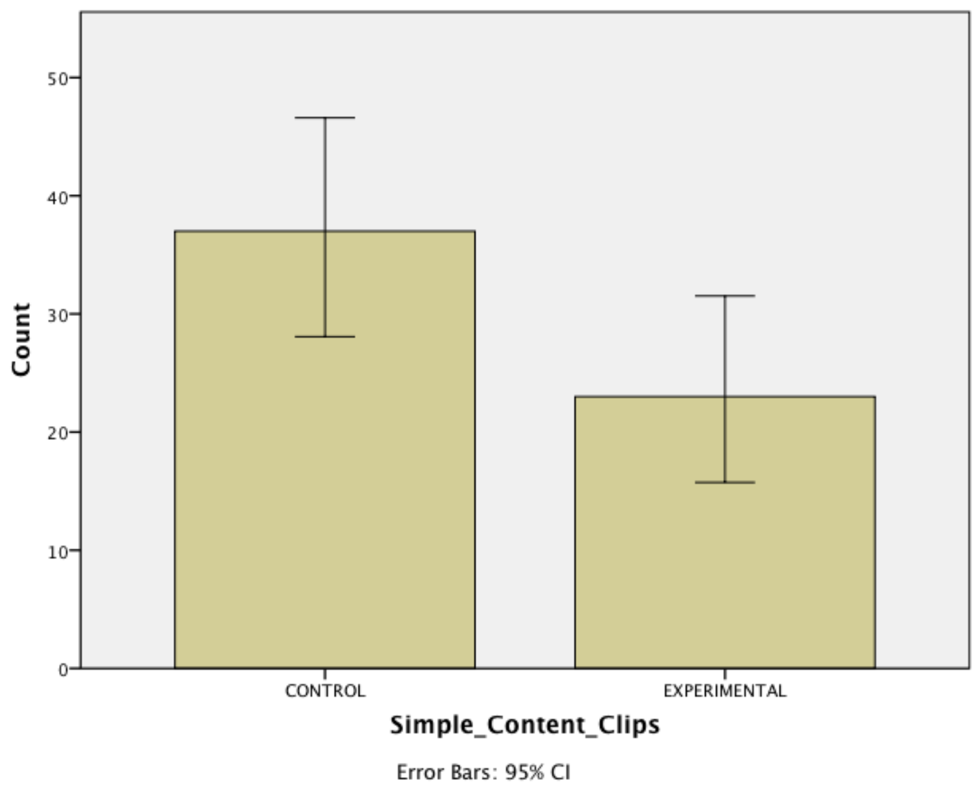


Figure 4-16: Participant response data graph for simple S3D content clips (Experiment 3Sb (LAHFF))

This was also the case for the analysis of the cumulative scores of all four clips (**Figure 4-17**). More specifically, the two-tailed significance level values obtained for the four clips were 0.58, 0.36, 0.36 and 0.20 for clips 1_R, 2_S, 3_R and 4_S respectively.

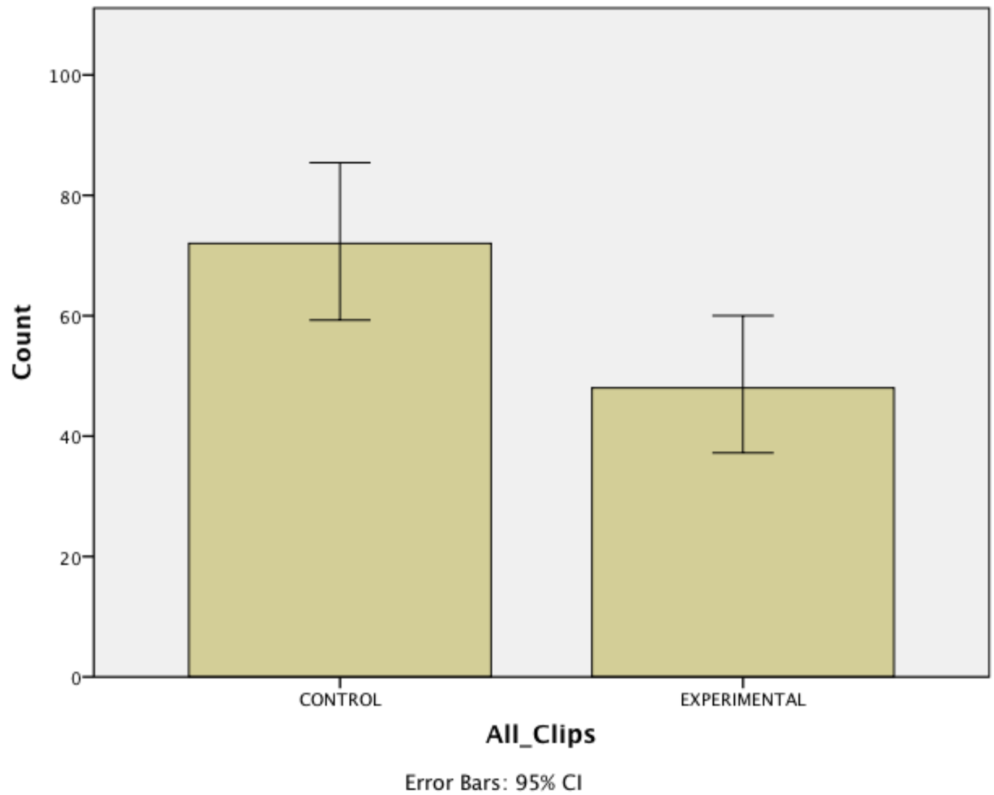


Figure 4-17: Participant response data graph for simple S3D content clips (Experiment 3Sb (LAHFF))

The calculated one-tailed significance level values were 0.29, 0.18, 0.18 and 0.10 for clips 1_R, 2_S, 3_R and 4_S respectively.

The values of the cumulative scores of clips 1_R and 3_R (rich S3D content), as well as the cumulative scores of clips 2_S and 4_S (simple S3D content) were as follows. The two-tailed significance level values obtained were 0.24 and 0.09 for the cumulative scores of clips 1_R / 3_R and 2_S / 4_S respectively. The calculated one-tailed significance level values were 0.12 and 0.04 for the cumulative scores of clips 1_R / 3_R and 2_S / 4_S respectively. The cumulative scores of all four clips returned a two-tailed significance level value of 0.03. The calculated one-tailed significance level value was 0.01. Detailed statistical data are provided in **Appendix 1 – Statistical Data**.

4.4.4 Discussion

Statistically significant results were observed in Experiments 3Sa (LAHFF) and 3Sb (LAHFF). In the case of Experiment 3Sa (LAHFF), these results correspond to the two clips that had *simple* S3D content, while no significant effect was observed in the two clips that had *rich* S3D content. This observation is in line with results from Experiment 2 (LA), where statistically significant results were also observed only for the two clips involving simple S3D content (clips 2_S and 4_S). This is also confirmed by the fact that the cumulative scores obtained for the clips with simple S3D content (clips 1_R and 3_R) in Experiment 3Sb (LAHFF) also returned statistically significant results. This is an interesting observation as it indicates that the effectiveness of auditory depth cues in S3D cinematic scenes may be related with the complexity of the S3D content. As all cases where statistically significant results were observed (Experiments 2 (LA), 3Sa/3Sb (LAHFF)) involved simple S3D content, it is possible that the absence of strong S3D visual cues may have caused the participants to search for perceptual depth cues in other sensory streams (e.g. available auditory cues). If this is confirmed, it may be an interesting piece of information for S3D filmmakers as they could exploit sound for increasing the perception of depth in cases where rich S3D visual information is either inconvenient to use or undesirable. An example of this is the use of strong auditory depth cues in parts of an S3D movie where the depth script dictates that S3D visual depth must be kept at a minimum in order to avoid visual stress and discomfort (Clark 2010; Mendiburu 2009; Autodesk 2008).

The fact that in Experiments 3Sa / 3Sb (LAHFF) the direction of the results was opposite to the one predicted may be an indication that in the cinematic context the auditory depth cues are not interpreted as they would in real life. In real life one would be inclined to perceive volume attenuation and high frequency loss as an indication that the sound emitting source (in this case the background of the clip/surroundings) is located further away. Thus, if this real life perceptual process was followed in cinema one could argue that a S3D clip involving sound emitting objects (the background objects or environment of the clip) that emit signals with attenuated high frequency and overall volume level signals could be perceived as having a greater depth. This would be due to the background of the clip being perceptually placed further away from the viewer/listener. However, the direction of the results indicates that louder sounds with a broader frequency range could cause the clips to appear as having a more pronounced S3D depth. It is suggested that the expectations and frame of mind of the viewer/listener when watching a cinematic presentation may be different to the ones employed in real life. It is possible, then, that perceptual cues work in different and more complex ways in the cinematic context. Further work on this may be necessary in order to investigate whether this observation is correct and how such knowledge can be used creatively for cinematic purposes.

The fact that no statistically significant results were observed for all four clips in Experiment 3NS (LAHFF), for the two clips with rich S3D content (clips 1_R and 3_R) in Experiment 3Sa (LAHFF) and all of the clips in Experiment 3Sb (LAHFF) when studied in isolation does not allow for firm conclusions and further future work towards this direction is necessary.

The results could be due to a number of different reasons.

1) It is possible that the combination of two different auditory cues in this experiment caused an unidentified side effect that compromised the effectiveness of the use of these auditory cues. This assumption is based on the fact that statistically significant results were observed in two cases in Experiment 2 (LA) where only one of the two auditory cues was used (i.e. volume attenuation), but no such results were observed in Experiment 3NS (LAHFF) where high frequency filtering was added. The addition of a second auditory cue was expected to magnify the perceptual effect, as it would confirm and solidify the perception of the apparent displacement of the sound source dictated by the volume level attenuation. The fact that this addition appears to have negated the effect of volume attenuation when used in isolation (i.e. Experiment 2 (LA)) may be an indication that high frequency filtering may have an unexpected perceptual effect in the cinematic context. More research towards this direction is needed in order to examine the validity of this claim.

2) Although in this experiment two auditory cues (i.e. selective filtering of high frequencies and overall volume attenuation) were used to increase the perception of S3D depth, the combination of cues was probably not strong enough to

overcome visual dominance (Shams and Kim 2010), similarly to Experiments 1 (HFF) and 2 (LA). It is possible that the available S3D visual information dominates the perception of depth of the viewer/listener.

3) The effect sought in this experiment may be dependent on the spectral and cognitive content of the sound. A large number of sounds and cinematic scenes should be studied in order to determine whether the audio content of the scene is affecting the effectiveness of the studied auditory depth cues (volume and high frequency attenuation).

4) Similarly to Experiments 1 (HFF) and 2 (LA), while the main characters were in the foreground (see: **Appendix 4 – Visual Clip Screenshots**) the experiment focused on the background sound. It is possible that the attention of the viewer/listener was taken to the foreground and any differences of distance of the background were not registered.

5) It is possible that the auditory effect was not strong enough to affect the overall sense of depth of the clip, as in Experiments 1 (HFF) and 2 (LA).
Based on the results of the test, the null hypothesis:

> *H_0: The perception of depth of a given S3D cinematic scene cannot be increased by attenuating both the high frequency content and the overall volume level of an accompanying broadband sound*

cannot be rejected.

4.5 General Discussion

With the exception of two clips in Experiment 2 (LA), in Experiments 1 (HFF), 2 (LA) and 3NS (LAHFF), when the subjects did not view two versions of a given clip in succession, altering the high frequency content and the volume of the background sound did not have an overall statistically significant effect on the perception of depth, in the presence of S3D cinematic visual content. There are two main points of interest directly related to this result. Firstly, these experiments seem to confirm findings of previous studies that, from the perspective of the viewer/listener, visual cues tend to dominate over auditory ones when both are available and perceptually assessed for the confirmation of a given audiovisual event (Shams et al. 2010, Wozczyk et al. 1995, Mastoropoulou 2006). The apparent ineffectiveness of the depth auditory cues could be explained by the fact that the visual cues present were trusted as the most reliable perceptual information between the two, potentially contrasting, types of sensory cues (visual and auditory) available in the clips. Secondly, the results of Experiments 1 (HFF), 2 (LA) and 3NS (LAHFF) could mean that the presence of an explicit auditory reference point may be crucial to make the auditory cues perceptually important. In these experiments the subjects were intentionally not given a direct reference against which they could judge the relative loss of volume and high frequency content of a given sound. This was done to maintain the reality of the cinematic experience, in which a scene is not repeated twice in succession.

Details on Experiment 1 (HFF)

In Experiment 1 (HFF), the effect of high frequency loss on the overall perception of depth in S3D clips was studied. One possible reason why the results are not significant is that the high frequency loss on its own, without level attenuation, is too weak a cue to be considered by the listener/viewer in the presence of S3D visuals. Indeed, high frequency attenuation is amongst the most changeable of the auditory cues the brain uses for depth/distance estimation (Coleman 1963) and it is possible that the brain only relies on it when in conjunction with other cues that confirm the estimated depth, such as level attenuation. Additionally, a determining factor for the effectiveness of this auditory cue alone is the spectral content of the sounds used. For instance, a background sound containing only low frequencies would not be affected by the high frequency attenuation applied to it and, thus, the auditory cue would not be relevant. Keeping this in mind, one has to consider that background sounds that accompany cinematic scenes greatly vary in terms of spectral content. In addition, the spectral content of such sounds more often than not also varies greatly within the same scene as the cinematic action unfolds. It is, thus, necessary to study the effect of high frequency loss of background sounds for use in S3D cinema on different frequency bands of the auditory spectrum as this might be greatly influencing the effectiveness of this cue.

Details on Experiment 2 (LA)

Experiment 2 (LA) studied the effect of overall volume attenuation on the perception of depth in S3D clips. Similarly to Experiment 1 (HFF), one possible reason why the majority of the results are not significant is that the level attenuation alone is too weak a cue to be considered by the listener/viewer in the presence of S3D visuals. Considering that volume level attenuation is a less variable, and thus more trustworthy, distance cue than high frequency loss (Coleman 1963), it is possible that one reason for its apparent ineffectiveness is that the focus of the viewer/listener is directed heavily towards the S3D visuals. In this case the, otherwise strong, auditory cue appears to have no effect in the presence of the S3D objects within the clips. This observation is also strengthened by the fact that the only significant results were observed in the two cases where clips with simple S3D visuals were used.

Details on Experiment 3NS (LAHFF)

In Experiment 3NS (LAHFF) no significant results were observed, despite the fact that two auditory depth cues (e.g. overall volume attenuation and high frequency loss) were used simultaneously. This suggests that even the supposedly stronger effect of the combined cues did not have the power to get the attention of the viewer/listener from the S3D visuals. The fact that no significant results were observed even in the case of clips with simple S3D visuals does not give us firm conclusions (for definitions of rich and simple clips in the context of this study see: **Chapter 3.3.4 - Apparatus and Material).** It is possible that the combination of these auditory cues functions differently in real life comparing to the cinematic context.

Overall it is possible that the subjects' inability to clearly perceive the differences of depth produced by the auditory cues was caused by the lack of a direct reference. Experiment 3S (LAHFF) was designed to verify this possibility.

Details on Experiment 3S (LAHFF)

The results of Experiment 3S (LAHFF) suggest that the perception of depth of the viewer/listener is affected when clips with simple S3D visuals are used and the two versions of these clips are presented in succession. This may be an indication that the viewer/listener needs a direct reference against which to evaluate the depth/distance of a scene. In this case, the direct reference is the clip presented first and against which the second version of the clip is evaluated. This observation is in line with findings of previous studies (Mershon and Bower 1979; Coleman 1963, Little et al. 1992).

The fact that significant results were observed again only in cases where clips with simple S3D content were used indicate that in the presence of strong S3D visuals the auditory cues may become ineffective. This is also in line with the results of Experiment 2 (LA).

Details on Experiment 3NS (LAHFF) in Relation to Experiment 3S (LAHFF)

Interesting observations can be made by examining Experiments 3NS / 3Sa / 3Sb (LAHFF) together. In these experiments identical S3D clips and experimental variables (the combination of high frequency and overall volume attenuation) were used. The only difference between the experiments was that in Experiment 3NS (LAHFF) the pairs of clips (control and experimental versions) were presented to the subject in a randomised order and separated one from the other by three other clips while in Experiments 3Sa / 3Sb (LAHFF) they were presented in succession. This means that in Experiments 3Sa / 3Sb, the first clip of the pair could be used as a reference for judging the depth of the second clip.

The results of Experiment 3NS (LAHFF) demonstrated that there was no statistically significant perceived difference between the perceived depth in the control and the experimental clips. The results of Experiment 3Sa (LAHFF) showed that in two out of four clips there was a statistically significant difference of perceived depth between the clips. It must be noted that the clips with statistically significant results were the ones with simple S3D content (i.e. clips 2_S, 4_S). This was in accordance with the cumulative results for clips with rich (1_R / 3_R) and simple (2_S / 4_S) S3D content in which the simple clips had statistically significant results. Finally, the cumulative results for all four clips were statistically significant. In Experiment 3Sb (LAHFF) although no statistically significant results were observed when studying each clip in isolation, the cumulative results for clips with simple (2_S / 4_S) S3D content were statistically significant. Similarly, the cumulative results for all four clips were also statistically significant. Based on these observations, it can be claimed that a genuine auditory perceptual depth effect may occur when the control and experimental versions of clips with simple S3D content are presented in succession to the viewer/listener.

In Experiments 3Sa / 3Sb (LAHFF), where statistically significant results were observed, subjects appeared to perceive the control clips as having a greater depth than the experimental ones. The direction of the results was unexpected, as the soundtrack of the experimental clips was constructed following realistic auditory depth cues. In other words, the hypothesis of the experiment is contradicted by this result. It is possible that the viewer/listener interprets perceptual cues differently when watching a cinematic performance as opposed to real life. When in the cinematic viewing mode the viewer/listener actively scans the environment for distinctive and strong audiovisual cues that are not necessarily realistic. When the louder sounds with richer frequency spectrum of the control group are present they stand out from the mixture of perceptual cues present in the S3D cinematic scene, while the quieter sounds with the narrower frequency spectrum of the experimental soundtrack may go unnoticed. As the control sounds become more noticeable they may contribute to the overall sense of S3D vastness, which could be also interpreted as S3D depth. In this case, the direction of the results may be explained by the fact that the control soundtrack involved louder sounds with richer frequency spectrum.

Discussion on Results for Richness of Scene

Interestingly, the only statistically significant results were observed for simple S3D clips where the visual stereoscopic cues of the backgrounds were either simple or weak, or completely absent (black). The significant results can be seen as an indication that the absence of strong stereoscopic visual background influenced the subjects' ability to judge the depth of scene.

Considering the total of the four experiments, in four out of eight occasions where the clips had simple S3D content, the subjects perceived some difference between control and experimental versions. In one case the result was in the predicted direction, while in the other three cases the result was in the opposite direction. This can be viewed as an indication that although there might be a genuine effect of auditory cues within the soundtrack in the perception of depth of the S3D clips, the direction of this effect might be affected by other perceptual and/or physical factors. The fact that the opposite direction of results was observed in Experiment 3Sa (LAHFF) may be taken as an indication that the perceptual cues provided in the context of a S3D cinematic viewing the viewer/listener uses a different interpretation mechanism when little S3D visual information is given. For example, in absence of clear visual depth cues, it could be possible that subjects associate depth and vastness of space with a louder background sound: a more powerful, imposing sound. This would be a metaphoric rather than physics-based association between sound and picture, a mechanism that might work better in these types of scenes. Overall these results suggest that visual complexity may need to be taken into account when assessing the role of auditory cues in this context.

4.6 Conclusion

This chapter presented a series of experiments studying the effectiveness of auditory depth cues as a means to increase the perceived sense of depth of S3D cinematic clips. The auditory cues used in the experiments were: high frequency loss, volume level attenuation (Coleman 1963) and the combination of the two in Experiments 1 (HFF), 2 (LA) and 3 (LAHFF) respectively. The idea behind these experiments was that by altering the high frequency content and/or the volume level of a background sound (corresponding to the background of the clip) the perception of depth of the S3D clip as a whole could be influenced without altering the S3D visual content.

Statistically significant results were observed for clips with simple S3D content in Experiments 2 (LA) and 3 (LAHFF). This suggests that the richness of S3D content within a scene may be an important factor for the effectiveness of auditory depth cues. Further work could involve the study of a wider range of S3D scenes with different levels of S3D visual complexity. The fact that the majority of these significant results are in the opposite direction to the one predicted does not allow for firm conclusions, however it suggests that the effect of auditory cues on the perception of depth may be different between S3D cinema and real life.

5 Experiments 4-6: Alteration of the Perceived Depth of a Moving S3D Animation Object with an Accompanying Sound

5.1 Introduction

This chapter presents a series of experiments studying the effectiveness of auditory depth cues as a means to increase the perceived sense of depth of a S3D animation object moving from within the S3D scene towards the participant. The idea behind these experiments is that by gradually increasing the high frequency content and/or the volume level of a sound accompanying the moving S3D object, the perception of depth of the S3D object and of the movement can be influenced. The auditory cues used in the experiments are: high frequency loss (Experiment 4), volume level alteration (Experiment 5) and the combination of the two (Experiment 6) (Coleman 1963). Details on the methodology and results of each experiment, followed by a discussion of the findings are presented in the following sections.

5.2 Experiment 4 (HFF)

5.2.6 Introduction

The aim of Experiment 4 (HFF) was to test whether the sole use of selective high frequency filtering of a sound that accompanies a moving S3D animation object can affect the perception of the distance this object covered within the S3D viewing space. As in previous experiments, this idea was tested on naive viewers. The hypothesis for this experiment was:

> *H_1: The perception of depth (distance covered) of a S3D cinematic object moving towards the viewer/listener can be increased solely by gradually*

increasing the high frequency content of an accompanying sound.

Participants were presented with a number of S3D clips involving cinematic objects moving towards the viewer/listener twice: once with the moving S3D object accompanied by an unfiltered corresponding sound (control soundtrack) and then accompanied by the same sound with its high frequencies gradually changing from filtered to unfiltered as the object approached the viewer/listener (experimental soundtrack, **Figure 5-1**).

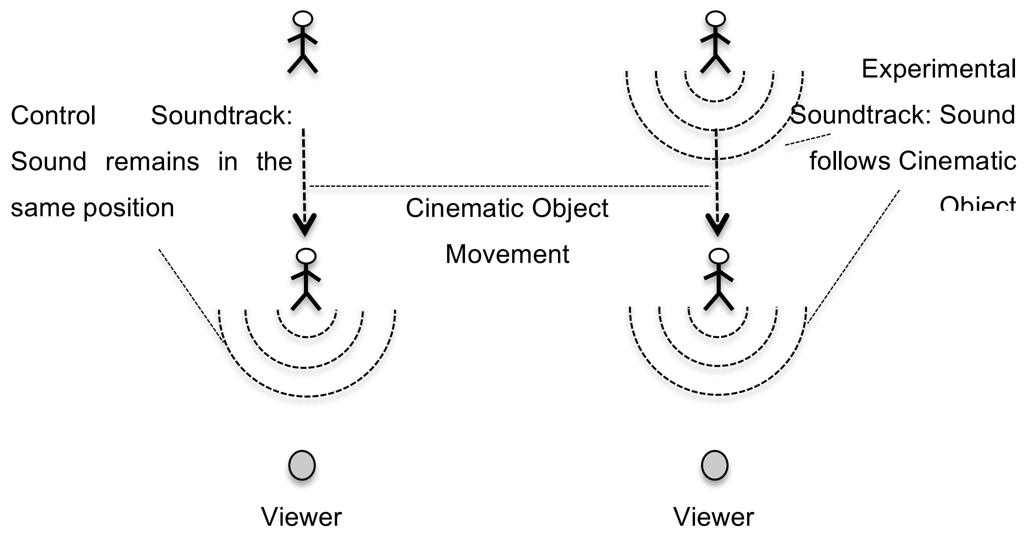


Figure 5-1: The movement of a S3D object is reflected in the alteration of the high frequency content of an accompanying sound

The order of presentation was randomised (see: **Chapter 4 - Methodology**).
After viewing each clip, participants were asked to rate the perceived level of S3D distance covered by the cinematic object on a 9-point scale (see: **Chapter 3.3.5 – Procedure**). If the hypothesis was correct, it was expected that the majority of participants would rate the objects accompanied by the experimental sound to have covered a greater distance (and, therefore, had a more pronounced perceived sense of depth) within the S3D cinematic space than those with the control soundtrack. More specifically it was expected that the gradual high frequency change in the experimental soundtrack (filtered-to-unfiltered gradual

high frequency change) would be attributed to a more pronounced movement of the cinematic object within the S3D space. This is because in real life the perceived high frequency content of a sound emitted by an approaching object gradually changes from the perspective of the viewer/listener (Coleman 1963, **Figure 5-1**). If the difference in ratings between experimental and control clips was statistically significant, this could be attributed to the auditory depth cue provided by the gradual change in the filtering of the high frequencies of the accompanying sound according to the proximity of the object from the participant. The background for this hypothesis is based on research and findings of auditory and multi-sensory distance perception in humans (Coleman 1963; Ecker and Laurie 2004; Turner *et al.* 2011).

5.2.7 Method

Design

The experiment employed a repeated measures design (Robson 1994: 100). The independent variable was the frequency content of the sound that accompanied the moving S3D cinematic objects. The variable had two levels: control (the original, static sound) and experimental (the same sound but with its high frequencies gradually changing from filtered/attenuated to unfiltered as the object moved towards the participant). The dependent variable was the perceived sense of depth of the moving object within the cinematic space. The test used an Absolute Rating Scale/ACR (ITU 1998) with a scale ranging from 1 to 9 (1=no significant movement to 9=very prominent movement). The scores obtained were analysed using the Wilcoxon Signed-Ranks test (ITU 1998: 5; Robson 1994; Harris 2008).

Participants

For the purposes of this experiment a sample of 35 participants were recruited. The sample consisted of 21 female and 14 male participants from the student and staff population of the University of York. The participants ranged from 18 to 62 years of age, with an average age of 24.3 (standard deviation = 10.42).

Apparatus and Material

A general description of the apparatus and material used in all the experiments presented here, as well as details regarding the creation and sourcing of the original S3D visual material used, are presented in **Chapter 3.3.4 - Apparatus and Material**. The source material used in this experiment was also used in Experiments 5 (LA) and 6NS (LAHFF) so that final results could be compared.

Figure 5-2: The movement of a S3D object towards the participant was reflected on the high frequency content level increase of an accompanying sound (Red Star Studio 2010)

The soundtracks for the clips were created in such a way as to reflect the overall context and the S3D objects visually present in the clips. More specifically, a background sound corresponding to the visual background of the clip was firstly created and added to the clip. Next, a sound corresponding to the cinematic object under investigation (the moving S3D object within the clip) was created and temporally assigned to the object (**Figure 5-2**). The background sound remained unchanged in both the control and the experimental version of each clip.

For the sound corresponding to the moving S3D object, two different versions were created: control and experimental. In the experimental version the sound started with its high frequencies attenuated in such a way as to roughly represent the high frequency attenuation caused by the initial distance of the sound source from the viewer/listener (**Figure 5-4**). As the S3D object moved towards the viewer/listener, the high frequency filtering was gradually decreased in order to reflect the movement. Once the S3D object reached the closest point to the viewer/listener the high frequency attenuation was completely decreased allowing the sound to reach the same (unfiltered) levels as in the control soundtrack (**Figure 5-3**, **Appendix 4 – Visual Clip Screenshots**).

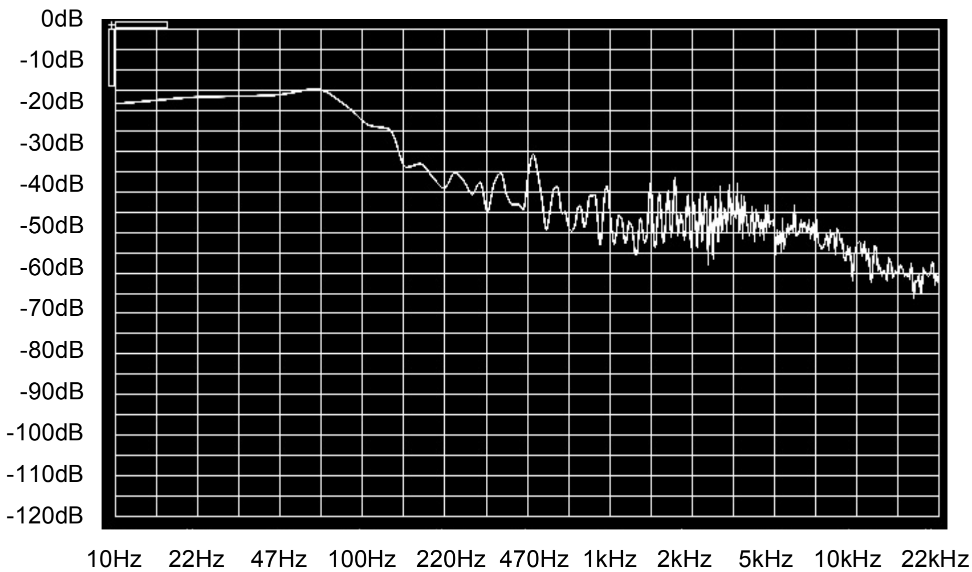


Figure 5-3: Spectrum Analysis for Control sound (Clip 3) for Experiments 4 (HFF), 5 (LA) and 6 (LAHFF)

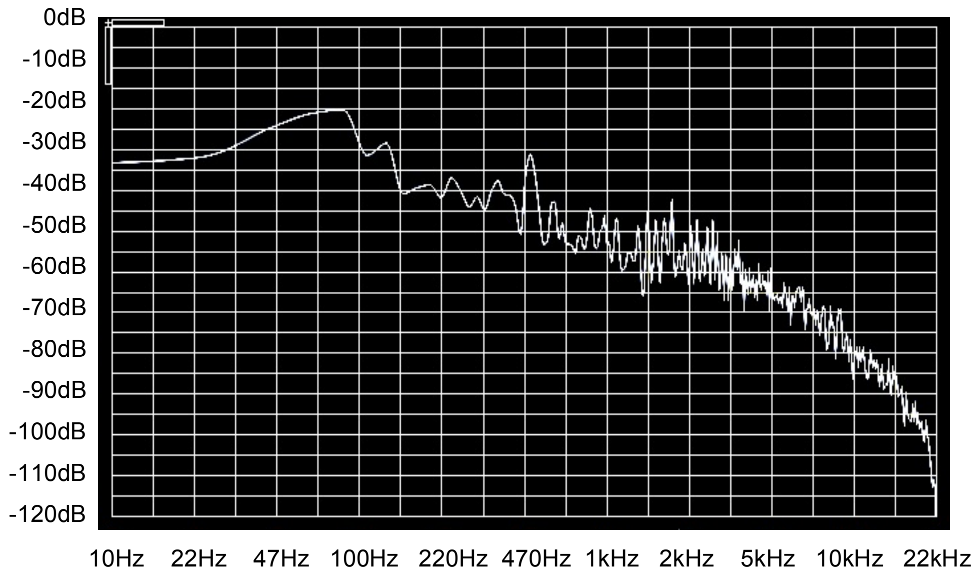


Figure 5-4: Spectrum Analysis for Experimental sound (Clip 3) for Experiment 4 (HFF - High Frequency Filtering) – Initial S3D object position

In order to filter the corresponding sound in such a way as to reflect the gradual high frequency content increase, the same table of absorption coefficients provided by Coleman was used (Coleman 1963, **Appendix 2 – Soundtrack Settings**) as in Experiments 1 (HFF) and 3 (LAHFF).

The frequency specific attenuation is caused as different frequency components are absorbed by the air at different rates and distances. The absorption coefficients were used to emulate how high frequencies are attenuated in the free field, and this sound alteration was produced using a digital multiband equalizer (EQ). The filtering was applied to the signal using the built-in digital EQs on a Logic Pro 9 digital audio workstation (Apple Inc. 2014).

In order to achieve the gradual change of the high frequencies filtering the EQ was appropriately automated. The initial filtering values for each EQ band (i.e. each frequency band) were calculated in such a way as to simulate the attenuation of frequencies of a sound emitting source that has been displaced away from the viewer by approximately twice the original distance of the sound from the viewer/listener (i.e. the apparent distance the sound source has on the control/unfiltered soundtrack). As the S3D object moved towards the viewer/listener the filtering was gradually decreased until the sound reached its original (unfiltered) high frequency content levels. The EQ settings used can be found in **Appendix 2 – Soundtrack Settings**. Details regarding the procedure followed in this experiment can be found in the relevant section (**Chapter 3.3.5 - Procedure**).

5.2.8 Results

The data were analysed using the Wilcoxon Signed Ranks test (Sani and Todman 2006; Harris 2008). The alpha level was set to 0.05 (Sani and Todman 2006), with a statistical power of 80% and effect size 0.5. The sample size was determined using the GPower software (Faul and Lang 2007; Heinrich-Heine-Universität 2013).

The results suggested that there was no statistically significant difference between the control and experimental versions of the clips in respect to the perception of distance (depth) covered by the moving S3D objects depicted in the clips. More specifically, the two-tailed significance level values obtained were 0.47, 0.38, 0.58 and 0.93 for pairs of clips 1, 2, 3 and 4 respectively. The calculated one-tailed significance level values were 0.23, 0.19, 0.29 and 0.46 for pairs 1, 2, 3, and 4 respectively. The cumulative scores of all four clips were also analysed. The two-tailed significance level value obtained was 0.57, while the one-tailed significance level value was 0.28. Detailed statistical data are provided in **Appendix 1 – Statistical Data**.

5.2.9 Discussion

Based on the results of the test, the null hypothesis:

> *H_0: The perception of depth (distance covered) of a S3D cinematic object moving towards the viewer/listener cannot be increased solely by gradually increasing the high frequency content of an accompanying sound*

cannot be rejected. This result could be due to a number of different reasons.

1) The effectiveness of the auditory cue may be dependent to the duration of the movement of the S3D visual object. It is possible that when the S3D object moves rapidly there is not enough time for the viewer/listener to detect and process the additional distance cues provided in the auditory domain. Further experimentation is required in this direction in order a conclusion to be reached regarding this claim.

2) Similarly to Experiments 1 (HFF) and 2 (LAHFF), in this experiment only one auditory cue was used to increase the perception of depth of a S3D object moving towards the viewer/listener. Since the brain seeks information for external events using more than one sensory cue (Holt 1997; Bolognini *et al.* 2005, Lippert *et al.* 2007) it is possible that the gradual change of the amount of filtering of high frequencies does not have the desired effect because it is not strong enough to influence perception. This possibility forms the basis of another experiment carried out during this research (Experiment 6 (LAHFF)) in which the cumulative effect of two auditory depth cues (high frequency filtering plus overall volume attenuation) is explored.

3) Visual dominance (Shams and Kim 2010) could be another reason for the ineffectiveness of the auditory cues in this experiment. It is possible that the gradual change of the amount of filtering of high frequencies is not effective in the presence of strong S3D visual depth cues that dominate perception. A range of auditory cues and different settings should be studied in order to verify the validity of this claim.

4) The effectiveness of the auditory cue could be dependent on the spectral content of the sound. It is possible that sounds with different spectral content could produce different results. In order to firmly conclude whether gradual high frequency increase can affect our perception of distance covered by moving S3D objects within S3D clips a large number of sounds of varied frequency content should be studied. It is proposed that this can form the basis for further work.

5) The auditory effect probably was not strong enough to affect the perception of depth of movement of S3D objects and more extreme differences between the sound in the initial and final position of the S3D object may be required in order to register as important audiovisual perceptual cues.

5.3 Experiment 5 (LA)

5.3.6 Introduction

The aim of Experiment 5 (LA) was to test whether increasing the volume of a sound that accompanies a S3D animation object moving towards the viewer/listener can affect the perception of depth of the object. This idea was tested on naive viewers.

The hypothesis for this experiment was:

> *H_1: The perception of depth (distance covered) of a S3D cinematic object moving towards the viewer/listener can be increased solely by gradually increasing the volume level of an accompanying sound.*

Participants were presented with a number of S3D animation clips involving animation objects moving towards the viewer/listener. The presentation was carried out twice: once with the moving S3D object accompanied by a corresponding sound of a given volume level (control soundtrack) and then accompanied by the same sound with its volume starting from a lower level and gradually increasing as the object approached closer to the viewer/listener (experimental soundtrack). The order of presentation was randomised (see: **Chapter 3 - Methodology**).

As in Experiment 4 (HFF), participants were asked to rate the perceived level of S3D distance covered by the cinematic object on a 9-point scale (see: **Chapter 3.3.5 – Procedure**). Based on the hypothesis, it was expected that most participants would rate the objects accompanied by the experimental sound as having covered a greater distance (and, therefore, had a more pronounced perceived sense of depth) within the S3D cinematic space than those with the control soundtrack. This is because in real life the volume of a sound emitted by an approaching object is expected to gradually increase from the perspective of the viewer/listener (Coleman 1963, **Figure 5-1**). If the difference in ratings between experimental and control clips was statistically significant in the predicted way it would be possible to attribute this difference to the auditory depth cue provided by the gradual change of the volume level of the accompanying sound according to the proximity of the object from the participant. The background for this hypothesis is based on research and findings of auditory and multi-sensory distance perception in humans (Coleman 1963; Ecker and Laurie 2004; Turner *et al.* 2011).

5.3.7 Method

Design

The experiment employed a repeated measures design (Robson 1994: 100). The independent variable was the volume level of the sound that accompanied the moving S3D cinematic objects. The variable had two levels: control (the original sound) and experimental (the same sound but with it volume starting at a lower level and gradually increasing to full volume as the object moved closer to the participant). The dependent variable was the perceived sense of distance covered by the moving S3D object within the cinematic space (and, thus, of the overall depth of the movement and the object). The test used an Absolute Rating Scale/ACR (ITU 1998) with a scale ranging from 1 to 9 (1 = no significant movement to 9 = very prominent movement). The scores obtained were analysed using the Wilcoxon Signed-Ranks test (ITU 1998: 5; Robson 1994; Harris 2008).

Participants

For the purposes of this experiment a sample of 35 participants were recruited. The sample consisted of 18 female and 17 male participants from the student and staff population of the University of York. The participants ranged from 18 to 39 years of age, with an average age of 24.5 (standard deviation = 6.07).

Apparatus and Material

A general description of the apparatus and material used in all the experiments of this study is given in **Chapter 3.3 - Method**. The source material used in this experiment was also used in Experiments 4 (HFF) and 6NS/6S (LAHFF) so that final results could be compared. Similarly to Experiment 4 (HFF), a background sound was firstly added to the clip. Next, a sound corresponding to the cinematic object under investigation (the moving S3D object within the clip) was created and temporally assigned to the object. The background sound remained unchanged in both the control and the experimental version. In the experimental version, the sound corresponding to the moving S3D object started with its volume level decreased in order to indicate an initial greater distance of the sound source from the viewer/listener (**Figure 5-6**). The volume level was gradually increased as the S3D object moved towards the viewer/listener until it reached the level of the control soundtrack once the S3D object was at its closest point to the viewer/listener (**Figure 5-5**, **Appendix 4 - Visual Clip Screenshots**). For the adjustments of the volume level of the corresponding sound, calculations based on the inverse square law were used (i.e. *6db volume attenuation for every doubling of the distance from a sound source*) (Coleman 1963: 306).

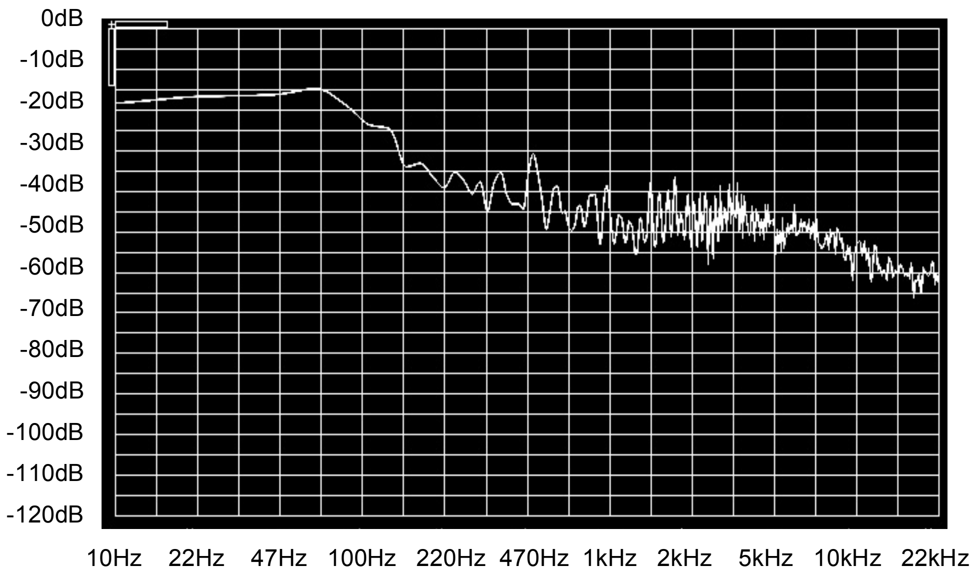


Figure 5-5: Spectrum Analysis for Control sound (Clip 3) for Experiments 4 (HFF), 5 (LA) and 6 (LAHFF)

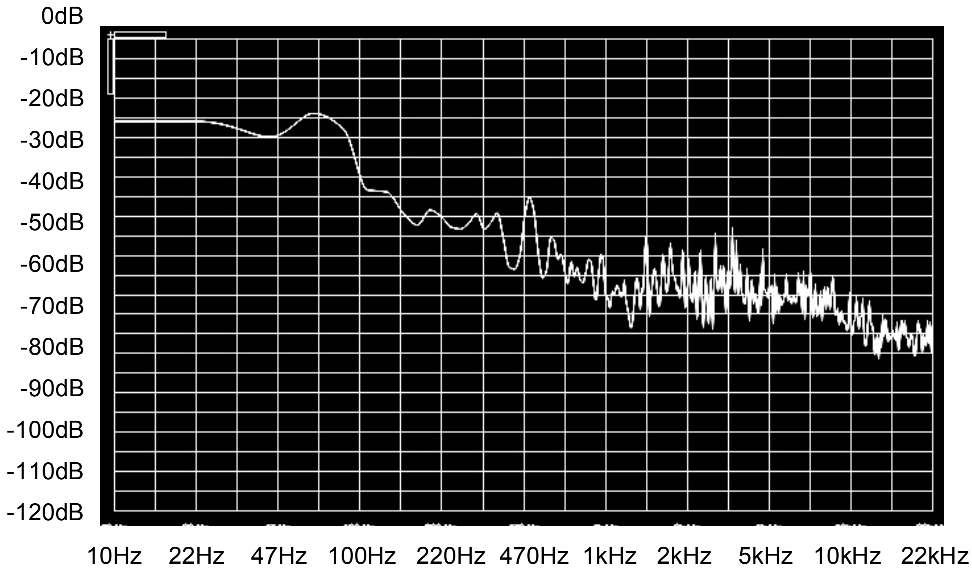


Figure 5-6: Spectrum Analysis for Experimental sound (Clip 3) for Experiment 5 (LA – Volume Level Alteration) – Initial S3D object position

Similarly to Experiment 4 (HFF), the volume at the initial position of the S3D object (experimental soundtrack) was adjusted in such a way as to simulate the attenuation of the overall volume of a sound emitting source that has been displaced away from the viewer by approximately twice the original distance of the sound from the viewer/listener (the apparent distance the sound source has on the control/unfiltered soundtrack). As the S3D object moved towards the viewer/listener the volume level was gradually increased. The gradual change of the volume levels was implemented by means of data automation on a digital audio workstation. Details regarding the procedure followed in this experiment can be found in the relevant section (**Chapter 3.3.5 - Procedure**).

5.3.8 Results

The data were collected using the experimental procedure described in the relevant chapter (**Chapter 3.3.5 - Procedure**) and analysed using the Wilcoxon Signed Ranks test (Sani and Todman 2006; Harris 2008). The alpha level was set to 0.05 (Sani and Todman 2006), with a statistical power of 80% and effect size of 0.5. The sample size was determined using the GPower software (Faul and Lang 2007; Heinrich-Heine-Universität 2013).

The results suggested that in three out of four cases (clips 1, 2 and 4) there was no statistically significant difference between the control and experimental versions of the clips in respect of the perception of depth of (distance covered by) the moving S3D objects depicted in the clips. In one case (clip 3) a statistically significant difference was observed between the experimental and control clip in the predicted direction (participants perceived increased S3D movement depth in the experimental version of the clip, **Figure 5-7**). It must be stressed that despite the fact that only one out of four clips returned statistically significant results when studied in isolation, the analysis of the cumulative scores obtained for all four clips returned significant results (**Figure 5-8**). More specifically, the two-tailed significance level values obtained were 0.26, 0.50, 0.02 and 0.26 for pairs of clips 1, 2, 3 and 4 respectively. The calculated one-tailed significance level values were 0.13, 0.25, 0.01 and 0.13 for pairs 1, 2, 3 and 4 respectively. The cumulative scores of all four clips returned results with a two-tailed significance level value of 0.009, and a one-tailed significance level value of 0.005. Detailed statistical data are provided in **Appendix 1 – Statistical Data**.

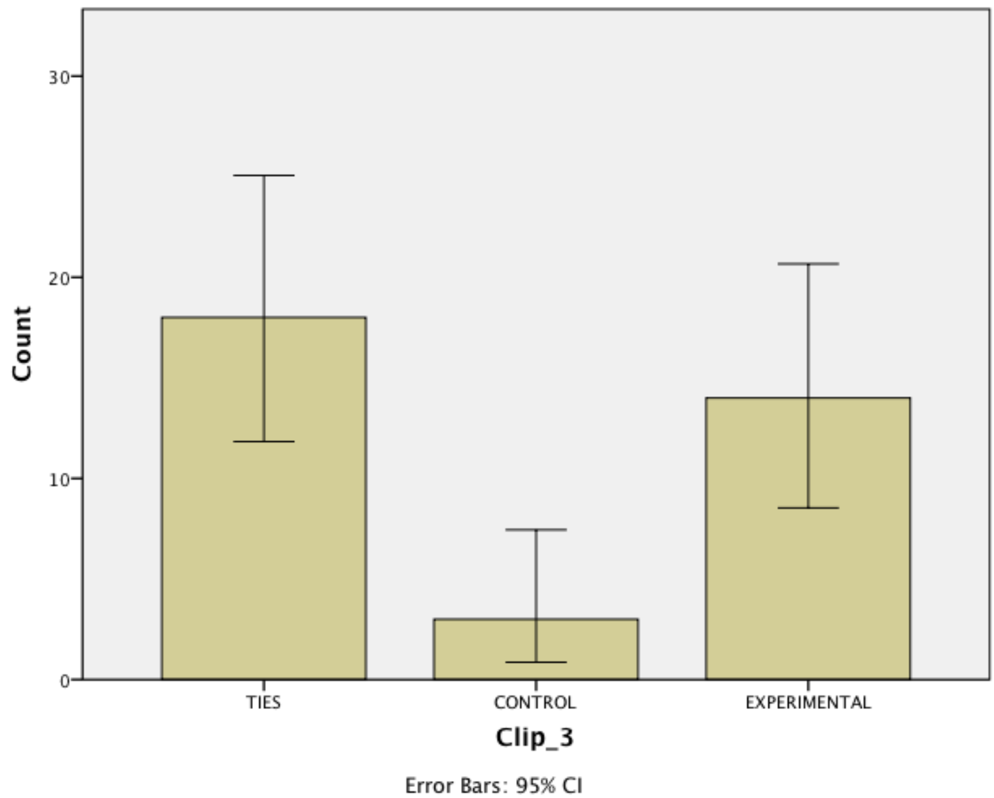


Figure 5-7: Participants response data graph for Clip 3 (Experiment 5 (LA))

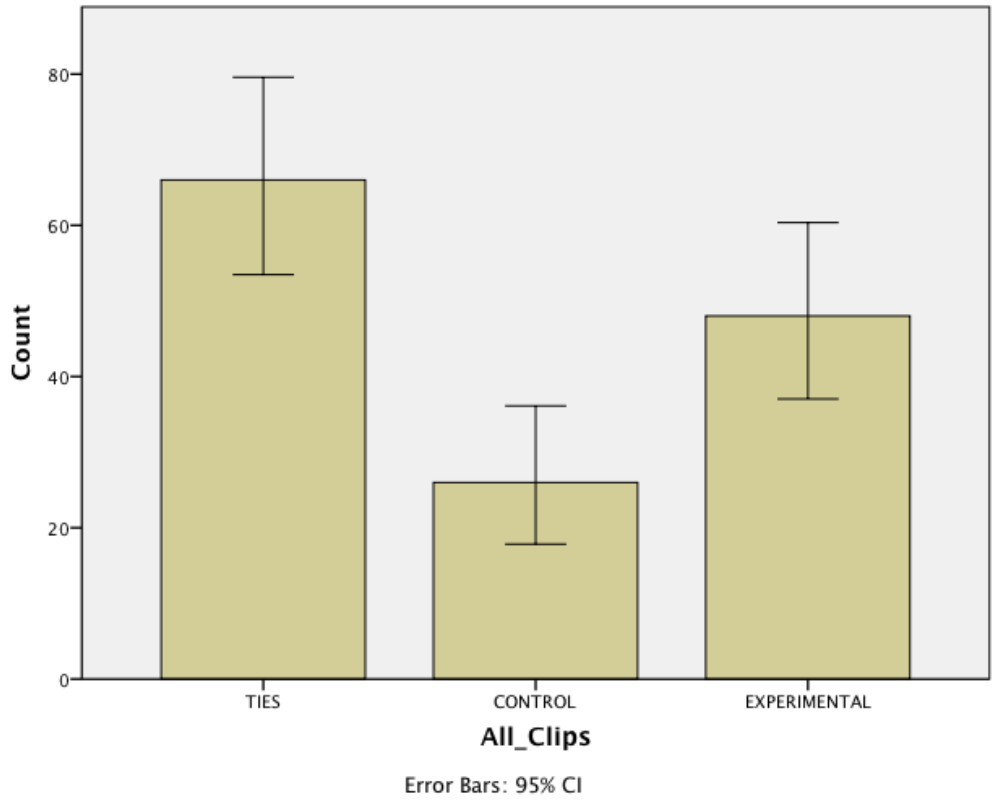


Figure 5-8: Participants response data graph for scores of all clips (Experiment 5 (LA))

5.3.9 Discussion

A statistically significant result was found when considering clip 3. This particular result makes the overall result for the test significant, however given that this significance is due to only one case out of four, it is appropriate to state that based on the results of the test, the null hypothesis:

> *H_0: The perception of the depth (distance covered) of a S3D cinematic object moving towards the viewer/listener cannot be increased solely by gradually increasing the volume level of an accompanying sound*

cannot be rejected without further evidence.

It should be noted that in this particular clip (clip 3) the movement of the S3D object was slower and lasted longer than the movements in the other three clips. In particular, the duration of movement of the S3D object in clip 3 was approximately 6 seconds, while in clips 1, 2 and 4 it was approximately 3, 4 and 2 seconds respectively. It is suggested that the duration of the movement could be an important factor when judging the effectiveness of level alteration as a means to increase the perception of movement depth in S3D. This result, although unique, cannot be dismissed without further evidence. The fact that the cumulative scores of all four clips returned statistically significant results is also relevant as it means that the significance for clip 3 was very high and should be taken into consideration. It is proposed that further experimentation in this direction is required, as the inconsistency of the results does not allow for firm conclusions.

The statistically insignificant results observed in three cases (clips 1, 2 and 4) could be due to a number of different reasons.

1) The effectiveness of the auditory cue may be dependent on the duration of the movement of the S3D visual object. It is possible that when the S3D object moves rapidly there is not enough time for the viewer/listener to detect and process the additional distance cues provided in the auditory domain. Further experimentation is required in this direction in order a conclusion to be reached regarding this claim.

2) As only one auditory cue was used it could have been not strong enough to influence the multimodal process of depth perception (Holt 1997; Bolognini *et al.* 2005, Lippert *et al.* 2007).

3) As in previous experiments (Experiments 1-4), the S3D visual content may have been too strong to allow the auditory cues to become noticeable, due to visual dominance (Shams and Kim 2010).

4) The cognitive content of the sound chosen for this experiment may have influenced the effectiveness of the auditory cue. Future work should include a wide range of sounds and scenes in order to fully explore the effectiveness of

volume level alteration as a cue that can affect our perception of distance covered by moving S3D objects within S3D clips.

5) The differences between the experimental and control versions of the soundtrack may have been too small to register. Experimentation with more extreme differences between sound in the initial and final positions of the S3D object may be helpful in order to determine the effectiveness of the auditory cue.

5.4 Experiments 6NS / 6S (LAHFF)

5.4.6 Introduction

The aim of Experiments 6NS / 6S (LAHFF) was to test whether the combination of gradual change of the high frequency content and the volume level of a sound that accompanies an S3D animation object moving towards the viewer can affect the perception of depth of (distance covered by) this object within the S3D cinematic space. As in previous experiments, this idea was tested on naive viewers

The hypothesis for this experiment was:

> *H_1: The perception of depth (distance covered) of a S3D cinematic object moving towards the viewer/listener can be increased solely by gradually increasing the high frequency content and the volume level of an accompanying sound.*

Participants were presented with S3D sequences firstly with the moving S3D object accompanied by a corresponding sound of a given high frequency content and volume level (control soundtrack) and then by the same sound with its high frequency content and volume starting from a lower level and gradually increasing following the object's movement towards the viewer/listener (experimental soundtrack).

In Experiment 6NS (LAHFF) the order of presentation was randomised (see: **Chapter 3 - Methodology**). After viewing each clip, participants were asked to rate the perceived level of depth of the S3D clip on a 9-point scale (see: **Chapter 3.3.5 – Procedure**).

In Experiment 6S (LAHFF) each pair of clips was presented to the participant in succession, but the order of the clips and the order of the presentation of the different versions within each pair of clips were randomised (see: **Chapter 3 - Methodology**). The participants viewed the two versions of each clip in succession and had to force-select the clip that appeared to have the most profound S3D depth effect among the two.

In both experiments, it was expected that the clips with the experimental soundtrack would be perceived to have a higher level of S3D depth for the moving object than those with the control soundtrack. This is because the combined effect of gradual change of high frequency content and volume level of the experimental soundtrack would be attributed to a more pronounced movement of the S3D object within the S3D space (**Figure 5-1**).

5.4.7 Method

Design

Experiment 6NS (LAHFF) employed a repeated measures design (Robson 1994: 100), with the independent variable being the combination of the gradual change of the high frequency and the volume level of a sound corresponding to the moving S3D object. As in previous experiments, the variable had two levels: control (the original, unfiltered background sound with original volume level) and experimental (the same background sound but with its high frequencies and overall volume starting from an appropriately decreased level and gradually being increased as the object approached the viewer/listener). The dependent variable was the perceived sense of depth of (distance covered by) the S3D object from the participants' viewpoint. Experiment 6NS (LAHFF) used an Absolute Rating Scale/ACR (ITU 1998) with a scale ranging from 1 to 9 (1=no depth to 9=very prominent depth). The scores obtained were analysed using the Wilcoxon Signed-Ranks test (ITU 1998: 5; Robson 1994; Harris 2008).

Experiment 6S (LAHFF) used the same clips and material as Experiment 6NS (LAHFF) but the control and experimental clips were presented to the subject in succession. The purpose of this this alternative method was to check whether successive viewings of the different versions of the clip would produce different results to non-successive viewings (Experiment 6NS (LAHFF)). The subjects were forced to choose between control and experimental clips and the data collected using this method were analysed using the Binomial Test analysis (McCluskey and Lalkhen 2007).

Participants

The sample population for both experiments was recruited from the student and staff population of the University of York.

Experiment 6NS (LAHFF) used a sample of 35 participants. The sample consisted of 21 female and 14 male participants ranging from 18 to 44 years of age, with an average age of 26.2 (standard deviation = 7.23).

For Experiment 6S (LAHFF) 30 participants were recruited, 15 of which were female and 15 male. The participants ranged from 18 to 38 years of age, with an average age of 23.6 (standard deviation = 4.90).

Apparatus and Material

A general description of the apparatus and material used in all the experiments of this study is given in **Chapter 3.3 - Method**.

As in Experiments 4 and 5, a background sound corresponding to the visual background was firstly created and added to the clip. Next, a sound corresponding to the moving S3D object was also added to the clip. The background sound remained unchanged in both the control and the experimental version of each clip and two different versions (control and experimental) were created for the moving S3D object sound. In the experimental version the sound started with its high frequency content and volume level decreased and then gradually increased in order to reflect the movement (**Figure 5-10**). The high frequency and volume level reached the same levels as in the control soundtrack when the S3D object reached the closest point to the viewer/listener (**Figure 5-9**, **Appendix 4 - Visual Clip Screenshots**).

The filtering of the high frequencies of the sound was based on the same absorption coefficients table as in previous experiments (Experiments 1, 3 and 4, **Appendix 2 – Soundtrack Settings**). The filtering was applied to the signal using EQ data automation on a Logic Pro 9 digital audio workstation (Apple Inc. 2014). Detailed EQ settings used can be found in **Appendix 2 – Soundtrack Settings**.

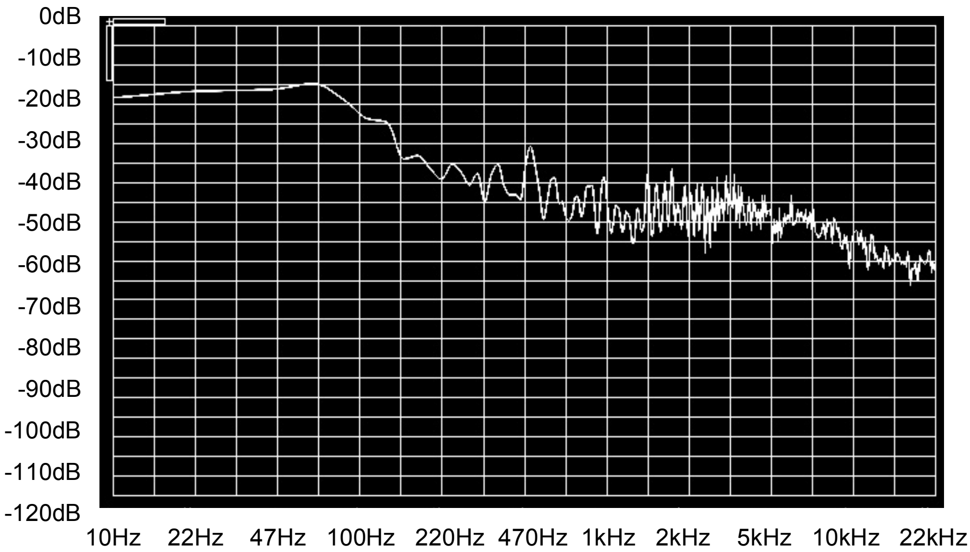


Figure 5-9: Spectrum Analysis for Control sound (Clip 3) for Experiments 4 (HFF), 5 (LA) and 6 (LAHFF)

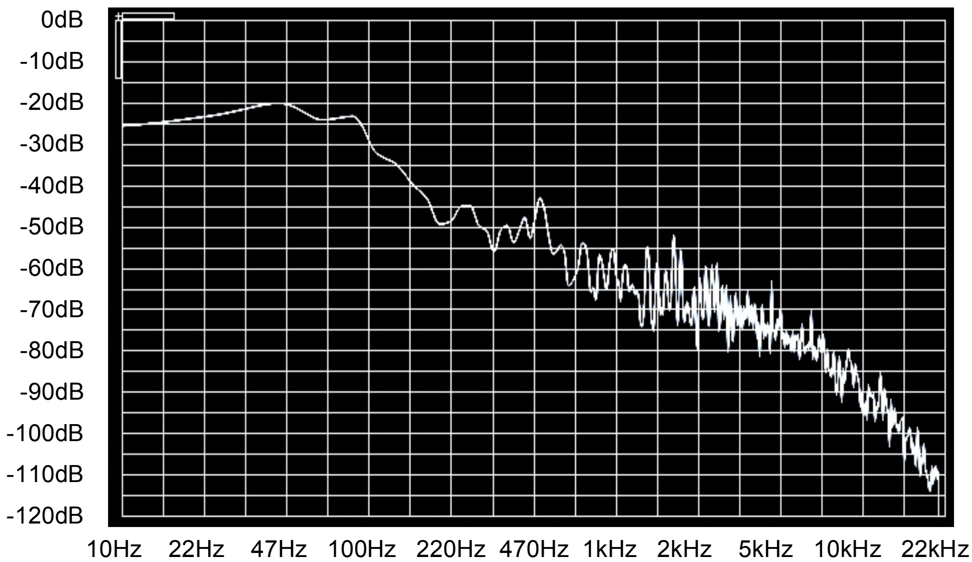


Figure 5-10: Spectrum Analysis for Experimental sound (Clip 3) for Experiment 6 (LAHFF – Volume Level Alteration and High Frequency Filtering) – Initial S3D object position

The volume level attenuation of the sound corresponding to the S3D object (in its starting position within the clip) was calculated based on the inverse square law (*6db volume attenuation for every doubling of the distance from a sound source*) (Coleman 1963: 306). The reference used for this calculation was the original unfiltered sound accompanying the S3D object in the control clip. Details regarding the procedure followed in this experiment can be found in the relevant section (**Chapter 3.3.5 - Procedure**).

5.4.8 Results

For Experiment 6NS (LAHFF), the data were collected using the experimental procedure described in **Chapter 3.3.5 - Procedure** and analysed using the Wilcoxon Signed Ranks test (Sani and Todman 2006; Harris 2008). The alpha level was set to 0.05 (Sani and Todman 2006), with a statistical power of 80% and effect size of 0.5. The sample size was determined using the GPower software (Faul and Lang 2007; Heinrich-Heine-Universität 2013). The results suggested that there was no statistically significant difference between the perception of depth of the S3D visual object in clips with the control soundtrack and clips with the experimental soundtrack. This was the case for all four different clips of the experiment. More specifically, the two-tailed significance level values obtained were 1.00, 0.94, 0.51 and 0.46 for pairs of clips 1, 2, 3 and 4 respectively. The calculated one-tailed significance level values were 0.50, 0.47, 0.25 and 0.23 for pairs 1, 2, 3 and 4 respectively. The cumulative scores of all four clips returned results with a two-tailed significance level value of 0.46, and a one-tailed significance level value of 0.23. Detailed statistical data are provided in **Appendix 1 – Statistical Data**.

For Experiment 6S, (LAHFF) the data were collected using the experimental procedure described in **Chapter 3.3.5 - Procedure** and analysed using the Binomial Test analysis (McCluskey and Lalkhen 2007). The alpha level was set to 0.05 (Sani and Todman 2006), with a statistical power of 80% and effect size to 0.5. THe sample size was determined using the GPower software (Faul and Lang 2007; Heinrich-Heine-Universität 2013). The results suggested that there was no statistically significant difference between the perception of depth of (distance covered by) the S3D visual object in clips with the control soundtrack and clips with the experimental soundtrack. This was the case for all four different clips of the experiment. More specifically, the two-tailed significance level values obtained were 0.85, 0.58, 0.85 and 0.58 for pairs of clips 1, 2, 3 and 4 respectively. The calculated one-tailed significance level values were 0.42, 0.29, 0.42 and 0.29 for clips 1, 2, 3 and 4 respectively. The cumulative scores of all four clips returned results with a two-tailed significance level value of 1.00, and a one-tailed significance level value of 0.50.

Detailed statistical data are provided in **Appendix 1 – Statistical Data**.

5.4.9 Discussion

Based on the results of the test, the null hypothesis:

> *H_0: The perception of depth (distance covered) by a S3D cinematic object moving towards the viewer/listener cannot be increased solely by gradually increasing the high frequency content and the volume level of an accompanying sound*

cannot be rejected. The results could be due to a number of different reasons.

1) As in previous experiments, the reason for the apparent ineffectiveness of the combination of high frequency filtering and volume level alteration to influence the sense of depth of the moving S3D object may be the strength of visual dominance (Shams and Kim 2010).

2) The spectral and/or cognitive content of the sound may be also related to the ineffectiveness of the combined auditory cue. The study of a wide range of sounds of different spectral and cognitive content exceeds the scope of this experiment but should be considered for further work.

3) Similarly to previous experiments, the effectiveness of the combination of the auditory cues may be dependent on the duration of the movement of the S3D visual object. Further experimentation with S3D movements of different duration may be helpful in order a conclusion to be reached.

4) It is possible that the combined auditory effect was not strong enough to affect the overall sense of depth of the clip and, thus, more extreme differences between the experimental and control versions of the soundtrack may be required.

5.5 General Discussion

In Experiments 4 (HFF), 5 (LA) and 6NS (LAHFF), when the subjects did not view two versions of a given clip in succession, gradually altering the high frequency content and the volume of the sound that accompanied the moving object did not have an overall statistically significant effect on the perception of depth of the object. It should be noted that these clips had *rich* S3D visual content. There are two main points of interest related to this result. Firstly, these experiments seem to confirm findings of previous studies that from the perspective of the viewer/listener, visual cues tend to dominate over auditory ones when both are available and perceptually assessed for the confirmation of a given audiovisual event (Shams *et al.* 2010, Wozczyk *et al.* 1995, Mastoropoulou 2006). The apparent ineffectiveness of the depth auditory cues could be possibly explained by the fact that the visual cues may have been trusted as the most reliable perceptual information between the two, potentially contrasting, types of sensory cues (visual and auditory) available in the clips. Secondly, the results of Experiments 4 (HFF), 5 (LA) and 6NS (LAHFF) could indicate that the presence of an explicit auditory reference point may be crucial to make the auditory cues perceptually important. In these experiments the subjects were intentionally not given a direct reference against which they could judge the relative loss of volume and high frequency content of a given sound. This was done to maintain the reality of the cinematic experience, in which a scene is not repeated twice in succession. It is possible that the subjects' inability to register the differences of depth produced by the auditory cues was caused by the lack of a direct reference.

Details on Experiment 4 (HFF)

Experiment 4 (HFF) studied the effect of gradual high frequency increase on the overall perception of depth of the moving object, as the corresponding S3D visual object approached the viewer/listener. One possible reason why the results were not statistically significant may be that the high frequency loss on its own, without level attenuation, is too weak a cue to be taken into consideration by the listener/viewer in the presence of strong S3D visual cues. Indeed, high frequency attenuation is amongst the most changeable of the auditory cues the brain uses for depth/distance estimation (Coleman 1963) and it is possible that the brain only relies on it when in conjunction with other cues that confirm the estimated distance, such as level attenuation. Additionally, a determining factor for the effectiveness of this auditory cue alone is the spectral content of the sounds used. For instance, a sound containing only low frequencies would not be affected by the gradual high frequency increase applied to it and, thus, the auditory cue would not be relevant. It is, thus, necessary to study the effectiveness of this auditory cue on different frequency bands of the auditory spectrum as this might be greatly influencing the effectiveness of this cue.

Details on Experiment 5 (LA)

Experiment 5 (LA) studied the effect of gradual volume increase on the overall perception of depth of the moving object, as the corresponding S3D visual object approached the viewer/listener. Similarly to Experiment 4 (HFF), one possible reason why the majority of the results were not significant could be that the level attenuation alone is too weak a cue to be considered by the listener/viewer in the presence of S3D visuals. Considering that volume level increase is a less variable, and thus more trustworthy, distance cue than high frequency increase (Coleman 1963), it is possible that one reason for its apparent ineffectiveness is that the focus of the viewer/listener is directed heavily towards the S3D visuals. In this case the, otherwise strong, auditory cue appears to have no effect in the presence of the S3D objects within the clips.

Details on Experiment 6NS (LAHFF) and 6S (LAHFF)

In Experiment 6NS (LAHFF) no significant results were observed, despite the fact that two auditory depth cues (overall volume and high frequency increase) were used simultaneously. This indicates that even the supposedly stronger effect of the combined cues did not have the power to get the attention of the viewer/listener from the S3D visuals. Experiment 6S (LAHFF), which employed successive viewings of the clips (and, thus, indirectly provided a depth reference in the form of the first clip) was designed to verify if providing a reference could change the results, but statistically significant results were not observed in this experiment either. This may be because either the combination of these two auditory cues were still not strong enough to alter the perception of the S3D visuals in general or that the specific combination of S3D visuals, sounds and auditory cue parameters were unaffected by it.

5.6 Conclusion

This chapter presented a series of experiments studying the effectiveness of auditory depth cues as a means to increase the perceived sense of depth of an S3D animation object movement going from within the S3D clip towards the participant. The idea behind these experiments was that by gradually increasing the high frequency content and/or the volume level of a sound accompanying the moving S3D object, the perception of depth of the S3D object and of the movement could be influenced. The auditory cues used in the experiments were: high frequency loss (Experiment 4), volume level alteration (Experiment 5), and the combination of the two (Experiment 6) (Coleman 1963).

Statistically significant results were observed in the case of clip 3 in Experiment 5 (LA), in which the S3D object movement had the greatest duration. This is an indication that the duration of the S3D movement may be related to the effectiveness of the auditory cue studied (gradual volume level alteration). This statistically significant result indicates that experimentation with a wider range of S3D object movements of different durations may be needed in order to fully investigate the effectiveness of gradual volume level alteration as a means to increase the perceived depth of moving S3D cinematic objects.

6 Experiment 7S (LAHFF): Alteration of the Perceived Depth of an S3D Object with an Accompanying Sound

6.1 Introduction

The aim of Experiment 7S (LAHFF) was to test whether the combination of selective high frequency filtering and appropriate volume attenuation of a sound that accompany specific objects within S3D animation clips can affect the perception of depth of the clips as a whole. The idea behind the experiment was related to Experiments 3Sa/3Sb (LAHFF) that also explored the combination of high frequency and volume alteration as a depth cue in S3D animation clips. The difference between Experiment 3Sa/Sb (LAHFF) and Experiment 7S (LAHFF) was that in Experiment 7S (LAHFF) a sound corresponding to a particular sound-emitting object within the S3D clip was studied, while in Experiment 3Sa/Sb the object of experimentation was the background sound of the clip as a whole. Therefore, the purpose of Experiment 7S (LAHFF) was to investigate the effectiveness of the studied auditory depth cues in a sound that is explicit and directly recognizable by the viewer/listener (the sound emitted by a visible and recognizable on-screen S3D object), in contrast to the generic, and potentially ambiguous, background sound of the clip studied in Experiments 3Sa/Sb (LAHFF).

Another difference of this experiment compared to Experiments 3Sa/3Sb was the presence of a background sound in addition to the explicit (studied) sound. It may be possible that the presence of two different sounds within the S3D clip is a factor that could affect the perceptual process of the participants. This is because the second sound may function as an additional indirect reference point against which S3D depth could be estimated.

Ultimately, the motivation for this experiment was to investigate whether the familiarity of the viewer/listener with the sound and the ambiguity of the sound

could affect its effectiveness as a depth cue in S3D scenes. Similarly to previous experiments, this idea was tested on naive viewers.
The hypothesis for this experiment was:

> *H_1: The perception of depth of a given S3D cinematic clip can be increased solely by selectively filtering the high frequencies and attenuating the overall volume level of a sound accompanying specific cinematic objects within the clip.*

Participants were presented with the S3D sequences once with a specific object within the clip accompanied by a corresponding sound of a given volume level and high frequency content (control soundtrack) and then accompanied by the same sound with its high frequencies filtered and its overall volume level decreased accordingly (experimental soundtrack, **Figure 6-1**).
Each pair of clips (control and experimental version) was presented to the participant in succession. The order of the clips and the order of the presentation of the different versions within each pair of clips were randomised, similarly to Experiments 3Sa/Sb (LAHFF) and 6S (LAHFF) (see: **Chapter 3 - Methodology**). Participants were asked to force-select the clip that appeared to have the most profound S3D depth.

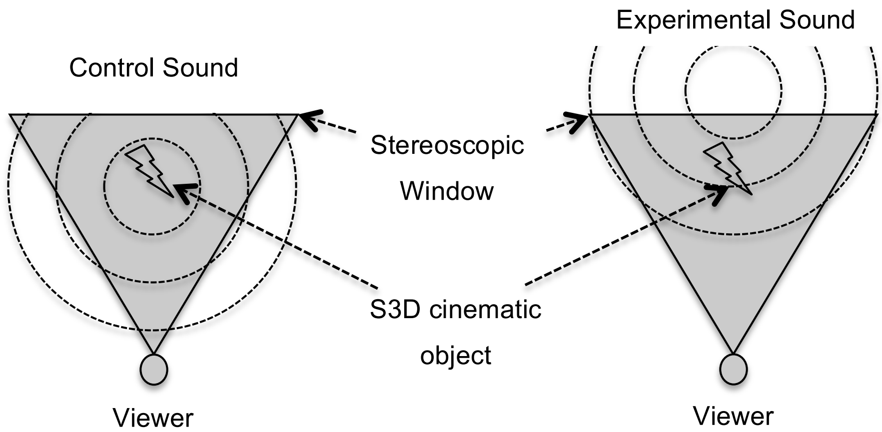


Figure 6-1: The high frequency content and overall volume level of a sound accompanying a S3D object is decreased in the experimental version

The expectation was that the majority of participants would rate the clips accompanied by the experimental soundtrack as having a higher level of perceived 3D depth than the clips with the control soundtrack, as the high frequency and volume level attenuation would be perceptually attributed to the sound-emitting object being located further out from the viewer/listener (**Figure 6-1**). The background for this was relevant research and findings on auditory and multi-sensory distance perception in humans (Coleman 1963; Ecker and Laurie 2004; Turner *et al.* 2011).

6.1.6 Method

Design

Experiment 7S (LAHFF) involved the successive viewings of the control and experimental versions of each clip. The independent variable was the combination of the high frequency content and the overall volume of the sound that accompanied the S3D visual object within the clip. The variable had two levels: control (the original, unfiltered sound with original volume level) and experimental (the same sound but with its high frequencies appropriately filtered and the overall volume level attenuated). The dependent variable was the perceived sense of depth of the S3D clip from the participants' viewpoint.
Experiment 7S (LAHFF) used successive viewings of the two versions of each clip, with the viewing order being randomized prior to the experiment. The subjects were forced to choose which clip (control or experimental) had the most pronounced S3D depth. The data were analysed using the Binomial Test analysis (McCluskey and Lalkhen 2007).

Participants

For Experiment 7S (LAHFF) a sample of 30 participants were recruited. The sample consisted of 15 female and 15 male participants from the student and staff population of the University of York, UK. The participants ranged from 18 to 38 years of age, with an average age of 23.6 (standard deviation = 4.90).

Apparatus and Material

A general description of the apparatus and material used in the experiment is given in **Chapter 3.3 - Method**.

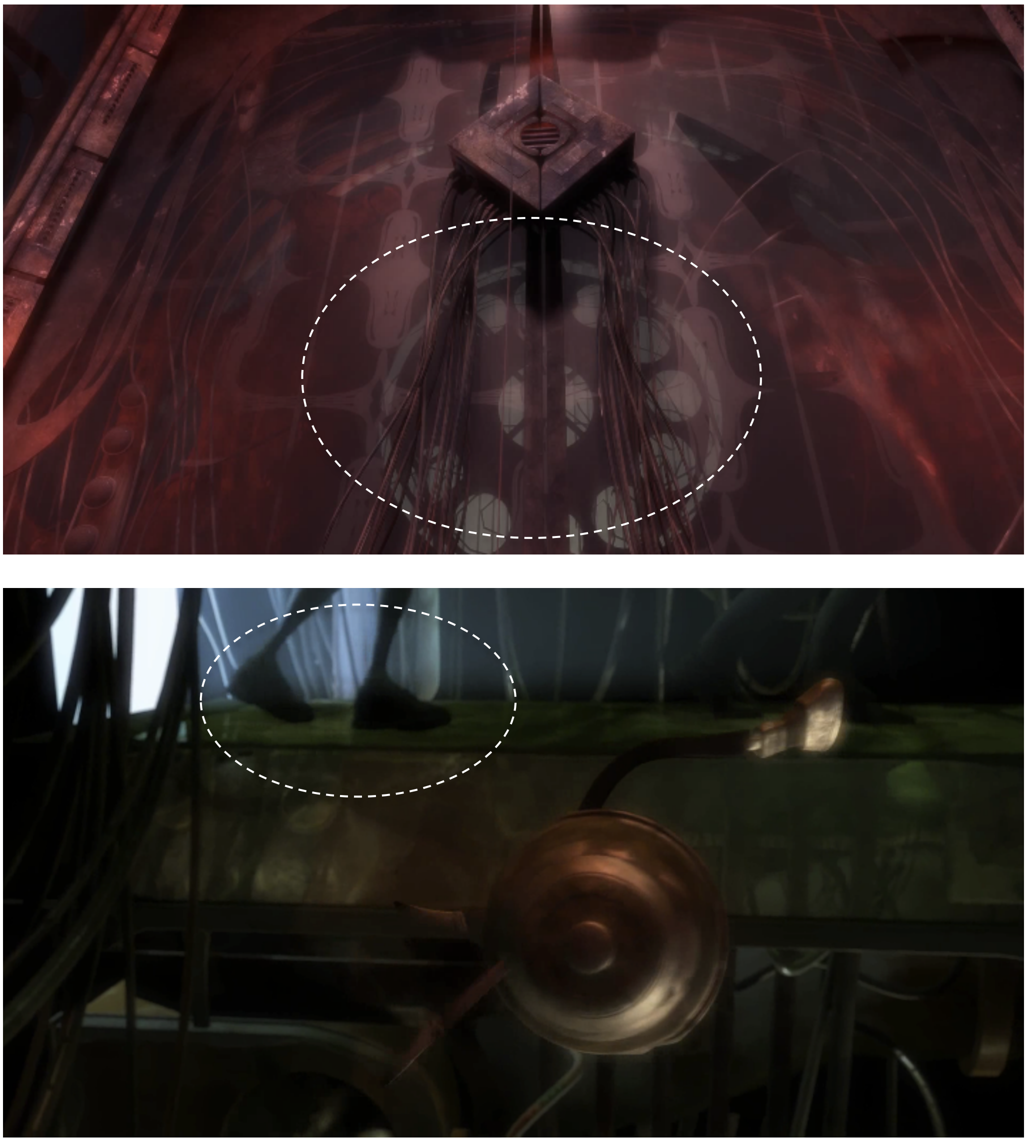


Figure 6-2: Appropriate sounds were added to S3D objects (i.e. clockwork mechanism and character's footsteps). In the experimental version of the clips the high frequency and volume level of these sounds were attenuated

For the creation of the soundtrack, a generic background sound was firstly created and added to the clip. The background sound was created in such a way as to reflect the surroundings depicted in the S3D clip (a ticking, mechanical sound for clip 1 and the humming of machinery for clip 2). Next, a corresponding sound was added to a distinct object or event within the clip (clockwork mechanism in clip 1, character's footsteps in clip 2, **Figure 6-2**).

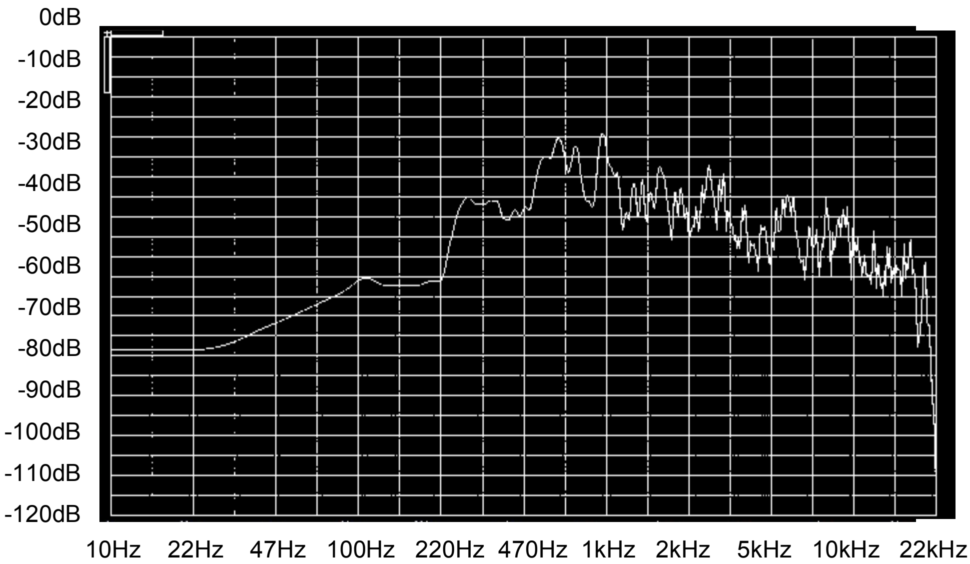


Figure 6-3: Spectrum Analysis for Control sound (Clip 1) for Experiment 7S (LAHFF)

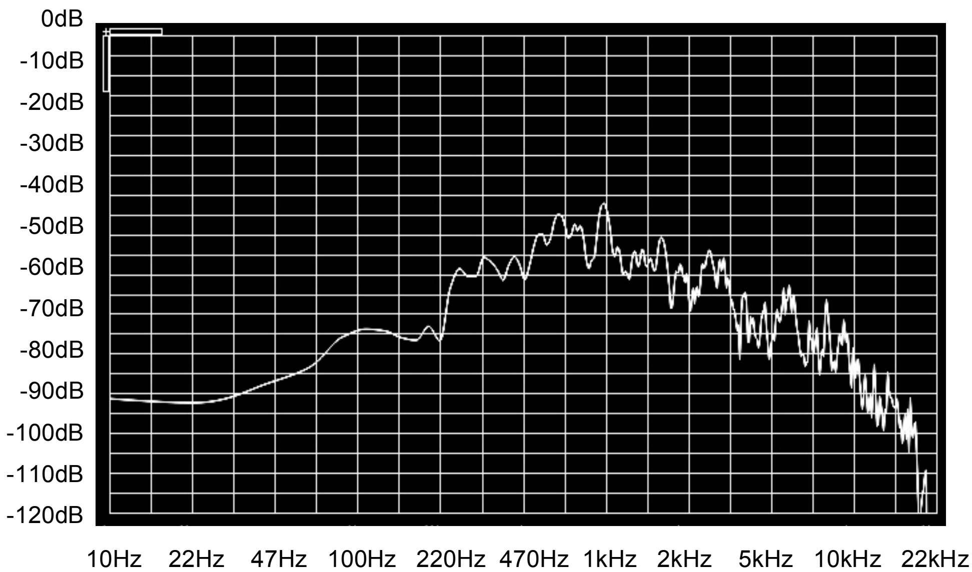


Figure 6-4: Spectrum Analysis for Experimental sound (Clip 1) for Experiment 7S (LAHFF - Volume Level Alteration and High Frequency Filtering)

For each clip, two different versions were created: control and experimental. Similarly to previous experiments, in the experimental version the high frequencies and the overall volume level of the sound accompanying the S3D object (clockwork mechanism, character's footsteps) were attenuated (**Figure 6-4**) as to roughly represent the respective attenuation caused by a displacement of the sound source by approximately twice the apparent distance of the original sound source from the viewer/listener (the sound in the control clip) (**Figure 6-3**). Calculations regarding the high frequency filtering required for the experimental versions of the soundtrack was based on Coleman's (1963) absorption coefficient table (**Appendix 2 – Soundtrack Settings**), similarly to Experiments 1 (HFF), 3 (LAHFF), 4 (HFF) and 6 (LAHFF). The volume level attenuation was calculated based on the inverse square law (Coleman 1963: 306). The calculated high frequency and volume level attenuation was applied to the soundtrack using EQs and volume level automation on a Logic Pro 9 digital audio workstation (Apple Inc. 2014). Details regarding the procedure followed in this experiment can be found in the relevant section (**Chapter 3.3.5 - Procedure**).

6.1.7 Results

For Experiment 7S (LAHFF), the data were collected using the experimental procedure described in **Chapter 3.3.5 - Procedure** and analysed using the Binomial Test analysis (McCluskey and Lalkhen 2007). The alpha level was set to 0.05 (Sani and Todman 2006), with a statistical power of 80% and effect size of 0.5. The sample size was determined using the GPower software (Faul and Lang 2007; Heinrich-Heine-Universität 2013).

The results showed that there was no statistically significant difference between the perception of depth in clips with the control soundtrack and clips with the experimental soundtrack. More specifically, the two-tailed significance level values obtained for the two clips were 0.36 and 0.85 for pairs of clips 1 and 2 respectively. The calculated one-tailed significance level values were 0.16, and 0.42 for pairs 1 and 2 respectively.

Detailed statistical data are provided in **Appendix 1 – Statistical Data**.

6.1.8 Discussion

Based on the results of the test, the null hypothesis:

> *H_0: The perception of depth of a given S3D cinematic clip cannot be increased solely by selectively filtering the high frequencies and attenuating the overall volume level of a sound accompanying specific cinematic objects within the clip*

could not be rejected. The fact that no statistically significant results were observed for both clips in Experiment 7S (LAHFF) could be due to a number of different reasons.

1) As in previous experiments, the effect of visual dominance (Shams and Kim 2010) could have influenced the effectiveness of the combined auditory depth cue.

2) The spectral and/or cognitive content of the sound may be related to its effectiveness as a depth cue. The study of a wide range of different sounds is required in order to fully investigate this possibility.

3) The effectiveness of the combined auditory depth cue may be related to the level of difference between the experimental and control versions of the soundtrack. Extreme differences between the experimental and control versions may register as important audiovisual perceptual cues in contrast to more subtle differences. A study of the thresholds at which such differences may become noticeable may form the basis for further work.

4) The existence of an additional sound (background sound) within the soundtrack may have had an impact on the perceptual effect of the studied sound. Further experimentation towards this direction may be required in order to clarify this.

6.2 Conclusion

This chapter presented an experiment studying the effectiveness of auditory depth cues as a means to increase the perceived sense of depth of a S3D animation object and, consequently, of the respective S3D clip as a whole. The idea behind the experiment was that by decreasing the high frequency content and/or the volume level of a sound accompanying a S3D object, the perception of depth of the S3D object from the participant could be influenced. This would encourage the viewer/listener to perceive a greater overall depth for the S3D clip, as the S3D sound emitting object would appear to be located at a more distant location. The auditory cue used in the experiment was the combination of high frequency loss and volume level attenuation (Coleman 1963). Details on the methodology and the results of the experiment were presented followed by a discussion of the findings.

The results suggested that there was no statistically significant difference between the control and experimental versions of the clips. The apparent ineffectiveness of the combined auditory depth cue to influence the perception of depth of the S3D clip could be due to the spectral and cognitive content of the selected sounds, the effect of visual dominance or the level of difference between the control and experimental versions. Further experimentation with a wide range of sounds and settings is required to fully explore the effectiveness of the studied auditory depth cues.

7 General Discussion, Summary and Conclusions

7.1 Introduction

This chapter re-states the research question of the thesis and summarises the methodology employed to address it. It outlines and discusses the most relevant findings of the series of experiments that were carried out to investigate the research question. It discusses a number of observations made based on these results and their potential relevance in the context of S3D filmmaking. Finally, it proposes further work on this area and includes a summary of the contents of the thesis.

7.2 Research Questions

The main research questions investigated in the thesis were the following:

> *Can S3D cinema be enhanced through the use of specific sound design techniques?*

And more specifically:

> *Can the perceived sense of depth of stereoscopic 3D cinematic scenes be increased by the introduction of auditory depth cues in the soundtrack?*

The research questions were based on the idea that although the novelty of stereoscopic 3D (S3D) cinema is related to vision, both human perception and cinema as a medium rely on cross-modal perceptual processes (see: **Chapter 2 – Background Overview**) that include auditory perception. Thus, it could be possible to enhance S3D cinema through the use of sound design techniques and to increase the perception of depth not only by adjusting the S3D visual content, but also by providing appropriate cross-modal auditory cues. If such sound design techniques and auditory cues were found to be effective they could provide an additional tool for S3D filmmakers.

The topic was approached with the following aims in mind:

- To investigate possible ways of constructing soundscapes that support and enhance the unique spatial characteristics (depth) of the S3D cinematic environment.
- To propose a number of sound design techniques and/or strategies that could enhance the perception of depth in S3D cinematic productions.

In order to achieve these aims, a number of objectives were set:

- To identify differences between the overall production cycle of 2D and S3D cinema.
- To identify the unique visual spatial characteristics of S3D cinema.
- To understand the impact and importance of sound in the perception of visual depth.
- To identify the possible impact of all the above on the production of the soundtrack.
- To propose a number of sound design techniques and strategies that could use the knowledge gained during the project to enhance the perception of depth in S3D cinema.
- To create a number of short S3D cinematic sequences that use the proposed ideas.
- To practically evaluate the proposed techniques through a number of audiovisual perceptual tests.
- To analyze and discuss the results of the tests and derive a series of sound design techniques that can be used to enhance the sense of depth in S3D.

7.3 Methodology and Experiments

A literature review was carried out outlining the relevant context stretching across the fields of cinema history, audiovisual perception, cinematic production and sound design.

Next, the research scope was narrowed down and focused on whether and how depth perception could be enhanced through sound design techniques.

The effectiveness of specific auditory cues to enhance S3D depth perception was investigated through several perceptual experiments testing specific hypotheses (see: **Chapters 4-6**). The experiments involved the study of the following auditory depth cues: high frequency filtering (i.e. HFF for High Frequency Filtering), level alteration (i.e. LA for Level Alteration) and the combination of the two (i.e. LAHFF for Level Alteration and High Frequency Filtering).

Experiments 1 (HFF), 2 (LA) and 3 (LAHFF) explored the effectiveness of high frequency and volume level alteration as a means to influence the overall sense of depth of a given S3D clip.

Experiments 4 (HFF), 5 (LA) and 6 (LAHFF) explored the effectiveness of altering the high frequency and volume level of the sound corresponding to a moving S3D object as a means to enhance the perception of depth of the object and its movement.

Experiment 7 (LAHFF) explored the effectiveness of altering the high frequency and volume level of the sound corresponding to a specific S3D object as a means to enhance the perception of depth of the S3D clip as a whole.

7.4 General Discussion

Simple versus Rich S3D Visual Content

In Experiment 2 (LA) significant results were observed in two cases. Both cases involved clips with simple rather than rich S3D visual content. Similarly, the results of Experiment 3S (LAHFF) suggest that the perception of depth of the viewer/listener is affected when clips with simple S3D visual content are used. This may be an indication that the absence of strong stereoscopic visual background influenced the subjects' ability to judge the depth of scene, and in particular that the presence of strong S3D visuals may render the auditory cues ineffective. If this observation proves to be correct, S3D filmmakers may use auditory depth cues when they are unable or unwilling to use extensive and rich S3D content but at the same time an increased sense of depth is desirable. Such cases may include parts of an S3D movie where the depth script dictates minimal use of S3D visual content (Autodesk 2008: 7; Mendiburu 2009: 93) or where the layout of a particular scene demands a visual background that is dark or relatively void of S3D objects. By and large, it is possible that visual complexity may need to be taken into consideration when assessing the role of auditory depth cues in the cinematic S3D context. From a practical perspective, this observation may be useful in cases S3D filmmakers require an enhanced sense of depth without having to achieve this necessarily through the use of S3D visual content.

3D Audio versus Cinematic Sound

It is also relevant to point out that the statistically significant results of the experiments were not always in the predicted direction. In Experiments 3Sa/3Sb (LAHFF) the direction of the results was opposite to the one predicted. This may be because the direction could be affected by other perceptual and/or physical factors, despite the fact that a genuine effect of auditory cues within the S3D soundtrack may exist. It is possible that the viewer/listener uses different perceptual mechanisms in real life and cinema. In the cinematic context the auditory depth cues may not be interpreted as they would in real life. In particular, in most cases in real life the viewer/listener would be inclined to perceive volume attenuation and high frequency loss as an indication that the sound emitting source is located further away. Thus, if this real life perceptual process were followed in cinema one could assume that a S3D scene involving cinematic objects that emit signals with attenuated high frequency and overall volume level signals would be perceived as having a larger depth as well. However, the contradictory direction of the results indicates that louder sounds with a broader frequency range may make scenes appear as having more pronounced S3D depth. It is possible that perceptual cues work in different and more complex ways in the cinematic context. In absence of clear visual depth cues the viewer/listener may be inclined to associate depth and vastness of space with a more powerful and imposing background sound. If true, this would be a metaphoric rather than physics-based association between sound and picture that may apply to cinema but not real life.

From the perspective of S3D cinema such an observation is relevant as it underlines another key aspect of cinematic sound: the audio cues and effects used for accompanying and supporting S3D cinematic visuals may not have to be necessarily realistic. This is also supported by the fact that unrealistic sound effects have been used systematically in cinema instead of the relevant recordings of the real sound events of the depicted objects or events (Sergi 1999, Bordwell and Thompson 2008: 287). It is widely known that frequently unnatural and exaggerated sound effects, like sounds representing gunshots and punches, can sound more convincing to the cinema viewer than the respective realistic sounds. Based on this, it is suggested that the use of audio cues in S3D filmmaking may not have to be viewed merely as another way to add realism to the S3D action, but as an additional cinematic tool with its own characteristics and qualities. It is possible that S3D filmmakers could use such audio cues in creative and unique ways unconstrained by the strict principles and expectations of real-life sound. Further work on this is necessary in order to investigate the validity of this observation and how such knowledge can be used creatively for cinematic purposes.

Reference for Comparison of Depth-Related Volume Attenuation

Experiment 3S (LAHFF) suggests that the perception of depth of the viewer/listener may be affected when clips with simple S3D visual content are used and the two versions of these clips are presented in succession. On the contrary, in Experiment 3NS (LAHFF) where the clips were presented in non-successive and random order no such perceptual effect was observed. This may be an indication that the viewer/listener needs a direct reference against which to evaluate the depth of an S3D scene. In the case of Experiment 3S (LAHFF), the direct reference is the clip presented first and against which the second version of the clip is evaluated. This observation is in line with findings of previous studies (Mershon and Bower 1979; Coleman 1963, Little et al. 1992).

In the S3D cinematic context this observation may be useful in a number of occasions. For instance, S3D filmmakers may introduce a direct depth reference point for a given cinematic object within the S3D space before altering the frequency content and/or volume of the sound that accompanies this object while the object visually remains at the same distance. This could assist in underlining a variation of the distance of the object from the viewer/listener without changing the actual visual depth. As an example, the story may dictate that a cinematic character should get closer to the viewer/listener from a psychological, but not necessarily from a physical, perspective. In such a case, it could be possible that the S3D visual position of the character (e.g. representing the physical location) could function as a reference for the depth of the character from the viewer while the audio (e.g. representing the psychological aspect) appears to move towards the viewer/listener by means of high frequency or volume alteration.

Alternatively, audio depth cues may be used in order to enhance the contrast of S3D depth during scene cuts or transitions. For instance, highlighting the contrast of depth of a given object between two different shots may be relevant to the story. In this case, a S3D filmmaker may decide to use the first shot as a reference point for the depth of the object against which the second would be evaluated. Although the obvious option for implementing such a transition would be arguably to use visual cues (i.e. use shots of the same objects from different distances) the second shot may also contain altered auditory depth cues that would be also judged against the initial (reference) depth point established during the first shot.

As a general observation, it may be relevant for S3D filmmakers to be aware that establishing a direct depth reference point may be crucial if they decide to use auditory depth cues, as the existence of such a reference may be necessary in order for such cues to be effective.

Experimental Conditions versus Cinematic Viewing

Another aspect that should be taken into account is the possibility that the experimental conditions used in this study may have influenced the perceptual effect. More specifically, it may be possible that the subjects were in a different perceptual mode in the experiments to that of a normal cinematic viewing. The experiments involved the supervised viewing of a single person at a time while in normal cinematic conditions many viewers are present in the same space and attend the cinematic show unsupervised. The perceptual mode may be affected by the sense of being part of a group viewing and being unsupervised and this may be an area where further experimentation may be required.

Additionally, it is true that in normal cinematic viewings the soundtrack is much more complex to the ones used in the experiments. In most occasions the soundtrack would contain a number of different simultaneous sounds, such as background sounds, dialogue or foreground sounds accompanying the various objects within a scene. As such, in many cases the viewer/listener may be able to establish reference points of depth partially based on comparisons between these different auditory cues within the soundtrack. A soundtrack containing only one or two distinct sounds, such as the ones used in the experiments, would not allow for such perceptual mechanisms to function. The introduction of multiple auditory cues within the S3D scenes that could function as reference points for depth judgment may be an interesting topic for further work in this field. If this idea is correct it would be a more useful piece of information for S3D filmmakers as it would resemble more closely the typical soundtrack of a real cinematic performance compared to the rather simplistic soundtracks used in controlled experiments like the ones carried out in this study.

Duration of Movement as a Factor for the Effectiveness of Audio Depth Cues

Experiment 5 (LA) studied whether increasing the volume of a sound that accompanied a S3D cinematic object moving towards the viewer/listener could affect the perception of depth of the object. A statistically significant result was found for one out of the four clips used in the experiment (clip 3). It should be noted that in this particular clip the movement of the S3D object was slower and lasted longer than the movements in the other three clips. In particular, the duration of movement of the S3D object in clip 3 was approximately 6 seconds, while in clips 1, 2 and 4 it was approximately 3, 4 and 2 seconds respectively. Based on this, it is suggested that the duration of the movement could be an important factor when judging the effectiveness of level alteration as a means to increase the perception of movement depth in S3D.

In case auditory depth cues are proven to be effective when used with slower movements within the S3D cinematic space they could be used as an additional tool for the creation of an enhanced sense of cinematic depth. This piece of information could be used in several ways by S3D filmmakers. For instance, a filmmaker may need to increase the sense of back-to-front movement of a given object for maximum dramatic effect (e.g. a ghost moving towards the viewer/listener in a horror movie). The filmmaker may choose to slow down this movement and add accompanying audio with its volume appropriately manipulated in order to achieve the maximum perceptual effect of the combined audio and visual depth cues.

On a different context, filmmakers may utilize this auditory effect in order to achieve an enhanced sense of movement of S3D cinematic objects without resorting in extreme, and possibly uncomfortable for the viewer, S3D visual movements. By slowing down the movement of an object moving towards (or away from) the viewer they could unlock an additional creative tool: the effective use of auditory depth cues in addition to (or instead of) the S3D visual cues. Such an approach may be useful in occasions where extensive S3D visual depth is undesirable, like parts of the movie where the depth script dictates so (Autodesk 2008: 7; Mendiburu 2009: 93) or where the filmmaker does not want to use extensive visual depth for creative or technical reasons.

Overall, the possibility of the effectiveness of auditory depth cues being influenced by the duration of the movement of the cinematic object they are connected to may be relevant in the context of S3D filmmaking. Further experimentation is needed in this direction and such observations could form the basis for further work.

Visual Dominance Considerations

Visual dominance over audition is a well-known effect and it has been the subject of multiple studies (Holt 1997: 115; Woszczyk et al. 1995: 4; Kitagawa and Ichihara 2002). It may be possible that in an environment rich in visual information, such as S3D cinema, the perceptual role of auditory depth cues is not as significant as in other occasions. For instance, if a sufficient amount of visual S3D depth information is provided within the cinematic scene the viewer/listener may be encouraged to focus and rely on the multitude of visual cues rather than the auditory ones. In such a case, the auditory depth cues may be less important in S3D cinema compared to other forms of cinema where less visual depth information is available. If true, this could be a relevant piece of information for S3D filmmakers. This is because the awareness of the decreased role of auditory depth cues in the S3D scene could allow them to focus more on other aspects of the soundtrack or the scene rather than on the creation of appropriate and detailed auditory depth cues. Further experimentation may be needed in order to test the validity of this claim and this may be the basis of further work.

Spectral Content

When using high frequency alteration as an auditory depth cue the sound spectral content may have to be taken into consideration. For instance, a background sound containing only low frequencies would not be affected by the high frequency alteration and, thus, the auditory depth effect would not be audible. This may be relevant in the cinematic context, as sounds that accompany the different cinematic scenes frequently also vary in terms of their spectral content. As such, in some occasions the auditory depth effect would be inaudible. This may be a relevant observation from the perspective of the S3D filmmaker, as it may affect the approach taken towards the composition of the sound textures when high frequency alteration is to be used as an auditory depth cue. Admittedly, there are indications that high frequency alteration on its own may not be a strong enough cue to affect perception in the rich sensory environment of S3D cinema. However, further experimentation is needed in order to thoroughly investigate the parameters that may affect the effectiveness of this auditory cue. Experimentation with different frequency bands of the audio spectrum may be useful as it could highlight the areas where the auditory cue is more effective for S3D cinematic purposes.

Auditory Depth Cues Outside the Field of View

The experiments in this study focused on high frequency and volume alteration as auditory depth cues inside the field of view of the viewer/listener. This is one of many possible areas that such auditory cues may be used. It may be possible that such auditory cues are more effective, and thus more relevant for S3D cinema from a practical perspective, when used to create a sense of depth outside the field of view. In this occasion they could function as a means to extend the cinematic world beyond the S3D viewing space rather than to support the depth conveyed by the visual information. More specifically, S3D filmmakers may find these auditory cues more useful as a tool to expand the cinematic space behind the viewer/listener rather than as a way to confirm the depicted depth of certain visible objects. Such a use of auditory cues could negate the perceptual effects of vision on sound, such as visual dominance. Experimentation with appropriate off-screen cinematic sounds that include auditory depth cues may be helpful in order to evaluate the effectiveness of this idea.

Animation versus Actual Cinematic Footage

The footage used for the experiments in this study was S3D animation excerpts and not actual cinematic footage with human actors/actresses and natural or realistic surroundings. This may be a factor that could influence the effectiveness of the auditory cues as participants may have had a better perception of the distances and depth within an actual cinematic scene than in an S3D animation one. This is because the participants may be able to estimate more accurately the relative size and distance of the various objects within the S3D scene based on real life experience, something that could be more challenging when using the clearly artificial S3D animation clips. Further experimentation including cinematic footage with shots of real-life objects and human actors/actresses, may be needed in order to investigate these possibilities.

Short Scenes versus Full Length Movies

In the majority of actual full-length cinematic productions the cinematic world and surroundings are progressively established throughout the movie either through the use of establishing shots or through the alternation of a number of shots that introduce the cinematic space to the viewer. This extensive exposure to the cinematic world may allow the spectator to have a better sense of the space and depth in which the cinematic action takes place or the characters and other objects are located. In contrast, the S3D animation clips used in this thesis was of a relatively short duration (under 20 seconds) and also consisted of a very small number of shots (or of a single shot). It is, thus, possible that the spatial layout of the scene was not very clear to the participant due to the small exposure time to the cinematic environment and the lack of relevant establishing shots that would assist in the familiarization with the cinematic surroundings under normal cinematic viewing conditions. For instance, in several cases the animation clip depicted one or two characters against a background of a large machine. Although the participant could possibly have a good idea of the size and distance of the human-like characters, the actual size and distance of the machine in the background could be more challenging to estimate without additional cues. In an actual cinematic viewing the filmmakers could have used either extensive shots of the machine or relevant establishing shots in order to make its size apparent to the audience, something that was not possible with the short animation clips used in the experiments.

If the actual size and depth of the background or of the cinematic objects were not very clear the participants may have had a more difficult time registering the auditory cue differences. This is because any alteration of volume or changes in the spectral content of the corresponding sounds may have been attributed to the background or object having different characteristics (e.g. size) rather than being closer to or further away from the participant. The possibility that the short duration and limited shot alternation of the clips used in this thesis may have been a limiting factor in making accurate decisions regarding the cinematic S3D depth should not be dismissed. Further experimentation with longer sequences and/or ones that are more complex in terms of shot composition could assist in evaluating this possibility.

7.5 Main Findings and Their Relation to the Research Aims and Questions

The first aim of this project was:

- to investigate possible ways of constructing soundscapes that support and enhance the unique spatial characteristics (depth) of the S3D cinematic environment.

This aim was reached by suggesting a number of possible ways the soundtrack can be used in the context of the unique spatial environment of S3D cinema. This was achieved by investigating and highlighting (through the literature review and the film case studies) the differences between 2D and S3D cinema both in terms of visual orientation and production. Based on these differences, a number of possible ways the soundtrack can reflect such differences in the audio domain were proposed.

The second aim of this project was:

- to propose a number of sound design techniques and/or strategies that could enhance the perception of depth in S3D cinematic productions.

This aim was reached by means of the experimental investigation of three possible ways in which the introduction of auditory depth cues could be used for altering the perception of depth in S3D scenes. These were the following:

- High frequency and/or volume level alteration of background sounds that accompany S3D clips.
- Gradual high frequency and/or volume level alteration of sounds that accompany S3D cinematic objects moving towards the viewer/listener.
- High frequency and/or volume level alteration of sounds that accompany an important static object within S3D clips.

This research attempted to answer two main research questions:

Can S3D cinema be enhanced through the use of specific sound design techniques?

And more specifically:

Can the perceived sense of depth of stereoscopic 3D cinematic scenes be increased by the introduction of auditory depth cues in the soundtrack?

Experiment results give a complex answer to both questions. In answer to the first question, evidence was found that the studied auditory cues could affect how we perceive S3D scenes and therefore could be used to enhance S3D cinema. In answer to the second question, these auditory cues can have an effect on how we perceive depth, however the effect may be weak and not simply related to how we perceive depth in real-life.

7.5.1 Experiment Results

The main results of the experiments presented in this thesis are:

- High frequency filtering and level attenuation are weak cues to create a sense of depth in S3D scenes.
- The effect of these auditory cues may be noticeable only when the visual content of S3D scenes is simple.
- The viewer/listener may not interpret auditory depth cues in cinematic S3D scenes with the same mechanisms used in real-life: in cinema a loud prominent sound could be interpreted as related to a more imposing and deeper environment.

It is possible that the auditory cues used in the experiments were not strong enough to consistently alter the perception of depth in the presence of strong S3D content. This is partly explained by the documented effect of visual dominance (Shams and Kim 2010). The results of Experiments 2 and 3S support this claim as statistical significance was observed only for clips with simple S3D content.
In addition, the sensory cues available in S3D environments may be different to real-life settings and to environments such as 2D cinema that use only monoscopic visual cues. In real life there is no screen restricting the visual field and events are expected to occur anywhere within the 360 degrees sphere surrounding the viewer/listener. In real-life the viewer/listener is expected to be very alert to sound sources located outside the field of view, as there is no restriction in the potential location of the sound-emitting object. In a classic 2D

cinematic environment, although the available monoscopic visual depth cues provide a means to calculate the depth of cinematic objects from viewer/listener, depth estimation is rather symbolic and abstract compared to the more realistic and detailed real life equivalent. Therefore, audiovisual distance perception in the context of S3D environments could be examined as a unique and separate topic with its own constraints and rules. Considering this, the use of realistic and detailed depth sonic representations of cinematic objects within the multi-channel soundtrack (e.g. the attempt to place the sounds of objects that appear on-screen at their correct and realistic supposed distance from the viewer/listener) might simply not be as important in the context of S3D filmmaking as it is in other 3D audiovisual applications, such as virtual reality, 3D simulators or video games. This could be because other, more symbolic and/or metaphorical, associations are at play in S3D filmmaking than simply realistic and causal relationships.

7.6 Further Work

Experiments 2 and 3S demonstrated that high frequency filtering and volume level attenuation can alter the perception of depth in clips with simple S3D visual content, but may be ineffective in clips with rich S3D content. This indicates that the complexity of the visual S3D content may be a factor for the effectiveness of the auditory cues studied. Experimentation with S3D scenes involving a wide range of S3D visual content of varied richness and complexity can be the subject of further work.

The results of Experiments 2 and 3S indicated that in S3D clips with simple visual content the auditory cues could affect the perception of depth. However, only one out of four cases in which statistically significant results were observed was in the predicted direction, while the other three was in the opposite direction. Further work can focus in fully explaining and confirming the reason for the direction of this effect.

In Experiment 5, statistically significant results were observed for the clip that had the greatest duration of movement among the four clips of the experiment. This is an indication that the duration of a movement may be a factor for the effectiveness of gradual volume level alteration as a depth cue. Experimentation with this auditory cue combined with S3D object movements of varied durations can be the subject of further work.

The cognitive content of the sounds used in the experiments is also possible to influence the effectiveness of the auditory cues studied. Experimentation with a wider range of sounds and S3D clips may be needed in order to verify this.

In Experiments 1-3, broadband sounds were used in an attempt to take into account the different effect of high frequency filtering and volume attenuation in different frequency bands. However, a larger variety of sounds should be tested to fully appreciate the importance of this auditory cue. This can form the basis of further experimentation.

When considering the results of Experiment 3NS in relation to Experiment 3S, it was observed that a genuine auditory perceptual depth effect occurs when the control and experimental versions of clips with simple S3D content are presented in succession to the viewer/listener (Experiment 3S). This effect was not observed when the experimental and control versions are presented to the viewer/listener in a non-successive manner. Further work in this direction is required in order to verify the importance of the presence of a direct reference. Similarly, experimentation with more complex soundtracks that contain a larger number of different background and foreground sounds may be needed. It is possible that such soundtracks would provide more auditory depth cues which could function as indirect reference points for the judgment of depth of other cues and, possibly, of the cinematic objects these cues accompany. This would be also more relevant to the cinematic context, as it would resemble more closely the complex soundtracks one may encounter in a typical cinematic viewing.

It is also suggested that other auditory depth cues, such as reverberation, could be studied in addition to high frequency and volume loss with distance. In particular, reverberation is one of the most commonly used auditory depth cues for cinematic purposes as it is highly effective in creating an increased sense of depth. The study of reverberation in S3D cinema may be an interesting topic for future work.

The results of Experiments 1, 3NS, 4, 6NS/S and 7S underline the possibility that the differences between the experimental and control versions of the soundtrack were not strong enough to affect the perception of depth in the presence of rich

S3D visual content. Further work could involve experimentation with clips in which more extreme differences between the two versions of the soundtrack are used.

The effect of visual dominance over audition may be also relevant. The possibility of auditory depth cues being less effective in the presence of strong S3D visual cues could not be dismissed without further experimentation. The study of visual dominance over audition in relation to depth perception in S3D cinema could form the basis of future work. In respect to the effects of visual dominance, in addition to studying the various auditory depth cues in isolation it may be interesting to also systematically study the effects of their combinations. It is possible that such combinations are more effective in creating a more pronounced sense of depth compared to using the various cues on their own, as they may create a perceptual effect strong enough to overcome visual dominance.

Another potential area for further study may be the exploration of the effectiveness of auditory depth cues when used with off-screen cinematic sounds. The fact that such sounds do not correspond to visible cinematic objects could affect the effectiveness of the auditory cues as it negates perceptual effects such as visual dominance. Experimentation with appropriate off-screen cinematic sounds that include auditory depth cues may be helpful in order to evaluate the effectiveness of this idea.

Finally, it is possible that the clips used in this thesis were too short and with very limited shot alternation in contrast to most actual cinematic viewings. This is something that could have influenced the ability of the participants to make accurate depth judgment decisions. Experimentation with longer sequences that resemble more closely the duration and complexity of actual cinematic viewings may be an interesting area for further work.

7.7 Conclusion

This chapter re-states the research questions and the aims of this work and summarises the research methodology employed to study them. It also summarises and discusses the results of the experiments that were carried out during this research. Finally, it presents a final discussion based on observations made based on the results and it presents possible areas for further work.

Appendix 1 – Statistical Data

Experiment 1 (HFF) – Statistical Data

Descriptive Statistics

	N	Mean	Std. Deviation	Minimum	Maximum
Clip_1_R_C	35	6.3714	1.69923	3.00	9.00
Clip_2_S_C	35	5.1429	1.86521	1.00	8.00
Clip_3_R_C	35	7.8857	1.40945	4.00	9.00
Clip_4_S_C	35	6.0000	1.49509	2.00	8.00
Rich_Content_Control	70	7.1286	1.72720	3.00	9.00
Simple_Content_Control	70	5.5714	1.73265	1.00	8.00
All_Clips_Control	140	6.3500	1.89253	1.00	9.00
Clip_1_R_E	35	6.4571	1.26823	4.00	9.00
Clip_2_S_E	35	5.0857	1.96096	1.00	9.00
Clip_3_R_E	35	7.9429	1.32716	3.00	9.00
Clip_4_S_E	35	5.8286	1.36092	2.00	8.00
Rich_Content_Exp	70	7.2000	1.49006	3.00	9.00
Simple_Content_Exp	70	5.4571	1.71680	1.00	9.00
All_Clips_Exp	140	6.3286	1.82486	1.00	9.00

Wilcoxon Signed Ranks Test

Test Statistics[a]

	1_R_E - 1_R_C	2_S_E - 2_S_C	3_R_E - 3_R_C	4_S_E - 4_S_C	Rich_E - Rich_C	Simple_E - Simple_C	All_Clips_E - All_Clips_C
Z	-.350[b]	.000[c]	-.847[b]	-.810[d]	-.715[b]	-.555[d]	-.017[d]
Asymp. Sig. (2-tailed)	.726	1.000	.397	.418	.475	.579	.986
Exact Sig. (2-tailed)	.769	1.000	.510	.438	.494	.586	.986
Exact Sig. (1-tailed)	.384	.505	.255	.219	.247	.293	.493
Point Probability	.026	.010	.034	.018	.007	.005	.002

a. Wilcoxon Signed Ranks Test

b. Based on negative ranks.

c. The sum of negative ranks equals the sum of positive ranks.

d. Based on positive ranks.

Ranks

		N	Mean Rank	Sum of Ranks
Clip_1_R_E - Clip_1_R_C	Negative Ranks	9[a]	10.67	96.00
	Positive Ranks	11[b]	10.36	114.00
	Ties	15[c]		
	Total	35		
Clip_2_S_E - Clip_2_S_C	Negative Ranks	12[d]	13.54	162.50
	Positive Ranks	13[e]	12.50	162.50
	Ties	10[f]		
	Total	35		
Clip_3_R_E - Clip_3_R_C	Negative Ranks	5[g]	8.00	40.00
	Positive Ranks	9[h]	7.22	65.00
	Ties	21[i]		
	Total	35		
Clip_4_S_E - Clip_4_S_C	Negative Ranks	12[j]	12.54	150.50
	Positive Ranks	10[k]	10.25	102.50
	Ties	13[l]		
	Total	35		
Rich_Content_E - Rich_Content_C	Negative Ranks	14[m]	18.43	258.00
	Positive Ranks	20[n]	16.85	337.00
	Ties	36[o]		
	Total	70		
Simple_Content_E - Simple_Content_C	Negative Ranks	24[p]	25.60	614.50
	Positive Ranks	23[q]	22.33	513.50
	Ties	23[r]		
	Total	70		
All_Clips_E - All_Clips_C	Negative Ranks	38[s]	43.79	1664.00
	Positive Ranks	43[t]	38.53	1657.00
	Ties	59[u]		
	Total	140		

a. Clip_1_R_E < Clip_1_R_C
b. Clip_1_R_E > Clip_1_R_C
c. Clip_1_R_E = Clip_1_R_C
d. Clip_2_S_E < Clip_2_S_C
e. Clip_2_S_E > Clip_2_S_C
f. Clip_2_S_E = Clip_2_S_C
g. Clip_3_R_E < Clip_3_R_C
h. Clip_3_R_E > Clip_3_R_C
i. Clip_3_R_E = Clip_3_R_C
j. Clip_4_S_E < Clip_4_S_C
k. Clip_4_S_E > Clip_4_S_C

l. Clip_4_S_E = Clip_4_S_C

m. Rich_Content_Exp < Rich_Content_Control

n. Rich_Content_Exp > Rich_Content_Control

o. Rich_Content_Exp = Rich_Content_Control

p. Simple_Content_Exp < Simple_Content_Control

q. Simple_Content_Exp > Simple_Content_Control

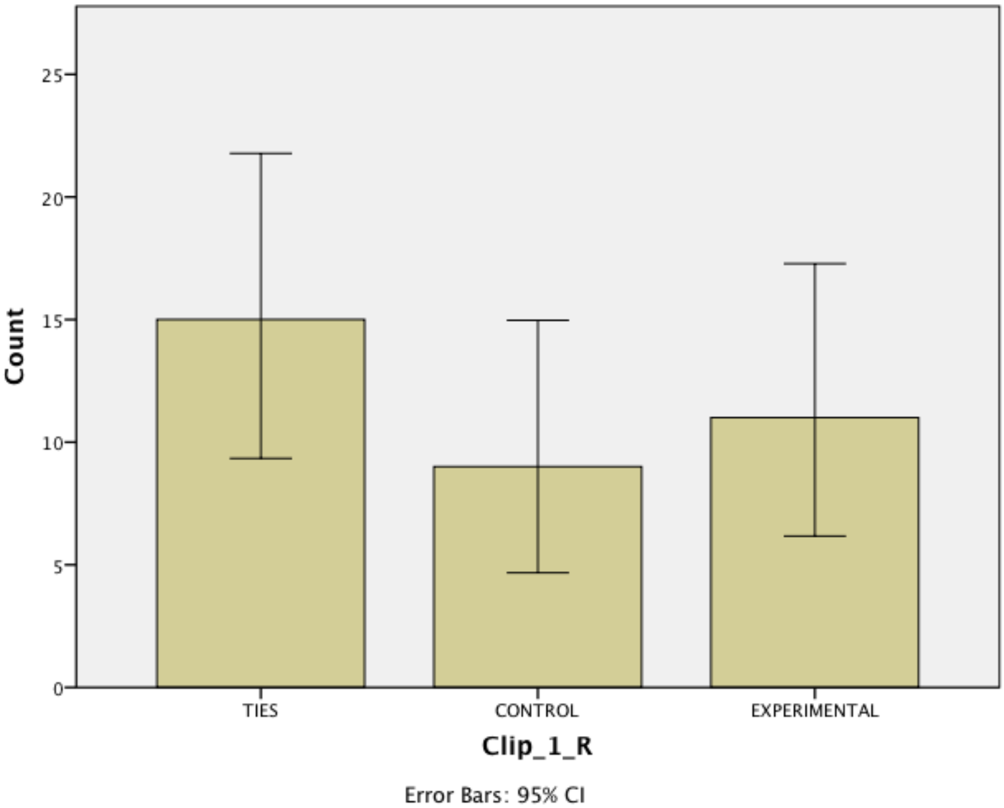
r. Simple_Content_Exp = Simple_Content_Control

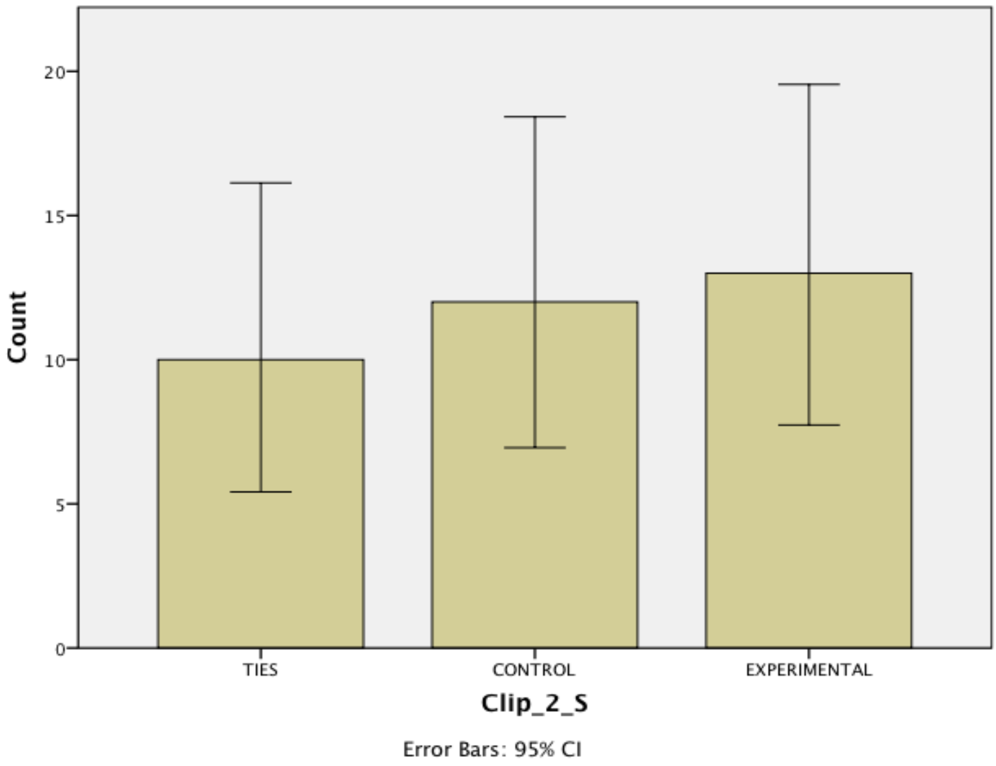
s. All_Clips_Exp < All_Clips_Control

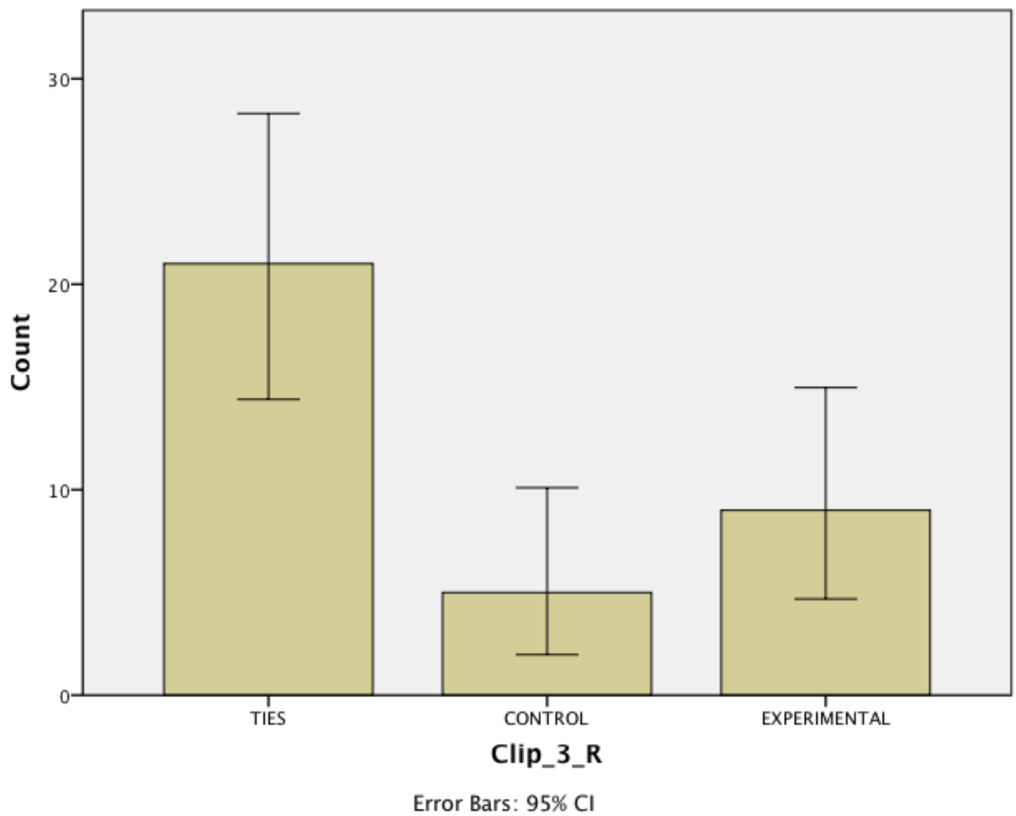
t. All_Clips_Exp > All_Clips_Control

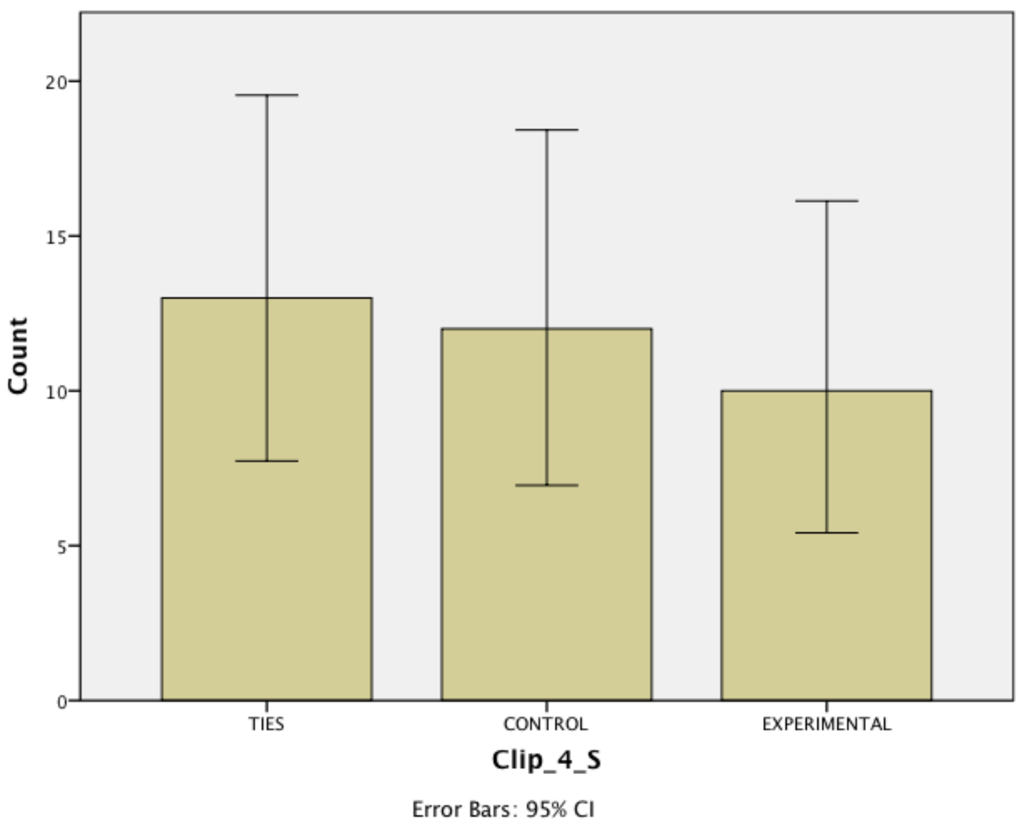
u. All_Clips_Exp = All_Clips_Control

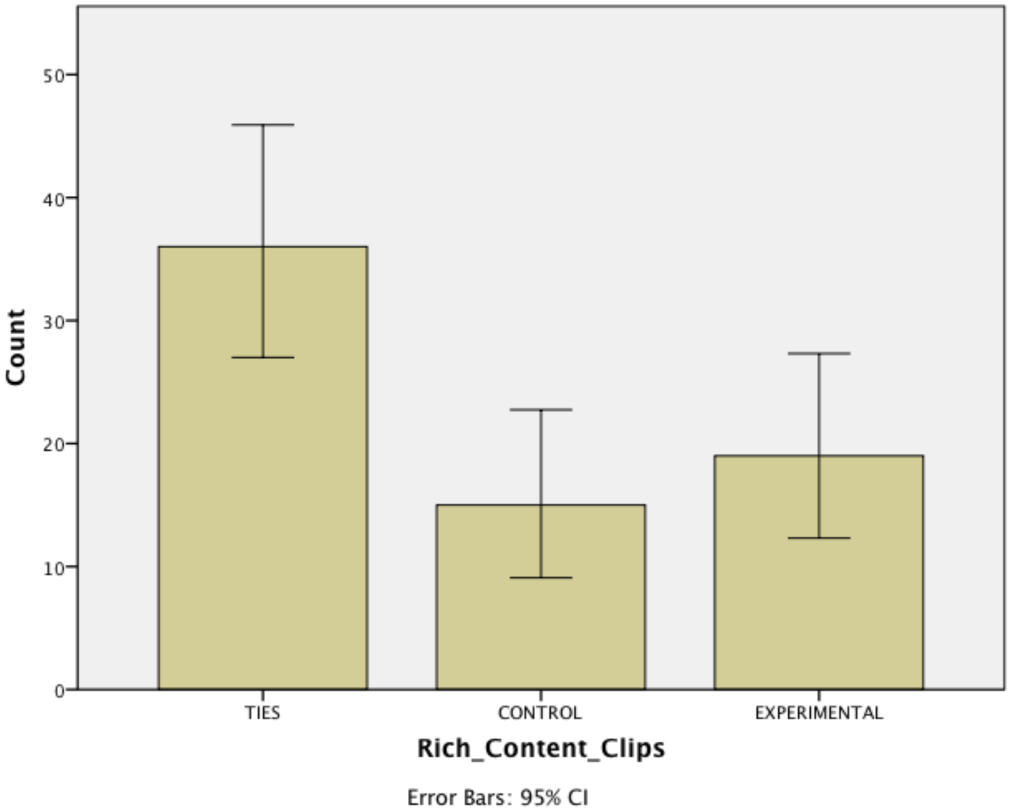
Participant Response Data Graphs



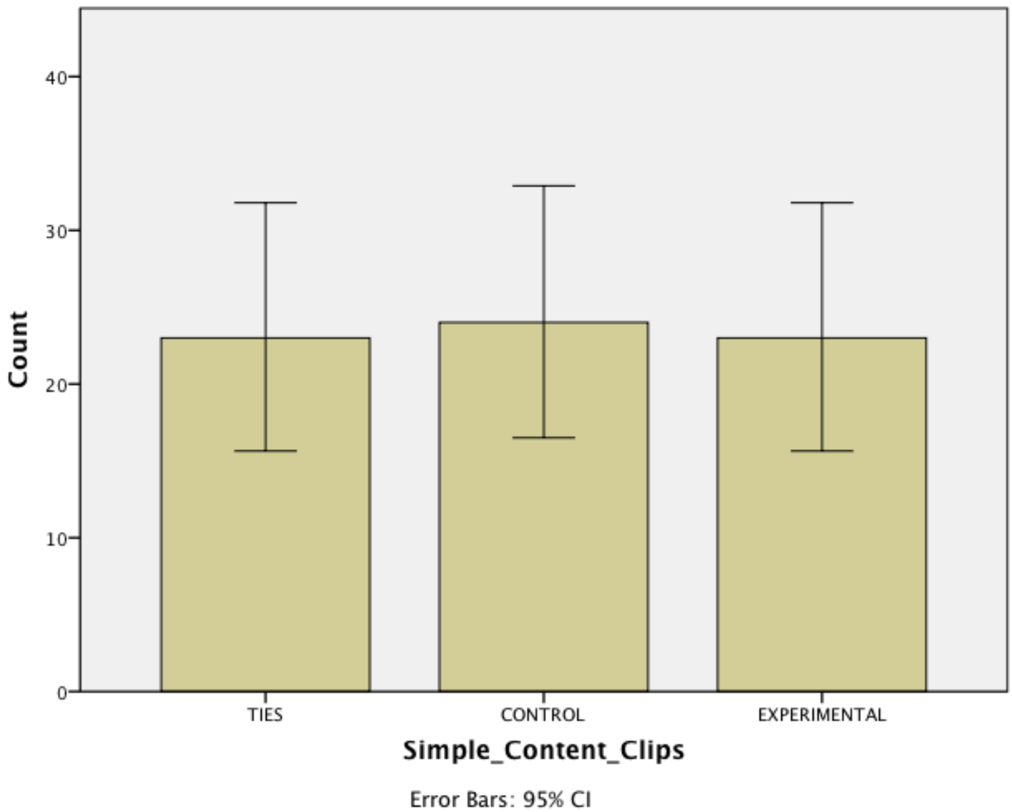



30
20
10
0
Count
TIES
CONTROL
EXPERIMENTAL
Clip_3_R
Error Bars: 95% CI

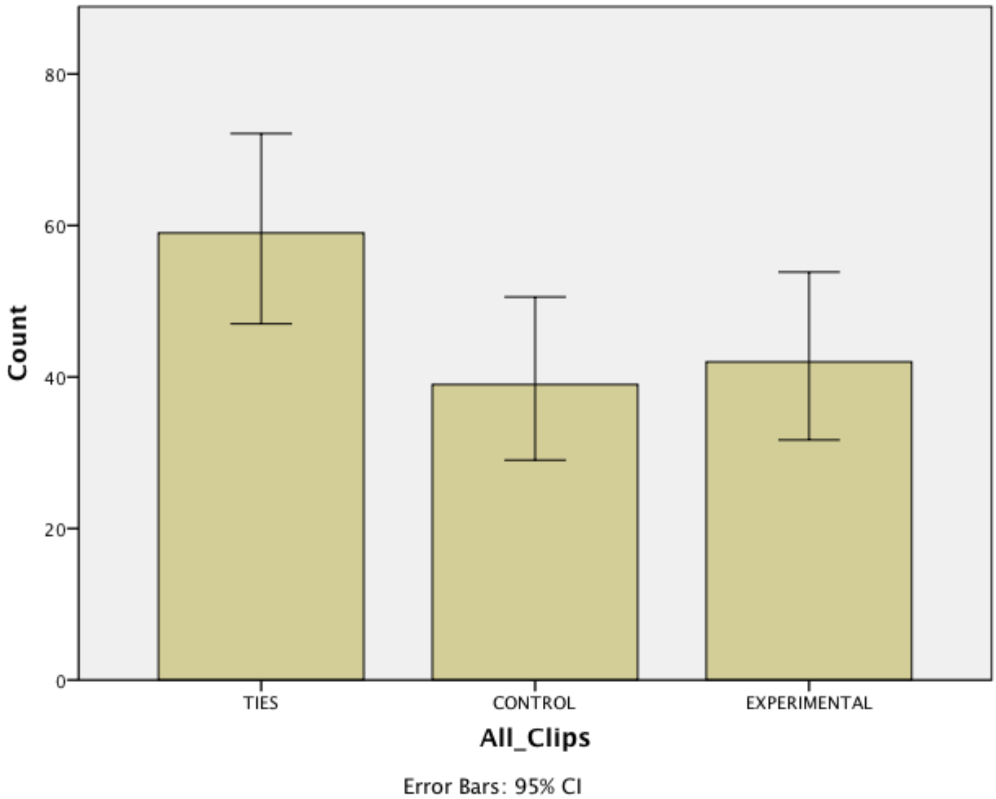

20
15
10
5
0
Count
TIES
CONTROL
EXPERIMENTAL
Clip_4_S
Error Bars: 95% CI



50
40
30
20
10
0
Count
TIES
CONTROL
EXPERIMENTAL
Rich_Content_Clips
Error Bars: 95% CI



40
30
20
10
0
Count
TIES
CONTROL
EXPERIMENTAL
Simple_Content_Clips
Error Bars: 95% CI


80
60
40
20
0
Count
TIES
CONTROL
EXPERIMENTAL
All_Clips
Error Bars: 95% CI

Experiment 2 (LA) - Statistical Data

Descriptive Statistics

	N	Mean	Std. Deviation	Minimum	Maximum
Clip_1_R_C	35	6.4857	1.50238	2.00	9.00
Clip_2_S_C	35	4.7429	1.93030	1.00	8.00
Clip_3_R_C	35	8.3429	1.13611	4.00	9.00
Clip_4_S_C	35	6.0857	1.44245	2.00	8.00
Rich_Content_Control	70	7.4143	1.61956	2.00	9.00
Simple_Content_Control	70	5.4143	1.82171	1.00	8.00
All_Clips_Control	140	6.4143	1.98913	1.00	9.00
Clip_1_R_E	35	6.3143	1.58618	2.00	8.00
Clip_2_S_E	35	5.1143	1.47072	2.00	8.00
Clip_3_R_E	35	8.4286	.88403	5.00	9.00
Clip_4_S_E	35	5.6571	1.51352	3.00	9.00
Rich_Content_Exp	70	7.3714	1.66090	2.00	9.00
Simple_Content_Exp	70	5.3857	1.50644	2.00	9.00
All_Clips_Exp	140	6.3786	1.86782	2.00	9.00

Wilcoxon Signed Ranks Test

Test Statistics[a]

	1_R_E - 1_R_C	2_S_E - 2_S_C	3_R_E - 3_R_C	4_S_E - 4_S_C	Rich_E - Rich_C	Simple_E - Simple_C	All_Clips_E - All_Clips_C
Z	-1.225[b]	-1.671[c]	-.632[c]	-2.146[b]	-.510[b]	-.325[b]	-.552[b]
Asymp. Sig. (2-tailed)	.221	.095	.527	.032	.610	.745	.581
Exact Sig. (2-tailed)	.256	.096	.621	.033	.649	.757	.586
Exact Sig. (1-tailed)	.128	.048	.311	.016	.324	.379	.293
Point Probability	.010	.002	.085	.004	.017	.006	.002

a. Wilcoxon Signed Ranks Test

b. Based on positive ranks.

c. Based on negative ranks.

Ranks

		N	Mean Rank	Sum of Ranks
Clip_1_R_E - Clip_1_R_C	Negative Ranks	11[a]	9.18	101.00
	Positive Ranks	6[b]	8.67	52.00
	Ties	18[c]		
	Total	35		
Clip_2_S_E - Clip_2_S_C	Negative Ranks	8[d]	9.63	77.00
	Positive Ranks	14[e]	12.57	176.00
	Ties	13[f]		
	Total	35		
Clip_3_R_E - Clip_3_R_C	Negative Ranks	6[g]	7.17	43.00
	Positive Ranks	8[h]	7.75	62.00
	Ties	21[i]		
	Total	35		
Clip_4_S_E - Clip_4_S_C	Negative Ranks	14[j]	12.54	175.50
	Positive Ranks	7[k]	7.93	55.50
	Ties	14[l]		
	Total	35		
Rich_Content_E - Rich_Content_C	Negative Ranks	17[m]	16.03	272.50
	Positive Ranks	14[n]	15.96	223.50
	Ties	39[o]		
	Total	70		
Simple_Content_E - Simple_Content_C	Negative Ranks	22[p]	22.68	499.00
	Positive Ranks	21[q]	21.29	447.00
	Ties	27[r]		
	Total	70		
All_Clips_E - All_Clips_C	Negative Ranks	39[s]	38.09	1485.50
	Positive Ranks	35[t]	36.84	1289.50
	Ties	66[u]		
	Total	140		

a. Clip_1_R_E < Clip_1_R_C

b. Clip_1_R_E > Clip_1_R_C

c. Clip_1_R_E = Clip_1_R_C

d. Clip_2_S_E < Clip_2_S_C

e. Clip_2_S_E > Clip_2_S_C

f. Clip_2_S_E = Clip_2_S_C

g. Clip_3_R_E < Clip_3_R_C

h. Clip_3_R_E > Clip_3_R_C

i. Clip_3_R_E = Clip_3_R_C

j. Clip_4_S_E < Clip_4_S_C

k. Clip_4_S_E > Clip_4_S_C

l. Clip_4_S_E = Clip_4_S_C

m. Rich_Content_Exp < Rich_Content_Control

n. Rich_Content_Exp > Rich_Content_Control

o. Rich_Content_Exp = Rich_Content_Control

p. Simple_Content_Exp < Simple_Content_Control

q. Simple_Content_Exp > Simple_Content_Control

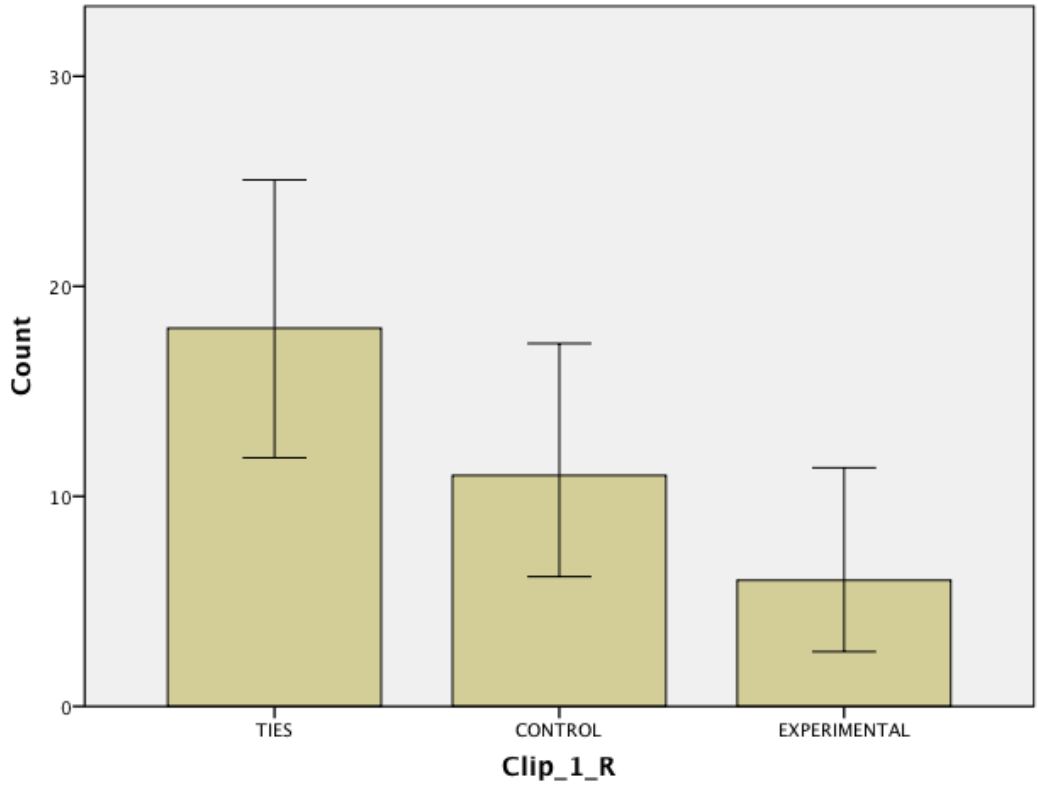
r. Simple_Content_Exp = Simple_Content_Control

s. All_Clips_Exp < All_Clips_Control

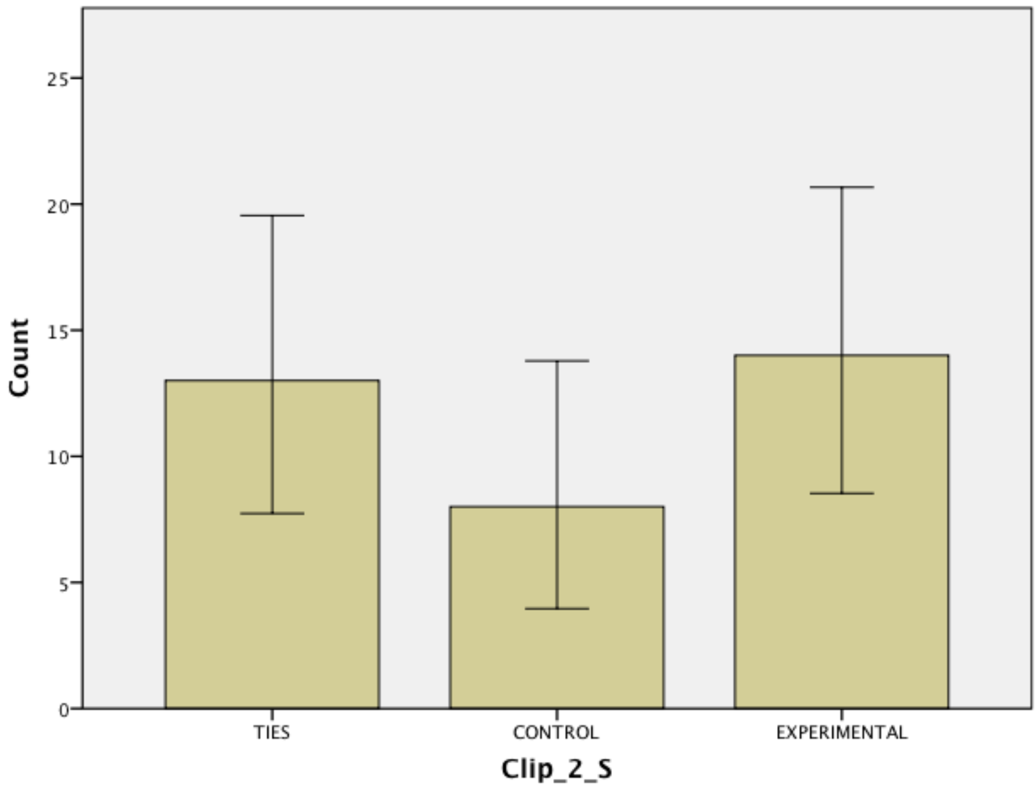
t. All_Clips_Exp > All_Clips_Control

u. All_Clips_Exp = All_Clips_Control

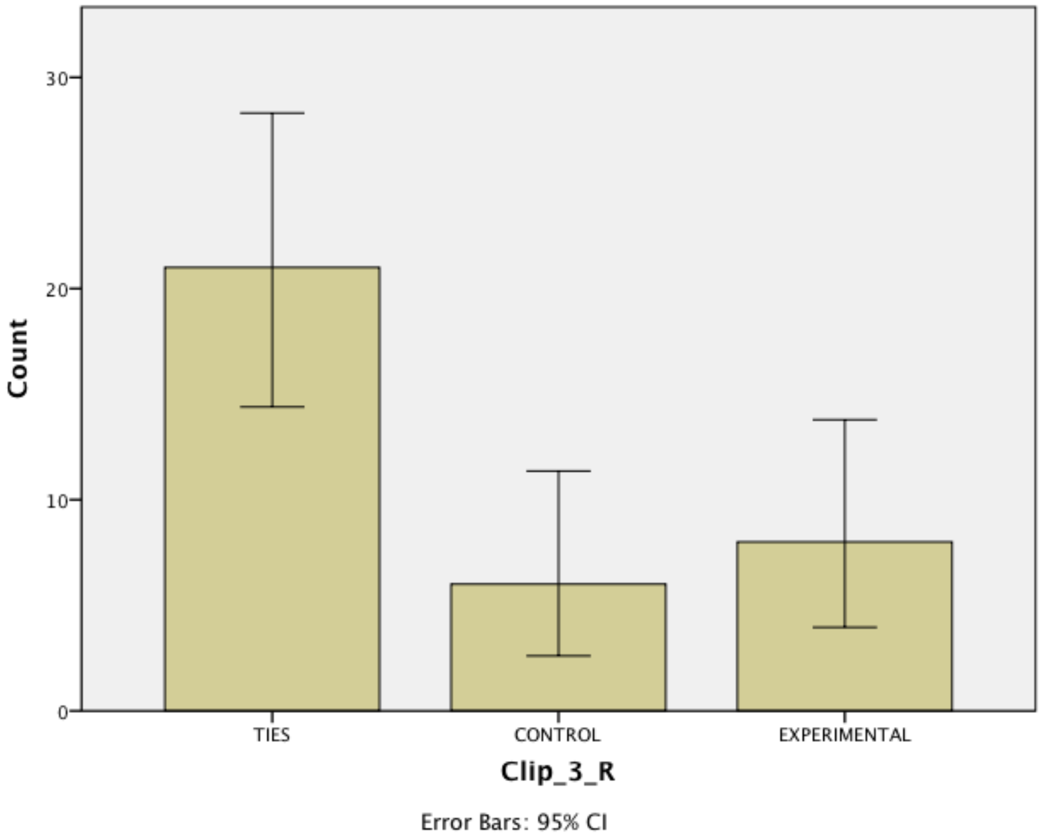
Participant Response Data Graphs

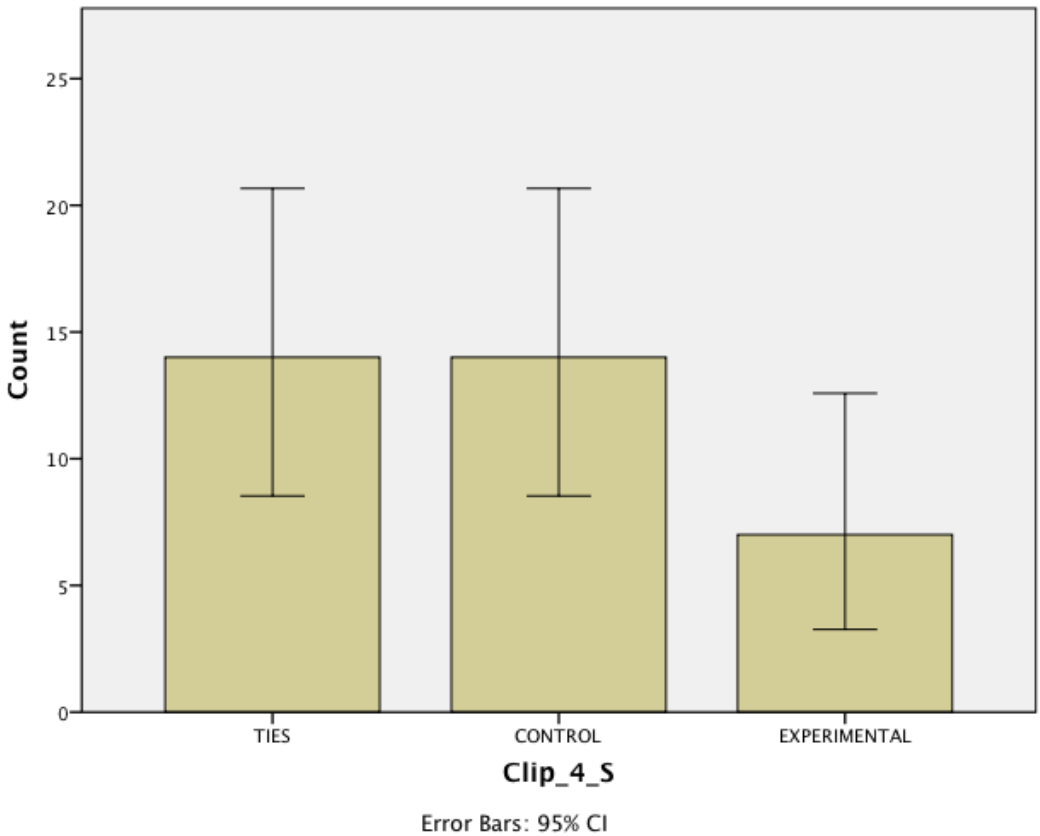


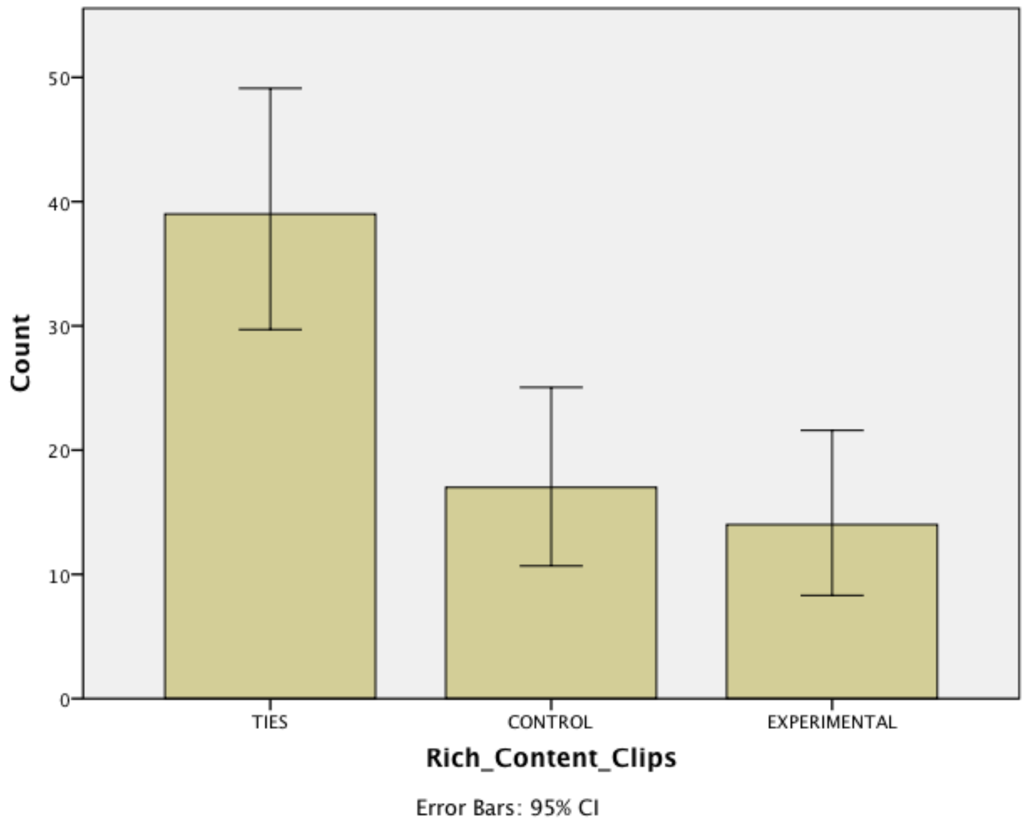
Error Bars: 95% CI

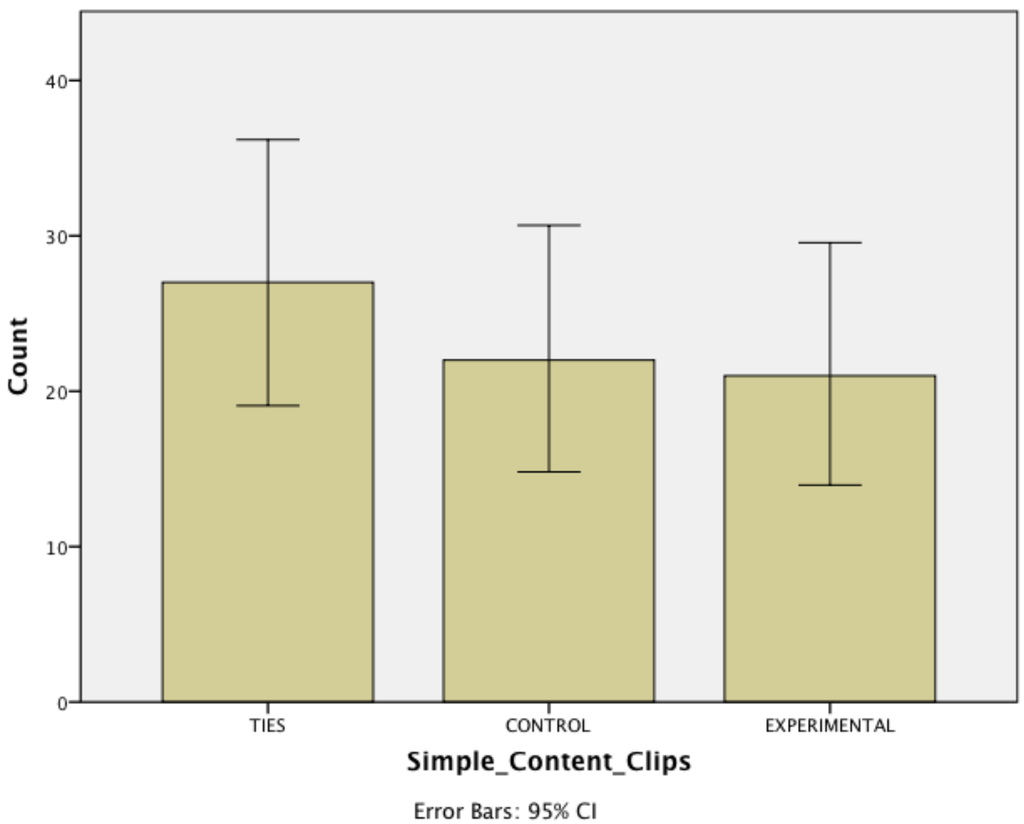


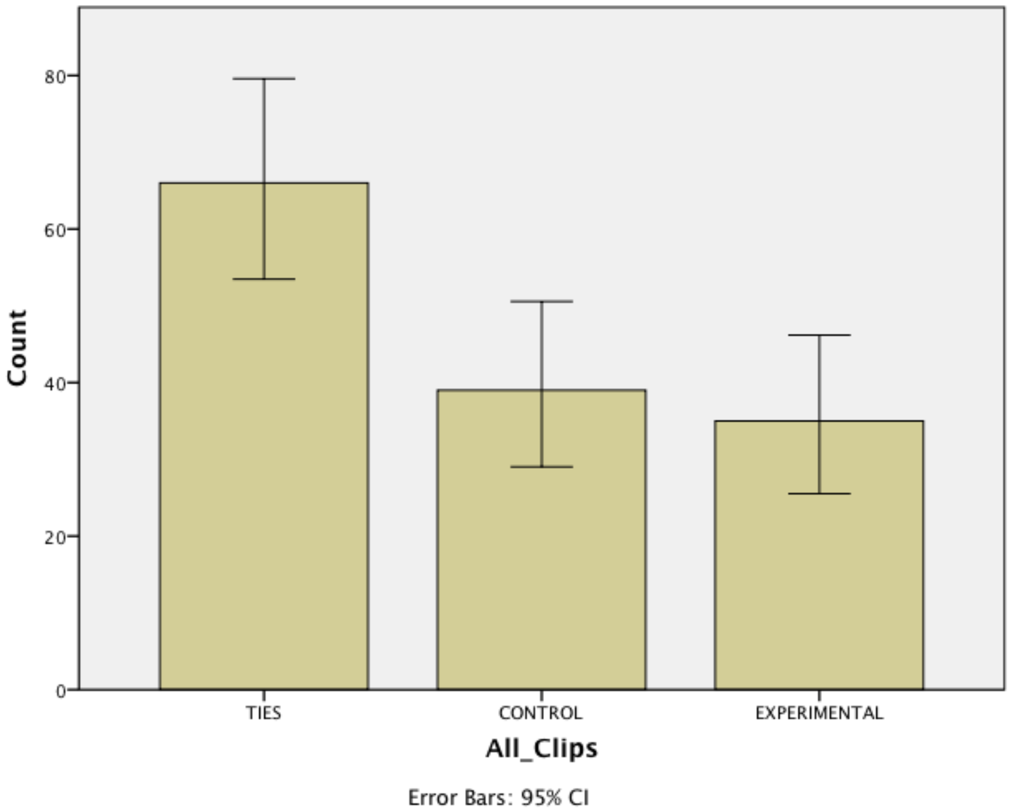
Error Bars: 95% CI


30
20
10
0
Count
TIES
CONTROL
EXPERIMENTAL
Clip_3_R
Error Bars: 95% CI


25
20
15
10
5
0
Count
TIES
CONTROL
EXPERIMENTAL
Clip_4_S
Error Bars: 95% CI


50
40
30
20
10
0
Count
TIES
CONTROL
EXPERIMENTAL
Rich_Content_Clips
Error Bars: 95% CI


40
30
20
10
0
Count
TIES
CONTROL
EXPERIMENTAL
Simple_Content_Clips
Error Bars: 95% CI


80
60
40
20
0
Count
TIES
CONTROL
EXPERIMENTAL
All_Clips
Error Bars: 95% CI

Experiment 3NS (LAHFF) - Statistical Data

Descriptive Statistics

	N	Mean	Std. Deviation	Minimum	Maximum
Clip_1_R_C	35	6.2000	1.62336	3.00	9.00
Clip_2_S_C	35	5.5429	1.44187	3.00	8.00
Clip_3_R_C	35	7.9714	1.04278	6.00	9.00
Clip_4_S_C	35	5.8857	1.43017	3.00	9.00
Rich_Content_Control	70	7.0857	1.62180	3.00	9.00
Simple_Content_Control	70	5.7143	1.43601	3.00	9.00
All_Clips_Control	140	6.4000	1.67418	3.00	9.00
Clip_1_R_E	35	6.2571	1.70368	2.00	9.00
Clip_2_S_E	35	5.6571	1.73108	2.00	9.00
Clip_3_R_E	35	7.8571	1.24009	5.00	9.00
Clip_4_S_E	35	5.7143	1.40527	3.00	9.00
Rich_Content_Exp	70	7.0571	1.68442	2.00	9.00
Simple_Content_Exp	70	5.6857	1.56541	2.00	9.00
All_Clips_Exp	140	6.3714	1.76024	2.00	9.00

Wilcoxon Signed Ranks Test

Test Statistics[a]

	1_R_E - 1_R_C	2_S_E - 2_S_C	3_R_E - 3_R_C	4_S_E - 4_S_C	Rich_E - Rich_C	Simple_E - Simple_C	All_Clips_E - All_Clips_C
Z	-.404[b]	-.683[b]	-.924[c]	-.845[c]	-.175[c]	-.188[c]	-.230[c]
Asymp. Sig. (2-tailed)	.686	.495	.356	.398	.861	.851	.818
Exact Sig. (2-tailed)	.671	.544	.484	.436	.893	.864	.817
Exact Sig. (1-tailed)	.335	.272	.242	.218	.447	.432	.409
Point Probability	.019	.049	.059	.031	.008	.016	.003

a. Wilcoxon Signed Ranks Test

b. Based on negative ranks.

c. Based on positive ranks.

Ranks

		N	Mean Rank	Sum of Ranks
Clip_1_R_E - Clip_1_R_C	Negative Ranks	12[a]	10.46	125.50
	Positive Ranks	11[b]	13.68	150.50
	Ties	12[c]		
	Total	35		
Clip_2_S_E - Clip_2_S_C	Negative Ranks	9[d]	9.72	87.50
	Positive Ranks	11[e]	11.14	122.50
	Ties	15[f]		
	Total	35		
Clip_3_R_E - Clip_3_R_C	Negative Ranks	8[g]	6.25	50.00
	Positive Ranks	4[h]	7.00	28.00
	Ties	23[i]		
	Total	35		
Clip_4_S_E - Clip_4_S_C	Negative Ranks	12[j]	10.58	127.00
	Positive Ranks	8[k]	10.38	83.00
	Ties	15[l]		
	Total	35		
Rich_Content_E - Rich_Content_C	Negative Ranks	20[m]	16.25	325.00
	Positive Ranks	15[n]	20.33	305.00
	Ties	35[o]		
	Total	70		
Simple_Content_E - Simple_Content_C	Negative Ranks	21[p]	20.17	423.50
	Positive Ranks	19[q]	20.87	396.50
	Ties	30[r]		
	Total	70		
All_Clips_E - All_Clips_C	Negative Ranks	41[s]	35.77	1466.50
	Positive Ranks	34[t]	40.69	1383.50
	Ties	65[u]		
	Total	140		

a. Clip_1_R_E < Clip_1_R_C
b. Clip_1_R_E > Clip_1_R_C
c. Clip_1_R_E = Clip_1_R_C
d. Clip_2_S_E < Clip_2_S_C
e. Clip_2_S_E > Clip_2_S_C
f. Clip_2_S_E = Clip_2_S_C
g. Clip_3_R_E < Clip_3_R_C
h. Clip_3_R_E > Clip_3_R_C
i. Clip_3_R_E = Clip_3_R_C
j. Clip_4_S_E < Clip_4_S_C
k. Clip_4_S_E > Clip_4_S_C

l. Clip_4_S_E = Clip_4_S_C

m. Rich_Content_Exp < Rich_Content_Control

n. Rich_Content_Exp > Rich_Content_Control

o. Rich_Content_Exp = Rich_Content_Control

p. Simple_Content_Exp < Simple_Content_Control

q. Simple_Content_Exp > Simple_Content_Control

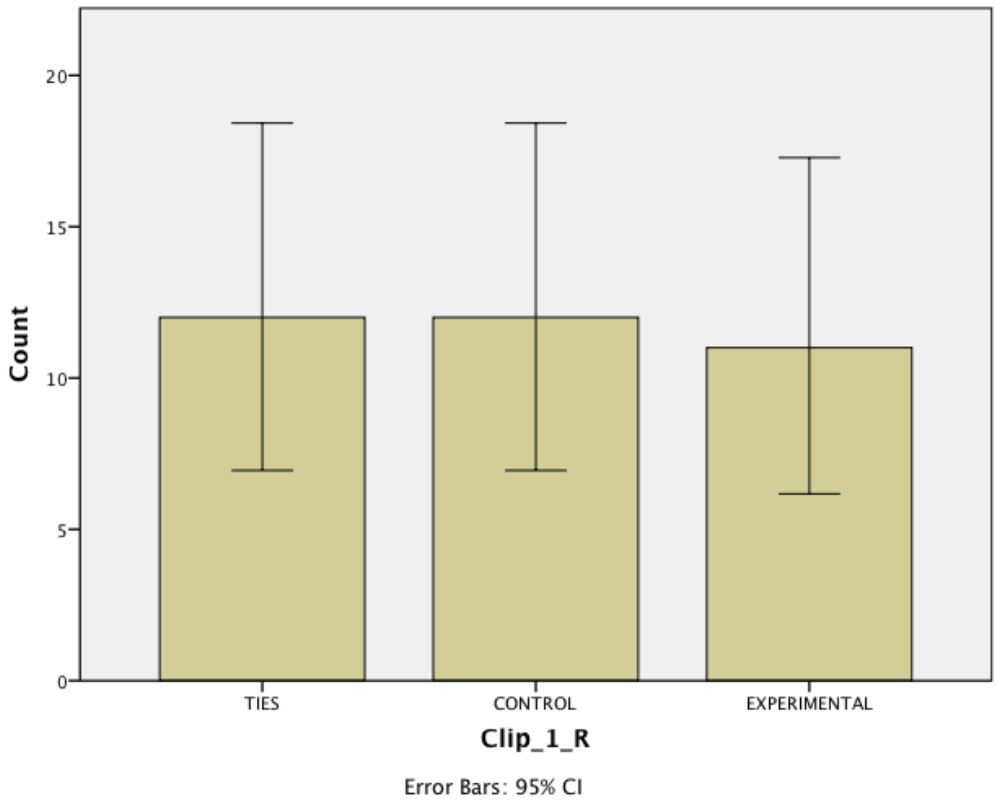
r. Simple_Content_Exp = Simple_Content_Control

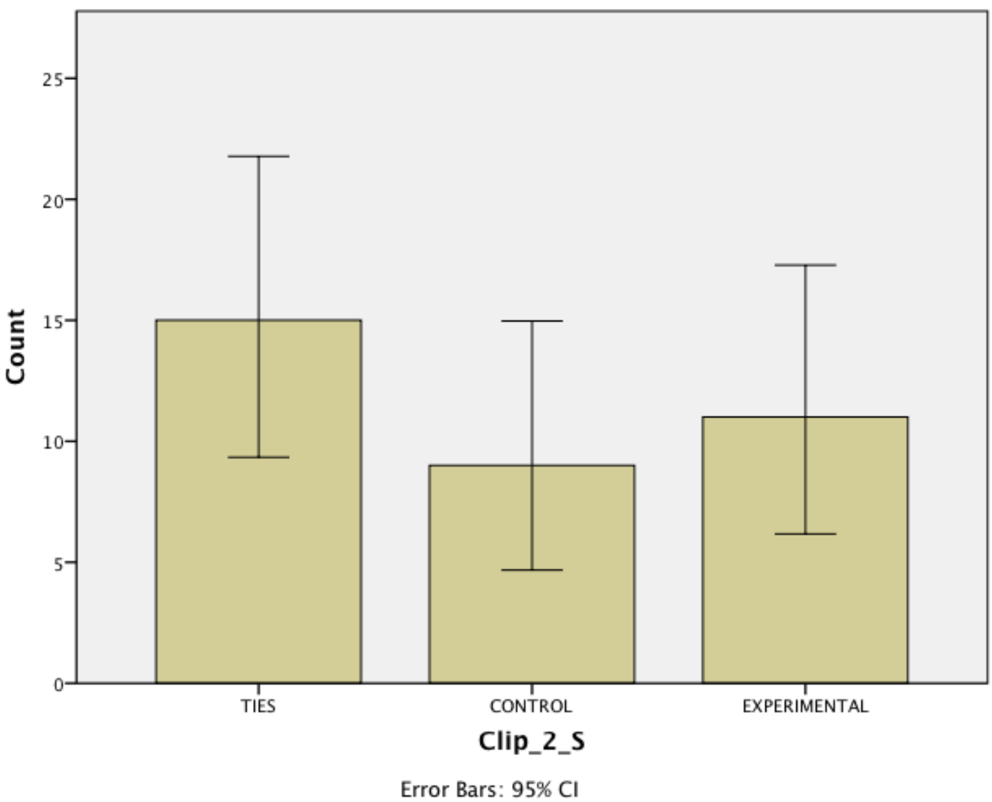
s. All_Clips_Exp < All_Clips_Control

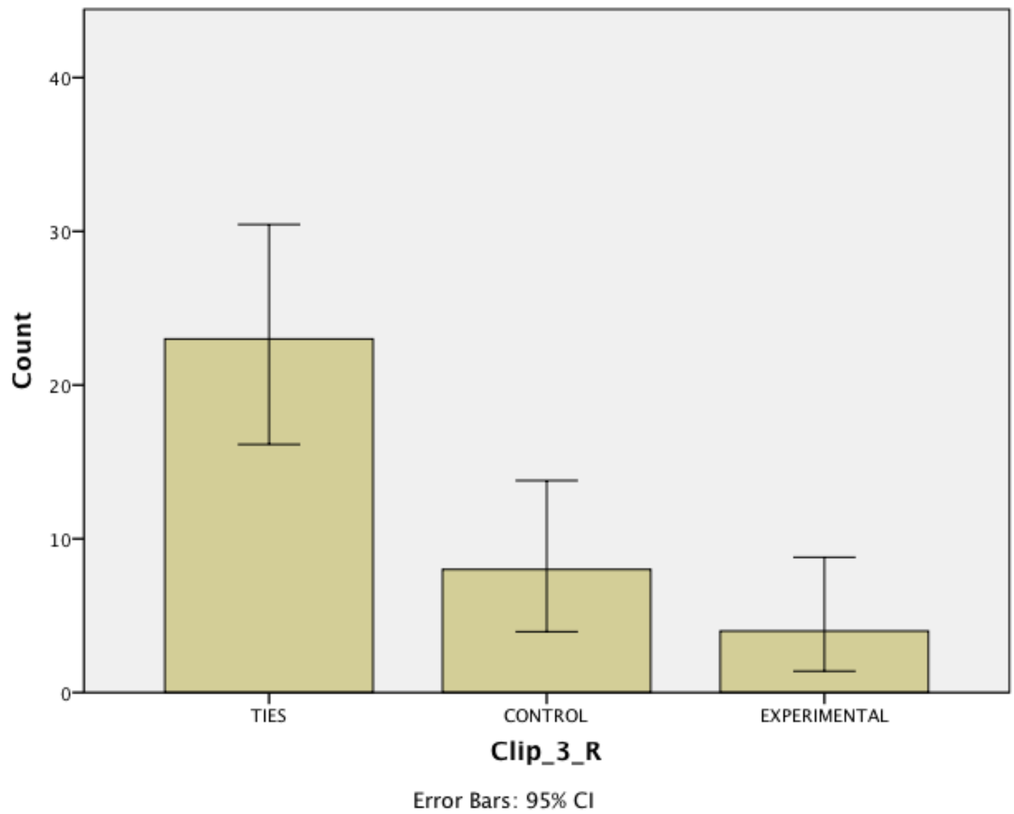
t. All_Clips_Exp > All_Clips_Control

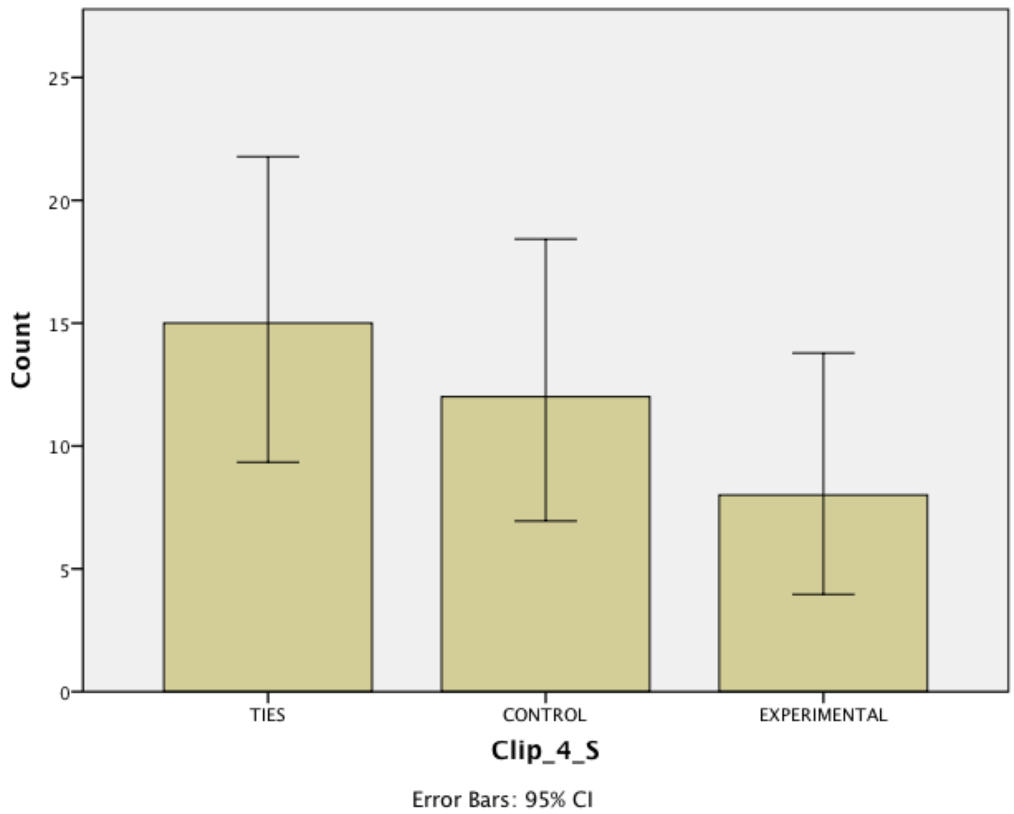
u. All_Clips_Exp = All_Clips_Control

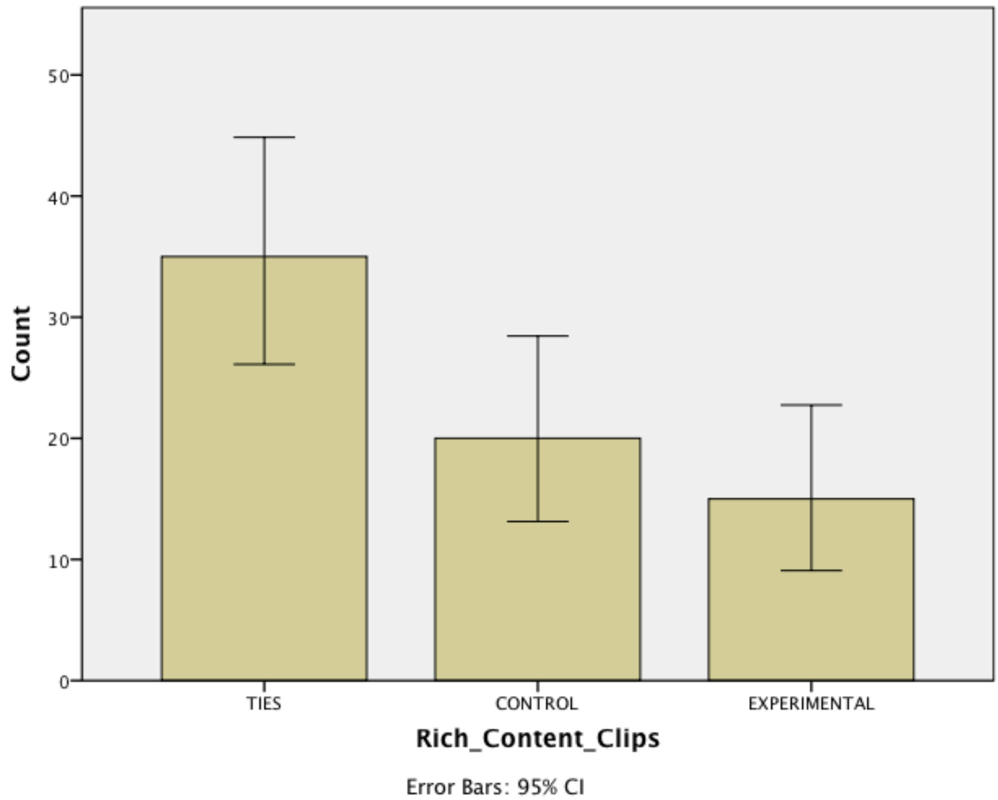
Participant Response Data Graphs

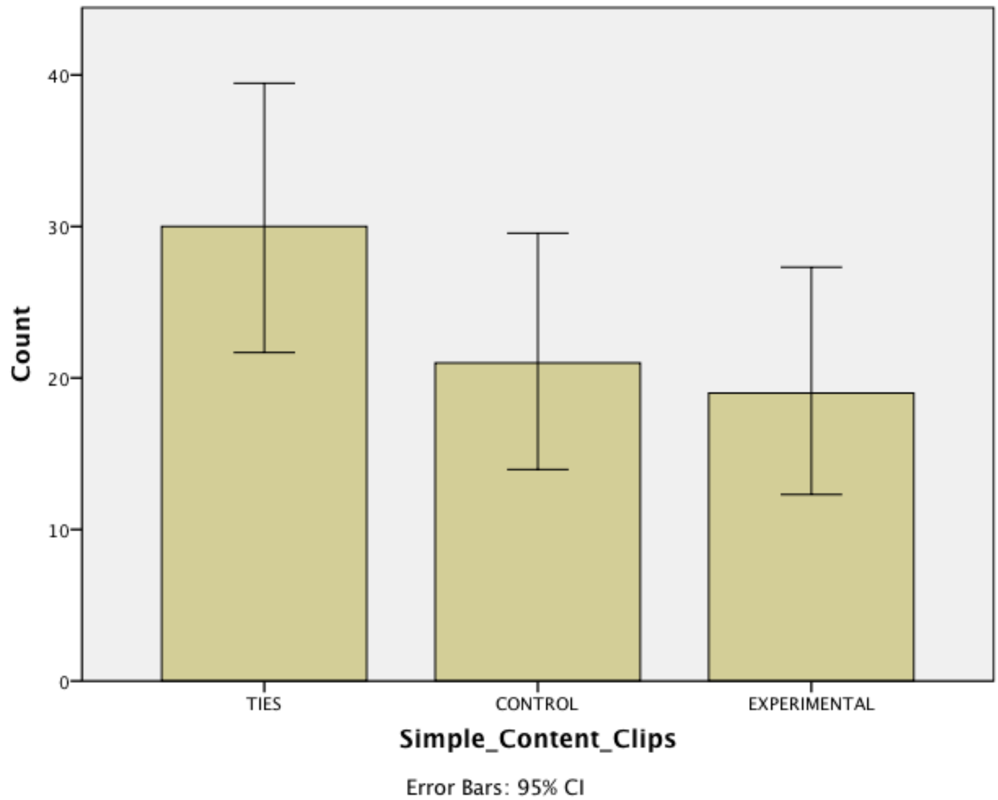


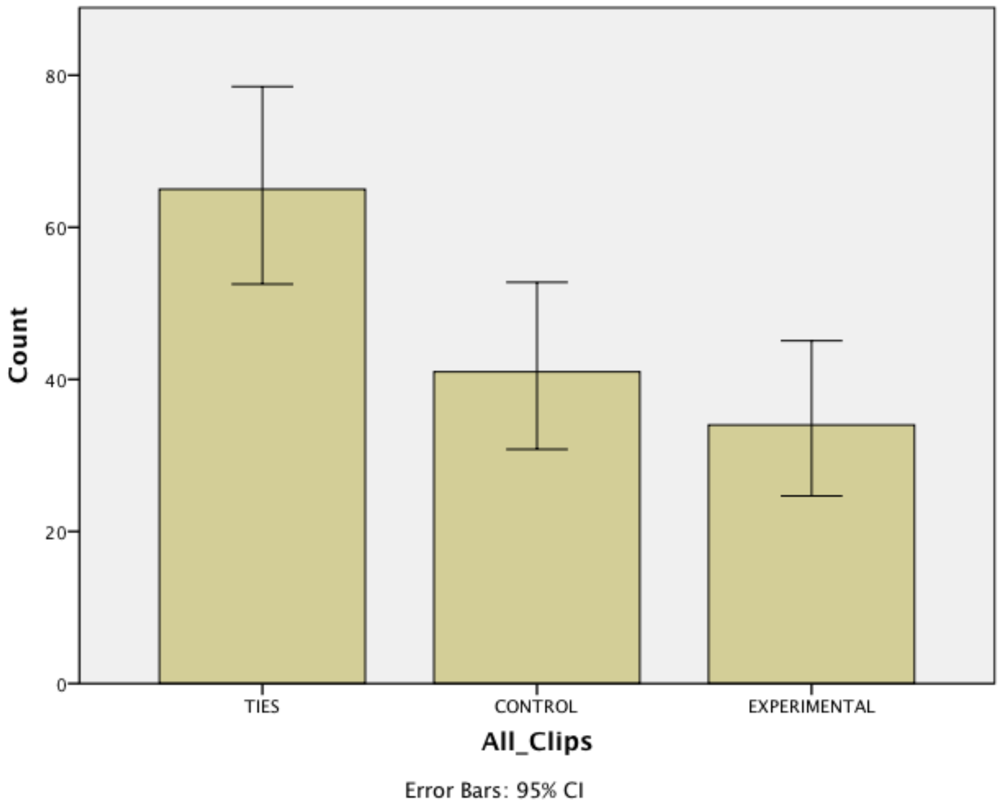



40
30
20
10
0
Count
TIES
CONTROL
EXPERIMENTAL
Clip_3_R
Error Bars: 95% CI


25
20
15
10
5
0
Count
TIES
CONTROL
EXPERIMENTAL
Clip_4_S
Error Bars: 95% CI


Count
50
40
30
20
10
0
TIES
CONTROL
EXPERIMENTAL
Rich_Content_Clips
Error Bars: 95% CI


Count
40
30
20
10
0
TIES
CONTROL
EXPERIMENTAL
Simple_Content_Clips
Error Bars: 95% CI


80
60
40
20
0
Count
TIES
CONTROL
EXPERIMENTAL
All_Clips
Error Bars: 95% CI

Experiment 3Sa (LAHFF) - Statistical Data

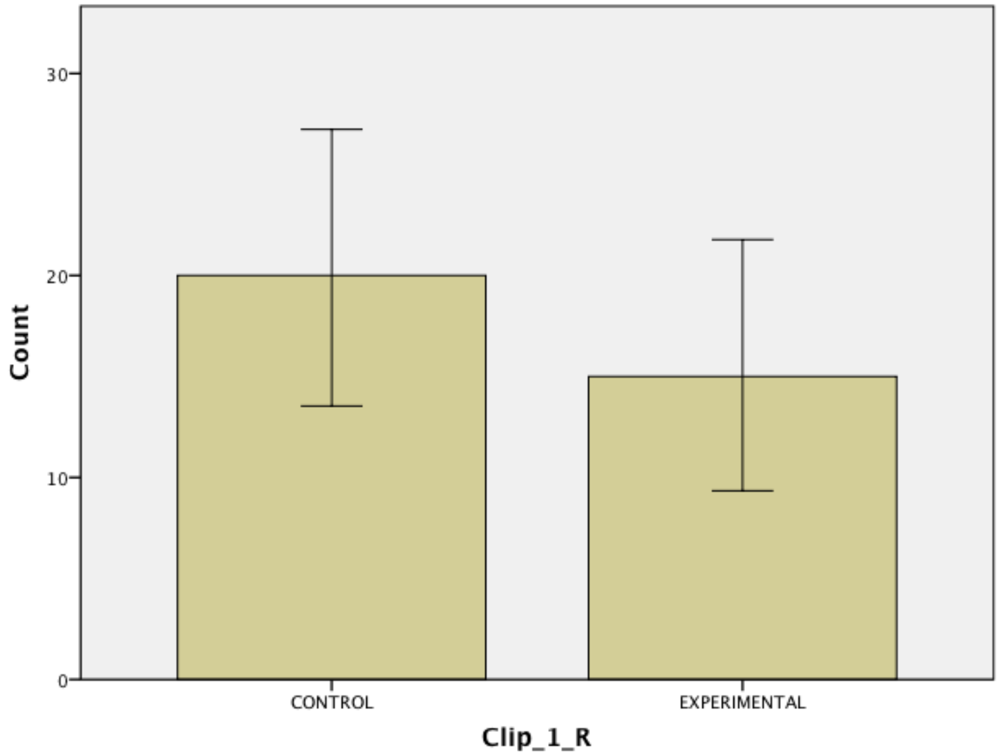
Descriptive Statistics

	N	Mean	Std. Deviation	Minimum	Maximum
Clip_1_R	35	.43	.502	0	1
Clip_2_S	35	.34	.482	0	1
Clip_3_R	35	.49	.507	0	1
Clip_4_S	35	.29	.458	0	1
Rich_Content	70	.46	.502	0	1
Simple_Content	70	.31	.468	0	1
All_Clips	140	.39	.489	0	1

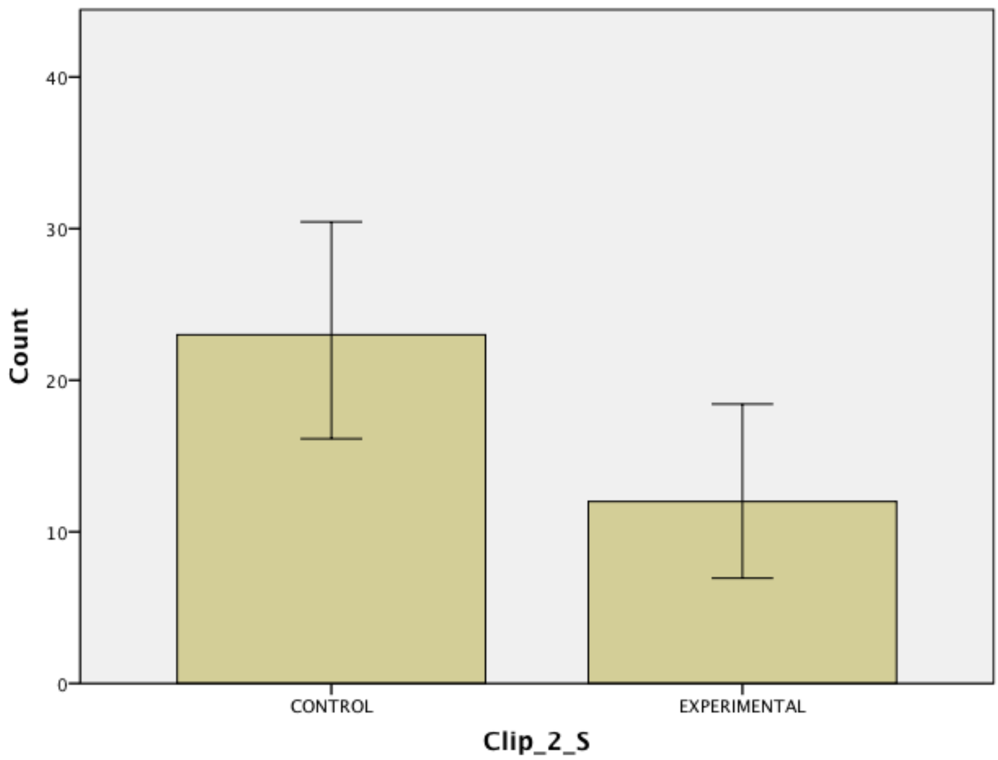
Binomial Test

		Category	N	Observed Prop	Test Prop	Exact Sig. (2-tailed)	Est. Sig. (1-tailed)
Clip_1_R	Group 1	EXP	15	.43	.50	.500	.250
	Group 2	CTRL	20	.57			
	Total		35	1.00			
Clip_2_S	Group 1	EXP	12	.34	.50	.090	.045
	Group 2	CTRL	23	.66			
	Total		35	1.00			
Clip_3_R	Group 1	EXP	17	.49	.50	1.000	.500
	Group 2	CTRL	18	.51			
	Total		35	1.00			
Clip_4_S	Group 1	EXP	10	.29	.50	.017	.008
	Group 2	CTRL	25	.71			
	Total		35	1.00			
Rich_Content	Group 1	EXP	32	.46	.50	.550	.280
	Group 2	CTRL	38	.54			
	Total		70	1.00			
Simple_Content	Group 1	EXP	22	.31	.50	.003	.001
	Group 2	CTRL	48	.69			
	Total		70	1.00			
All_Clips	Group 1	EXP	54	.39	.50	.009	.005
	Group 2	CTRL	86	.61			
	Total		140	1.00			

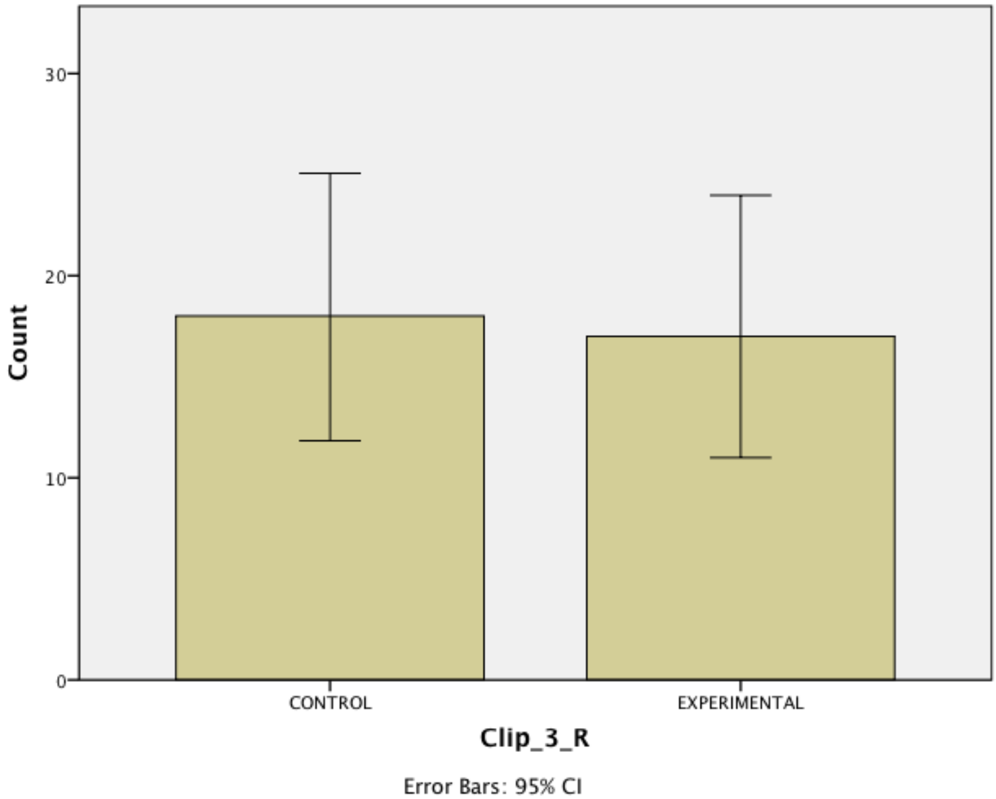
Participant Response Data Graphs

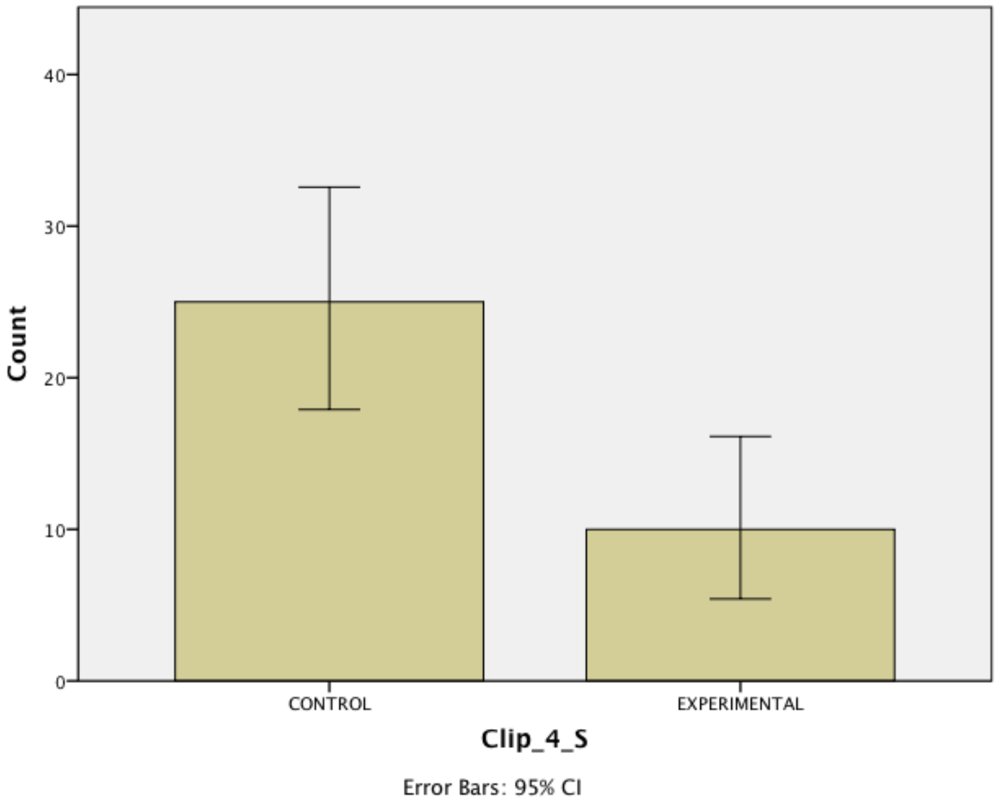


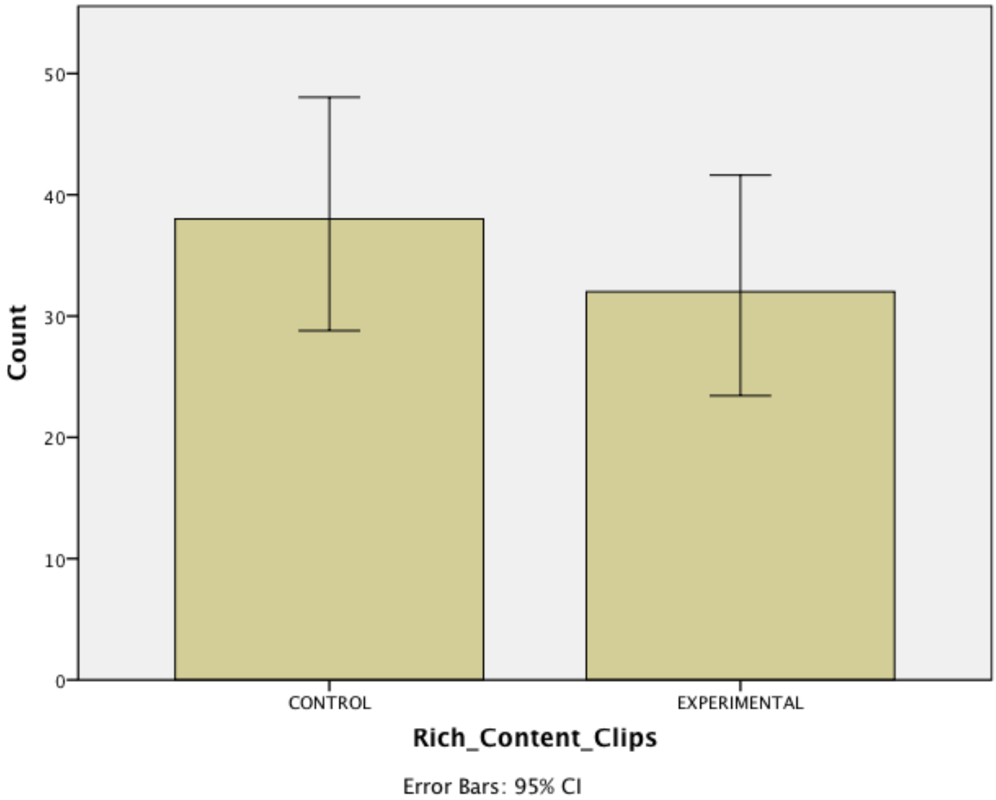
Error Bars: 95% CI

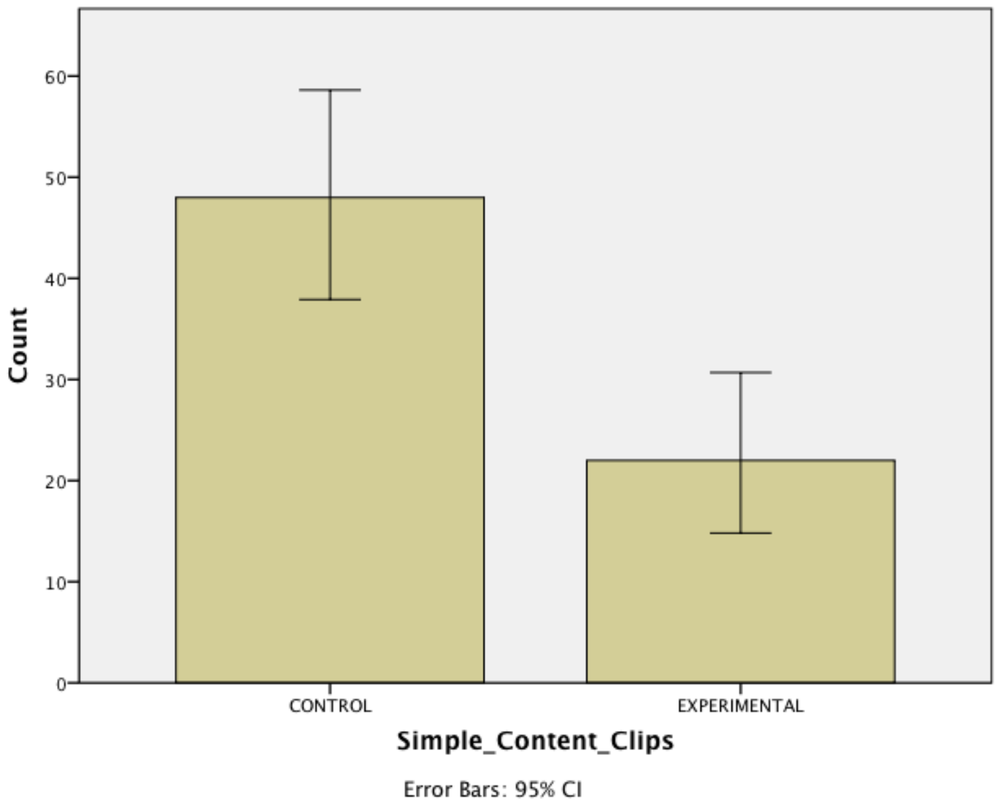


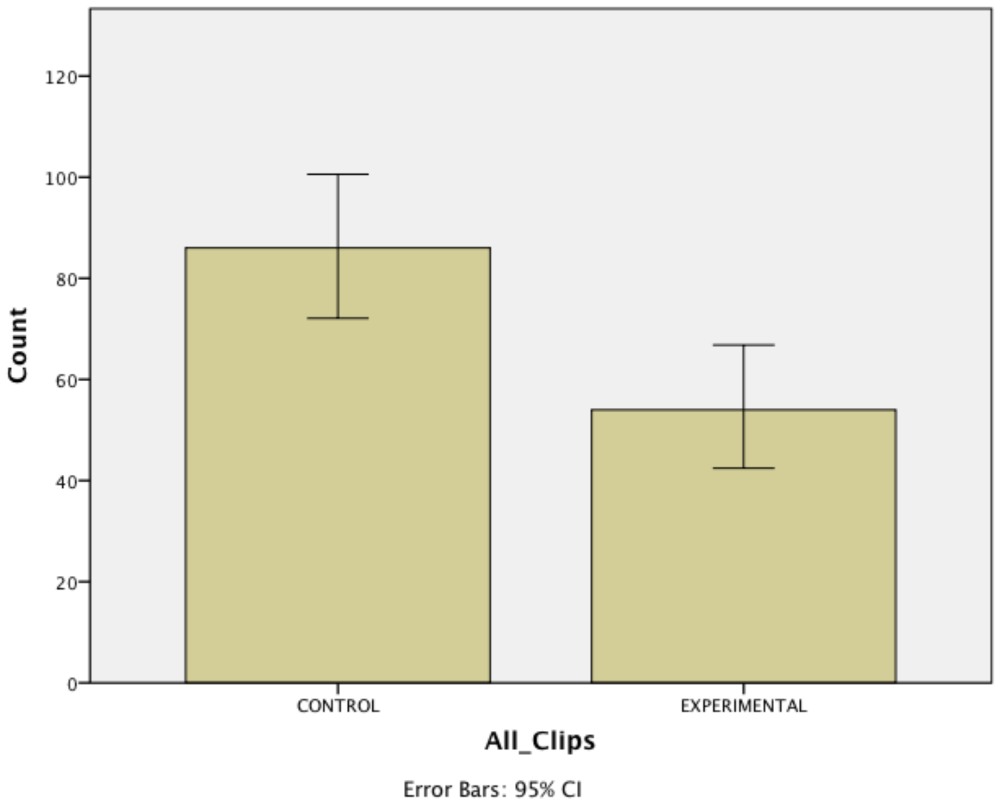
Error Bars: 95% CI


30
20
10
0
Count
CONTROL
EXPERIMENTAL
Clip_3_R
Error Bars: 95% CI


40
30
20
10
0
Count
CONTROL
EXPERIMENTAL
Clip_4_S
Error Bars: 95% CI


50
40
30
20
10
0
Count
CONTROL
EXPERIMENTAL
Rich_Content_Clips
Error Bars: 95% CI


60
50
40
30
20
10
0
Count
CONTROL
EXPERIMENTAL
Simple_Content_Clips
Error Bars: 95% CI


120
100
80
60
40
20
0
Count
CONTROL
EXPERIMENTAL
All_Clips
Error Bars: 95% CI

Experiment 3Sb (LAHFF) - Statistical Data

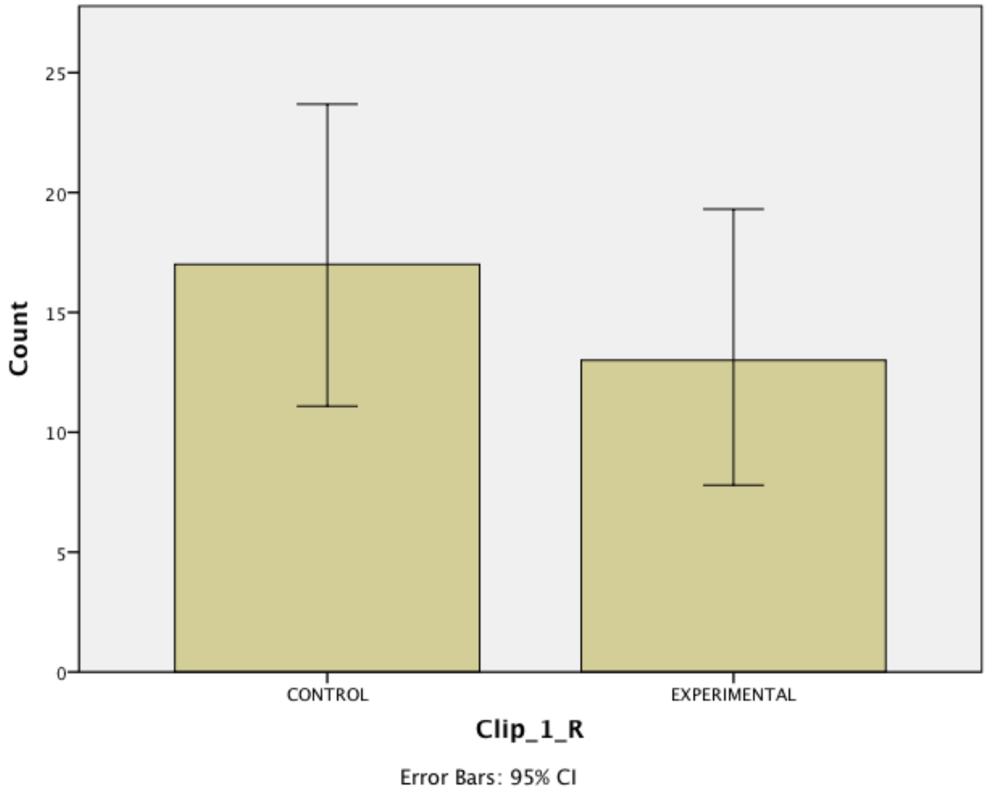
Descriptive Statistics

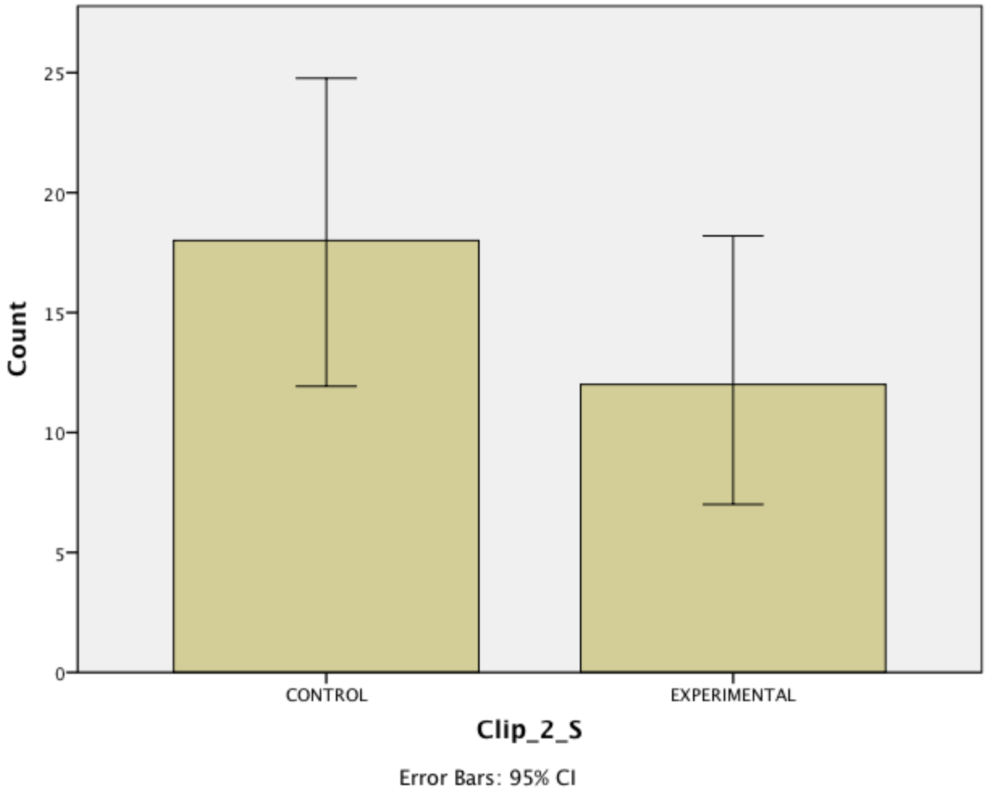
	N	Mean	Std. Deviation	Minimum	Maximum
Clip_1_R	30	.43	.504	0	1
Clip_2_S	30	.40	.498	0	1
Clip_3_R	30	.40	.498	0	1
Clip_4_S	30	.37	.490	0	1
Rich_Content	60	.42	.497	0	1
Simple_Content	60	.38	.490	0	1
All_Clips	120	.40	.492	0	1

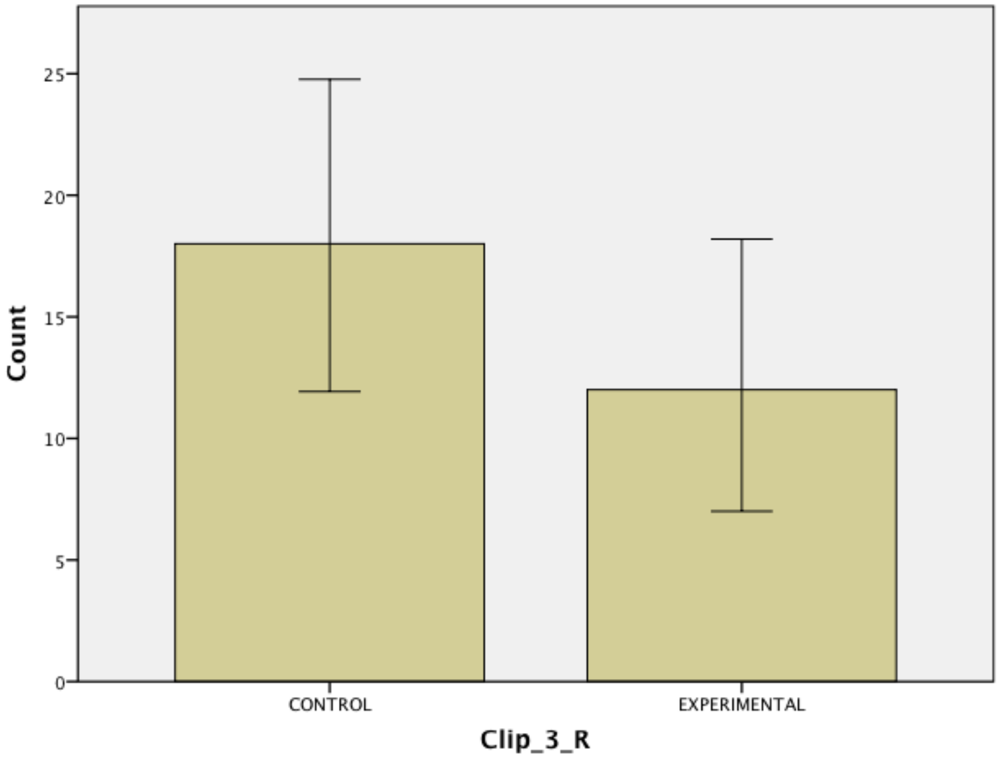
Binomial Test

		Category	N	Observed Prop	Test Prop	Exact Sig. (2-tailed)	Est. Sig. (1-tailed)
Clip_1_R	Group 1	EXP	13	.43	.50	.585	.293
	Group 2	CTRL	17	.57			
	Total		30	1.00			
Clip_2_S	Group 1	EXP	12	.40	.50	.362	.181
	Group 2	CTRL	18	.60			
	Total		30	1.00			
Clip_3_R	Group 1	EXP	18	.60	.50	.362	.181
	Group 2	CTRL	12	.40			
	Total		30	1.00			
Clip_4_S	Group 1	EXP	19	.63	.50	.200	.100
	Group 2	CTRL	11	.37			
	Total		30	1.00			
Rich_Content	Group 1	EXP	25	.42	.50	.245	.122
	Group 2	CTRL	35	.58			
	Total		60	1.00			
Simple_Content	Group 1	EXP	23	.38	.50	.092	.046
	Group 2	CTRL	37	.62			
	Total		60	1.00			
All_Clips	Group 1	EXP	48	.40	.50	.035	.017
	Group 2	CTRL	72	.60			
	Total		120	1.00			

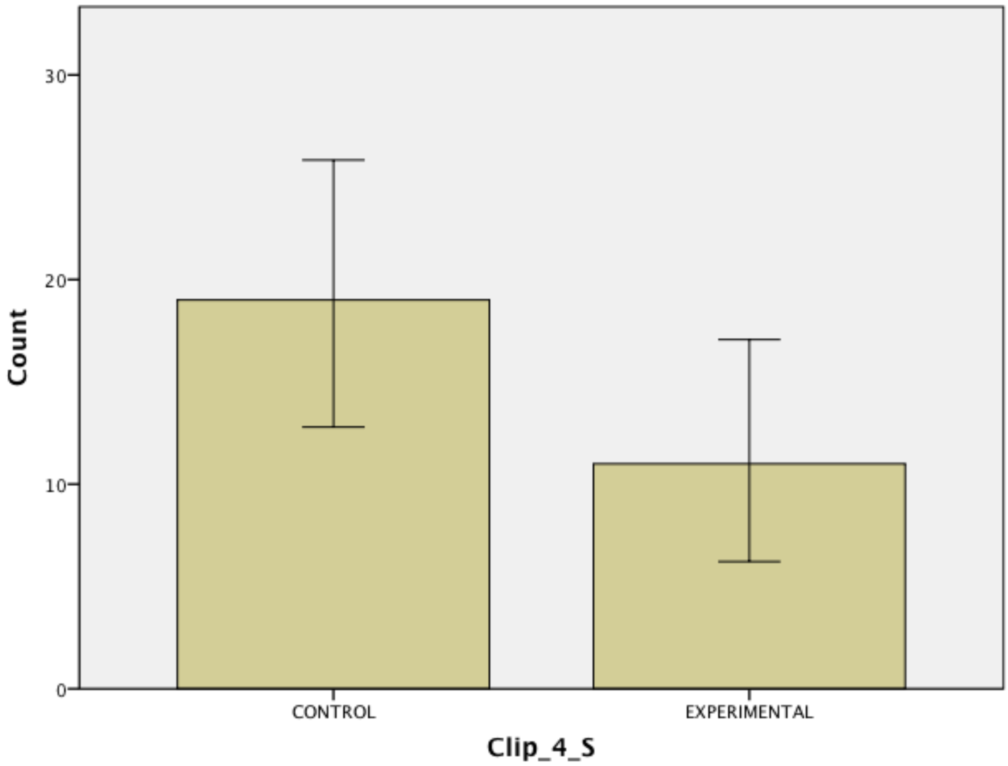
Participant Response Data Graphs



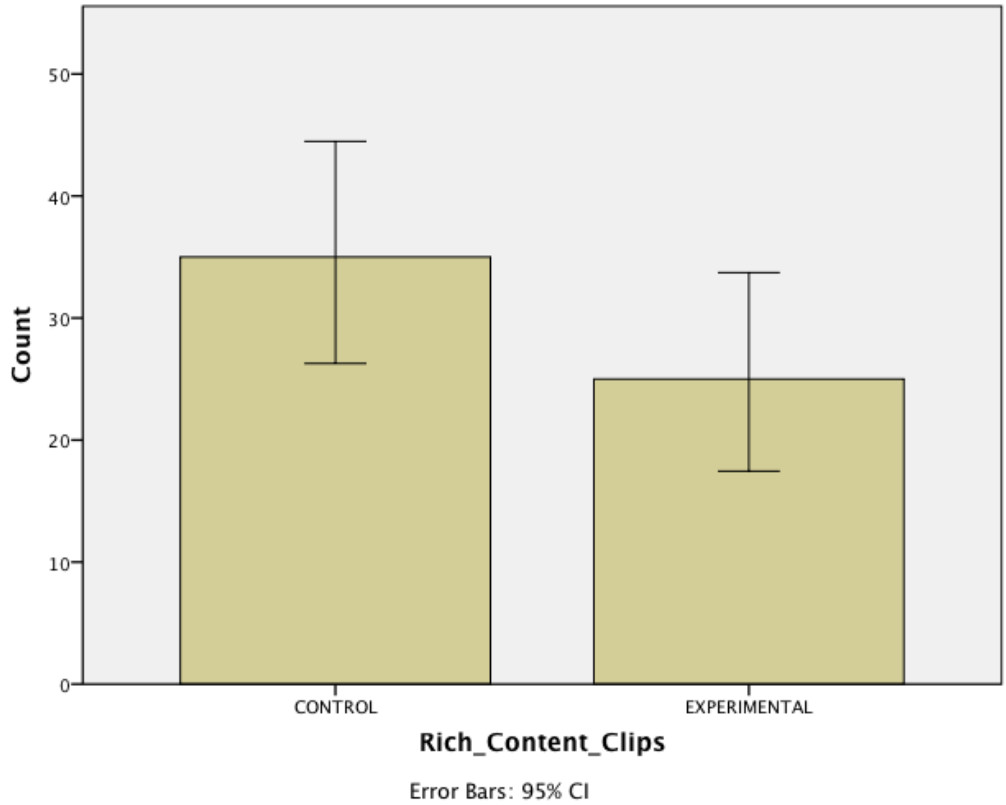


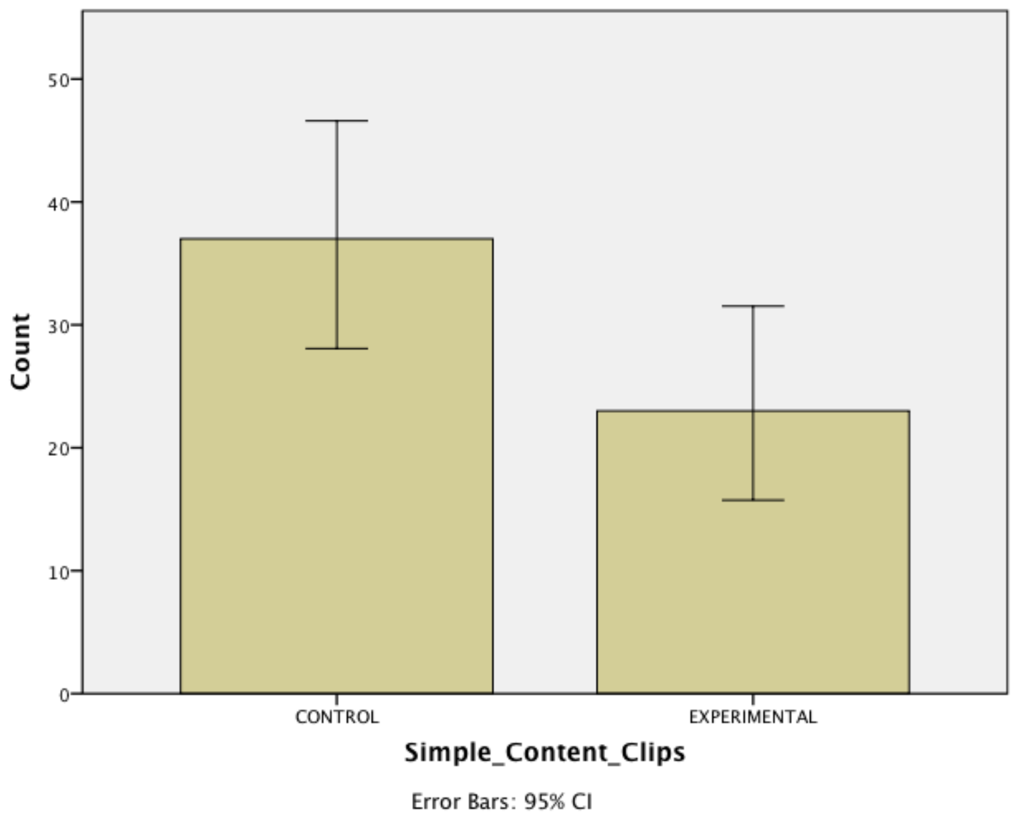


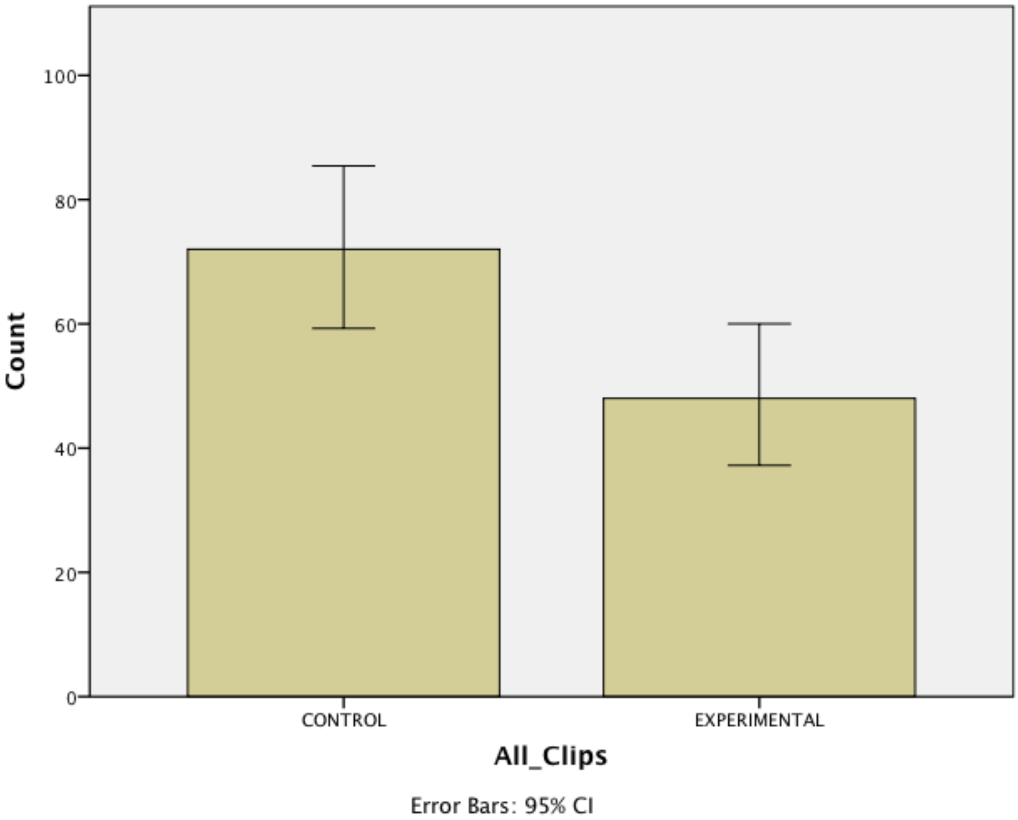
Error Bars: 95% CI



Error Bars: 95% CI


Count
0
10
20
30
40
50
CONTROL
EXPERIMENTAL
Rich_Content_Clips
Error Bars: 95% CI


Count
0
10
20
30
40
50
CONTROL
EXPERIMENTAL
Simple_Content_Clips
Error Bars: 95% CI



100
80
60
40
20
0
Count
CONTROL
EXPERIMENTAL
All_Clips
Error Bars: 95% CI

Experiment 4 (HFF) - Statistical Data

Descriptive Statistics

	N	Mean	Std. Deviation	Minimum	Maximum
Clip_1_C	35	6.9714	1.29446	3.00	9.00
Clip_2_C	35	6.9143	1.73835	3.00	9.00
Clip_3_C	35	7.5143	1.37993	4.00	9.00
Clip_4_C	35	6.4571	1.63316	2.00	9.00
All_Clips_Control	140	6.9643	1.55203	2.00	9.00
Clip_1_E	35	7.0857	1.33662	3.00	9.00
Clip_2_E	35	6.8000	1.62336	3.00	9.00
Clip_3_E	35	7.7429	1.33599	4.00	9.00
Clip_4_E	35	6.5143	1.54104	2.00	9.00
All_Clips_Exp	140	7.0357	1.51923	2.00	9.00

Wilcoxon Signed Ranks Test

Test Statistics[a]

	1_E - 1_C	2_E - 2_C	3_E - 3_C	4_E - 4_C	All_Clips_E - All_Clips_C
Z	-.722[b]	-.919[c]	-.684[b]	-.238[b]	-.375[b]
Asymp. Sig. (2-tailed)	.471	.358	.494	.812	.708
Exact Sig. (2-tailed)	.477	.387	.512	.854	.711
Exact Sig. (1-tailed)	.239	.193	.256	.427	.356
Point Probability	.006	.019	.012	.042	.001

a. Wilcoxon Signed Ranks Test

b. Based on negative ranks.

c. Based on positive ranks.

Ranks

		N	Mean Rank	Sum of Ranks
Clip_1_E - Clip_1_C	Negative Ranks	8[a]	11.94	95.50
	Positive Ranks	13[b]	10.42	135.50
	Ties	14[c]		
	Total	35		
Clip_2_E - Clip_2_C	Negative Ranks	12[d]	11.75	141.00
	Positive Ranks	9[e]	10.00	90.00
	Ties	14[f]		
	Total	35		
Clip_3_E - Clip_3_C	Negative Ranks	9[g]	7.83	70.50
	Positive Ranks	9[h]	11.17	100.50
	Ties	17[i]		
	Total	35		
Clip_4_E - Clip_4_C	Negative Ranks	7[j]	8.00	56.00
	Positive Ranks	8[k]	8.00	64.00
	Ties	20[l]		
	Total	35		
All_Clips_E - All_Clips_C	Negative Ranks	36[m]	37.69	1357.00
	Positive Ranks	39[n]	38.28	1493.00
	Ties	65[o]		
	Total	140		

a. Clip_1_R_E < Clip_1_R_C

b. Clip_1_R_E > Clip_1_R_C

c. Clip_1_R_E = Clip_1_R_C

d. Clip_2_S_E < Clip_2_S_C

e. Clip_2_S_E > Clip_2_S_C

f. Clip_2_S_E = Clip_2_S_C

g. Clip_3_R_E < Clip_3_R_C

h. Clip_3_R_E > Clip_3_R_C

i. Clip_3_R_E = Clip_3_R_C

j. Clip_4_S_E < Clip_4_S_C

k. Clip_4_S_E > Clip_4_S_C

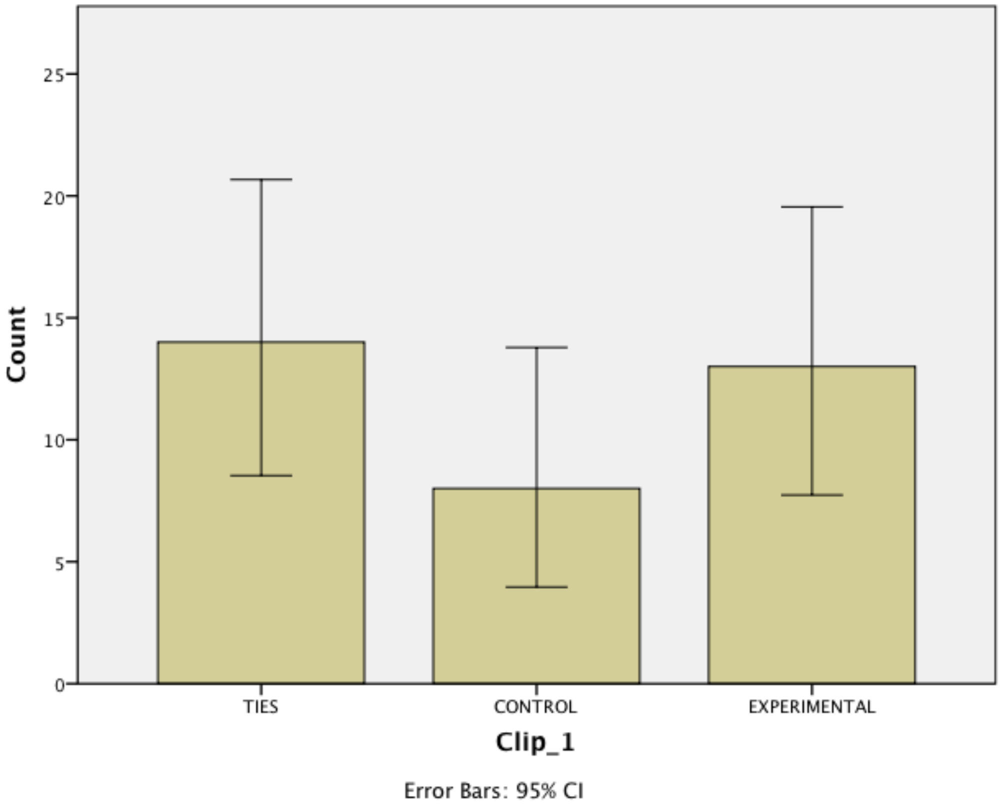
l. Clip_4_S_E = Clip_4_S_C

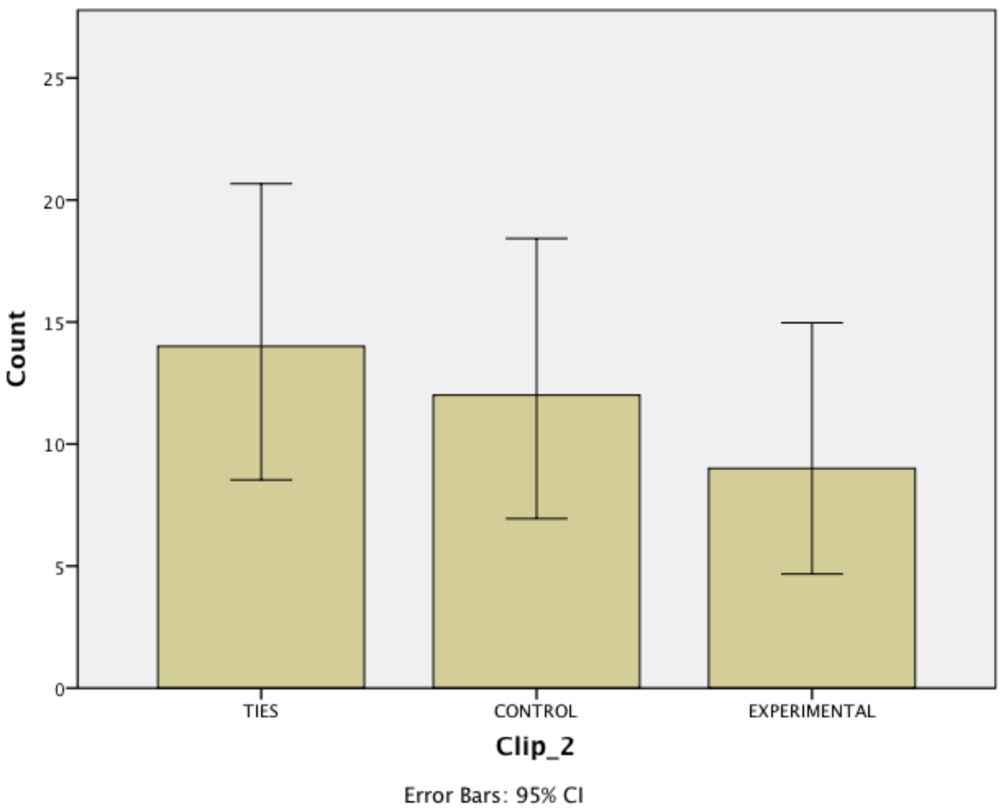
m. All_Clips_Exp < All_Clips_Control

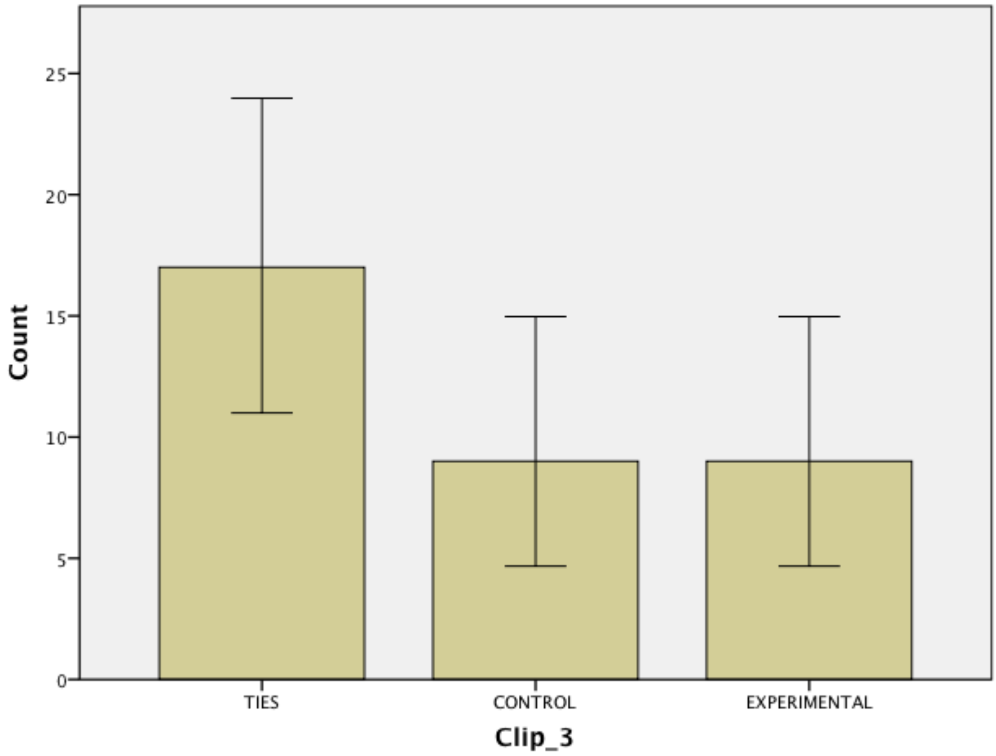
n. All_Clips_Exp > All_Clips_Control

o. All_Clips_Exp = All_Clips_Control

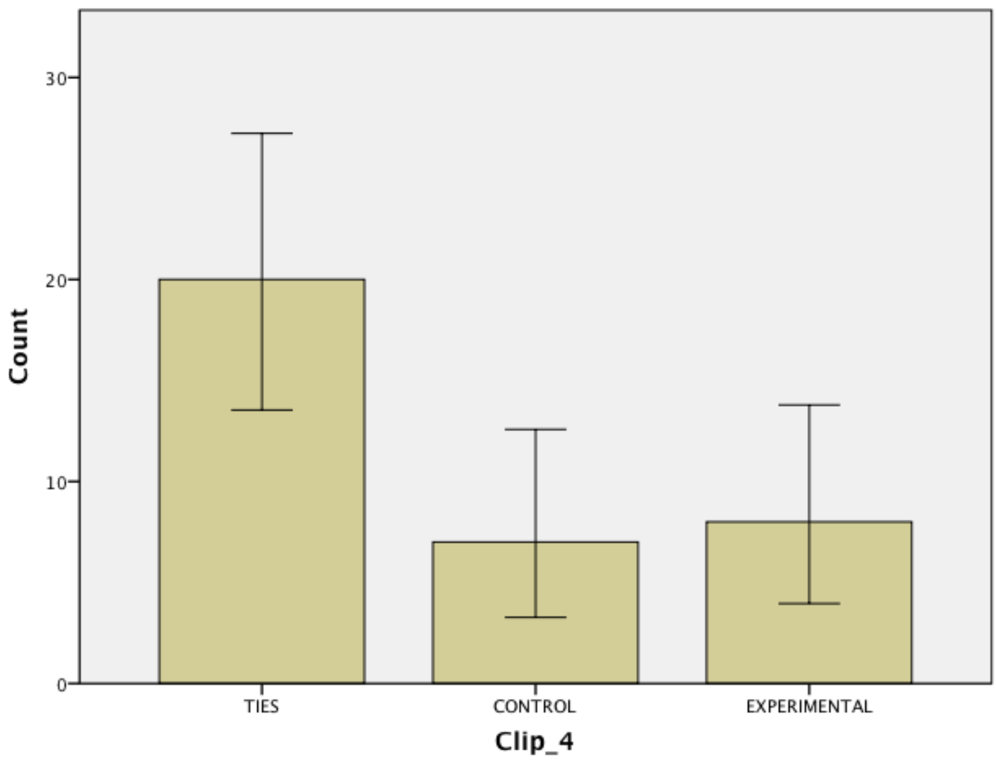
Participant Response Data Graphs



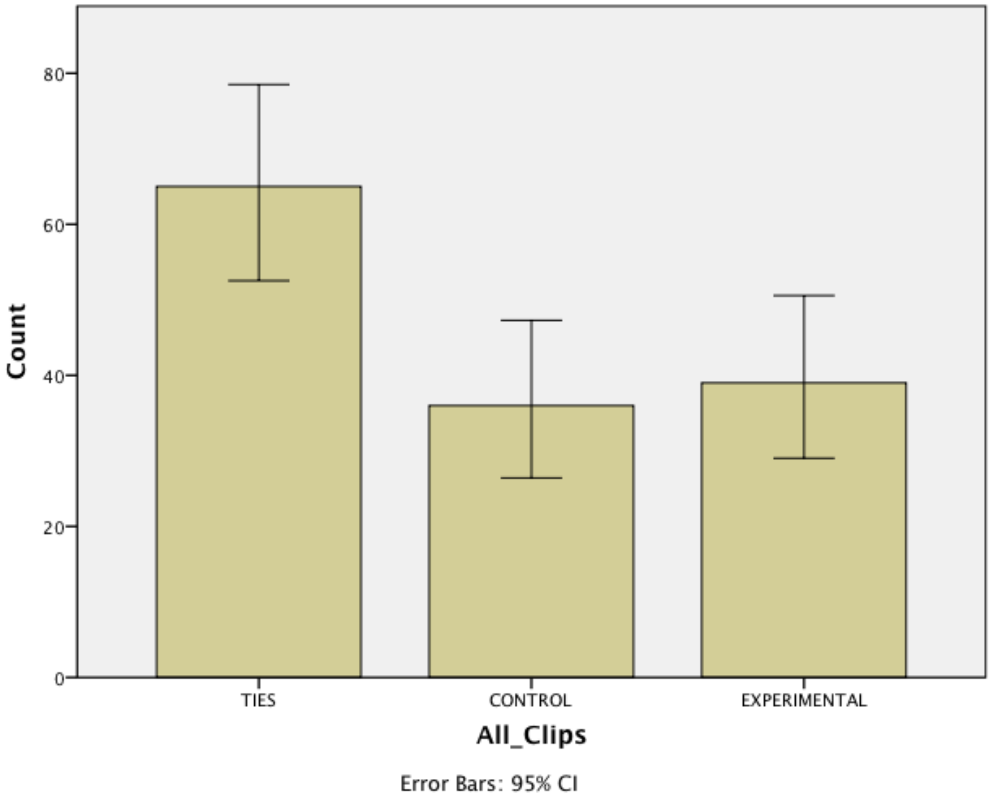




Error Bars: 95% CI



Error Bars: 95% CI


80
60
40
20
0
Count
TIES
CONTROL
EXPERIMENTAL
All_Clips
Error Bars: 95% CI

Experiment 5 (LA) - Statistical Data

Descriptive Statistics

	N	Mean	Std. Deviation	Minimum	Maximum
Clip_1_C	35	6.8000	1.67683	3.00	9.00
Clip_2_C	35	6.4286	1.95968	2.00	9.00
Clip_3_C	35	7.4571	1.59674	3.00	9.00
Clip_4_C	35	6.1143	2.04035	2.00	9.00
All_Clips_Control	140	6.7000	1.87640	2.00	9.00
Clip_1_E	35	7.0000	1.41421	4.00	9.00
Clip_2_E	35	6.5714	1.59569	4.00	9.00
Clip_3_E	35	7.8000	1.34602	4.00	9.00
Clip_4_E	35	6.3143	1.89071	2.00	9.00
All_Clips_Exp	140	6.9214	1.65806	2.00	9.00

Wilcoxon Signed Ranks Test

Test Statistics[a]

	1_E - 1_C	2_E - 2_C	3_E - 3_C	4_E - 4_C	All_Clips_E - All_Clips_C
Z	-1.259[b]	-.678[b]	-2.307[b]	-1.176[b]	-2.592[b]
Asymp. Sig. (2-tailed)	.208	.498	.021	.240	.010
Exact Sig. (2-tailed)	.264	.504	.027	.262	.009
Exact Sig. (1-tailed)	.132	.252	.014	.131	.005
Point Probability	.041	.007	.002	.010	.000

a. Wilcoxon Signed Ranks Test

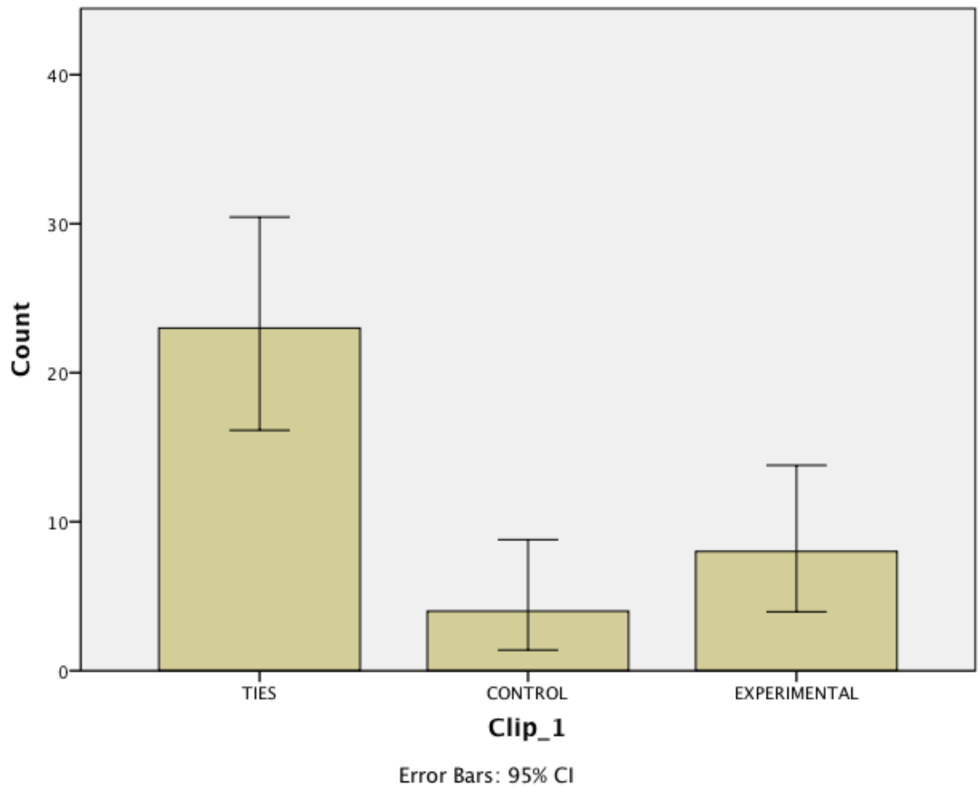
b. Based on negative ranks.

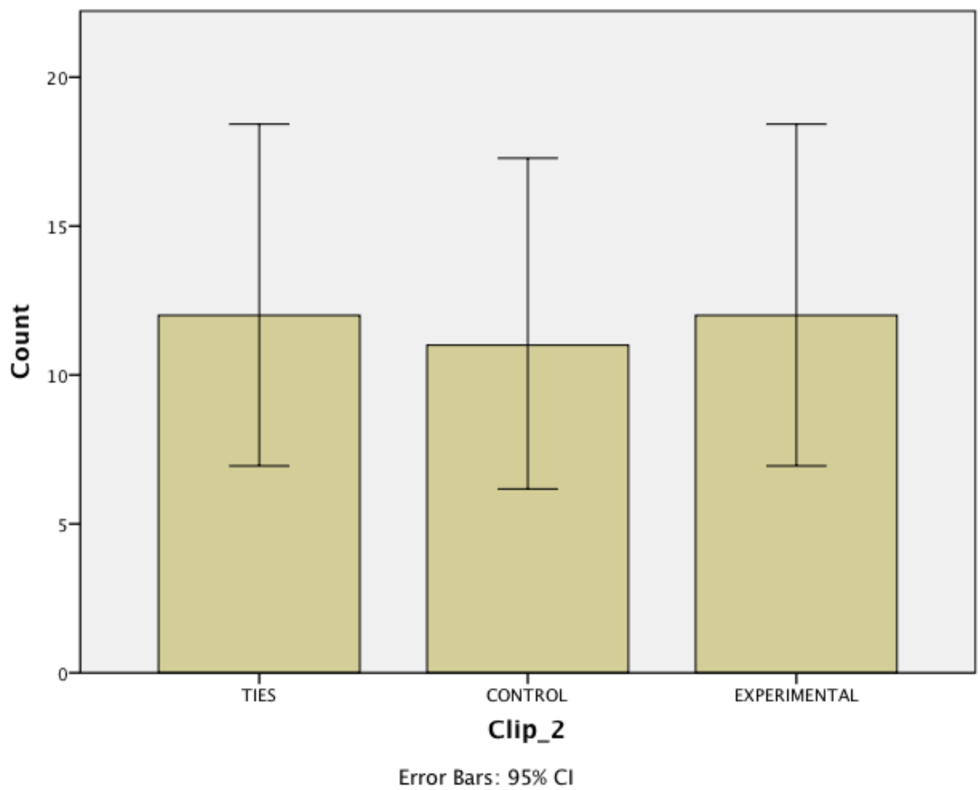
Ranks

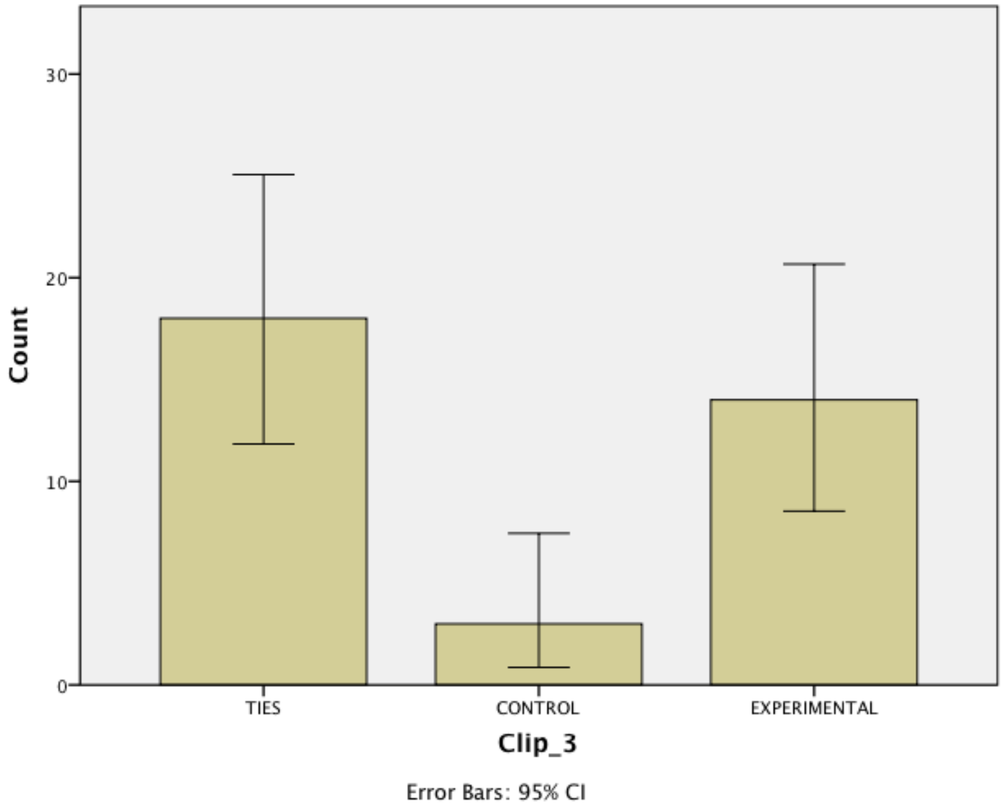
		N	Mean Rank	Sum of Ranks
Clip_1_E - Clip_1_C	Negative Ranks	4[a]	5.88	23.50
	Positive Ranks	8[b]	6.81	54.50
	Ties	23[c]		
	Total	35		
Clip_2_E - Clip_2_C	Negative Ranks	11[d]	10.64	117.00
	Positive Ranks	12[e]	13.25	159.00
	Ties	12[f]		
	Total	35		
Clip_3_E - Clip_3_C	Negative Ranks	3[g]	10.33	31.00
	Positive Ranks	14[h]	8.71	122.00
	Ties	18[i]		
	Total	35		
Clip_4_E - Clip_4_C	Negative Ranks	8[j]	11.50	92.00
	Positive Ranks	14[k]	11.50	161.00
	Ties	13[l]		
	Total	35		
All_Clips_E - All_Clips_C	Negative Ranks	26[m]	35.88	933.00
	Positive Ranks	48[n]	38.38	1842.00
	Ties	66[o]		
	Total	140		

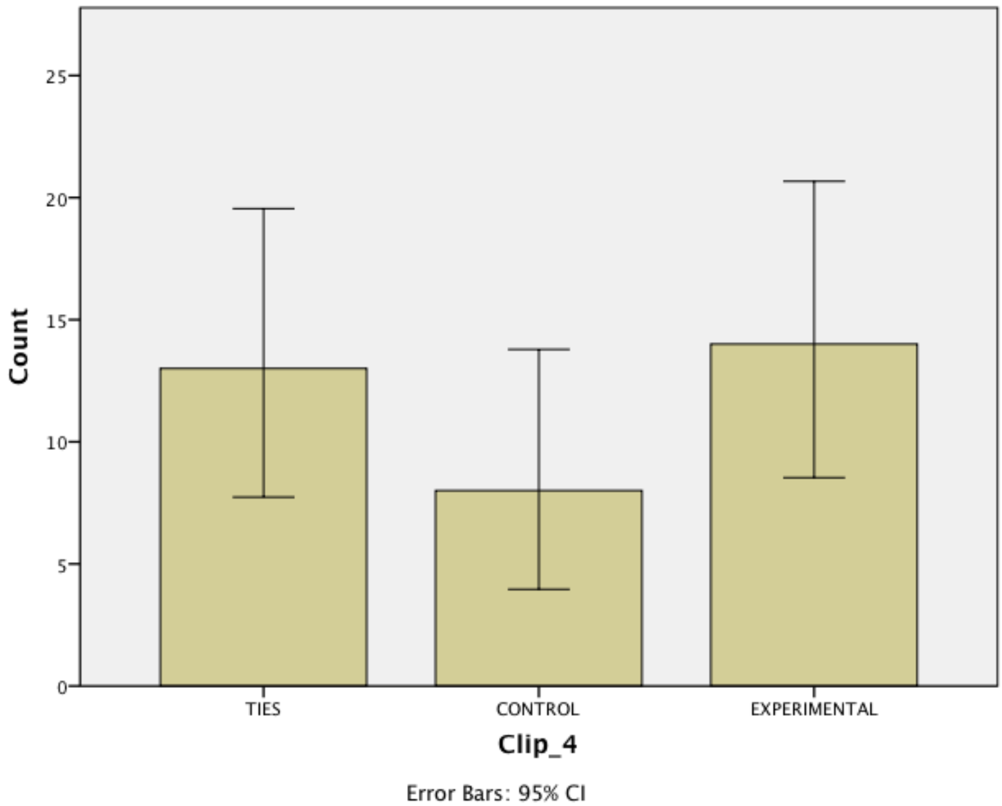
a. Clip_1_R_E < Clip_1_R_C
b. Clip_1_R_E > Clip_1_R_C
c. Clip_1_R_E = Clip_1_R_C
d. Clip_2_S_E < Clip_2_S_C
e. Clip_2_S_E > Clip_2_S_C
f. Clip_2_S_E = Clip_2_S_C
g. Clip_3_R_E < Clip_3_R_C
h. Clip_3_R_E > Clip_3_R_C
i. Clip_3_R_E = Clip_3_R_C
j. Clip_4_S_E < Clip_4_S_C
k. Clip_4_S_E > Clip_4_S_C
l. Clip_4_S_E = Clip_4_S_C
m. All_Clips_Exp < All_Clips_Control
n. All_Clips_Exp > All_Clips_Control
o. All_Clips_Exp = All_Clips_Control

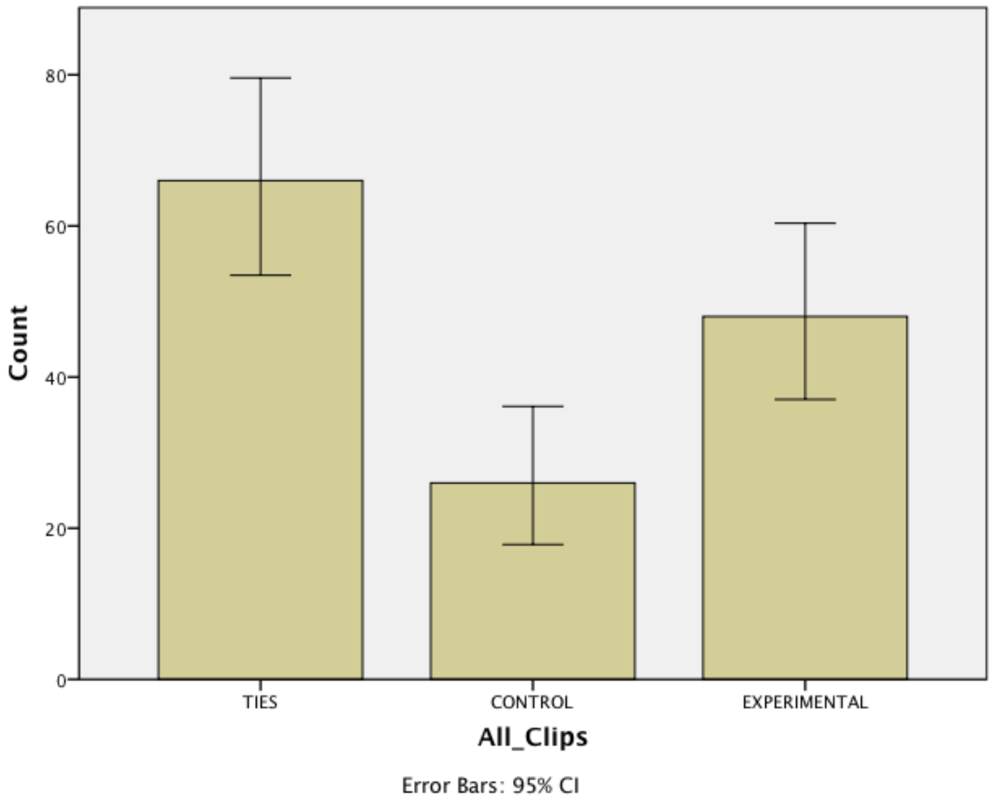
Participant Response Data Graphs






30
20
10
0
Count
TIES
CONTROL
EXPERIMENTAL
Clip_3
Error Bars: 95% CI


25
20
15
10
5
0
Count
TIES
CONTROL
EXPERIMENTAL
Clip_4
Error Bars: 95% CI


80
60
40
20
0
Count
TIES
CONTROL
EXPERIMENTAL
All_Clips
Error Bars: 95% CI

Experiment 6NS (LAHFF) - Statistical Data

Descriptive Statistics

	N	Mean	Std. Deviation	Minimum	Maximum
Clip_1_C	35	6.9429	1.32716	4.00	9.00
Clip_2_C	35	6.8000	1.49115	4.00	9.00
Clip_3_C	35	7.8571	1.00419	6.00	9.00
Clip_4_C	35	6.3714	1.78368	2.00	9.00
All_Clips_Control	140	6.9929	1.51490	2.00	9.00
Clip_1_E	35	6.9143	1.54104	4.00	9.00
Clip_2_E	35	6.8000	1.69428	3.00	9.00
Clip_3_E	35	7.9714	1.12422	5.00	9.00
Clip_4_E	35	6.5143	1.72134	3.00	9.00
All_Clips_Exp	140	7.0500	1.61969	3.00	9.00

Wilcoxon Signed Ranks Test

Test Statistics[a]

	1_E - 1_C	2_E - 2_C	3_E - 3_C	4_E - 4_C	All_Clips_E - All_Clips_C
Z	.000[b]	-.097[c]	-.842[c]	-.811[c]	-.740[c]
Asymp. Sig. (2-tailed)	1.000	.923	.400	.417	.460
Exact Sig. (2-tailed)	1.000	.946	.518	.469	.464
Exact Sig. (1-tailed)	.507	.473	.259	.235	.232
Point Probability	.014	.025	.061	.046	.003

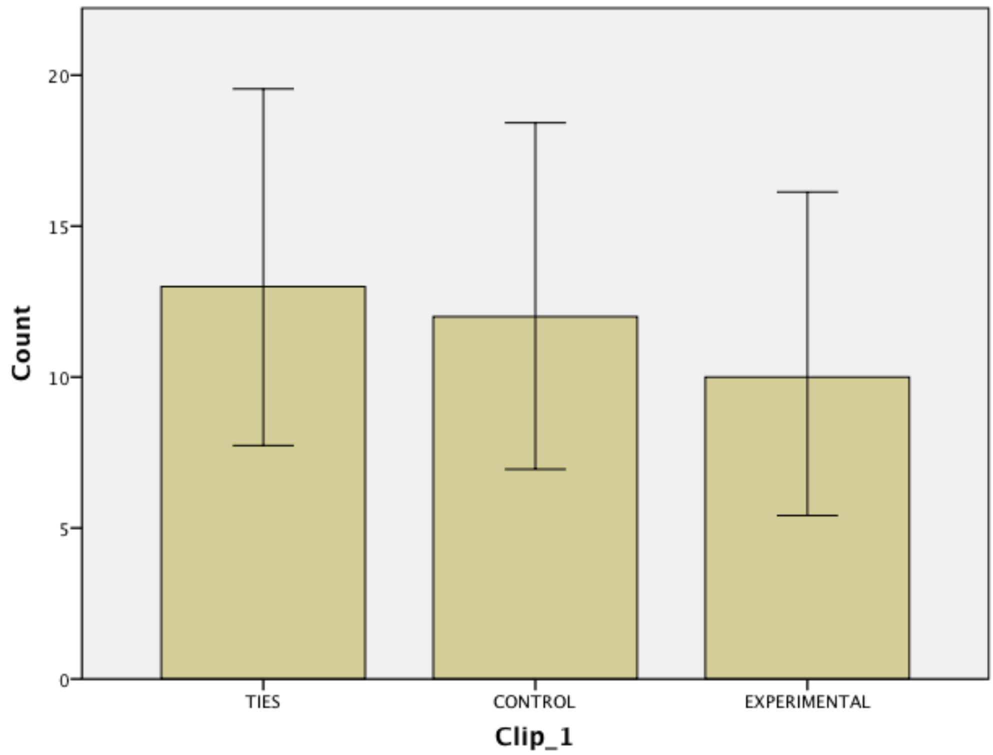
a. Wilcoxon Signed Ranks Test
b. The sum of negative ranks equals the sum of positive ranks.
c. Based on negative ranks.

Ranks

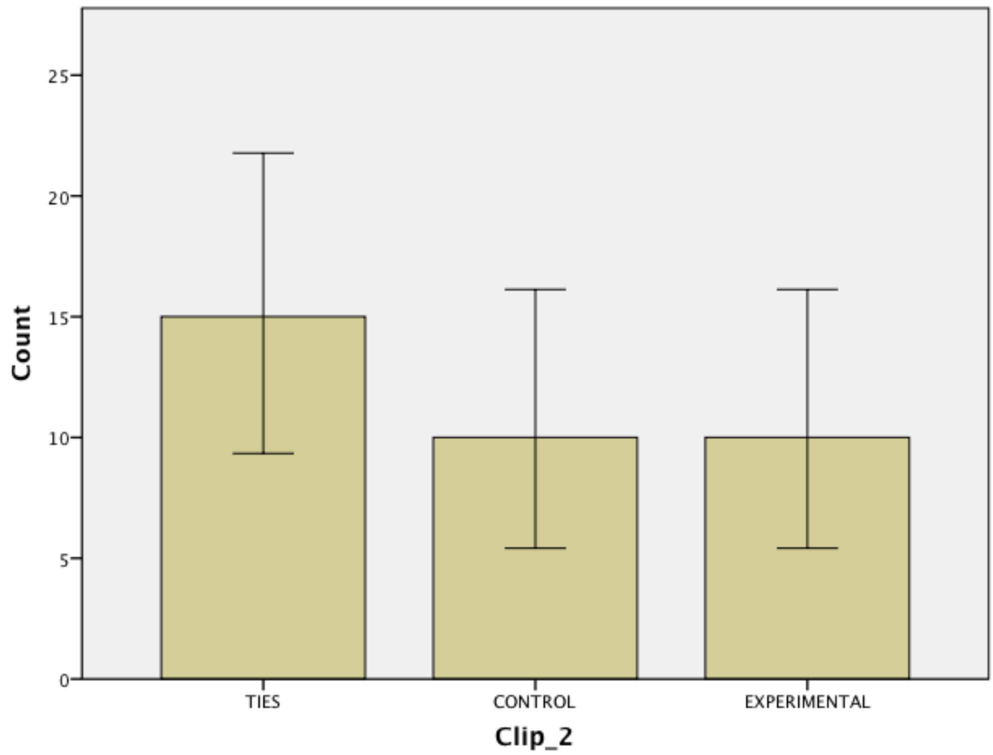
		N	Mean Rank	Sum of Ranks
Clip_1_E - Clip_1_C	Negative Ranks	12[a]	10.54	126.50
	Positive Ranks	10[b]	12.65	126.50
	Ties	13[c]		
	Total	35		
Clip_2_E - Clip_2_C	Negative Ranks	10[d]	10.25	102.50
	Positive Ranks	10[e]	10.75	107.50
	Ties	15[f]		
	Total	35		
Clip_3_E - Clip_3_C	Negative Ranks	6[g]	8.83	53.00
	Positive Ranks	10[h]	8.30	83.00
	Ties	19[i]		
	Total	35		
Clip_4_E - Clip_4_C	Negative Ranks	7[j]	8.64	60.50
	Positive Ranks	10[k]	9.25	92.50
	Ties	18[l]		
	Total	35		
All_Clips_E - All_Clips_C	Negative Ranks	35[m]	36.91	1292.00
	Positive Ranks	40[n]	38.95	1558.00
	Ties	65[o]		
	Total	140		

a. Clip_1_R_E < Clip_1_R_C
b. Clip_1_R_E > Clip_1_R_C
c. Clip_1_R_E = Clip_1_R_C
d. Clip_2_S_E < Clip_2_S_C
e. Clip_2_S_E > Clip_2_S_C
f. Clip_2_S_E = Clip_2_S_C
g. Clip_3_R_E < Clip_3_R_C
h. Clip_3_R_E > Clip_3_R_C
i. Clip_3_R_E = Clip_3_R_C
j. Clip_4_S_E < Clip_4_S_C
k. Clip_4_S_E > Clip_4_S_C
l. Clip_4_S_E = Clip_4_S_C
m. All_Clips_Exp < All_Clips_Control
n. All_Clips_Exp > All_Clips_Control
o. All_Clips_Exp = All_Clips_Control

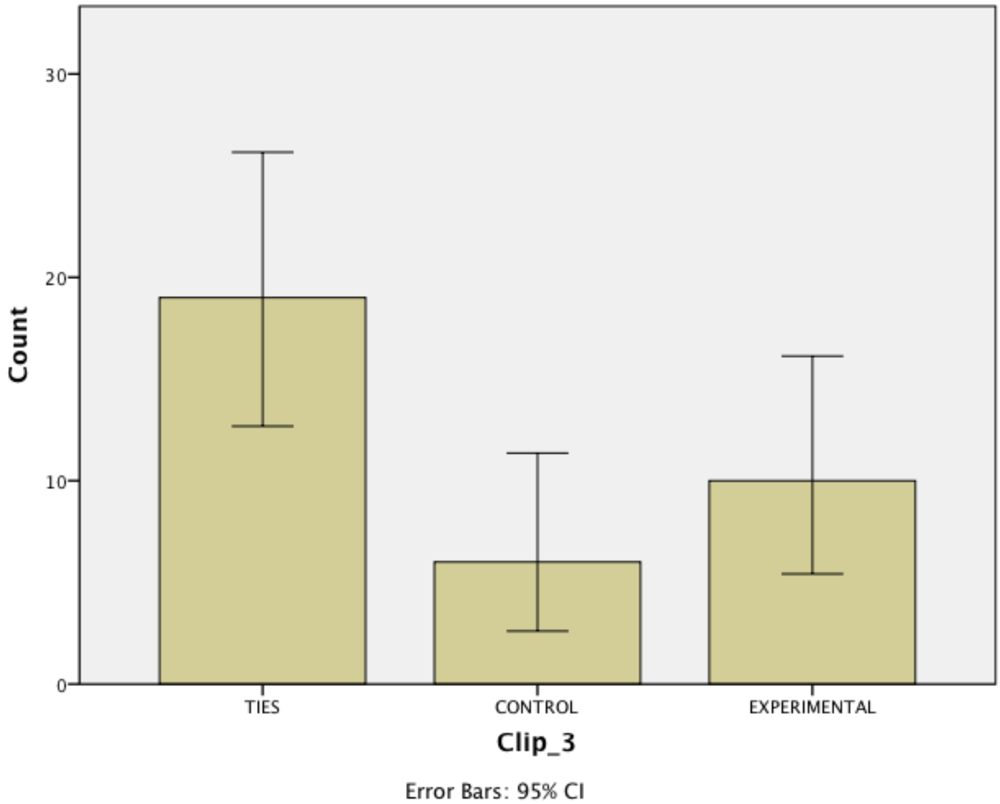
Participant Response Data Graphs

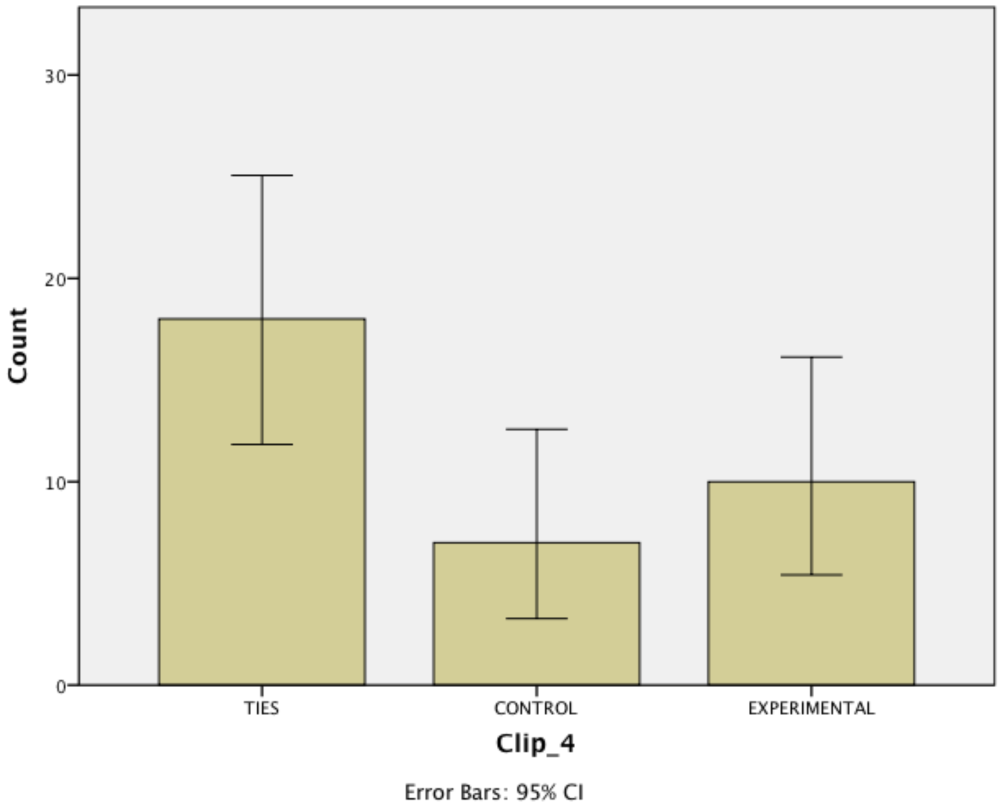


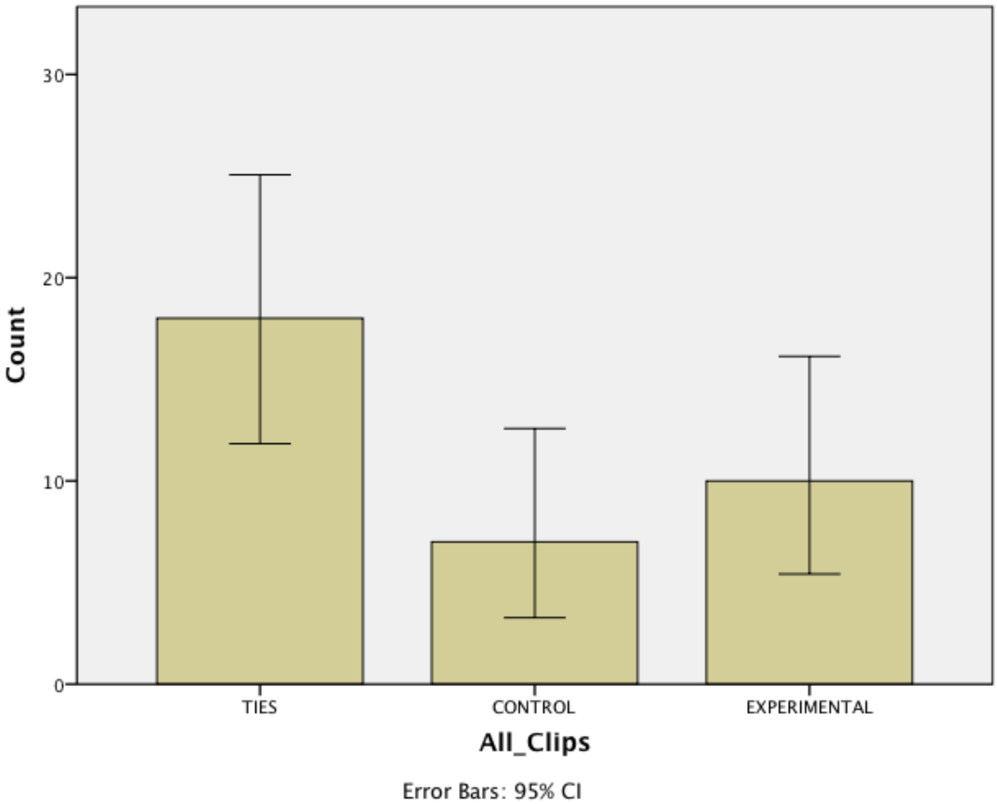
Error Bars: 95% CI



Error Bars: 95% CI


30
20
10
0
Count
TIES
CONTROL
EXPERIMENTAL
Clip_3
Error Bars: 95% CI


30
20
10
0
Count
TIES
CONTROL
EXPERIMENTAL
Clip_4
Error Bars: 95% CI


30
20
10
0
Count
TIES
CONTROL
EXPERIMENTAL
All_Clips
Error Bars: 95% CI

Experiment 6S (LAHFF) - Statistical Data

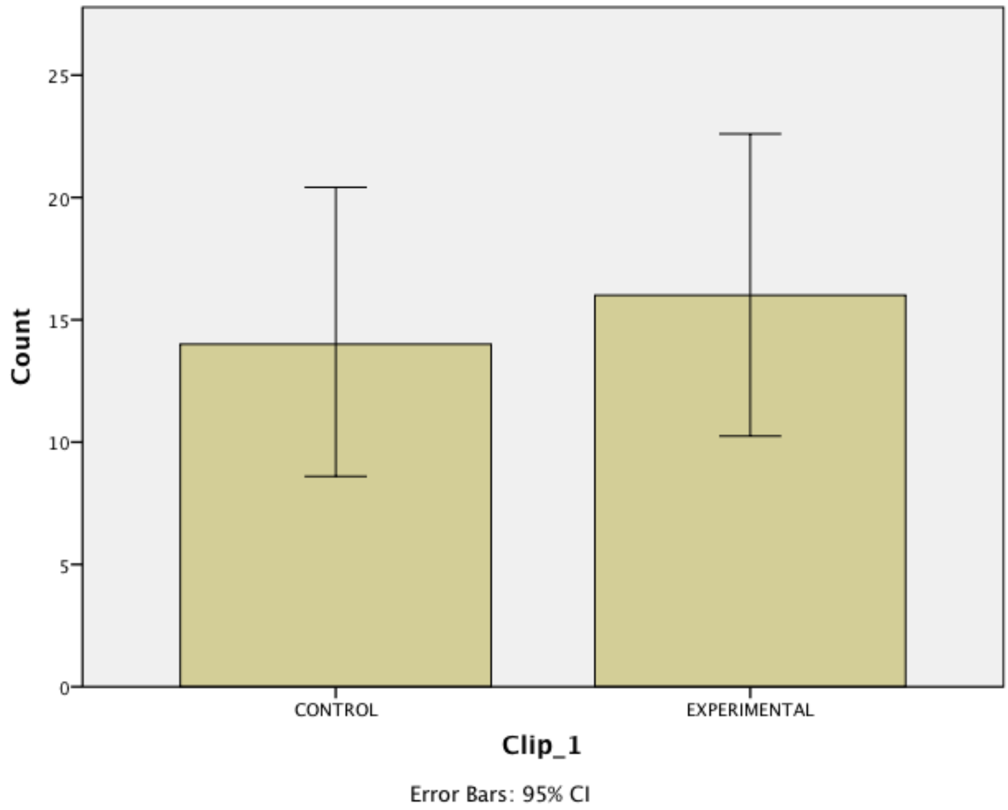
Descriptive Statistics

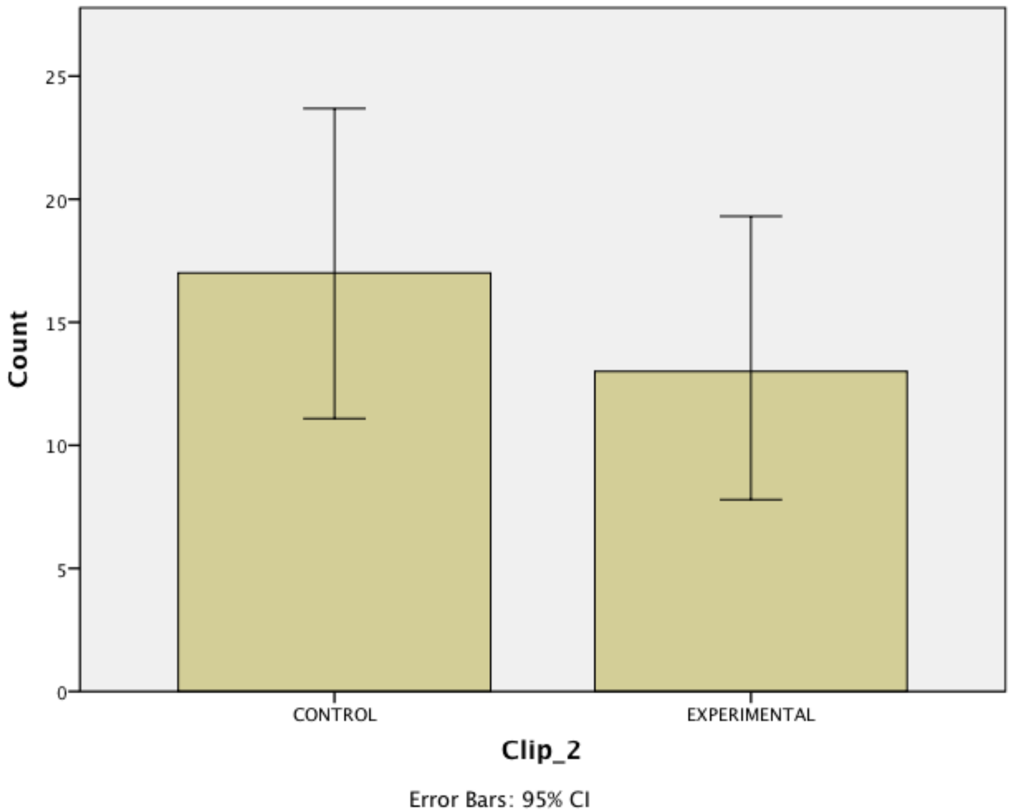
	N	Mean	Std. Deviation	Minimum	Maximum
Clip_1	30	.53	.507	0	1
Clip_2	30	.43	.504	0	1
Clip_3	30	.47	.507	0	1
Clip_4	30	.57	.504	0	1
All_Clips	120	.50	.502	0	1

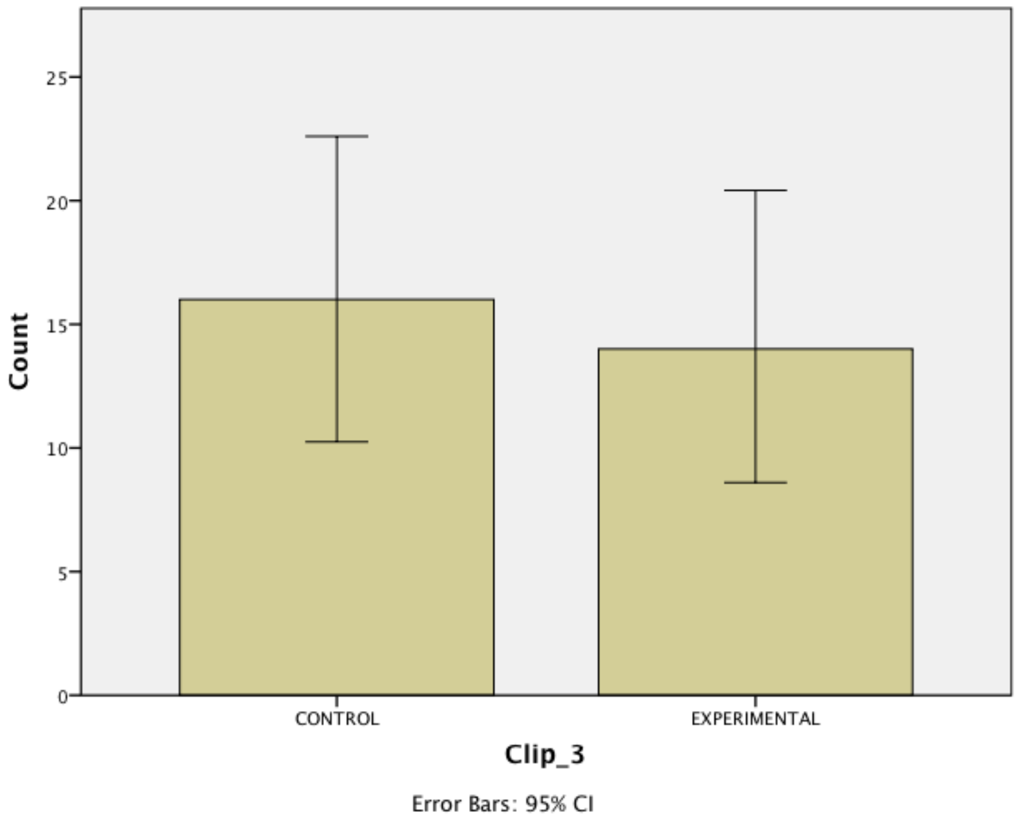
Binomial Test

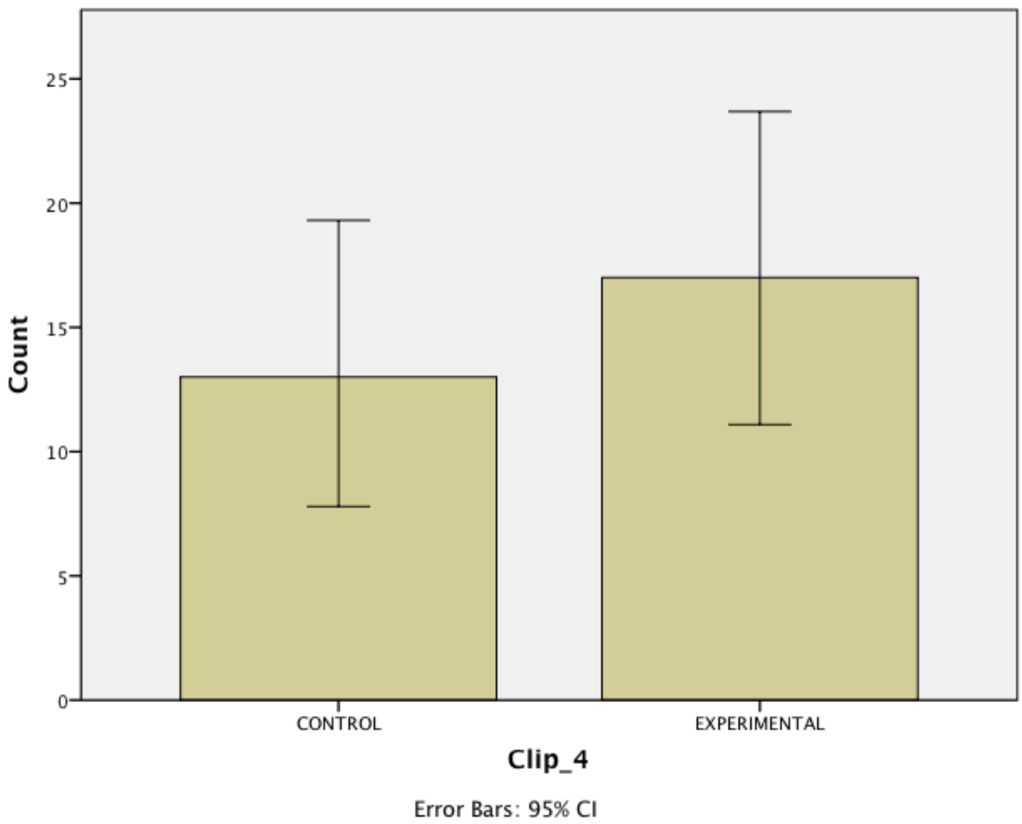
		Category	N	Observed Prop	Test Prop	Exact Sig. (2-tailed)	Est. Sig. (1-tailed)
Clip_1	Group 1	EXP	14	.47	.50	.856	.428
	Group 2	CTRL	16	.53			
	Total		30	1.00			
Clip_2	Group 1	EXP	17	.57	.50	.585	.292
	Group 2	CTRL	13	.43			
	Total		30	1.00			
Clip_3	Group 1	EXP	14	.47	.50	.856	.428
	Group 2	CTRL	16	.53			
	Total		30	1.00			
Clip_4	Group 1	EXP	13	.43	.50	.585	.292
	Group 2	CTRL	17	.57			
	Total		30	1.00			
All_Clips	Group 1	EXP	60	.50	.50	1.000	.500
	Group 2	CTRL	60	.50			
	Total		120	1.00			

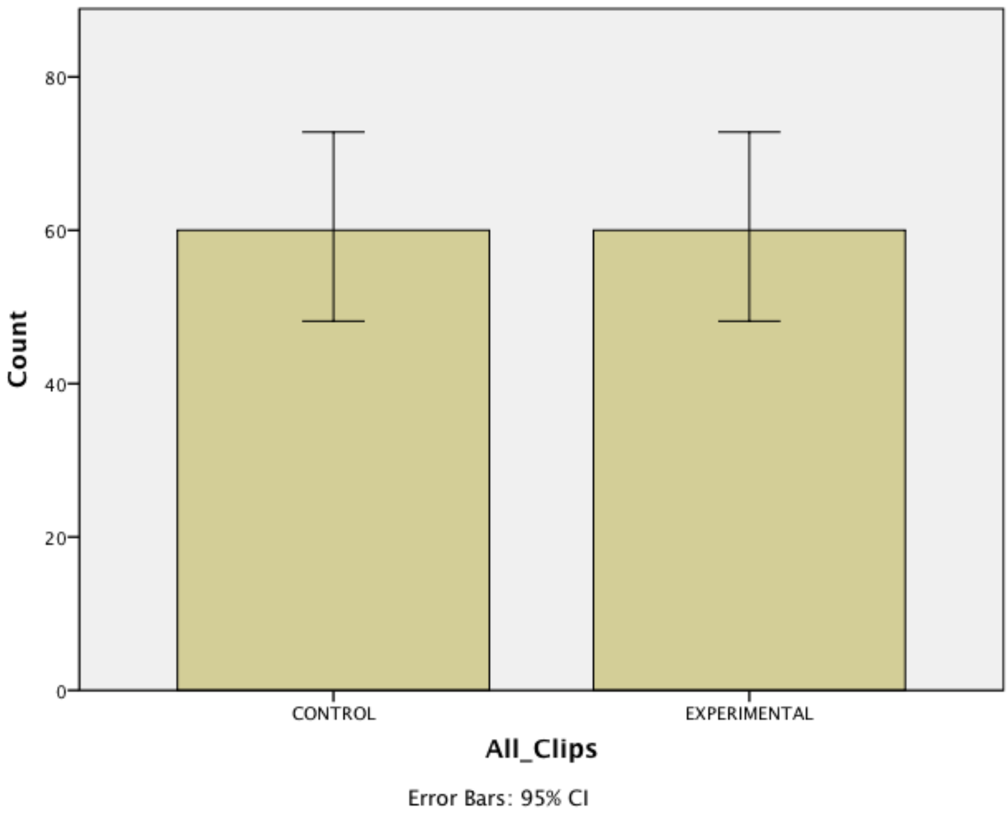
Participant Response Data Graphs






25
20
15
10
5
0
Count
CONTROL
EXPERIMENTAL
Clip_3
Error Bars: 95% CI


25
20
15
10
5
0
Count
CONTROL
EXPERIMENTAL
Clip_4
Error Bars: 95% CI



80
60
40
20
0
Count
CONTROL
EXPERIMENTAL
All_Clips
Error Bars: 95% CI

Experiment 7 (LAHFF) - Statistical Data

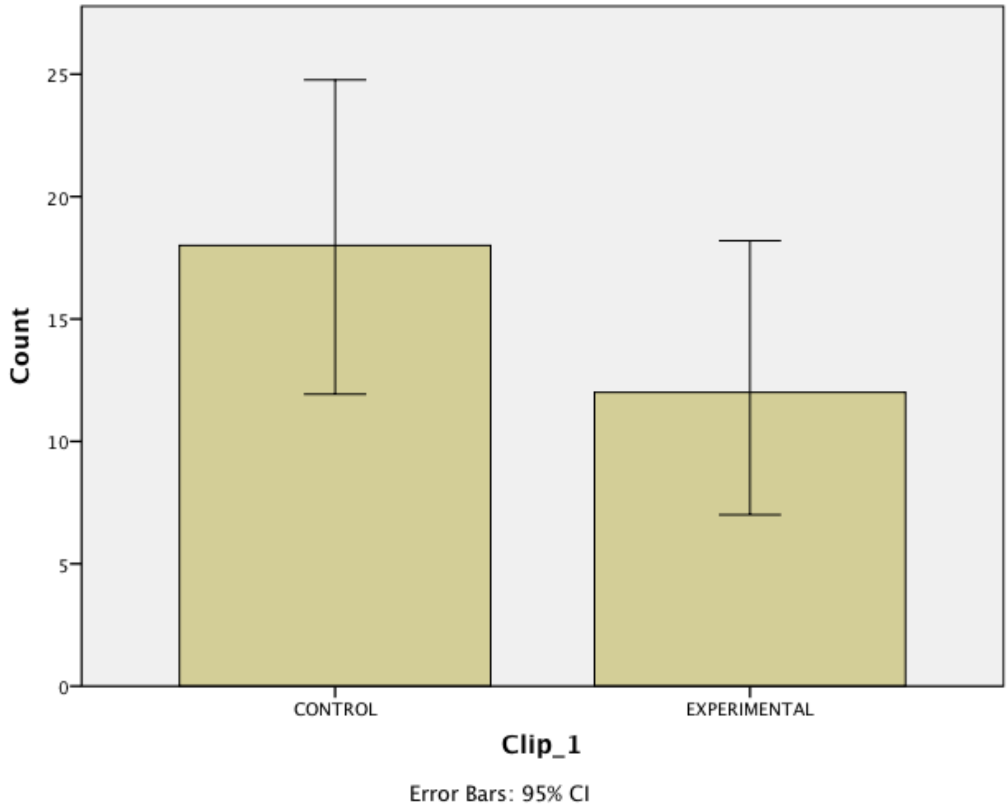
Descriptive Statistics

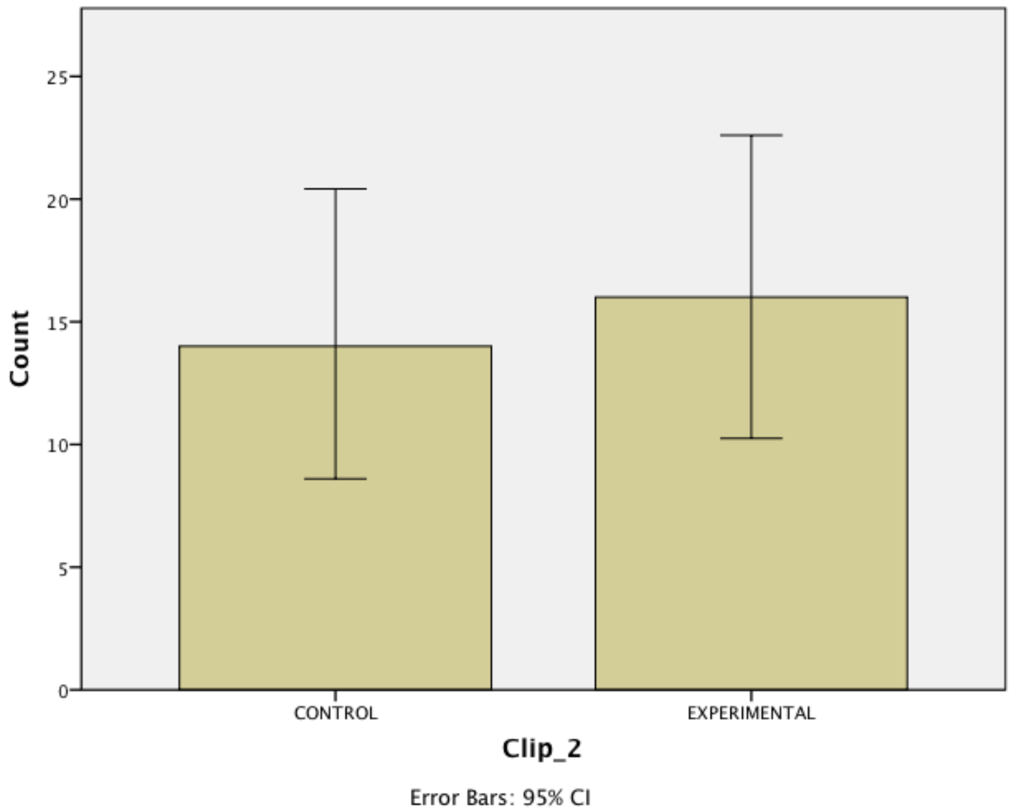
	N	Mean	Std. Deviation	Minimum	Maximum
Clip_1	30	.40	.498	0	1
Clip_2	30	.53	.507	0	1
All_Clips	60	.47	.503	0	1

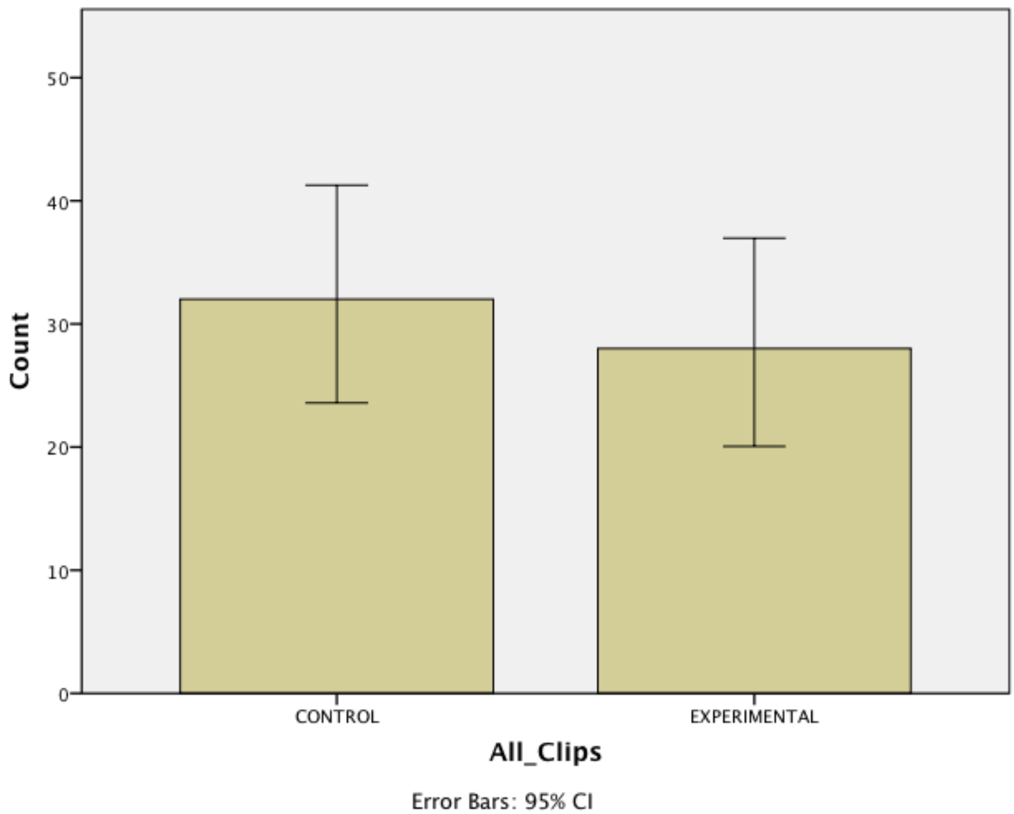
Binomial Test

		Category	N	Observed Prop	Test Prop	Exact Sig. (2-tailed)	Est. Sig. (1-tailed)
Clip_1	Group 1	EXP	18	.60	.50	.362	.181
	Group 2	CTRL	12	.40			
	Total		30	1.00			
Clip_2	Group 1	EXP	16	.53	.50	.856	.428
	Group 2	CTRL	14	.47			
	Total		30	1.00			
All_Clips	Group 1	EXP	32	.53	.50	.699	.349
	Group 2	CTRL	28	.47			
	Total		60	1.00			

Participant Response Data Graphs






50
40
30
20
10
0
Count
CONTROL
EXPERIMENTAL
All_Clips
Error Bars: 95% CI

Appendix 2 – Soundtrack Settings

Logic Pro 9 EQ Settings for High Frequency Filtering

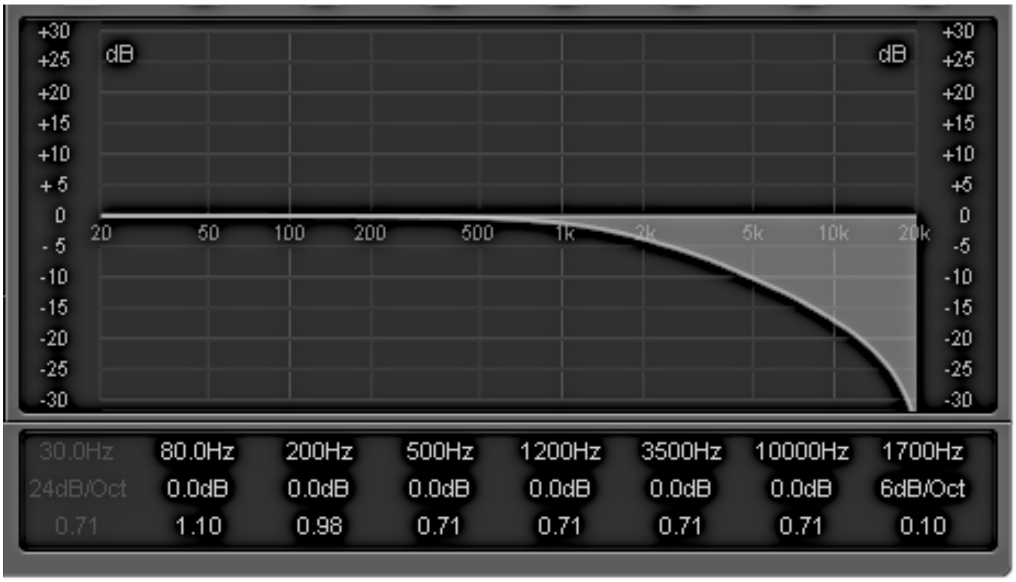


Figure Logic Pro 9 EQ Settings for High Frequency Filtering in Experiments 1 (HFF), 3NS/3S (LA), 4 (HFF), 6NS/6S (LAHFF) and 7 (LAHFF). Values were based on sound absorption coefficients in free space provided by Coleman (Coleman 1963).

TABLE 1

ABSORPTION COEFFICIENT IN DB/100 FT AT 50% RELATIVE HUMIDITY AND 20° C.

Frequency in cps	a_1	a_2
1000	.2	.2
2000	.3	.3
3000	.5	.6
4000	.7	1.0
5000	1.1	1.5
6000	1.5	2.1
7000	2.0	2.6
8000	2.4	3.3
9000	2.9	4.0
10000	3.5	4.7

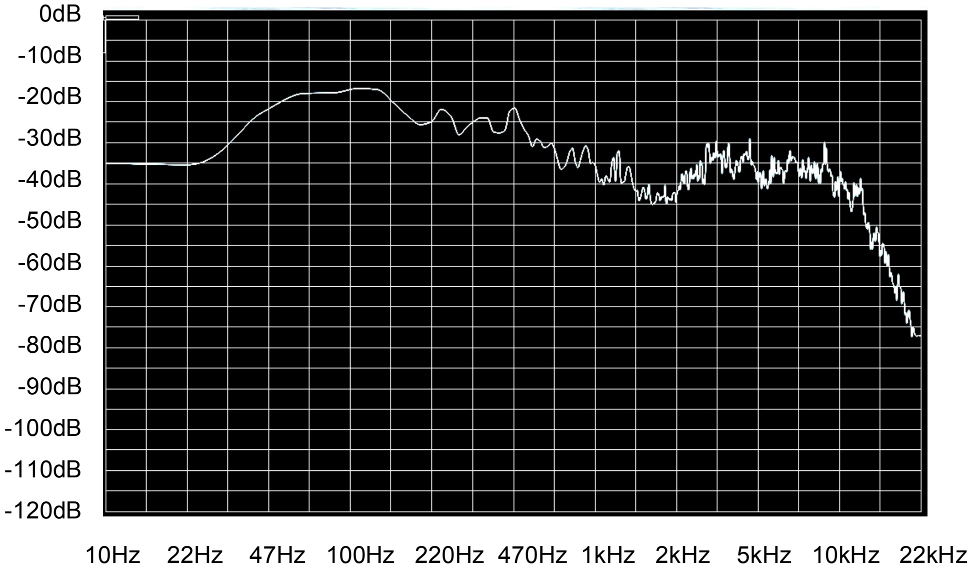
Note.—*a* is absorption coefficient.

Indicative absorption coefficients in DB per FT (table reproduced from Coleman (Coleman 1963: 307)

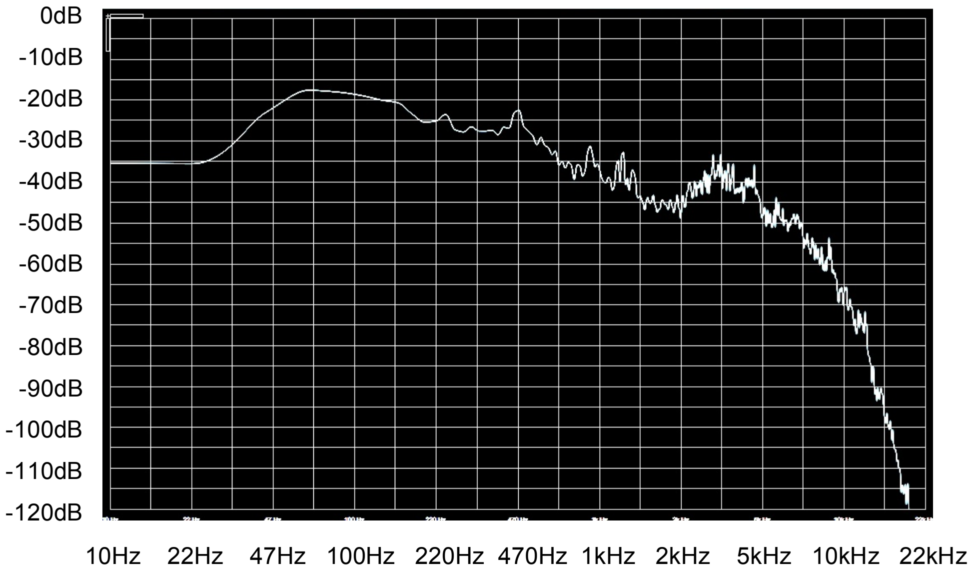
Spectrum Analysis Graphs for Sounds Used in the Experiments

Experiments 1 (HFF), 2 (LA), 3 (LAHFF)

Clips 1_R and 2_S



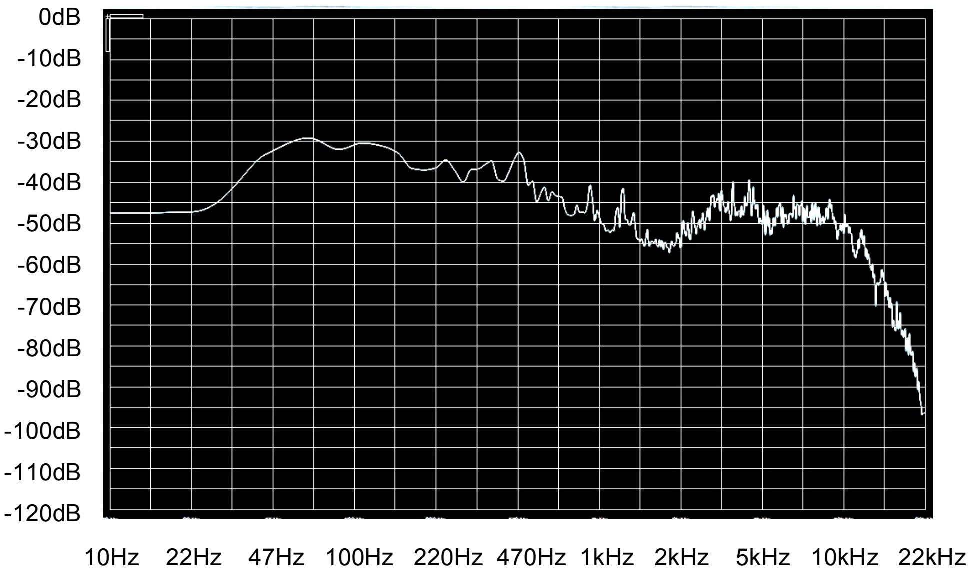
Control Sound for Experiments 1 (HFF), 2 (LA) and 3 (LAHFF)



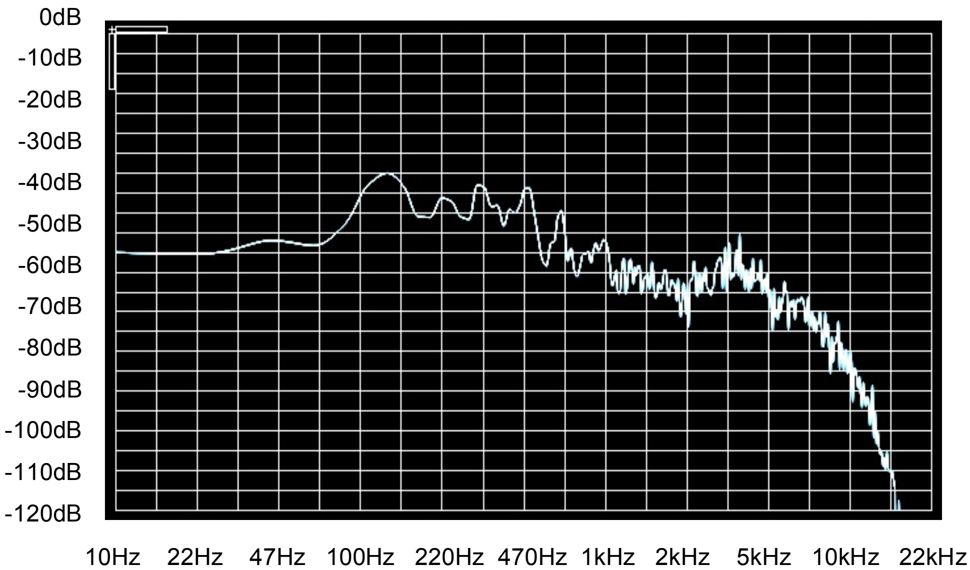
Experimental Sound for Experiment 1 (HFF - High Frequency Filtering)

Experiments 1 (HFF), 2 (LA), 3 (LAHFF)

Clips 1_R and 2_S



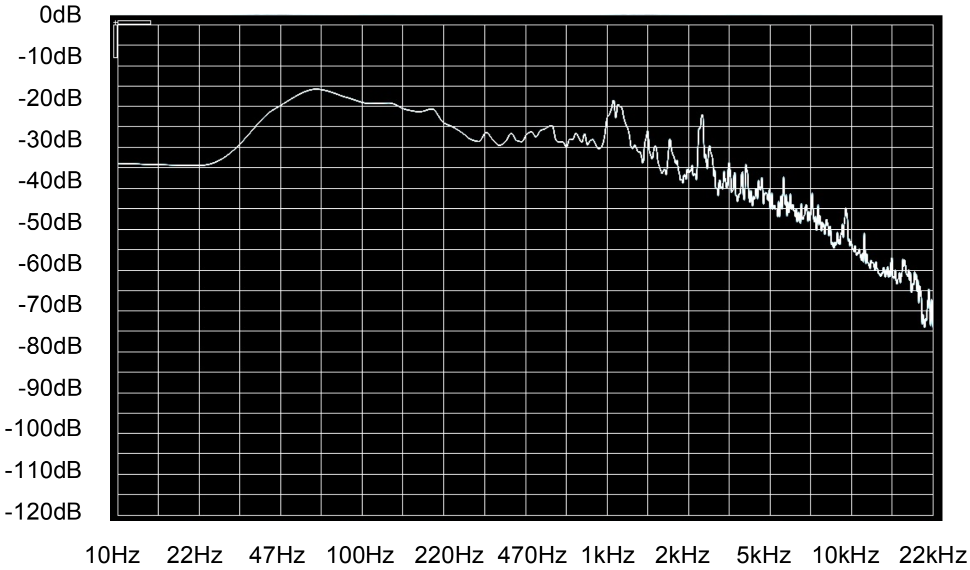
Experimental Sound for Experiment 2 (LA – Volume Level Alteration)



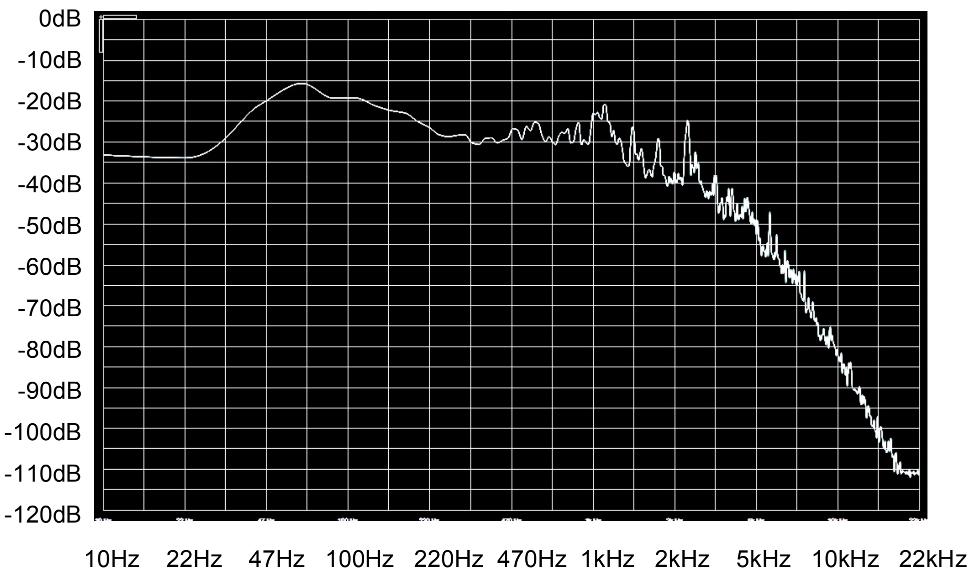
Experimental Sound for Experiment 3 (LAHFF – Volume Level Alteration and High Frequency Filtering)

Experiments 1 (HFF), 2 (LA), 3 (LAHFF)

Clips 3_R and 4_S



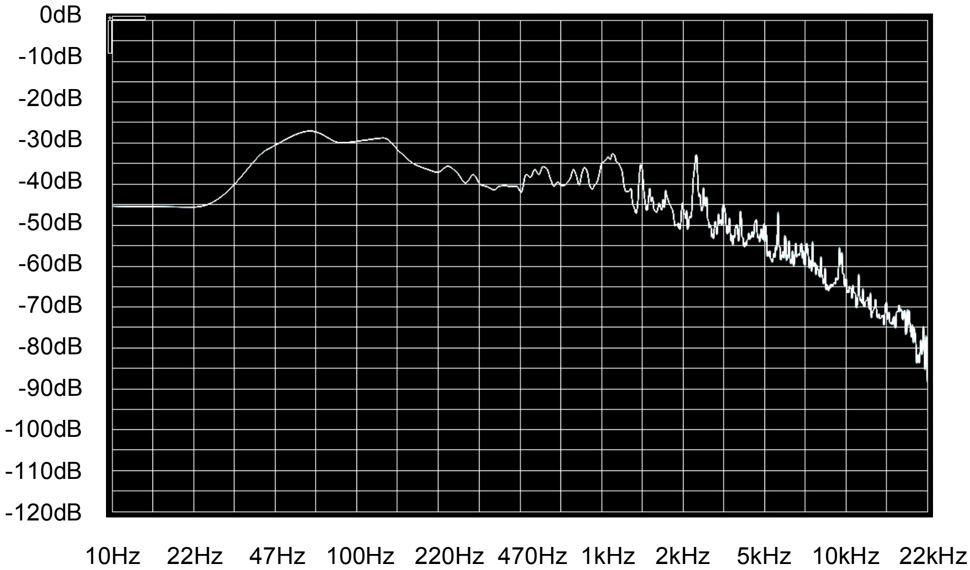
Control Sound for Experiments 1 (HFF), 2 (LA) and 3 (LAHFF)



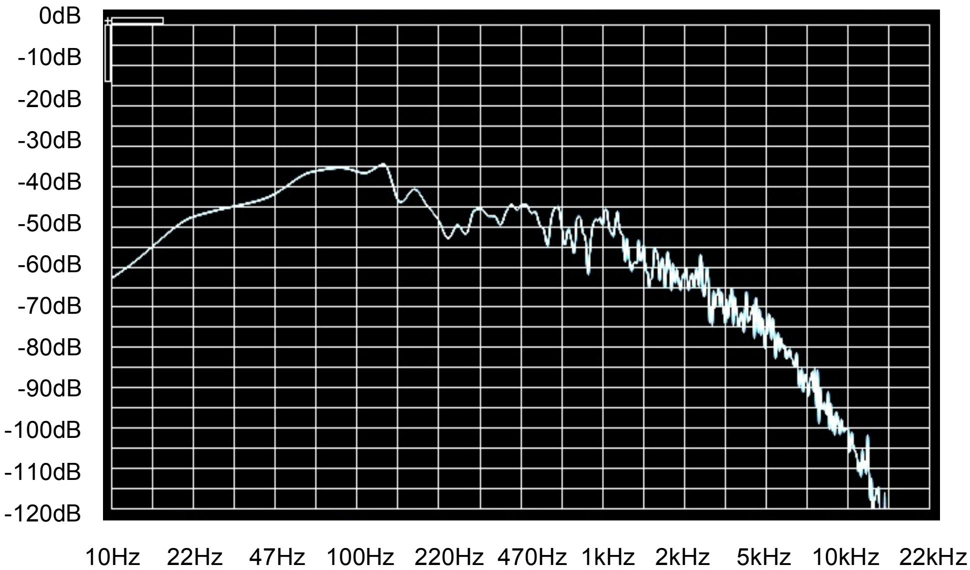
Experimental Sound for Experiment 1 (HFF - High Frequency Filtering)

Experiments 1 (HFF), 2 (LA), 3 (LAHFF)

Clips 3_R and 4_S



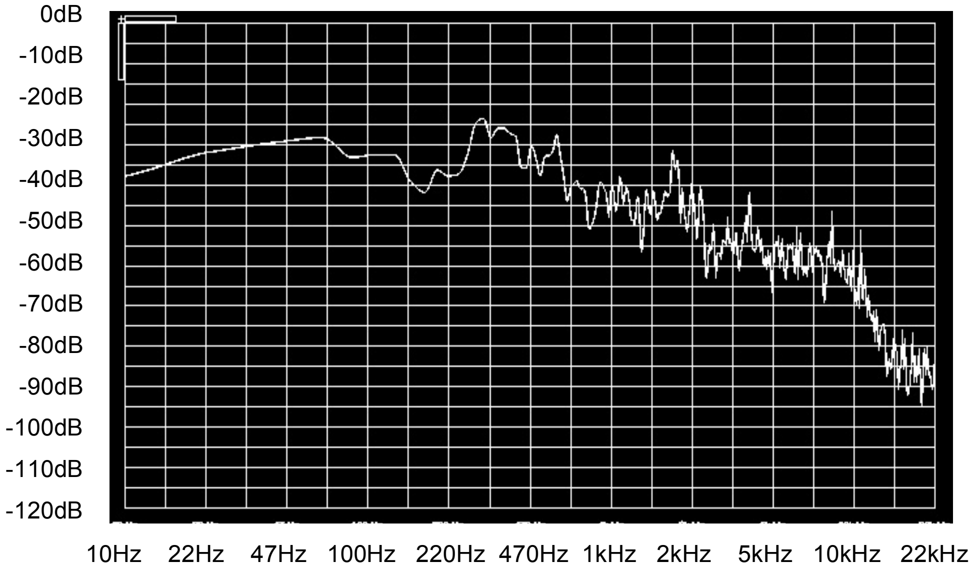
Experimental Sound for Experiment 2 (LA – Volume Level Alteration)



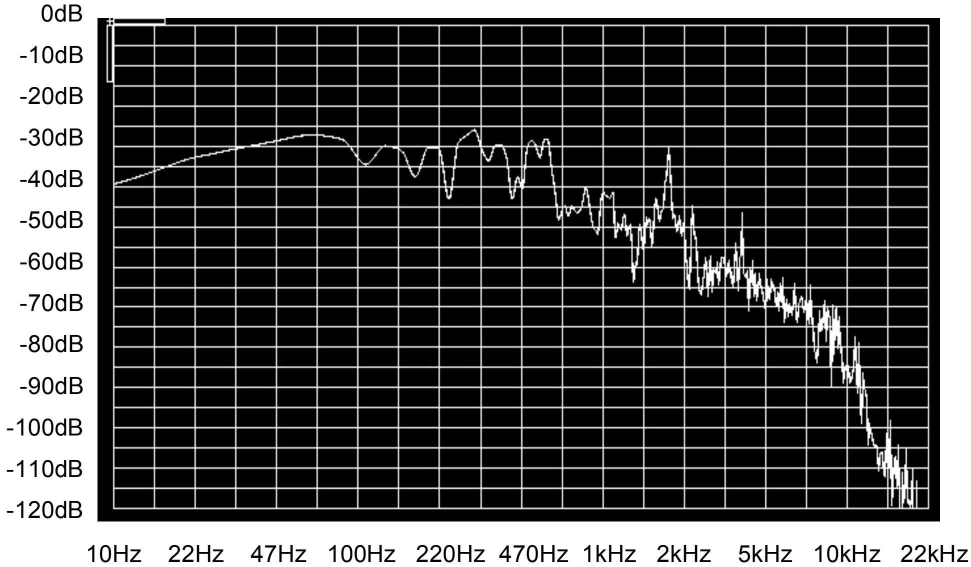
Experimental Sound for Experiment 3 (LAHFF – Volume Level Alteration and High Frequency Filtering)

Experiments 4 (HFF), 5 (LA), 6 (LAHFF)

Clip 1



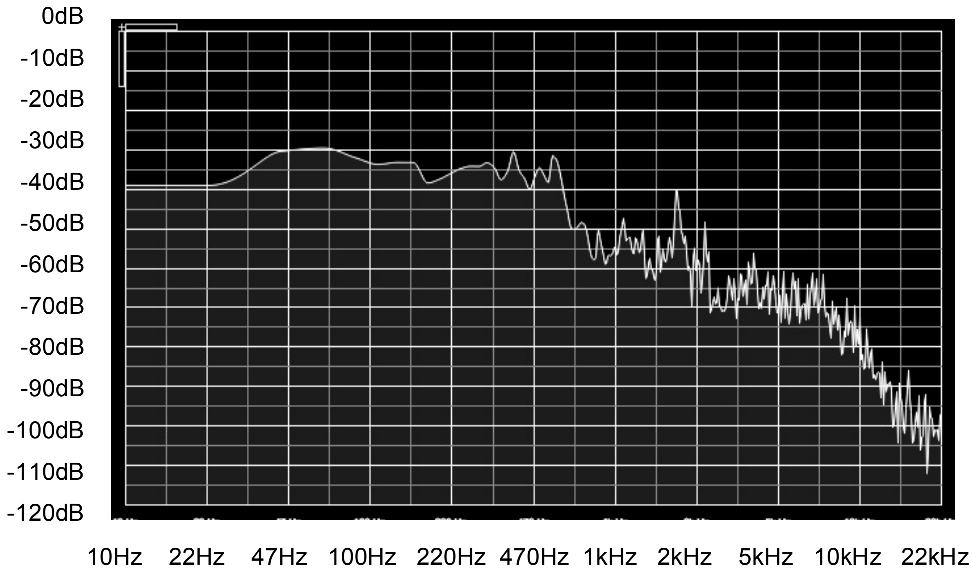
Control Sound for Experiments 4 (HFF), 5 (LA) and 6 (LAHFF)



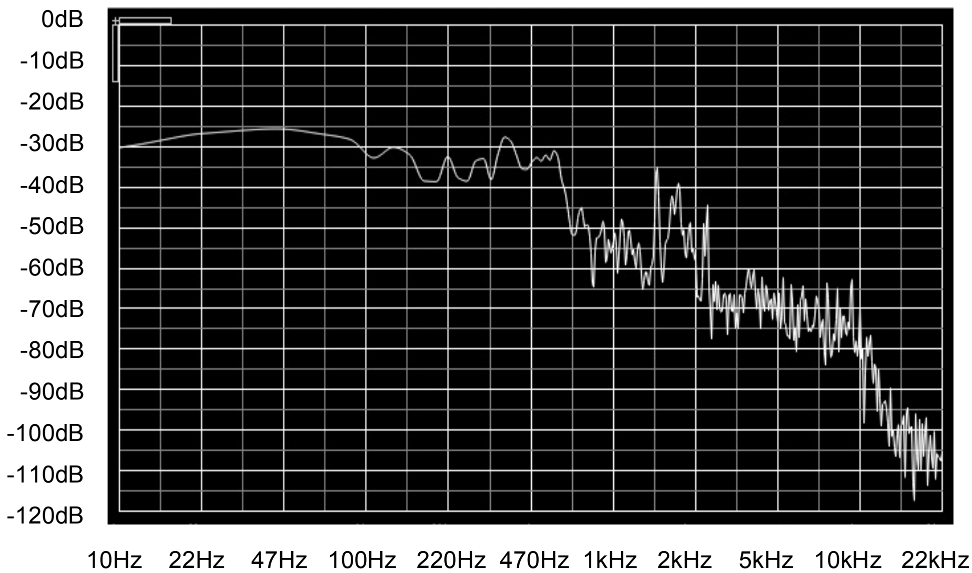
Experimental Sound for Experiment 4 (HFF - High Frequency Filtering) – Initial S3D object position

Experiments 4 (HFF), 5 (LA), 6 (LAHFF)

Clip 1



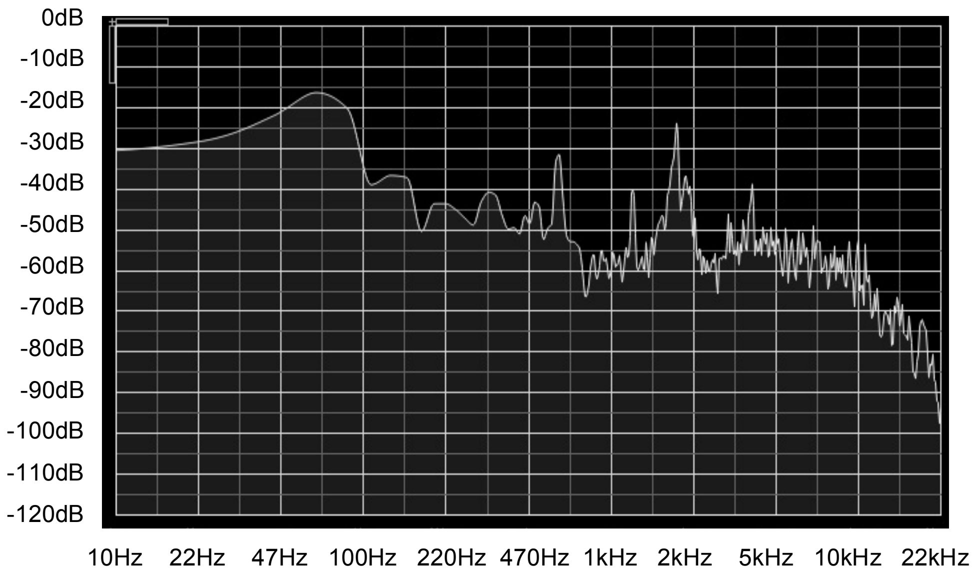
Experimental Sound for Experiment 5 (LA – Volume Level Alteration) – Initial S3D object position



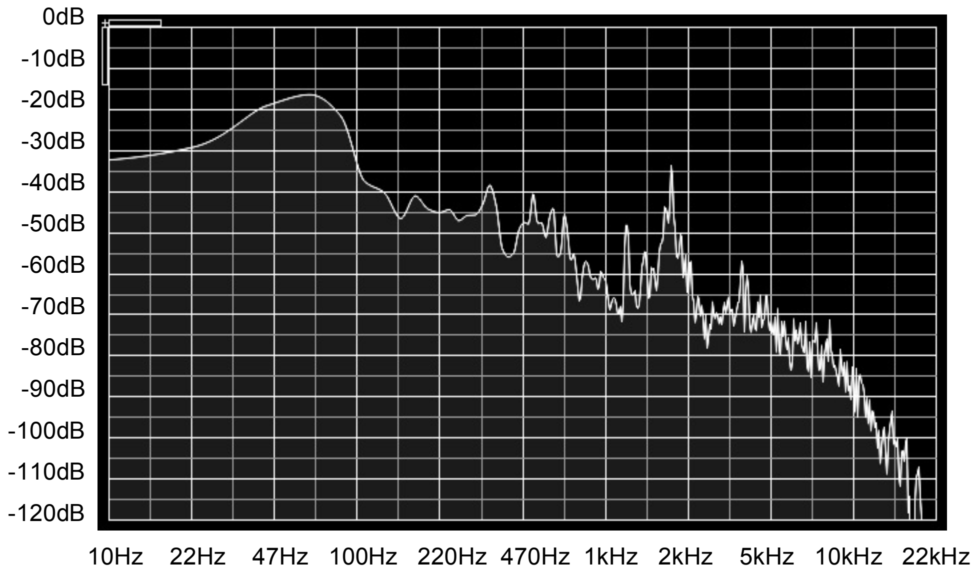
Experimental Sound for Experiment 6 (LAHFF – Volume Level Alteration and High Frequency Filtering) – Initial S3D object position

Experiments 4 (HFF), 5 (LA), 6 (LAHFF)

Clip 2



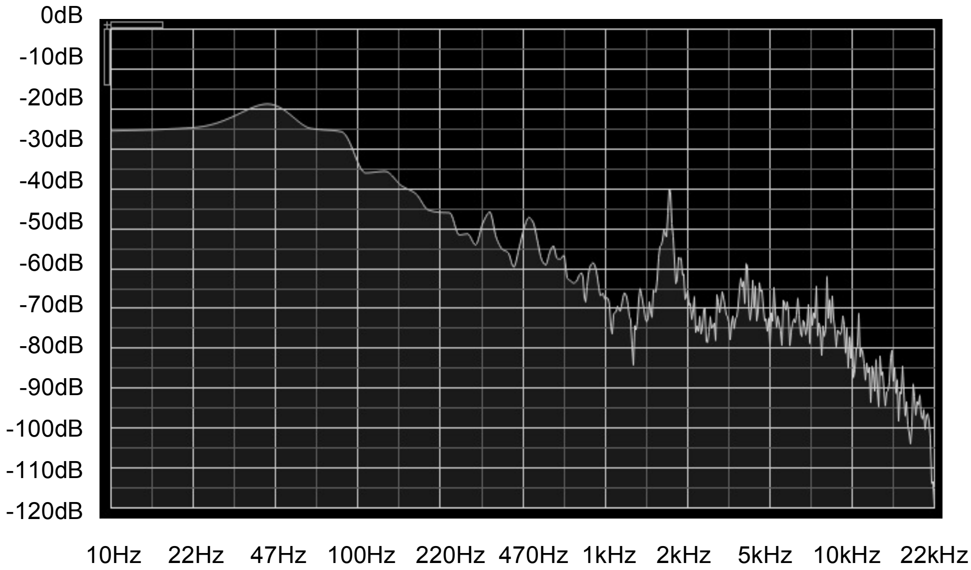
Control Sound for Experiments 4 (HFF), 5 (LA) and 6 (LAHFF)



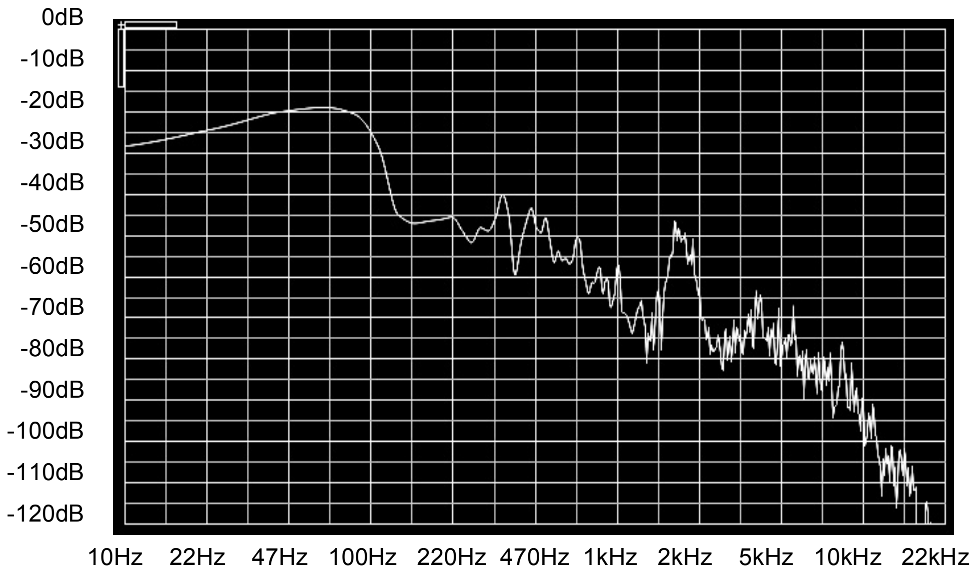
Experimental Sound for Experiment 4 (HFF - High Frequency Filtering) – Initial S3D object position

Experiments 4 (HFF), 5 (LA), 6 (LAHFF)

Clip 2



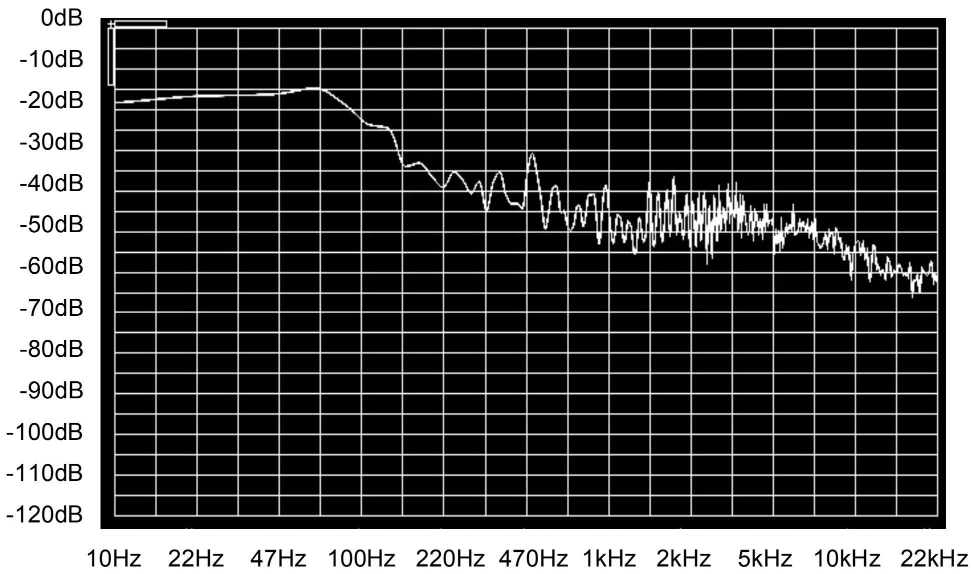
Experimental Sound for Experiment 5 (LA – Volume Level Alteration) – Initial S3D object position



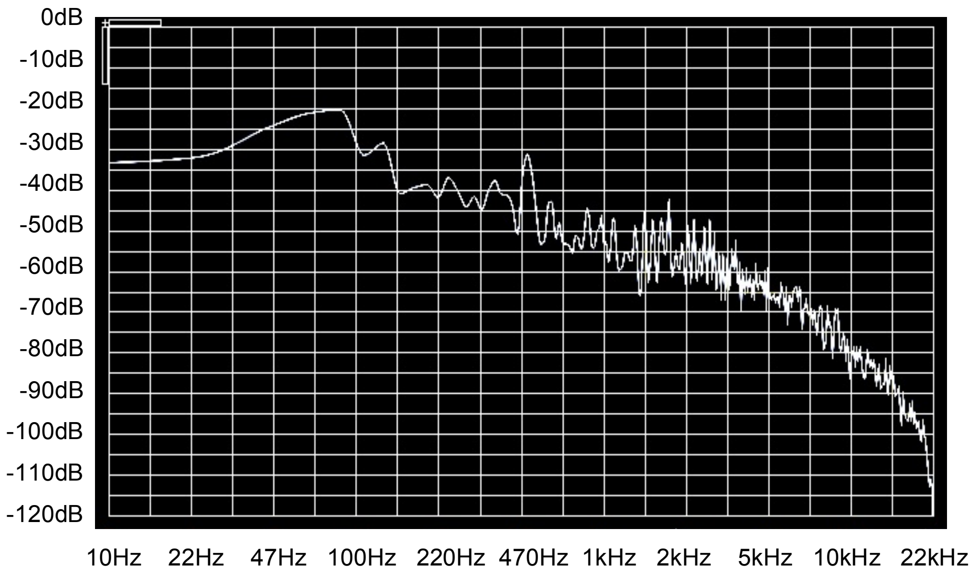
Experimental Sound for Experiment 6 (LAHFF – Volume Level Alteration and High Frequency Filtering) – Initial S3D object position

Experiments 4 (HFF), 5 (LA), 6 (LAHFF)

Clip 3



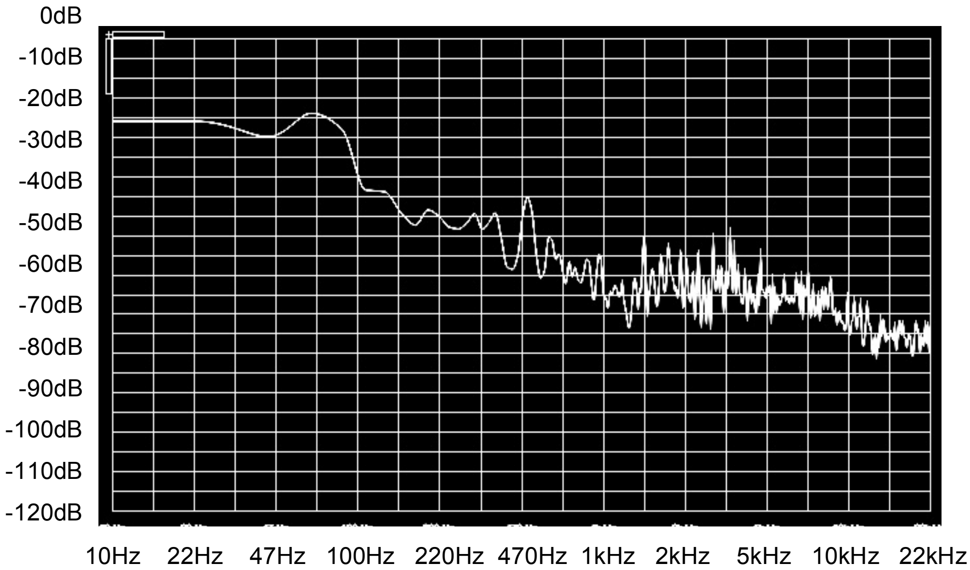
Control Sound for Experiments 4 (HFF), 5 (LA) and 6 (LAHFF)



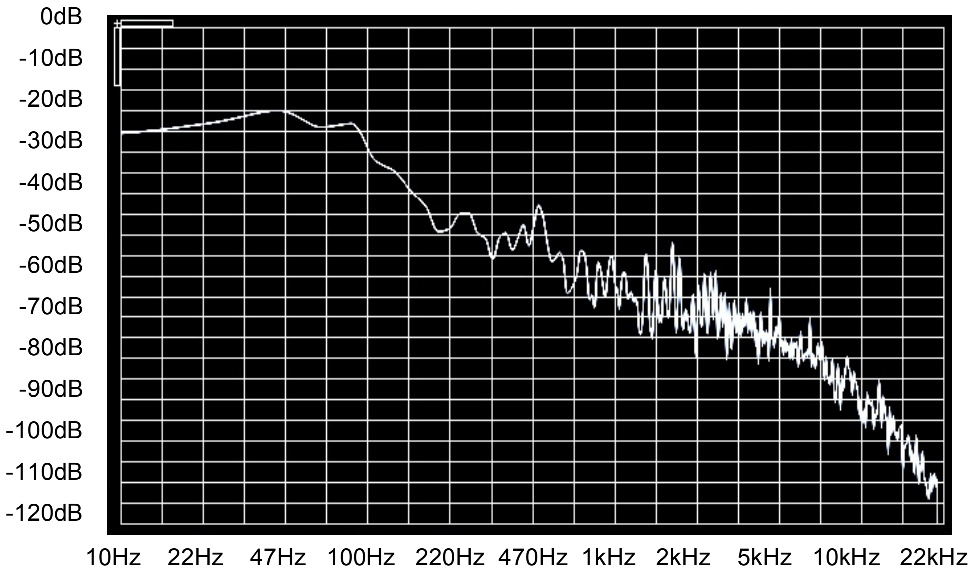
Experimental Sound for Experiment 4 (HFF - High Frequency Filtering) – Initial S3D object position

Experiments 4 (HFF), 5 (LA), 6 (LAHFF)

Clip 3



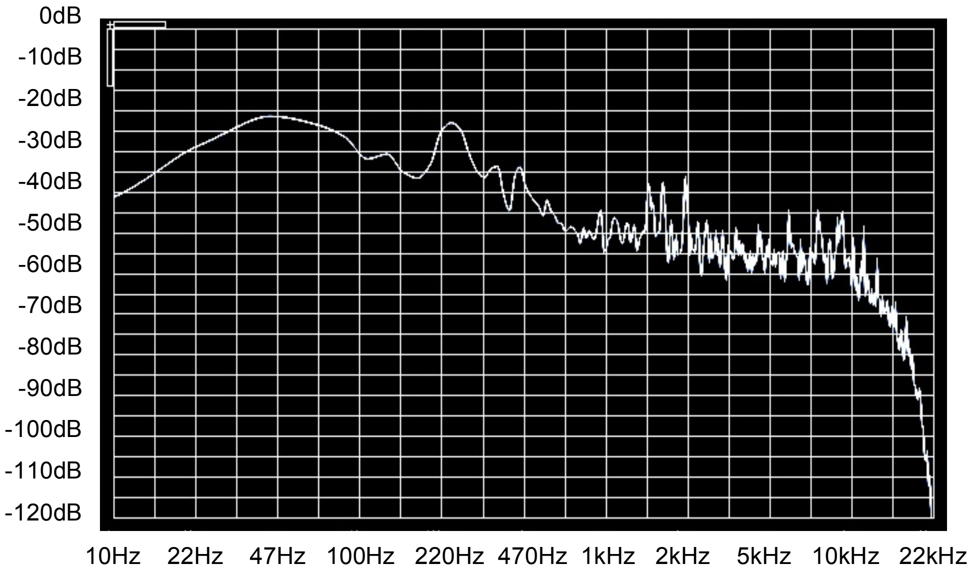
Experimental Sound for Experiment 5 (LA – Volume Level Alteration) – Initial S3D object position



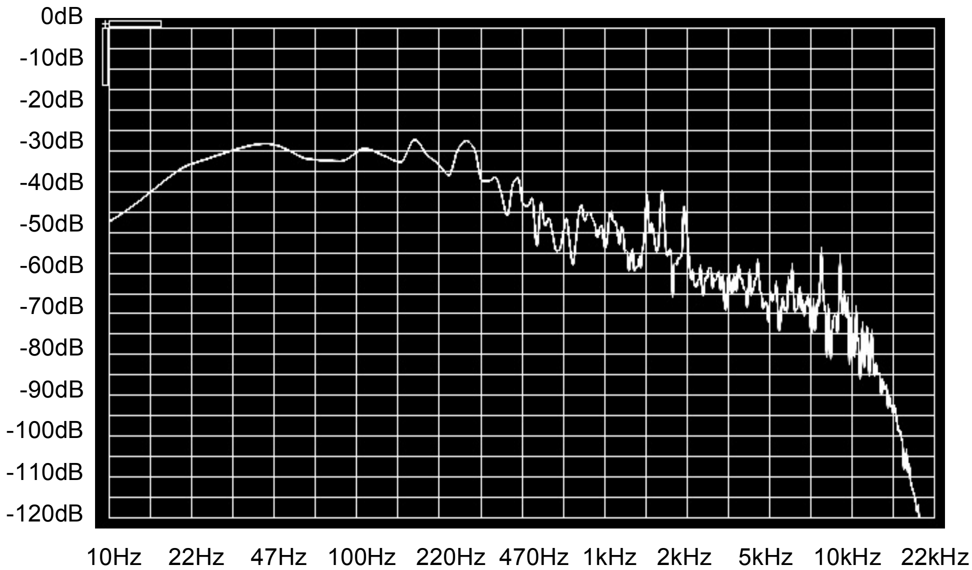
Experimental Sound for Experiment 6 (LAHFF – Volume Level Alteration and High Frequency Filtering) – Initial S3D object position

Experiments 4 (HFF), 5 (LA), 6 (LAHFF)

Clip 4



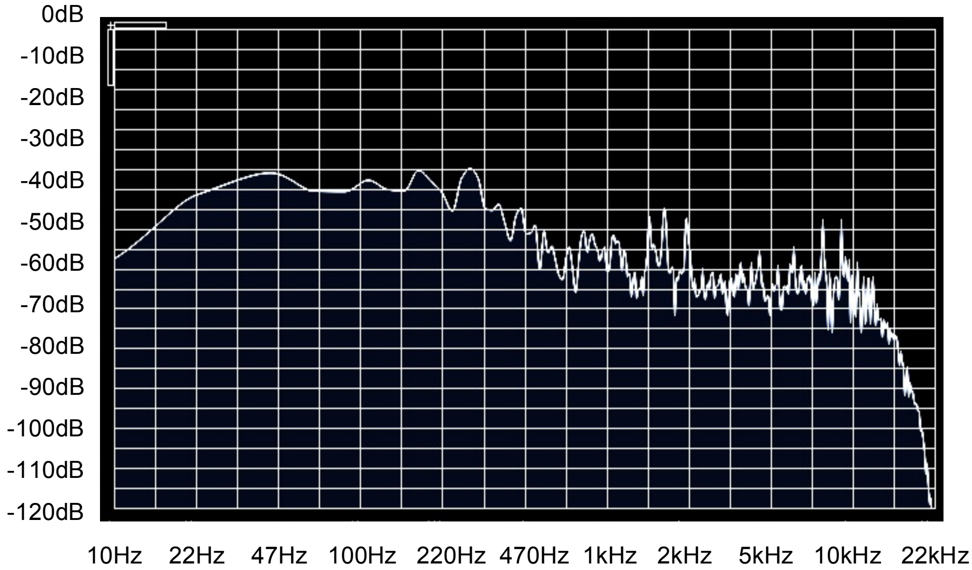
Control Sound for Experiments 4 (HFF), 5 (LA) and 6 (LAHFF)



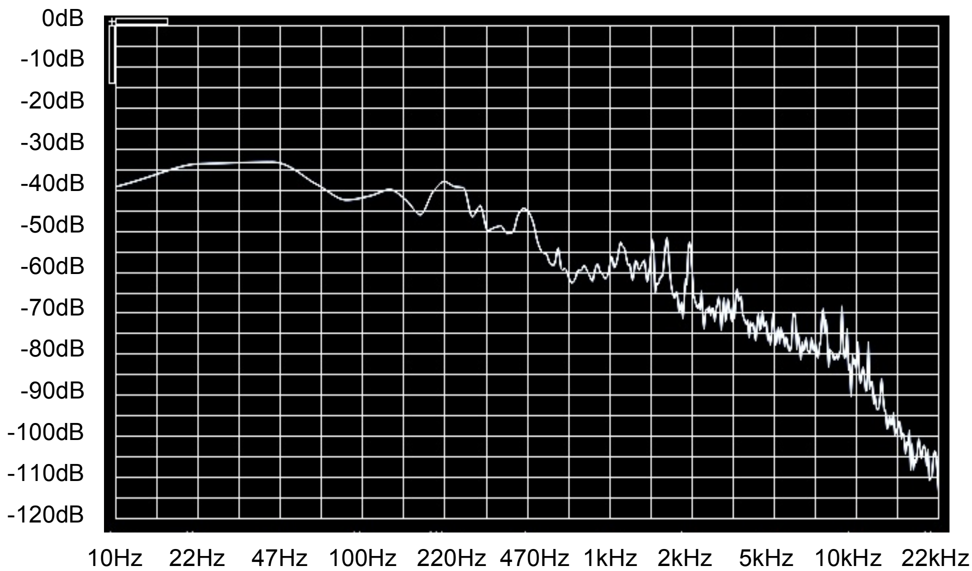
Experimental Sound for Experiment 4 (HFF - High Frequency Filtering) – Initial S3D object position

Experiments 4 (HFF), 5 (LA), 6 (LAHFF)

Clip 4



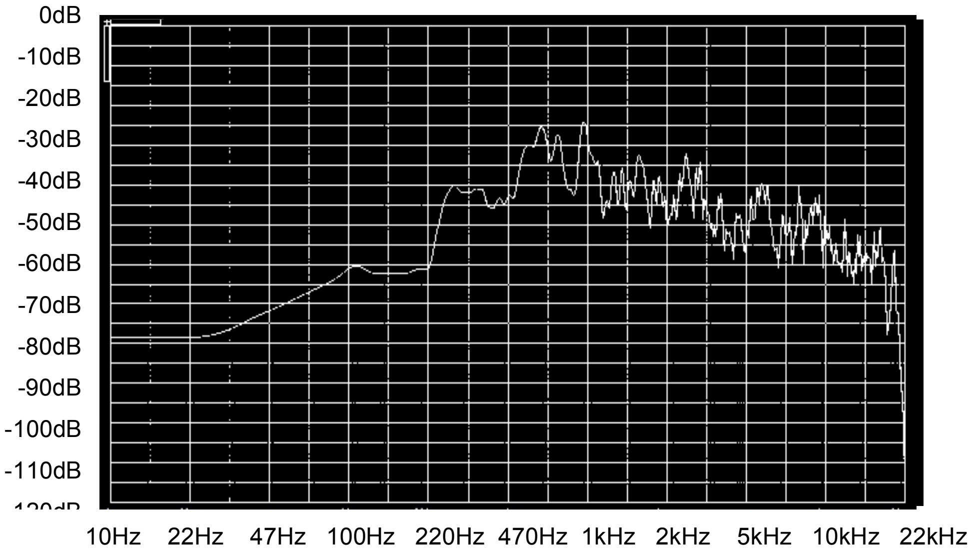
Experimental Sound for Experiment 5 (LA – Volume Level Alteration) – Initial S3D object position



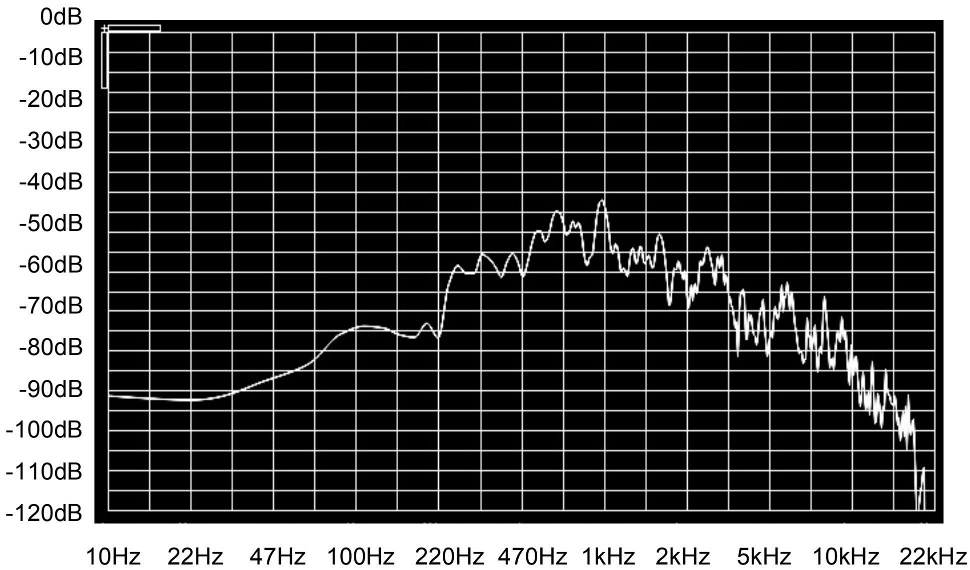
Experimental Sound for Experiment 6 (LAHFF – Volume Level Alteration and High Frequency Filtering) – Initial S3D object position

Experiment 7 (LAHFF)

Clip 1



Control Sound for Experiment 7 (LAHFF)

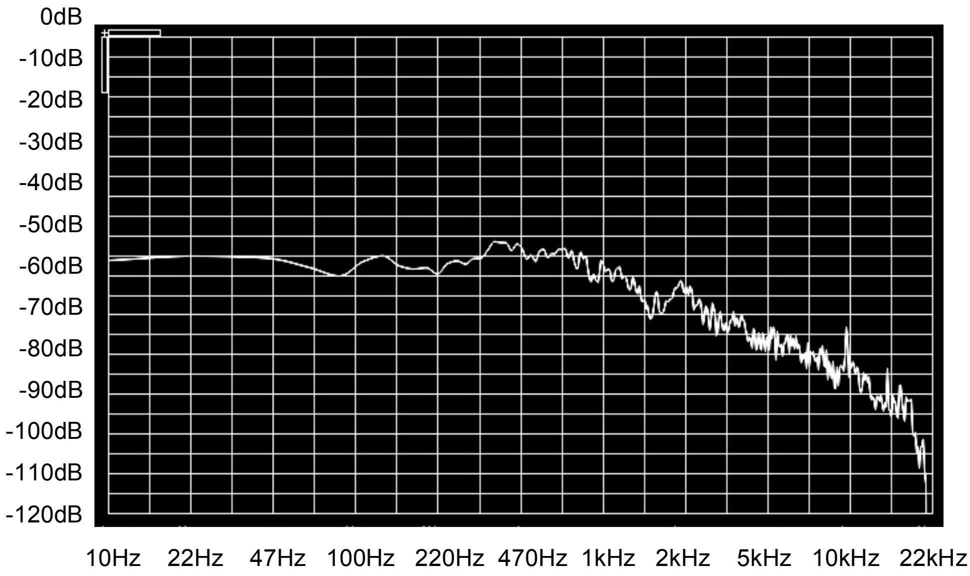


Experimental Sound for Experiment 7 (LAHFF - Volume Level Alteration and High Frequency Filtering)

Experiment 7 (LAHFF)

Clip 2

Control Sound for Experiment 7 (LAHFF)



Experimental Sound for Experiment 7 (LAHFF - Volume Level Alteration and High Frequency Filtering)

Appendix 3 –Experiment Documents and Forms

1. Experiment Participant Instructions

Experiments 1 (HFF), 2 (LA), 3NS (LAHFF)

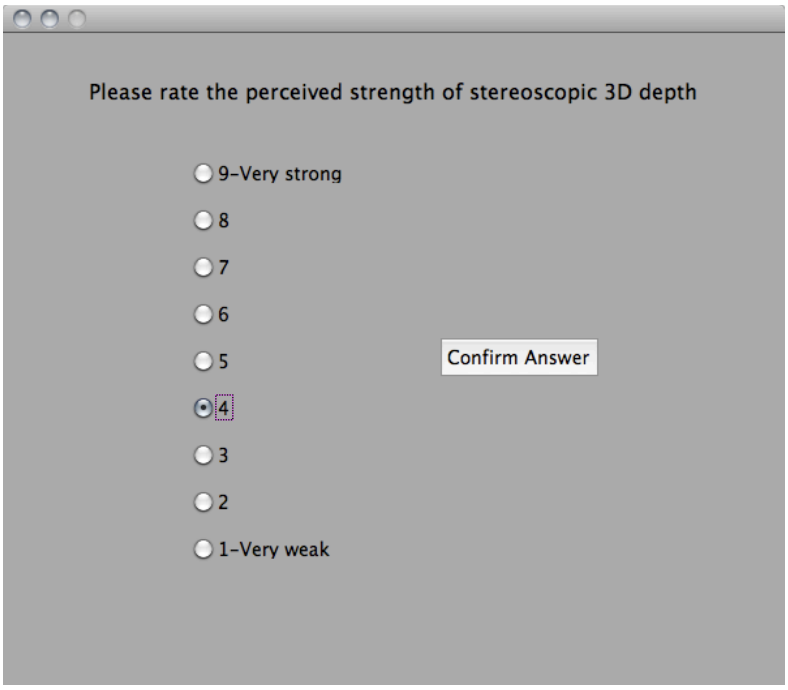
In stereoscopic (3D) cinema, the sense of depth (e.g. the stereoscopic 3D effect) can be manipulated at will in order to make a given scene appearing deeper or shallower.

In the current test you will be presented with a series of short stereoscopic 3D clips accompanied by appropriate audio stimuli (soundtrack).

After each clip is presented, you will be asked to rate the overall perceived 3D depth of scene by selecting the desired score on a rating scale between 1 and 9.

User input will be accepted through a specially designed digital answering form (**Picture 1**).

Picture 1



Once a selection has been made, you will have to press the ‘Confirm Answer’ button in order the supervisor of the experiment to load the next 3D clip. Once all clips have been viewed and rated you will receive a relevant notification on screen.

Thank you for your participation.

Experiments 3Sa / 3Sb / 7 (LAHFF)

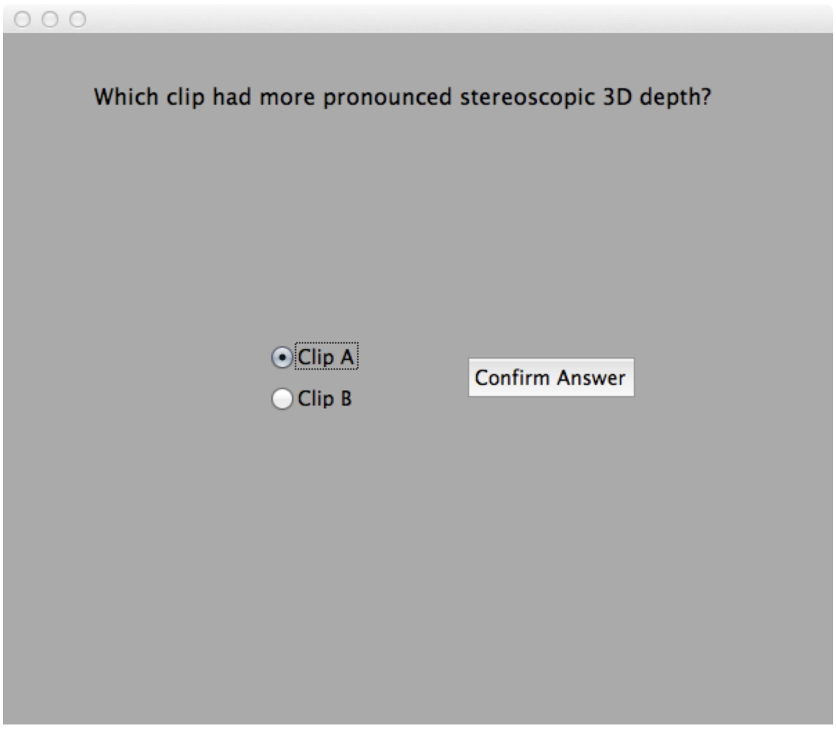
In stereoscopic (3D) cinema, the sense of depth (e.g. the stereoscopic 3D effect) can be manipulated at will in order to make a given scene appearing deeper or shallower.

In the current test you will be presented with a series of pairs of short stereoscopic 3D clips accompanied by appropriate audio stimuli (soundtrack).

Each pair consists of two versions of the same clip. After each pair of clips is presented, you will be asked to select which of the two versions (i.e. A or B) had the more pronounced overall stereoscopic 3D depth.

User input will be accepted through a specially designed digital answering form (**Picture 1**).

Picture 1



Once a selection has been made, you will have to press the 'Confirm Answer' button in order the supervisor of the experiment to load the next 3D clip. Once all clips have been viewed and rated you will receive a relevant notification on screen.

Thank you for your participation.

Experiments 4 (HFF), 5 (LA), 6NS (LAHFF)

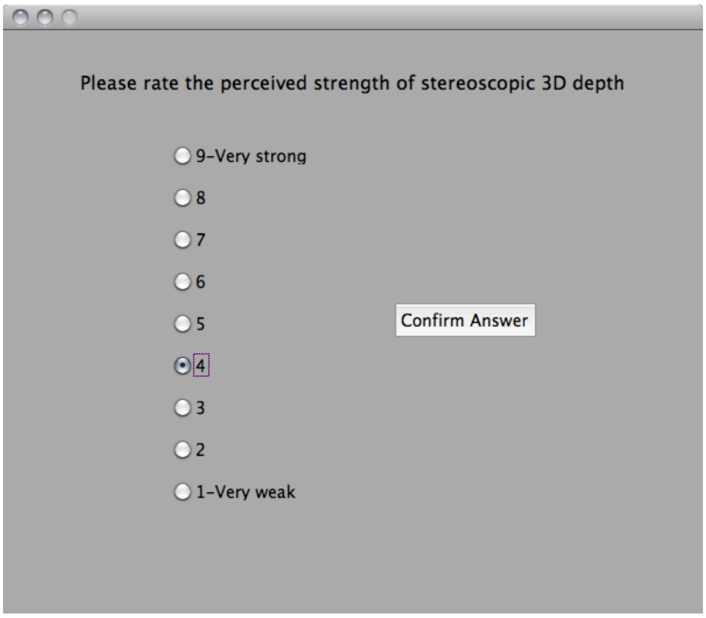
In stereoscopic (3D) cinema, the sense of depth (e.g. the stereoscopic 3D effect) of a given cinematic object can be manipulated at will in order to make a given scene appearing deeper or shallower.

In the current test you will be presented with a series of short stereoscopic 3D clips accompanied by appropriate audio stimuli (soundtrack).

The clips include a distinct movement of a cinematic object or character (e.g. a rocket) from inside the screen and towards you. After each clip is presented, you will be asked to rate the perceived depth of the movement of the S3D object by selecting the desired score on a rating scale between 1 and 9.

User input will be accepted through a specially designed digital answering form (**Picture 1**).

Picture 1



Once a selection has been made, you will have to press the 'Confirm Answer' button in order the supervisor of the experiment to load the next 3D clip. Once all clips have been viewed and rated you will receive a relevant notification on screen.

Thank you for your participation.

Experiment 6S (LAHFF)

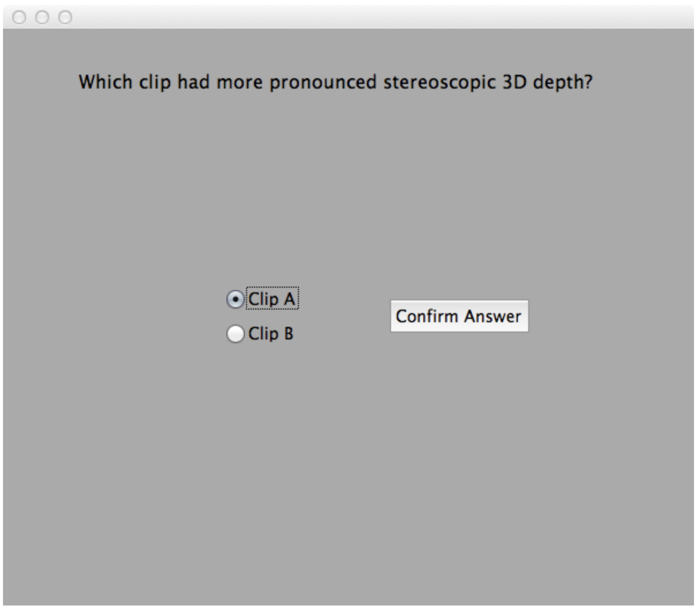
In stereoscopic (3D) cinema, the sense of depth (e.g. the stereoscopic 3D effect) of a given cinematic object can be manipulated at will in order to make a given scene appearing deeper or shallower.

In the current test you will be presented with a series of pairs of short stereoscopic 3D clips accompanied by appropriate audio stimuli (soundtrack).

Each pair consists of two versions of the same clip. The clips include a distinct movement of a cinematic object or character (e.g. a rocket) from inside the screen and towards you. After each pair of clips is presented, you will be asked to select in which of the two versions (i.e. A or B) the perceived depth of the movement of the S3D object was greater.

User input will be accepted through a specially designed digital answering form (**Picture 1**).

Picture 1



Once a selection has been made, you will have to press the 'Confirm Answer' button in order the supervisor of the experiment to load the next 3D clip. Once all clips have been viewed and rated you will receive a relevant notification on screen.

Thank you for your participation.

2. Informed Consent Form

The current project is related with the perceived sense of stereoscopic 3D depth in 3D cinema. The project is part of a postgraduate research study in the Department of Theatre, Film and TV in the University of York.

The data collected during the test, will be kept strictly confidential and will be available only to members of the research team. Excerpts from the results may be made part of the final research report, but under no circumstances will any identifying characteristics be included in the report.

Before we proceed with the test, we would like to emphasise the following:

- Your participation is entirely voluntary
- You are free to refuse to answer any question
- You are free to withdraw at any time
- You MUST ask to withdraw from the test if you, or a member of your family, have a history of epilepsy
- You MUST ask to withdraw from the test if any of the symptoms described in the following paragraphs are detected

EPILEPSY WARNING: A small percentage of the population may experience epileptic seizures when viewing certain types of TV images that contain flashing patterns of light. **The following people should consult a doctor before viewing in stereoscopic 3d:**

- Anyone with a history of epilepsy, or who has a family member with a history of epilepsy
- Anyone who has ever experienced epileptic seizures or sensory disturbances triggered by flashing light effects

You SHOULD NOT participate in the current test if any of the above applies.

GENERAL HEALTH WARNING: Some light patterns may induce seizures in persons with no prior history of epilepsy. **Discontinue stereoscopic 3d use if you experience any of the following symptoms while viewing stereoscopic 3d images:**

- Involuntary movements, eye or muscle twitching
- Muscle cramps
- Extensive nausea, dizziness, or queasiness
- Convulsions
- Disorientation, confusion, or loss of awareness of your surroundings

You SHOULD WITHDRAW from the test if any of the above is detected. Do not engage in any potentially hazardous activity (for example, driving a vehicle) until your symptoms have completely gone away.

By signing the current form and completing the test it is assumed that you have read and understood the above, and you agree to participate in the study.

Signature

3. Personal/Demographic Information Form

Gender:

Age:

Nationality:

Contact Email (optional):

Appendix 4 –Visual Clip Screenshots

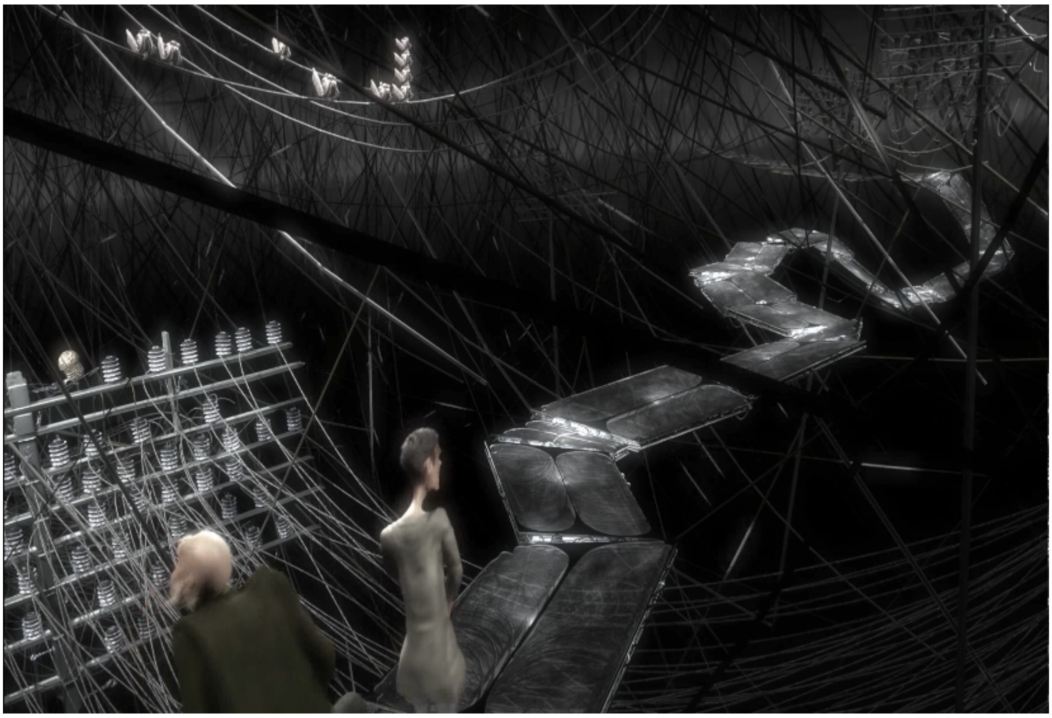
Experiments 1 (HFF), 2 (LA), 3 (LAHFF)

Clip 1_R (Rich S3D Content)

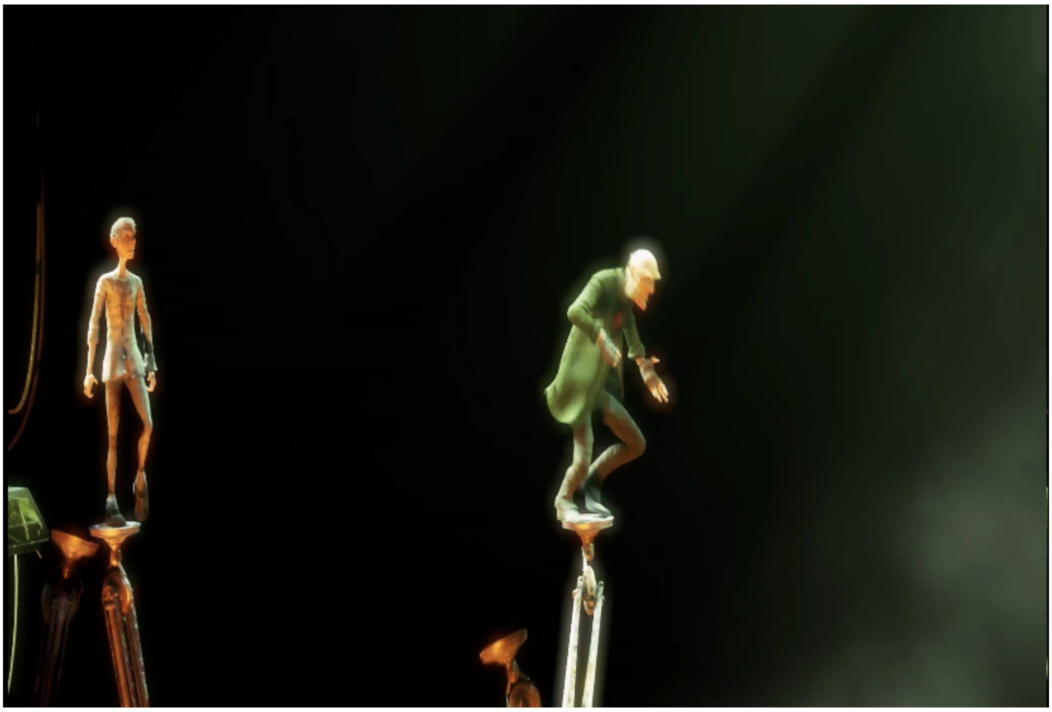
Clip 2_S (Simple S3D Content)

Experiments 1 (HFF), 2 (LA), 3 (LAHFF)

Clip 3_R (Rich S3D Content)



Clip 4_S (Simple S3D Content)



Permission to use original audiovisual material for the purposes of Experiments 4 (HFF, 5 (LA) and 6 (LAHFF) was obtained by the respective companies and copyright holders (nWave Pictures 2007; Red Star Studio 2010; Lightspeed Design 2014).

Experiments 4 (HFF), 5 (LA), 6 (LAHFF)

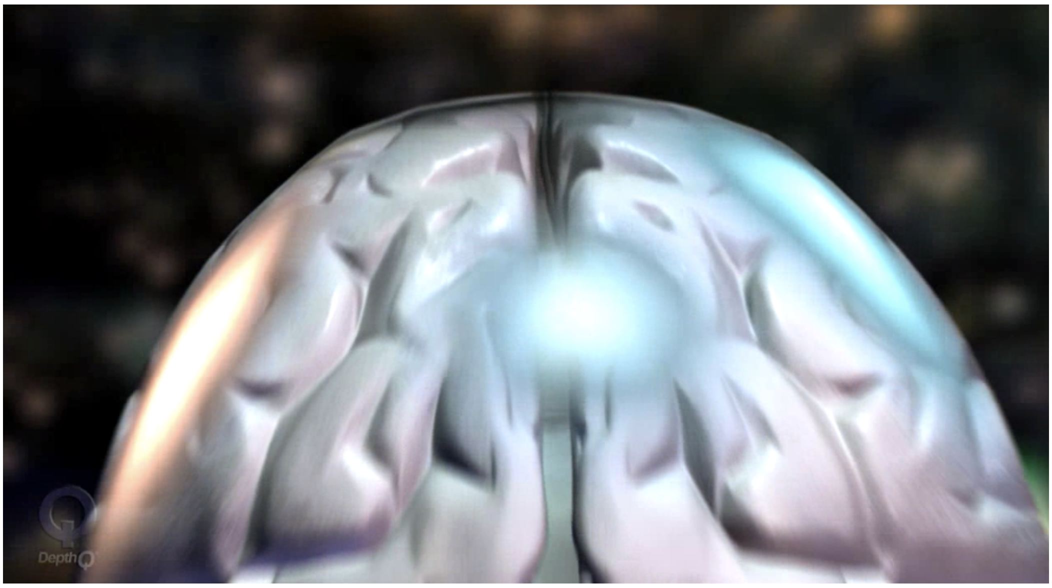
Clip 1 (Initial S3D object position)

Clip 1 (Final S3D object position)

Experiments 4 (HFF), 5 (LA), 6 (LAHFF)

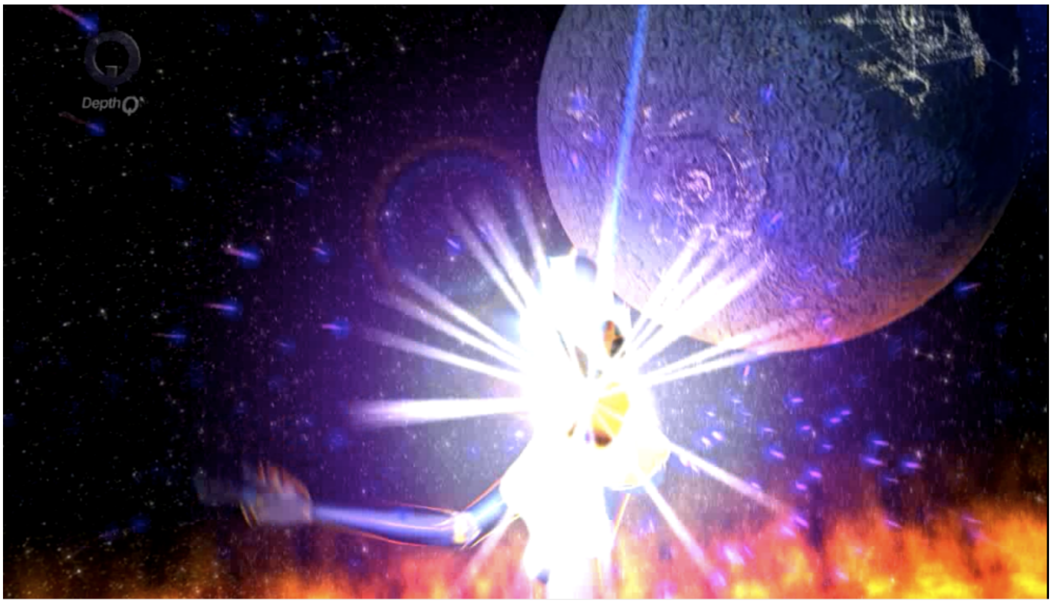
Clip 2 (Initial S3D object position)

Clip 2 (Final S3D object position)

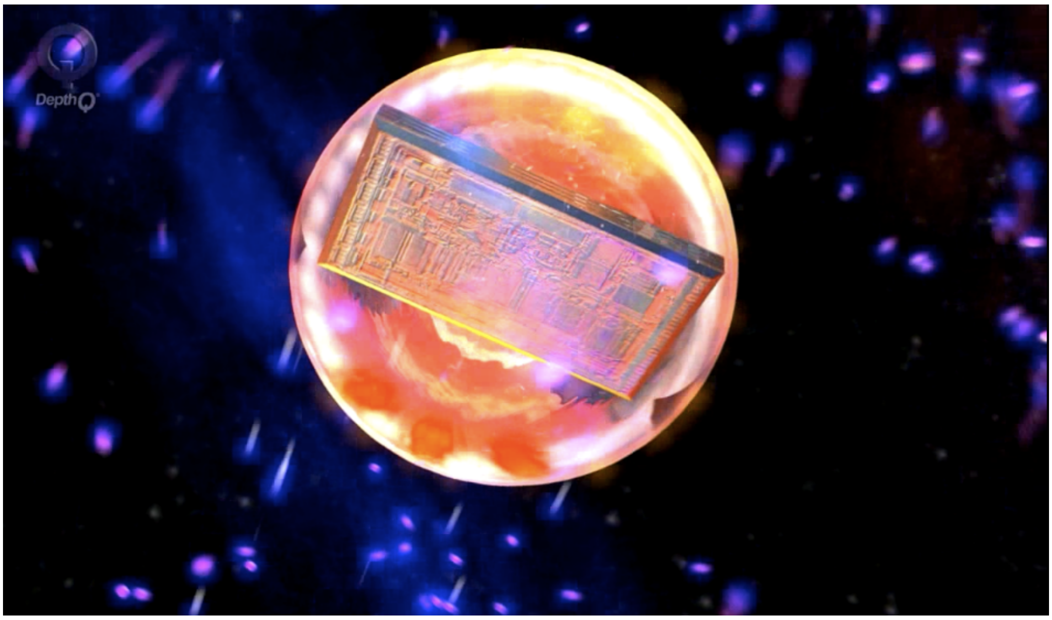


Experiments 4 (HFF), 5 (LA), 6 (LAHFF)

Clip 3 (Initial S3D object position)

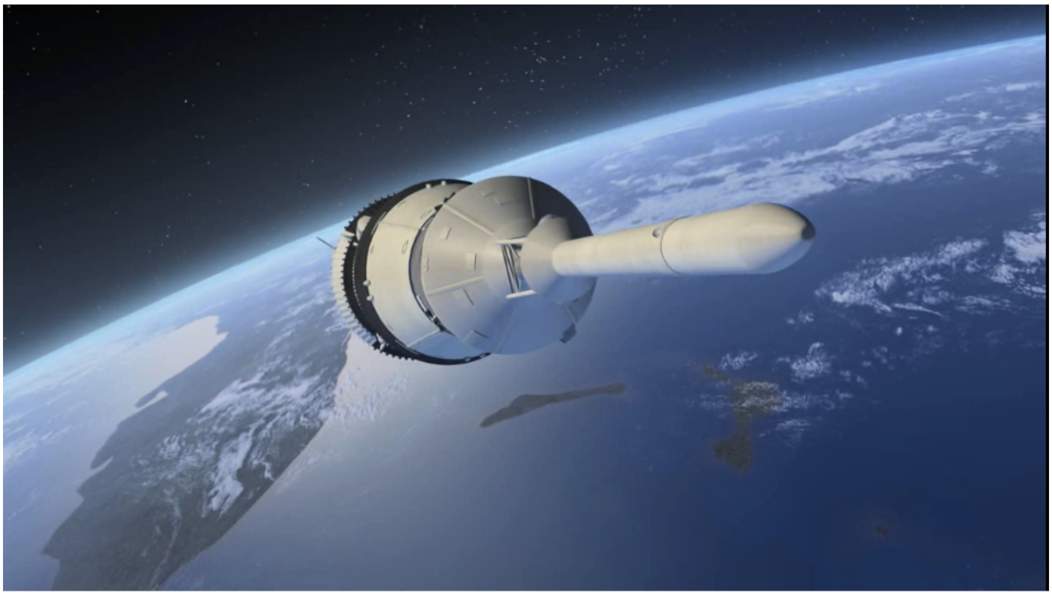


Clip 3 (Final S3D object position)



Experiments 4 (HFF), 5 (LA), 6 (LAHFF)

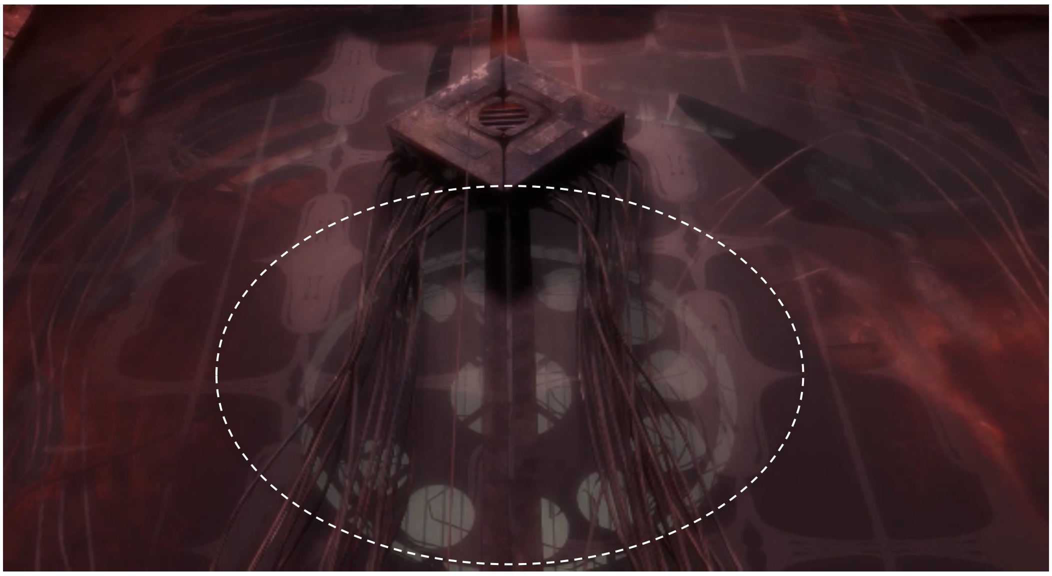
Clip 4 (Initial S3D object position)



Clip 4 (Final S3D object position)

Experiment 7 (LAHFF)

Clip 1



Clip 2

References

- 2020 3D Media (2008). *2020 3D Media – Spatial Sound and Vision*. 2020 3D Media. http://www.20203dmedia.eu [Last accessed on 31/08/2014].
- 20th Century Fox (2009). *Avatar.* 20th Century Fox Film Corporation and Dune Entertainment LLC.
- 3D Stereo Media (2010). *3D Stereo Media Summit.* 3D Stereo Media - Palais De Congress, Liege, Belgium, 8-10 December 2010.
- 3DIIM (2008). *3D, Immersive, Interactive Media Cluster*. 3dmedia-cluster.eu. http://www.3dmedia-cluster.eu/projects.html [Last accessed on 02/02/2011].
- 3DInYourFace (2010). *3D In Your Face: The Business of 3D In Film.* 3DInYourFace - Games, TV and Interactive Media Conference. 21-22 October 2010. Bradford, UK.
- Abrams, Nathan; Bell, Ian; Udris, Jan (2001). *Studying Film*. Arnold.
- André, Cédric; Embrechts, Jean-Jacques; Verly, Jacques G. (2010). *Adding 3D Sound to 3D Cinema: Identification and Evaluation of Different Reproduction Techniques.* Proceedings of the 2010 International Conference on Audio, Language and Image Processing, Shanghai, China, 23-25 November 2010.
- Apple Inc. (2014). *Logic Pro X.* Apple Inc. http://www.apple.com/uk/logic-pro [Last accessed on 20/04/2014].
- Audio Engineering Society (2010). *An Audio Timeline*. Audio Engineering Society. http://www.aes.org/aeshc/docs/audio.history.timeline.html [Last accessed 31/08/2014].

- AURO3D (2010). *Auro 3D: The Future in 3D Sound*. AURO Technologies. http://www.auro-3d.com [Last accessed on 31/08/2014].
- Autodesk (2008). *The Business and Technology of Stereoscopic Filmmaking.* Stereoscopic Filmmaking Whitepaper. Autodesk Inc.
- Back, Maribeth; Des, D. (1996). *Micro-Narratives in Sound Design: Context, Character, and Caricature in Waveform Manipulation.* Proceedings of the 3rd International Conference on Auditory Display, ICAD96.
- Bayon, David (2010). *Why 3D and modern filmmaking techniques don't mix*. Seminar by Buzz Hays, chief instructor for the Sony 3D Technology Center in Culver City, California. PC Pro: Real World Computing. Dennis Publishing. http://www.pcpro.co.uk/blogs/2010/08/11/3d-filmmaking-depth-of-field-lighting-and-editing [Last accessed on 31/08/2014].
- BEAST (2014). *Birmingham ElectroAcoustic Sound Theatre.* Electro Acoustic Studios, University of Birmingham.
- Beauchamp, Robin (2005). *Designing Sound For Animation*. Elsevier.
- Bekesy, Georg Von (1961). *Concerning the pleasures of observing and the mechanics of the inner ear*. Nobel Lecture, December 11, 1961.
- Belton, John (2004). *The Curved Screen*. Film History, Volume 16, pp. 277-285. John Libbey Publishing.
- Blauert, Jens (ed. 1997). *Spatial Hearing: The Psychophysics of Human Sound Localization*. The MIT Press, Massachusetts Institute of Technology.
- Blender Foundation, Netherlands Media Art Institute (2006). *Elephants Dream*. Blender Foundation / NMI. www.elephantsdream.org
  [Last accessed on 20/04/2014].

- Bolognini, Nadia; Frassinetti, Francesca; Serino, Andrea; Làdavas, Elisabetta (2005). *Acoustical vision of below threshold stimuli: interaction among spatially converging audiovisual inputs*. Experimental Brain Research, 2005, 160: 273–282.
- Boone, Marinus; Verheijen, Edwin; van Tol, Peter (1995). *Spatial Sound-Field Reproduction by Wave-Field Synthesis.* JAES, Volume 43, Issue 12 pp. 1003-1012. December 1995.
- Bordwell, David; Thompson, Kristin (ed. 2008*). Film Art An Introduction*. Eighth Edition. University of Wisconsin, McGraw Hill.
- Boyes, Christopher (2010a) *Sound Design: Christopher Boyes.* Dolby Labs. http://www.dolby.co.uk/consumer/experience/interviews/movie/christopher-boyes-dead-mans-chest.html [Last accessed on 02/02/2011].
- Boyes, Christopher (2010b). *Sound Design in Avatar.* Synthgear. http://www.synthgear.com/2010/video/sound-design-in-avatar [Last accessed on 31/08/2014].
- Braudy, Leo; Cohen, Marshall (ed. 2009). *Film Theory and Criticism*. Oxford University Press.
- Bregman, S. Albert (1994). *Auditory Scene Analysis: The Perceptual Organization of Sound*. The MIT Press, Massachusetts Institute of Technology.
- British Film Institute (2013). *BFI IMAX*. British Film Institute. http://www.bfi.org.uk/bfi-imax [Last accessed on 31/08/2014].
- BSHAA (2011). *Online Hearing Test*. British Society of Hearing Aid Audiologists. http://www.bshaa.com [Last accessed on 31/08/2014].

- Cameron, James (2009). *James Cameron interview for Avatar*. The Telegraph. http://www.telegraph.co.uk/culture/film/6720156/James-Cameron-interview-for-Avatar.html [Last accessed on 31/08/2014].
- Chion, Michel (1994). *Audio-Vision: Sound on Screen*. Columbia University Press.
- Chion, Michel (1999). *The Voice in Cinema*. Columbia University Press.
- Chueng, Priscilla; Marsden, Phil (2002). *Designing Auditory Spaces to Support Sense of Place: The Role of Expectation*. Interactive Technology Research Group, School of Computing and Engineering, University of Huddersfield. Position Paper for The Role of Place in On-line Communities Workshop. CSCW2002, New Orleans, November 2002. http://www.acm.org/conferences/cscw2002 [Last accessed on 31/08/2014].
- CircleVision Group (2008). *Digital Circlevision Vs Disney Circlevision360*. MediaMind Digital Inc. www.circlevisiongroup.com [Last accessed on 02/02/2011].
- CJ 4Dplex (2014). *Corporate Profile*. CJ 4Dplex. http://www.cj4dx.com/corporate/corporateProfile.asp [Last accessed on 31/08/2014].
- Clark, Barry (2010). *3D Production and Post*. Entertainment Technology Center, Basic 3D Concepts. http://www.etcenter.org/etc-activities/projects/consumer-3d-experience-project/basic-3d-concepts [Last accessed on 02/02/2011].
- Cochran, Paul; Throop, Janet; Simpson, W.E. (1968). *Estimation of Distance of a Source of Sound.* The American Journal of Psychology, Vol. 81, No. 2, pp. 198-206, June 1968.

- Coleman, Paul D. (1963). *An Analysis Of Cues To Auditory Depth Perception in Free Space*. Psychological Bulletin, 1963, Vol. 60, No. 3, 302-315.
- Courville, Daniel (2010). *Daniel Courville: Ambisonic studio*. Ambisonic Studio. www.radio.uqam.ca/ambisonic [Last accessed on 31/08/2014].
- Cousins, Mark (2004). *The Story of Film*. Pavilion Books.
- Creative Commons (2013). *About the Licenses*. Creative Commons. http://creativecommons.org/licenses [Last accessed on 31/08/2014].
- CTIE Hargrave (2002). *Étienne-Jules Marey*. Aviation and Aeromodelling – Interdependent Evolutions and Histories. http://www.ctie.monash.edu.au/hargrave/marey.html [Last accessed on 31/08/2014].
- D-BOX Technologies Inc. (2014). *D-BOX Technology*. D-BOX Technologies Inc. http://www.d-box.com/en/industrial/technology/index.html [Last accessed on 01/05/2014].
- De Bruijn, Werner P.J.; Boone, Marinus M. (2002). *Subjective experiments on the effects of combining spatialized audio and 2D video projection in audio-visual systems*. Audio Engineering Society - 112th Convention, Munich, Germany, 2002 May 10–13.
- De Lancie, Philip (2000). *Surround for Picture: Beyond the Cineplex*. Millimeter - The Magazine of Motion Picture and Television Production, 28:1, January 2000, pp. 113-114, 116.

- Dickson, William Kennedy-Laurie; Dickson, Antonia (1970). *History of the kinetograph, kinetoscope and kinetophonograph*. New York Arno Press.
- Dimension3 (2011). *Dimension 3 EXPO*. Dimension 3, Paris, 24-26 May 2011.
- Dirks, Tim (2010). *Film History by Decade*. American Movie Classics LLC - Filmsite. http://www.filmsite.org [Last accessed on 31/08/2014].
- Disney/Pixar (2010). *Toy Story 3*. The Walt Disney Company. http://disney.go.com/toystory [Last accessed on 31/08/2014].
- DisneyWorld (2010). *Attractions at Epcot.* The Walt Disney Company. http://disneyworld.disney.go.com/parks/epcot/attractions [Last accessed on 31/08/2014].
- Dix, Andrew (2008). *Beginning Film Studies.* Manchester University Press.
- Dolby Laboratories (1999). *Surround Sound Past, Present, and Future: A history of multichannel audio from mag stripe to Dolby Digital.* Dolby Laboratories Inc.
- Dolby Laboratories (2000). *5.1 Channel Production Guidelines.* Issue 1. S00/12957. Dolby Laboratories Inc.
- Dolby Laboratories (2005). *Dolby Surround Mixing Manual* - Issue 2. Dolby Laboratories Inc.
- Dolby Laboratories (2010). *Dolby Surround 7.1.* Dolby Laboratories Inc. http://www.dolby.com/us/en/technologies/dolby-surround-7-1.html [Last accessed on 02/02/2011].
- DTS Digital Entertainment (2010). *Milestones.* DTS Digital Entertainment. http://www.dts.com/Corporate/About_Us/Milestones.aspx [Last accessed on 02/02/2011].

- Dykhoff, Klas (2008). *About the perception of sound*. Dramatiska Institutet, University College of Film, Radio, Television and Theatre, Stockholm.
- Ecker, J. Adam; Heller, M. Laurie (2004). *Audio-Visual Cue Combination in Depth Perception*. Journal of Vision, August 13, 2004 Vol. 4, No. 8, Article 699.
- Elen, Richard (2007). *Ambisonic.net, where surround sound comes to life*. Ambisonic.net. www.ambisonic.net [Last accessed on 31/08/2014].
- Faul F, Erdfelder E, Lang AG, Buchner A (2007). *G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences*. Behavior Research Methods, 39, 175-191.
- Field, Ambrose (2002). *Breaking and making the rules: sound design in 6.1*. Audio Engineering Society - 113th Convention, Los Angeles, Ca, Usa, 2002 October 5–8.
- Field, Andy (ed. 2005). *Discovering Statistics Using SPSS*. SAGE Publications.
- Freeman, Elliot; Driver, Jon (2008). *Direction of Visual Apparent Motion Driven Solely by Timing of a Static Sound.* Current Biology, Volume 18, Issue 16, August 26, 2008, pp. 1262–1266.
- G.R.A.S. Sound & Vibration (2013). *Kemar-rejuvenated*. G.R.A.S. Sound and Vibration. kemar.us [Last accessed on 31/08/2014].
- Gardner, Brian (2009). *Perception and The Art of 3D Storytelling*. Creative Cow. http://magazine.creativecow.net/article/perception-and-the-art-of-3d-storytelling [Last accessed on 10/03/2011].
- Gardner, William (2005). *Spatial Audio Reproduction: Towards Individualized Binaural Sound*. Wave Arts Inc. Arlington, Massachusetts.

- Garity, W.M.E; Hawkins, J.N.A. (1941). *Fantasound*. Journal of the Society of Motion Picture Engineers. The American Widescreen Museum, 1998.
- Genelec (2014). *Genelec 8050A Bi-Amplified Loudspeaker System*. Genelec Oy. http://www.genelec.com/products/previous-models/8050a [Last accessed on 31/08/2014].
- Gernsback, Hugo (ed. 1934). *Three Dimensional Sound Created*. Science and Mechanics, April 1934.
- Glasgal, Ralph (2010). *The Blumlein Conspiracy*. Ambiophonics.org. http://www.ambiophonics.org/209.238.98.180:21/blumlein_conspiracy.htm [Last accessed on 10/03/2011].
- Groupe Archisoft (2013). *Zappiti Player.* Groupe Archisoft. http://zappiti.com/zappiti_player.html [Last accessed on 31/08/2014].
- Guichardan, Thibaud (2001). *Interactions between surround sound level and the immersive feeling in the multichannel movie experience.* AES - 111th Convention, 2001 September 21–24, New York, NY, USA.
- Haines, Thomas; Hooker, Michael (1997). *Multichannel Audio Dramas: A Proposal*. AES - 102nd Convention, 1997 March 22-25, Munich, Germany.
- Harris, Peter (ed. 2008). *Designing and Reporting Experiments in Psychology.* Open University Press, McGraw-Hill Education.
- Heinrich-Heine-Universität (2013). *G*Power: Statistical Power Analyses for Windows and Mac.* Heinrich-Heine-Universität Düsseldorf. http://www.gpower.hhu.de/en.html. [Last accessed on 31/08/2014].
- Hirst, Jonathan Mark (2006). *Spatial Impression in Multichannel Surround Sound Systems*. Research Institute for the Built and Human Environment. University of Salford, UK.

- Holocinema (2005). *Holographic And 3-D Digital Cinema: From Dream To Reality.* Holocinema.com. Holographic and 3-D Digital Technologies. http://www.holocinema.com/3dcinema_oview.html [Last accessed on 10/03/2011].
- Holography (2008). *Holographic Studio of NIKFI.* Holography.ru. http://www.holography.ru/nikfieng.htm [Last accessed on 31/08/2014].
- Holt, Nigel James (1997). *Perceptual Lateralisation Of Audio-Visual Stimuli.* Dissertation - The University of York, September 1997. Department of Psychology, The University of York.
- IMAX Corporation (2014). *IMAX Corporate Overview*. https://www.imax.com/corporate [Last accessed on 31/08/2014].
- Institut de Recherche et Coordination Acoutique/Musique (2014). *Research*. IRCAM. http://www.ircam.fr/recherche.html?L=1 [Last accessed on 31/08/2014].
- Institute of Electronic Music and Acoustics (2014). *Research*. University of Music and Performing Arts Graz. http://iem.kug.ac.at/en/research.html [Last accessed on 31/08/2014].
- IOSONO (2010). *IOSONO - The Future of Spatial Audio*. http://www.iosono-sound.com [Last accessed on 31/08/2014].
- ITU (1998). *ITU-T Recommendation P.911.* Subjective audiovisual quality assessment methods for multimedia applications. ITU Telecommunication Standardization Sector (ITU-T).
- Jackson, Blair (2010). *Avatar.* Mix Magazine. New Bay Media, LLC. http://mixonline.com/post/features/avatar-0110/index3.html [Last accessed on 31/08/2014].

- Kaye, Deena; Lebrecht, James (ed. 2009). *Sound and Music for the Theatre*. Focal Press.
- Kearney, Gavin; Gorzel, Marcin; Boland, Frank; Rice, Frank (2010). *Depth Perception in Interactive Virtual Acoustic Environments Using Higher Order Ambisonic Soundfields*. Proceedings of the 2nd International Symposium on Ambisonics and Spherical Acoustics, May 6-7 2010, Paris France.
- Kerins, Mark (2006). *Narration in the Cinema of Digital Sound.* The Velvet Light Trap - Number 58, Fall 2006, pp. 41-54.
- Kerins, Mark (2007). *Constructing the Diegesis in a Multi-Channel World.* Discourses on Diegesis. OFFSCREEN, Vol. 11, Nos. 8-9, Aug/Sept 2007.
- Kerins, Mark (2011). *Beyond Dolby Stereo: Cinema in the Digital Sound Age*. Indiana University Press.
- Kermode, Mark (2010). *Interview with Martin Scorsese*. The Guardian - The Observer. Sunday, 21 November 2010. http://www.guardian.co.uk/film/2010/nov/21/martin-scorsese-3d-interview-kermode [Last accessed on 19/02/2011].
- Kitagawa, Norimichi; Ichihara, Shigeru (2002). *Hearing visual motion in depth*. Nature, Vol. 416, 14 March 2002.
- Kubrick, Stanley (1968). *2001: A Space Odyssey*. Metro-Goldwyn-Meyer (MGM), Stanley Kubrick Productions.
- Langkjær, Birger (1997). *Spatial Perception and Technologies of Cinema Sound.* Convergence, 1997, 3: 92.
- Lantz, Ed (2003). *Large Format Digital Cinema: Medium of the Future?* Trends in Leisure Entertainment (TiLE) Conference, Berlin, Germany, 2003.

- Lantz, Ed (2006). *Digital Domes and the Future of Large-Format Film.* LF Examiner, Vol. 9, No. 8, Summer 2006.
- Lelyveld, Phil (2009). *Executive Briefing: Basic Visual Perception Concepts Related to 3D Movies*. Consumer 3D Experience Lab, Entertainment Technology Center at USC.
- Lightspeed Design (2014). *DepthQ Player - S3D Demo Movies*. Lightspeed Design Inc. http://www.depthq.com/playerdemos.html
  [Last accessed on 21/05/2014].
- Lippert, Michael; Logothetis, Nikos; Kayser, Christoph (2007). *Improvement of visual contrast detection by a simultaneous sound*. Brain Research 1173, 2007, 102–109. Elsevier.
- Little DA, Mershon HD, Cox HP (1992). *Spectral content as a cue to perceived auditory distance*. Perception, 1992, Vol. 21, pp. 405-416.
- Loughney, Patrick (1999). *Domitor Witnesses the First Complete Public Presentation of the Dickson Experimental Sound Film in the Twentieth Century*. Film History, Volume 11, pp. 400-403. John Libbey & Company.
- Malham, D.G. (1998). *Spatial Hearing Mechanisms and Sound Reproduction.* The University of York. http://www.ambisonic.net/
  [Last accessed on 21/05/2014].
- Mastoropoulou, Georgia (2006). *The Effect of Audio on the Visual Perception of High-Fidelity Animated 3D Computer Graphics*. Department of Computer Science, University of Bristol, UK.
- Mastoropoulou, Georgia; Debattista, Kurt; Chalmers, Alan; Troscianko, Tom (2005). *Auditory Bias of Visual Attention for Perceptually-Guided*

Selective Rendering of Animations. GRAPHITE 2005 (ACM SIGGRAPH), Dunedin, New Zealand, December 2005.

- McCluskey A, Lalkhen AG (2007). *Statistics III: Probability and Statistical tests*. Continuing Education in Anaesthesia, Critical Care & Pain, Vol. 7, No. 5, pp. 167-170.
- Meares, D.J. (1993). *Perceptual Attributes of Multichannel Sound.* Proceedings of the AES 12th International Conference: Perception of Reproduced sound, Copenhagen, Denmark, 1993 June 28-30.
- Media College (2011). *Depth Perception Test*. MediaCollege.com. http://www.mediacollege.com/3d/depth-perception/test.html. [Last accessed on 20/04/2014].
- Mendiburu, Bernard (2009). *3D Movie Making: Stereoscopic Digital Cinema from Script to Screen*. Elsevier/Focal Press.
- Merlin Entertainments (2014). *Our Company*. Merlin Entertainments. http://www.merlinentertainments.biz/company [Last accessed on 31/08/2014].
- Mershon, Donald H; Bower, John N (1979). *Absolute and relative cues for the auditory perception of egocentric distance*. Perception, 1979, Volume 8, pp. 311 -322.
- Microsoft (2014). *Xbox One*. http://www.xbox.com/en-GB/xbox-one/meet-xbox-one?xr=shellnav [Last accessed on 31/08/2014].
- Miller, Michael (2004). *The History of Surround Sound.* Informit. Pearson Education. http://www.informit.com/articles/article.aspx?p=337317
- [Last accessed on 31/08/2014].

- Mitchell, Doug (2010). *Timeline of Sound-on-Film: development of film and sound on film.* Middle Tennessee State University. http://web.archive.org/web/20050320171006/http://www.mtsu.edu/~dsmitche/rim458/Timeline/timeline.html [Last accessed on 31/08/2014].
- Moore CJ Brian (ed. 2003). *An Introduction to the Psychology of Hearing - Fifth Edition*. Emerald Group Publishing.
- Morton, Robert; Edwardz, David (2010). *Dead Boring 3d: Guerilla Stereo*. Australian Cinematographer Magazine, March 2010, pp. 30-37.
- Murch, M. Gerald (1973). *Visual and Auditory Perception*. The Bobbs-Merrill Company, Inc. Indianapolis, New York.
- Murch, Walter (2000). *Dickson Experimental Sound Film 1895*. Excerpt from The Cinema Audio Society Discussion Board, June 3, 2000. http://www.filmsound.org/murch/dickson.htm [Last accessed on 31/08/2014].
- Music Research Centre, The University of York (2014). *About the Music Research Centre*. Department of Music, The University of York. http://www.york.ac.uk/music/mrc/about [Last accessed on 01/05/2014].
- National Media Museum (2009). *A Very Short History of Cinema.* Nationalmediamuseum.org.uk. www.nationalmediamuseum.org.uk/pdfs/cinehistory.pdf [Last accessed on 10/03/2011].
- Nobelprize (2010). *The Nobel Prize in Physics 1971*. Nobelprize.org. http://nobelprize.org/nobel_prizes/physics/laureates/1971 [Last accessed on 31/08/2014].

- Nowell-Smith, Geoffrey (edited by, 1996). *The Oxford History of World Cinema*. Oxford University Press.
- Nudds, Matthew (2007). *Auditory Perception and Sounds*. Edinburgh Research Archive, The University of Edinburgh.
- Nvidia Corporation (2014). *Nvidia 3D Vision: Nvidia 3DTV Play Overview.* Nvidia Corporation. http://www.nvidia.co.uk/object/3dtv-play-uk.html [Last accessed on 31/08/2014].
- Nwave Pictures (2007). *Fly Me to the Moon 3D.* Nwave Pictures. http://www.nwave.com/films/3d-feature-films/13-films-cat/feature/7-fly-me-to-the-moon-3d [Last accessed on 31/08/2014].
- Oellers, Helmut (2010). *Wave Field Synthesis*. Holophony.net. http://www.holophony.net/Wavefieldsynthesis.htm [Last accessed on 31/08/2014].
- Optoma Europe Ltd. (2012). *Optoma Projectors*. Optoma Europe Ltd. http://www.optoma.co.uk [Last accessed on 31/08/2014].
- Orpen, Valerie (2003). *Film Editing: The Art of the Expressive*. Wallflower Press.
- Payatagool, Chris (2008). *First Live and Interactive 3D Hologram Transmitted Across Europe, From London to Berlin*. Telepresence Options. http://www.telepresenceoptions.com/2008/12/first_live_and_interactive_3d [Last accessed on 31/08/2014].
- Red Star Studio (2010). *Dracula 4D*. Red Star Studio Ltd.
- Riehle, Jake (2010). *TOY STORY 3: Exclusive Interview with Tom Myers, Michael Semanick, and Al Nelson.* Designing Sound: Art and Technique of

Sound Design. http://designingsound.org/2010/06/toy-story-3-exclusive-interview-with-tom-myers-michael-semanick-and-al-nelson
[Last accessed on 31/08/2014].

- Robson, Colin (ed. 1994). *Experiment, Design and Statistics in Psychology*. Penguin Books.
- Rumsey, Francis (2001). *Spatial Audio*. Focal Press.
- Rydstrom, Gary (2010). *Interview with Mixer Gary Rydstrom of Star Wars: Episode I*. Dolby Laboratories Inc.
- Sani, Fabio; Todman, John (2006). *Experimental Design and Statistics for Psychology: A First Course*. Blackwell Publishing.
- Scientific American Supplement (1900). *Illustration of the Cineorama balloon simulation at the 1900 Paris Exposition*. Scientific American Supplement No.1287.
- Sergi, Gianluca (1999). *The Sonic Playground: Hollywood Cinema and its Listeners.* Filmsound.org. http://filmsound.org/articles/sergi/index.htm
[Last accessed on 31/08/2014].
- Sergi, Gianluca (2004). *The Dolby Era: Film Sound in Contemporary Hollywood.* Manchester University Press.
- Shams, Ladan; Kim, Robyn (2010). *Crossmodal influences on visual perception*. Physics of Life Reviews. Volume 7, Issue 3, September 2010, Pages 269-284.
- Shatnoff, Judith (1967). *Expo 67: A Multiple Vision.* Film Quarterly, Vol. 21, No. 1, 2-13. The University of California Press.
- Shaw, Jon; Lantz; Ed (1998). *Dome Theaters: Spheres of Influence*.
Published in TiLE Proceedings, 1998, pp. 59-65.

- Shelton, B; Searle, C. (1980). T*he Influence of Vision on the Absolute Identification of Sound-Source Position*. Perception & Psychophysics, Volume 28, Issue 6, 1980, pp. 589-596.
- Smith, Dave (1996). *Disney A to Z: the official encyclopedia*. Hyperion - The University of Michigan.
- Snowden, Robert; Thompson, Peter; Troscianko, Tom (2006). *Basic Vision: An Introduction to Visual Perception*. Oxford University Press.
- Sokol, Mike (2005). *Surround Sound Mixing Techniques*. PRIMEDIA, Business Magazines & Media Inc. http://www.digifreq.com/digifreq/article.asp?ID=23 [Last accessed on 31/08/2014].
- Solomon, Kate (2013). *The sound of silence: How Dolby Atmos brought gravitas to Gravity.* Techradar. Future Publishing Limited. http://www.techradar.com/news/audio/the-sound-of-silence-how-dolby-atmos-brought-gravitas-to-gravity-1201069 [Last accessed on 31/08/2014].
- Sony Computer Entertainment Europe (2014). *Playstation 4*. Sony Computer Entertainment Europe. http://uk.playstation.com/ps4/. [Last accessed on 31/08/2014].
- Strauss Jr., Johann (1867). *The Blue Danube*. Op.314
- Studio for Electro-Instrumental Music (2014). *Education*. Studio for STEIM. http://steim.org/education [Last accessed on 31/08/2014].
- The Academic Server, Cleveland State University (2010). *Who invented movies?* The Academic Server, Clevelant State University.

- The Bill Douglas Cinema Museum (2014). *Optical Toys.* The Bill Douglas Cinema Museum, The University of Exeter.
- The Harry Ransom Center (2007). *The First Photograph*. The Harry Ransom Centre, The University of Texas at Austin.
- The Hearing Care Centre (2014). *Hearing Loss.* The Hearing Care Centre. http://www.skeelshearing.co.uk/hearing_loss.html [Last accessed on 31/08/2014].
- The Library of Congress (2009). *Dickson experimental sound film*. American Memory. The Library of Congress.
- ThinkQuest (2011). *ThinkQuest Library: The Eye Exam test*. ThinkQuest.org. http://library.thinkquest.org/C005949/fun/eyechart.htm [Last accessed on 31/08/2014].
- Thompson, Kristin; Bordwell, David (ed. 2003). *Film History: An Introduction*. McGraw Hill.
- THX Limited (2010). *Cinema Certification*. THX Limited. http://www.thx.com/professional/cinema-certification [Last accessed on 31/08/2014].
- Turner, Amy; Berry, Jonathan; Holliman, Nick (2011). *Can the perception of depth in stereoscopic images be influenced by 3D sound?* Stereoscopic Displays and Applications XXII, Proceedings of SPIE, Vol. 7863, 15 February 2011.
- Urbaniak, C. Geoffrey; Plous, Scott (2008). *Research Randomizer.* www.randomizer.org [Last accessed on 31/08/2014].

- Vanhoutte, Kurt; Joris, Eric; Debackere, Brecht; Wynants, Nele (2010). *Cinematography of surround video, assuming a passive spectator*. 2020 3D Media. http://www.20203dmedia.eu/index.htm
  [Last accessed on 31/08/2014].
- VIEW (2010). *VIEW Conference.* VIEW, International Computer Graphics Conference 2011. 25-28 October 2011, Turin, Italy.
- Vroomen, Jean; de Gelder, Beatrice (2000). *Sound Enhances Visual Perception: Cross-modal Effects of Auditory Organization on Vision*. Journal of Experimental Psychology: Human Perception and Performance, Volume 26, Issue 6, 1583-1590.
- Walters, Jonathan (2002). *Technology*. Earlycinema.com – An Introduction to Early Cinema. http://www.earlycinema.com/technology/index.html
  [Last accessed on 31/08/2014].
- Warner Bros (2014). *Gravity*. Warner Bros Entertainment Inc. http://gravitymovie.warnerbros.com/#/home [Last accessed on 31/08/2014].
- White, Paul and Robjohns, Hugh (2001). *Surround Sound Mixing*. Sound on Sound. SOS Publications Group.
  http://www.soundonsound.com/sos/jan01/articles/surround.htm
  [Last accessed on 31/08/2014].
- Wierzbicki, James (2009). *Film Music: A History*. Taylor & Francis/Routlege.
- Wiggins, Bruce (2010). *Ambisonic Audio in Codemasters DiRT 2 Game.* School of Technology, University of Derby.
- Wilkinson, Scott - Source Interlink Media (2009). *True Stereo*. Sound & Vision. TEN: The Enthusiast Network.

http://www.soundandvision.com/content/true-stereo
[Last accessed on 31/08/2014].

- Winslow, Lance (2007). *Holographic Projection Technologies of the Future - "Killer Applications*". The Online Think Tank. www.WorldThinkTank.net [Last accessed on 31/08/2014].
- Woszczyk, Wieslaw; Bech, Soren; Hansen, Villy (1995). *Interaction between audio-visual factors in a home-theater system*. Audio Engineering Society - 99th Convention, October 6-9 1995, New York.
- Wyatt, Hilary; Amyes, Tim (ed. 2005). *Audio Post Production for Television and Film: An introduction to technology and techniques.* Third Edition. Focal Press / Elsevier.
- Xscape (2014). *XD Experience*. XScape. http://www.xscape.co.uk/yorkshire/concessions/xdexperienceyk [Last accessed on 31/08/2014].
- Zmoelnig MJ; Ritsch, W; Sontacchi, A (2003). *The IEM-Cube*. Institute of Electronic Music and Acoustics University of Music and Applied Arts, Graz, Austria. Proceedings of the 2003 International Conference on Auditory Display, Boston, MA, USA, 6-9 July.
- Zone Ray (2007). *Stereoscopic Cinema and the Origins of 3-D Film*. The University Press of Kentucky.
- Zyber, Josh (2013). *Cinerama v2.0*. High-def Digest, The Bonus View. http://www.highdefdigest.com/blog/screenx-theaters [Last accessed on 31/08/2014].