The Effect of Audio Cues and Sound Source Stimuli on Looming Perception

Sonia Wilkie

Submitted in partial fulfilment of the requirements of the Degree of Doctor of Philosophy

School of Electronic Engineering and Computer Science
Queen Mary, University of London

January 28, 2015

Statement of Originality

I, Sonia Wilkie, confirm that the research included within this thesis is my own work or that where it has been carried out in collaboration with, or supported by others, that this is duly acknowledged below and my contribution indicated. Previously published material is also acknowledged below.

I attest that I have exercised reasonable care to ensure that the work is original, and does not to the best of my knowledge break any UK law, infringe any third partys copyright or other Intellectual Property Right, or contain any confidential material.

I accept that the College has the right to use plagiarism detection software to check the electronic version of the thesis.

I confirm that this thesis has not been previously submitted for the award of a degree by this or any other university.

Signature:

Date:

Abstract

Objects that move in depth (looming) are ubiquitous in the real and virtual worlds. How humans interact and respond to these approaching objects may affect their continued survival in both the real and virtual words, and is dependent on the individual's capacity to accurately interpret depth and movement cues.

In computer-generated environments, including hyper and virtual reality, film, and gaming, these cues are often complex sounds with multiple audio cues that are creatively designed for maximum effect.

To accurately generate a dynamic and rich perception of looming objects, the design of such complex stimuli should be based on a firm scientific foundation that encompasses what we know about how people visually and aurally perceive events and interactions.

Conversely, many psychological studies investigating auditory looming depict the object's movement using simple audio cues, such as an increase in the amplitude, which are applied to tones that are not regularly encountered in the natural world, such as sine, triangle, or square waves. Whilst the results from these studies have provided important information on human perception and responses, technological advances now allow us to present complex audiovisual stimuli and to collect measurements on human perception and responses to real and hyper-real stimuli.

The research in this thesis begins to address the gap that exists between the research corpus and industry usage. This is initially accomplished by conducting a feature analysis of the audio cues and complex sounds constructed by sound designers for film scenes presenting objects moving in depth. This is followed by a perceptual study measuring human responses, both physical and emotional, to the complex audio cues designed for the film scenes.

Using physical models, we then select a number of audio cues for closer inspection and introduce the parameter of 'room reflections' as an audio cue. We investigate whether or not human responses to various audio cues differ when they are presented individually or in combination, or when they are applied to an artificial (square wave) sound source or a real world sound source. Finally, we test the capacity of these audio cues to bias multimodal auditory-visual perception of an approaching object.

For
Murdoch, Netta, and Peggy

Acknowledgements

First and foremost, I would like to thank my supervisor Dr. Tony Stockman for his years of sound advice, his cheerful disposition, for providing me with a great deal of freedom to explore the topics of my choice and work on the research areas that interest me the most.

Thank you to my examiners Dr. Peter Lennox and Prof. Peter McOwen. Your rigorous scrutiny of the thesis, grilling at the viva, and advice has been greatly appreciated.

Thank you to my thesis readers, Brandy Monteith & Brenton Wilkie, Greg Wilkie, Dogu Erdener, Alice Clifford, Nerissa Haby, Farzana Rita, and Prof. John Neuhoff.

To the past and present researchers of the Centre for Digital Music and the Cognitive Science research groups, you have made these four years easily my most pleasant research experience. It has been a privilege to work along side of you. Special thank you to Mark Plumbley, Marcus Pearce, Andrea Halpern, Josh Reiss, Anssi Klapuri, Mark Sandler, Simon Dixon, Elaine Chew, Panos Kudumakis, Andrew McPherson, Alice Clifford, Emmanouil Benetos, Michael Terrell, Georgy Fazekas, Peter Foster, Andy Simpson, Daniele Barchiesi, Mathieu Barthet, Magdalena Chudy, Joachim Ganseman, Steven Hargreaves, Robert Tubb, Boris Mailhe, Martin Morrell, Ken OHanlon, Thomas Wilmering, Adam Stark, Matthias Mauch, Asterios Zacharakis, Holger Kirchhoff, Dan Stowell, Luis Figueira, Andrew Robertson, Chunyang Song, Yading Song, Becky Stewart, Andrew Nesbit, Ireti Olowe, Oussama Metatla, Mi Tian, Tian Cheng, Bob Sturm, David Ronan, Joachim Ganseman, Duncan Menzies, Jordan Smith, Saoirse Finn, Pollie Barden, Ioana Dalca, Nela Brown, Doon McDonald, Sebastian Ewert, Chris Harte. The support staff Melissa Yeo, Sharon Cording, Hayley Cork, Colin Powell, Ho Huen, Nicole Gozna, Joan Hunter, Carly Wheeler, Tracy Bussoli, Stephen Hutchinson.

Thank you to Claire Luu, Rai Manku, Asma Rafiq, Ailar Javadi, Lefke Andreou, Arman Abrahamyan, Netta Wilkie, Peg McCarthy, and Murdoch Wilkie.

The Buildings Management crew and Hainaught Services for keepin' it real.

The support staff from the School of Geography, Andrew Loveland, Dan Israel, Ayesha Sabri, Elenor Schriber, Emma Schapcott, Helen Mclurg, Lis Parnham, and Anna Dullic-Sills. And the team of The E-Learning Unit, Stella Ekebuisi, Brett Lucas, Alysa Bramble, Gil Ritchie, Manoj Singh, Anisa Patel, Richard Chantler, Rume Begum, and Shelly Reader.

This research was funded by Queen Mary, The University of London.

Contents

List of Figures

List of Tables

List of Abbreviations and Symbols

2D	Two-dimensional
3D	Three-dimensional
Amp	Amplitude Increase
α	Alpha Level for Significance
$CI_{.95}$	Confidence Interval
dB	Decibel
F	F-test value
FFT	Fast Fourier Transform
Hz	Hertz
IAD	Inter-aural Differences
ILD	Inter-aural Level Differences
ITD	Inter-aural Temporal Differences
kph	Kilometres per Hour
M	Arithmetic Mean
m	Slope m value
MAX	MAX / MSP / Jitter
max	Maximum
ms	Millisecond
m/s	Metres per Second
min	Minute
min	Minimum
n	Number (in sample)
p	Probability
r	Pearson Product-moment Correlation Coefficient
r^2	Coefficient of Determination
Ref	Direct-to-Reflections Sound Energy Ratio
SD	Standard Deviation
SE	Standard Error
SPL	Sound Pressure Level
x^2	Chi-squared

Chapter 1

Introduction

One feature of computer-generated environments, including hyper and virtual reality, film, gaming, and simulators, is interacting with objects that move in space, particularly objects that move in depth on an approaching trajectory toward the viewer. Examples can be seen in 3-D presentations where objects appear to leap out of the screen towards the viewer, and in gaming where judgements are made to either avoid or attack approaching objects.

The extent to which a user can perceptually immerse in a multidimensional world and interact with moving objects is reliant on many elements. These include the effect of the simultaneous presentation of multimodal sensory information, and the degree to which algorithms can integrate the sensory stimuli parameters, such as the duration of both audio and visual presentation, speed and magnitude of movement, depth and spatialisation, and temporal synchronisation, all of which individually vary in real time.

The programming of such systems should be based on a firm scientific foundation which captures as accurately as possible our knowledge of how we visually and aurally perceive events and interactions in order to accurately generate a dynamic and rich perception of the objects and events being represented.

Whilst the domain of physics has provided us with various laws to model the propagation of sound and how the sound changes when objects move in depth, the sound effects often presented in virtual environments are rarely generated according to the physical models with mathematically and physically correct parameters. Usually the sounds are enhanced creatively by the sound designers and post production technicians for maximum affect. However, which parameters of sound, the amount of their manipulation, and the intended and resulting effects, are often not documented and remain something of a dark art.

Furthermore, psychological research on human perception of approaching objects, particularly in regard to auditory perception and the parameters of sound which act as audio cues for movement in depth often investigate auditory looming using simple audio cues such as an increase in amplitude. These simple audio cues are often applied to artificial tones such as sine, triangle, or square waves that are not regularly encountered in the natural world. So whilst the results from such research has provided important information on human perception and responses, can the conclusions drawn from these results, which are based on artificial conditions, transfer to real world, or hyper-real scenarios? How can these conclusions, which have been drawn from results based on artificial conditions, be used to predict and manipulate human perception and response in the real or hyper-real world? Can we use conclusions that have been drawn from results based on artificial conditions to design audio cues for use as acoustic models in virtual environments and devices with the capacity to predict precise human responses and reactions to the stimuli?

We now have the technological ability to generate and manipulate complex auditory stimuli in precise detail. We are also able to collect measurements on human perceptions and responses to complex and often real world stimuli. This, then, is a direction in which experimental design should proceed. This would enable the collection of information on how humans respond in real world scenarios and would help to bridge the gap to the wealth of information from which conclusions have be drawn from artificial stimuli administered under experimental conditions that have a robust internal validity.

This thesis summarises the research that was conducted investigating the presentation of an object moving in depth on an approaching trajectory (looming), focusing on which parameters of sound acted as audio cues for movement in depth. The research examined the effect of audio cues on human perception, when those cues are presented in combination as multiple audio cues, compared to their effect when presented as single audio cues. The question of whether the effectiveness of the audio cues differs when they are presented with real world stimuli compared to their effectiveness when presented with artificial stimuli is also examined.

1.1 Research Questions

The questions that are the focus for this research are:

- Which audio cues are important for generating a perception of looming, and are any cues more important than others;
- Are there any correlations between the observers perceived time-to-contact, emotion, and engagement ratings, and the audio cues presented;
- Is the effect of the audio cues on human perception greater when they are presented in combination as multiple audio cues compared to their effect when they are presented as single audio cues;
- Does human looming perception differ when presented with real world stimuli compared to artificial stimuli;
- Does the audio cues effectiveness differ when presented with real world stimuli compared to artificial stimuli.

1.2 Thesis Structure

We start the thesis with a brief introduction, outlining our motivation for the research, the research questions, the contributions that our research has made in the process of this PhD research, and list our publications.

In order to answer our core research questions, we begin by conducting a review of three areas that are associated with looming research and application. Firstly we remind ourselves of the various laws of acoustics and psychoacoustics that explain how sound changes when an object moves in depth. This gives us a scientific foundation to understand of how sound behaves in an ideal scenario and perfect environment. Secondly, we investigate the psychological research on looming, identifying which parameters of sound and their manipulation are currently used to generate auditory-visual looming scenarios. Finally, we examine industries application of looming stimuli to understand where, and how, people interact with looming stimuli in the real and virtual worlds.

After reviewing these three areas that are fundamental to undertaking robust research, we then conduct our own experiments. We conduct a total of four experiments. The first is a feature analysis study of the audio cues used to present movement in depth as designed by sound designers and post production technicians in the film industry. The second experiment is a psychoacoustic study evaluating human responses to these same film samples used in the feature analysis study. The third experiment takes a closer inspection of the audio cues, introduces the audio cue of 'direct-to-reflections sound energy ratio', and evaluates human responses to these audio cues. The fourth experiment applies these audio cues to visual stimuli (moving images) to determine if these audio cues bias auditory-visual perception of the looming stimuli. In the final chapter we discuss our key findings from our experiments and conclude the thesis with directions for future research and possible real world applications.

Chapter 1 Introduction: establishes the motivation for the research and outlines our questions. The contributions and publications resulting from the project are also listed.

Chapter 2 Background: reviews three key areas fundamental to conducting robust research on Auditory Looming. Firstly, we begin by reminding ourselves of the laws of acoustics which describe the propagation of sound and how sound changes when objects move in depth, and the psychoacoustic factors that underpin human auditory perception of an approaching object. Secondly, we review psychological studies on auditory looming, highlighting key results and conclusions from previous experiments, whilst examining the experimental design, auditory stimuli, and parameters that may have affected human perception and the overall outcome or wider application of the results. Thirdly, we explore the application of auditory looming in various industries to understand where, and how, people interact with looming stimuli in the real and virtual worlds.

Chapter 3 A Feature Analysis Study of the Audio Cues in Film Looming Scenes: is the first of our studies. We present the results from an analysis of the audio cues from 27 film looming scenes, to understand how objects (moving on an approaching trajectory, in hyper-real situations) are represented, how the parameters of sound are manipulated and designed as audio cues for maximum effect.

Chapter 4 Responses to Designed Film Looming Stimuli: presents our second study, a novel psychoacoustic experiment measuring human perception of, and response to, the film looming stimuli analysed in Chapter 3. This study provides us with information on human responses to complex stimuli that contain multiple auditory cues, which have been designed for hyper-real scenarios and to generate emotional (valence and arousal) responses in observers.

Chapter 5 The Effect of Audio Cues and Sound Source Stimuli on the Perception of Approaching Objects: is our third study, and is a closer inspection of a selection of audio cues for movement in depth. We introduce the new cue of 'direct-to-reflections sound energy ratio' and compare its effectiveness to other cues for presenting an object moving on a frontal mid-line plane. We also consider the complexity of the cues, comparing single versus multiple cues, to see if single cues are as effective as multiple cues. And lastly, we also investigate the sound source, comparing responses to an artificial sound source versus a real world sound source, to determine if the cues applied to artificial sounds are as effective when transferred to real world sounds.

Chapter 6 Responses to Complex Auditory-Visual Looming: is our forth and final study. We take the audio cues investigated in Chapter 5, and apply them to visual stimuli in order to measure human responses to the multimodal audiovisual presentation. We examine whether or not particular audio cues affect the overall audiovisual perception of the approaching object. We consider if the number and complexity of the cues (i.e. multiple versus single cues) affect the overall audiovisual perception of the approaching object, and if the sound source (artificial versus real world stimuli) affects the overall audiovisual perception of the approaching object.

Chapter 7 Conclusions and Future Perspectives: discusses our research findings, draws comparisons and contrasts between the studies, compares our results with previous findings, outlines further ideas and questions for future research, and propose real world applications for the research.

1.3 Contributions

The principal contributions of this research and thesis are:

- The provision of new information on the human perception of, and responses to, auditory(-visual) looming that use multiple audio cues and complex sound sources.
- The provision of new information on human perception of, and responses to, ecologically valid stimuli using complex real world, and hyper-real stimuli. This bridges the gap to the results and conclusions drawn from auditory looming studies that use artificial stimuli, whilst providing a foundation for future experiments to incorporate real world stimuli and parameters into their design.
- Introduction of the sound parameter of 'direct-to-reflections sound energy ratio' as an audio cue for auditory(-visual) looming, and the measurement of human responses, both physical and emotional, to this cue.
- Measurement of human emotional responses (valence, arousal, and engagement) to auditory(-visual) looming, thereby increasing the limited information collected on this aspect of human perception and action.
- The development of experimental design and implementation of measurement techniques to evaluate human responses to complex looming stimuli, therefore strengthening the foundations for more ecologically valid experiments with greater external validity, bridging the gap between experimental looming research in laboratory conditions and real world applications.

1.4 Publications

Book Chapter:

An early version of Chapter 5 was published as a (peer reviewed) book chapter:

- Wilkie, S. and Stockman, T. "The Perception of Auditory-Visual Looming in Film." From Sounds to Music and Emotions. Springer Berlin Heidelberg, 2013. 378-386. (Wilkie and Stockman [2013]).

Peer-reviewed Conference Papers:

Portions of the research from this thesis were published and presented at international conferences:

- Wilkie, S., Stockman, T., and Reiss, J.D. "Amplitude Manipulation For Perceived Movement In Depth." Audio Engineering Society Convention 132. Audio Engineering Society, 2012. (Wilkie et al. [2012]).
- Wilkie, S. and Stockman, T. "The Perception of Auditory-visual Looming in Film" CMMR Symposium, London, June 2012. (Wilkie and Stockman [2012]).

Conference Posters and Demos:

Portions of the research from this thesis were presented as posters and demonstrations at the following conferences:

- DMRN +5, QMUL, Dec 2010.
- EECS Postgraduate Conference, May 2011.
- C4DM 10th Birthday Celebrations, Sept 2011.
- C4DM at AES, AES headquarters, Oct 2011.
- DMRN +6, QMUL, Dec 2011.
- EECS Postgraduate Conference, May 2012.
- IEEE International Conference on Multimedia & Expo, Melbourne, July 2012.
- DMRN +7, Dec 2012.
- EECS Postgraduate Conference, April 2013.

Chapter 2

Background

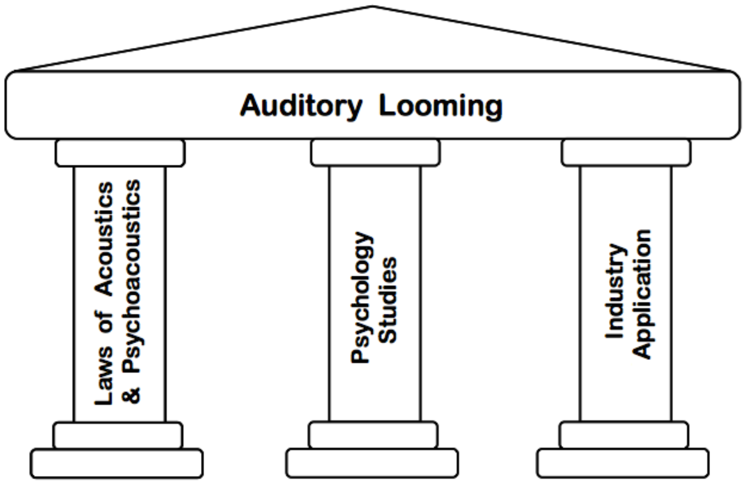

Figure 2.1: Three Key Areas For Research On Auditory Looming
Three key areas fundamental to conducting robust research on Auditory Looming are investigated in this section. They are 1. The laws of Acoustics and Psychoacoustics; 2. Psychology Studies; 3. Industry Application.

In this chapter, we review three key areas fundamental to conducting robust research on auditory looming, and establish the foundation for which the thesis questions and arguments are based. We begin by reminding ourselves of the laws of acoustics which describe the propagation of sound and how sound changes when objects move in depth, and the psychoacoustic factors that underpin human perception of, and response to, the sound of a moving object. This provides us with a scientific basis to understand how sound operates in an ideal scenario and perfect environment. We then investigate psychology's research corpus on looming to understand human perception of, and response to looming stimuli. We identify which parameters of sound and their variables have been explored in auditory-visual looming experiments, investigate human perception and response to the stimuli, and consider the conclusions drawn from these studies. Finally, we examine the application of looming stimuli in various industries to understand where, and how, people interact with looming stimuli in the real and virtual worlds.

Reviewing these three areas provides a solid foundation from which we form our research questions, and equips us with an understanding of the factors essential for the experimental design in the subsequent chapters.

2.1 Acoustics and Psychoacoustics of a Moving Object

In this section we present the technical knowledge needed to understand subsequent chapters. We remind ourselves of the laws of physics which describe the propagation of sound and how sound changes when an object moves in depth, and the psychoacoustic factors that explain human perception of, and response to, a moving object.

Understanding the laws that explain how sound changes when an object moves, and the resulting affect on human perception and action, provides a scientific foundation from which we can compare the acoustic features of, and human responses to, sound behaving according to a physics based perfect environment, with the sound stimuli used in the psychoacoustic experiments, and the sound effects designed for film looming scenes.

By ascertaining how the acoustic features from sound behaving according to a physics based perfect environment, differs to the features of artificial stimuli used in the psychoacoustic experiments, and the features of the hyper-real stimuli of the film looming scenes, enables us to understand how we can bias human perception of an approaching object, and to predict the possible response and action when the audio cues are presented at levels greater or lesser than the physical reality.

More detailed and technical explanations of the following laws can be found in standard physics dictionaries, with the Peters et al. [2011] textbook providing a comprehensive and clear explanation, and the Nave [2012a] website providing online calculators with detailed interactive examples.

2.1.1 Amplitude Level and the Inverse Square Law

It is a well known and experienced sensation that a sound which is transmitted a constant level will be louder when its proximity is closer to the observer and softer when farther away. When an object moves towards an observer, the amplitude increases according to the Inverse Square law (inverse proportional), which for a point source in a free field is at a rate of approximately 6dB per halving of distance, and is calculated using equation 2.1 [Illingworth, 2004].

$$I = \frac{P}{2\ \pi r^2} \tag{2.1}$$

where:
I = Intensity (dB) at observer,
P = Power (dB) emitted by the sound source,
r = Distance of the sound source to the observer.

As the proximity of the object nears the observer, the amplitude increases on a non-linear slope, and at a much greater rate at the closest proximities. This slope is illustrated in Figure 2.2 plotting the amplitude level over distance to the observer.

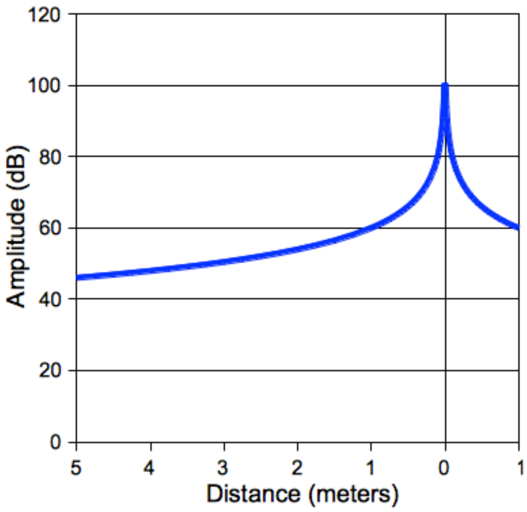

Figure 2.2: Amplitude Change Over Distance According To The Inverse Square Law
The amplitude of the sound source increases non-linearly at a rate of approximately 6dB per halving of distance.

Neuhoff [2001] (p.102) notes that the absolute loudness level does not provide meaningful information about an object's approach, but rather it is the rate of change that provides meaningful information as it can indicate an arrival and potential intercept time. It is also claimed by Sheeline [1983] that for the amplitude level to function as an audio cue for object localisation, distance perception, or movement, familiarity with the object and having some prior knowledge of the object's normal transmission level at a given distance is required, thereby providing a reference point for which judgements on its change in amplitude level and rate of change can be made.

2.1.2 The Doppler Shift

The pitch of the sound produced by a moving object will change according to the Doppler shift. According to this law, a stationary observer will initially experience the pitch of an approaching object (that is moving at a constant velocity) at a frequency higher than is actually being transmitted by the source. When the object reaches the observer, the pitch drops to the actual source frequency, then continues to decrease as the object moves farther away. The perceived frequency and the rate of change are dependent on a number of factors, including the object's actual transmitted frequency, the object's velocity, the object's angle of trajectory, as well as environmental factors such as temperature and humidity.

The frequency of an object as it approaches a stationary observer on a frontal midline path, is calculated using equation 2.2 [Illingworth, 2004].

$$f_{observed} = f_{source} \left[\frac{V}{V - V_{observer}} \right] \tag{2.2}$$

where:
$f_{observed}$ = Frequency (Hz) perceived by observer,
f_{source} = Frequency (Hz) transmitted by the object / sound source,
V = Velocity (the speed of sound (343.2 m/s)),
$V_{observer}$ = Velocity of the object when it reaches the observer.

2.1.3 Surface Reflections

The reflections of a sound and its spectral content off (wall) surfaces provides spatial information that describes the environment, the distance of an object in relation to any surrounding walls or obstacles, and in relation to the observer. Further, the spectral content of the reflected sound may be modified over distance and time by the material properties of the surface and its reflective or absorbent propensity. As Sheeline [1983] notes, the absorptive properties of reflecting surfaces and obstacles in the sound path significantly modify the frequency spectrum of the reverberant energy over time.

The use of surface reflections for echo-location navigation has been well documented for various species of bats and dolphins (for more detail, see Thomas et al. [2004]). It has also been found that human listeners can echolocate the presence of walls and obstacles, as well as echolocating the distance and shape of the walls and obstacles, by the surface reflections from a sound [Rosenblum et al., 2000a].

2.1.3.1 Reverberation

Reverberation is the term given to the reflections that reach the observer in the first $\leq$ 50ms after a sound is made, and may take any duration of time to decay. Reverberation (RT_{60}) is measured using the Sabine Formula (equation 2.3) giving the amount of time it takes for the amplitude level of the direct sound (of a point source) to decay by 60dB, with the duration of the decay being dependent on the size of the space, shape and position of reflective surfaces, and the material composition and properties of the surfaces.

The duration that a sound reverberates provides architectural information to an observer about the surrounding space, if walls or obstacles are present, their surface materials and how reflective or absorbent the properties are, the distance of walls, and overall size of the space. As the surfaces in the space become more reflective (and less absorbent) the duration of the reverberation becomes longer. A long reverberation time may suggest that the sound source (and observer) are placed in an enclosed space with highly reflective surfaces, whereas little or no reverberation time may suggest that the sound source is placed in either a free field (perhaps outside) with no surfaces to reflect the sound,or that the sound source is placed in a room with very absorbent surfaces (i.e. anechoic chamber).

If the room dimensions and absorption properties are known, the reverberation time can be calculated according to the Sabine Formula [Illingworth, 2004].

$$RT_{60} = K \ \frac{V}{A} \tag{2.3}$$

where:
RT_{60} = reverberation time (seconds),
K = 0.161 meters,
V = Volume of the room (cubic meters),
A = Total Absorption (square meters),

Computational models can accurately generate room reflections for virtual environments (examples include Catt Acoustic [Dalenbäck, 2006] and SLAB 3D [NASA, 2013]), however are highly dependent on many known factors and presets including, the

- Room Dimensions, surface materials and absorption properties,
- Distance of the observer to the walls,
- Distance of source to observer,

- Distance of the source to the walls (and any obstacles / reflective surfaces),
- Velocity and trajectory of the moving object.

2.1.3.2 Direct-to-Reverberant Energy Ratio

The direct-to-reverberant energy ratio is the proportion of the total sound energy which is comprised of the direct sound, versus the proportion of the total sound energy which is comprised of the reflections.

As an audio cue, the direct-to-reverberant energy ratio has been shown to provide information about the distance of the object, and the surrounding space in relation to the observer [Von Békésy and Wever, 1960; Mershon and King, 1975]. If an object (in a reverberant space) is at a great distance, the ratio between the direct sound and the reverberant energy levels is small, that is to say, both the direct sound and the reflected sound have similar intensity levels. As the proximity of the object becomes closer to the observer, this ratio between the direct and reverberant energy levels change, with the level of the direct sound increasing until it completely masks the intensity level of the reverberant energy.

Bronkhorst [1995; 1999] demonstrated that increases to the number of, and amplitude level of, surface reflections (as compared to the level of the sound source) resulted in a perceived increase of the sources distance.

2.1.4 Environmental Attenuation

Environmental effects such as atmospheric air absorption, wind and temperature gradients, and ground absorption, attenuate the amplitude level of the spectral components of a sound, with the amount of attenuation increasing with distance. The ISO standards 96131:1993 and 96132:1996 [ISO, 1993; Norma, 1996] specify the methods for calculating the attenuation of sound at a range of distances and sound sources (both artificial and real world sound sources), as well as different temperatures (-20°c to +50°c) and humidity levels (10 % to 100 %), whilst Harris [1966] publishes detailed graphs plotting the level of attenuation in dB according to temperature and humidity, for frequency bands ranging from 125 Hz to 4000 Hz.

For a sound source that is comprised of a wide range of frequencies that are all transmitted at the same amplitude level, the high frequencies are attenuated more than the lower frequencies. When an object is at a great distance from the observer, the overall pitch heard by the observer will appear lower, and the spread of the spectrum will be narrower. As the proximity of the object becomes closer to the observer, the higher frequencies with shorter wavelengths become apparent, causing the overall spread of the frequencies to also increase and resulting in a broader frequency spectrum. Ingård [1953] documented that for a sound source placed at distances of ≥ 15 meters from an observer, air absorption significantly reduced the spectral content of the sound source, attenuating the high frequencies most with 4kHz losing 3 to 4dB per 100 meters. Whilst this reduction in amplitude level is small, Coleman [1968] found that sound source's with attenuated high frequency spectral content, were perceived to be at a greater distance from the observer, than was physically presented.

Ground absorption attenuates the sound according to the surface properties of the ground. If the surface is hard, very little attenuation occurs, however if the surface is very absorbent (such as thick grass) a 2kHz sound can be attenuated by 10dB per 100 meter distance [Wiener and Keast, 1959].

2.1.5 Inter-aural Differences

Inter-aural Differences are the differences in a sound between an observers two ears. If the sound source is placed at an angle on the azimuth plane to the observers head (as illustrated in Figure 2.3) the sound wave will arrive at one ear before the other. This causes a difference in the arrival time and is known as the Inter-aural Temporal Difference (ITD). The duration of the delay is dependent on the angle of the sound source, with the 90° angle providing the maximum difference between the two ears, an ITD of approximately 1.5ms [Zwislocki and Feldman, 1956; Howard, 2012].

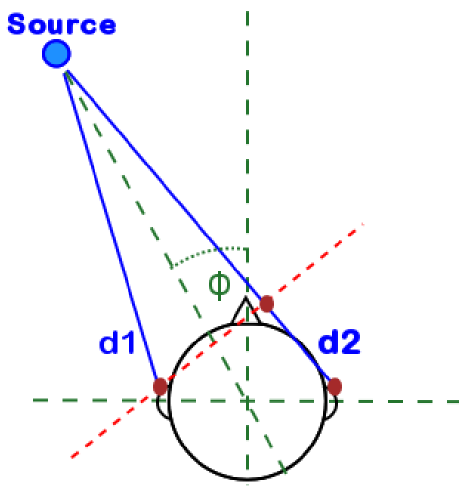

Figure 2.3: Inter-Aural Differences
For D1 (distance 1) the sound travels a shorter distance than D2 (distance 2). These differences result in a temporal delay (ITD) and a level difference (ILD).

The inter-aural difference also means that the sound travels more distance to the ear which is farthest to the source. This results in a difference in amplitude level between the two ears with the ear that is closest to the source presented a greater amplitude level, and is known as the inter-aural level difference (ILD).

The inter-aural differences (ITD and ILD) both work in combination to facilitate judgements on distance and localisation, and there effectiveness as audio cues are dependent on a number of factors, including the angle and trajectory of the sound source, frequency and wavelength of the sound source ITD's provide more beneficial information for low frequencies (below $1,500$Hz) than at higher frequencies [Licklider and Webster, 1950], whereas ILD's are more beneficial at higher frequencies [Peters et al., 2011] (pg 58). Because the inter-aural differences are greater when the object is placed at $90°$ angle, than for lesser angles, and theoretically non-existent when at $0°$ (although in reality a person's ears are not precisely symmetrically spaced, so some inter-aural difference are present at the $0°$ frontal midline plane), humans are more accurate at judging the distance of a sound source when the object is placed nearer the $90°$ extremities, than for smaller angles [Zahorik et al., 2005].

2.1.6 Summary of Acoustics and Psychoacoustics

In this section, we reminded ourselves of the acoustic laws that describe the propagation of sound and how sound changes when an object moves towards an observer, and the psychoacoustic factors that underpin human auditory localisation and distance perception of an object. This technical knowledge was provided as a simple foundation to understand how sound operates in an ideal scenario and perfect environment, and as a basis to understand and compare the parameters of sound and audio cues discussed in subsequent chapters. More detailed and comprehensive accounts of these laws can be found in standard physics textbooks, for example Howard [2012].

In the next section, we review psychology's investigations on looming in order to understand human perception of, and response to, looming stimuli. We identify which parameters of sound and their variables have been explored in auditory-visual looming experiments, investigate human perception and response to the stimuli, and consider the conclusions drawn from these studies.

2.2 Psychological Investigations of Looming

Auditory-visual looming involves the presentation of an object moving in depth towards an observer. In both experimental conditions and virtual environments, the object itself is often computer generated (CGI), or is complemented with elements that are computer generated such as sound effects. As it is a simulation of the movement and not physical movement, understanding the parameters of sound which act as audio cues for movement in depth requires psychological investigations into human perception. In this section, we review the corpus of psychological research on looming.

2.2.1 Ecological Validity

We will first address the concept of ecological validity as it is a fundamental component for the motivation of this research and the design of the experiments.

Brunswik initially proposed the theoretical framework of representative design [1943; 1953; 1955] advocating that the design of an experimental task, environment, or condition should be representative of the natural situation or environment, and to not exclude certain conditions or variables for the creation of a neat experimental design. One way to achieve representative design, is to sample conditions from the organisims natural environment outside of the laboratory.

Whilst the concept was a radical proposal for psychological research in the 1940's and 1950's, it did gain a number of supporters, notably James J Gibson with his investigations into visual perception and advocating 'reasons for realism' [Gibson et al., 1982; Gibson, 2013], and the development of ecological psychology whereby "the mind directly perceives environmental stimuli without additional cognitive construction or processing" [Rutherford and Fancher, 2012].

An important component of Brunswik's representative design, was the ecological validity which indicated "the degree of correlation between a proximal cue and the distal variable to which its is related" [Hammond, 1998].

Hammond [1998] however notes the erosion of Brunswik's original meaning of ecological validity, with the term often being confused with, or used interchangeably with, representative design.

As this thesis is building upon the work of previous looming researchers (in particular, the work conducted by John Neuhoff and colleagues [2004]) we will also adopt their use of the term ecological validity, which although incorrect, has been regularly used when actually meaning representative design. That is, the design of experiments and the use of stimuli experienced in the real-world, so that the results have external validity, generalisation of, and application in, the world outside the laboratory.

2.2.2 Unimodal Looming

Limited research has been published on multimodal auditory-visual looming as the notion of different sensory systems integrating information is a recent theory [Driver and Spence, 2000]. Previously, sensory systems were thought to process information independently and as such, the majority of psychological looming studies are unimodal - being auditory only or visual only looming studies.

By providing information on modality specific parameters and variables, these unimodal studies form a useful foundation for the development of multimodal studies. Therefore, we briefly review the results from studies on auditory looming and visual looming in order to build a solid understanding of, and foundation for, parameters that can be used in auditory-visual looming studies.

2.2.2.1 Visual Looming

The focus of this research is on the auditory cues for looming and their effectiveness in both auditory and auditory-visual looming. However, it is important that we understand the visual looming cues and perception, the visual stimuli used in experiments, and the conclusions drawn from research on visual looming, as it will allow us to comprehend the visual component in auditory-visual looming, and how the audio cues may affect visual perception. Examining visual looming studies may also reveal factors that are applicable to auditory looming studies, therefore, we will briefly review the visual looming corpus of research.

Visual looming involves the visual presentation of an object moving towards an observer. In many studies, the visual stimuli is often an expanding white disc or square on a black background [Schiff et al., 1962; Schiff, 1965; Gray and Regan, 1998, 1999; Regan and Beverley, 1978; Hong and Regan, 1989]. Recently, a number of studies interested in increasing the external validity of experiments, have presented image sequences of moving vehicles as the approaching object [Schiff and Oldak, 1990; Caird and Hancock, 1994; Hancock and Manster, 1997; Horswill et al., 2005; Terry et al., 2008; Rodrigues et al., 2012]. It is the object's area expansion over time that represents an object moving towards an observer.

The measurement of human responses to looming (and also often receding) stimuli are made via the prediction of time-to-contact, whereby observers are presented the stimuli and are asked to predict the contact / arrival time of the object. It is a demonstrated response attribute that looming stimuli prompts people to underestimate the contact time of an approaching (looming) object, which does not occur with receding stimuli. This underestimation of the contact time provides an advantage to the observer, giving them more time to prepare (an increased safety margin) for the objects arrival, and to initiate the appropriate response (being fight or flight) therefore increasing the chance of survival. An object that is moving away from the observer poses less threat, therefore does not require the increased safety margin of additional time which is gained by underestimating the arrival time.

Overall, many more studies have been conducted on visual looming, than auditory looming. This is evident by the sheer number of visual looming studies that were relevant enough to be referenced in this thesis, and is by no means an exhaustive list [Ball and Tronick, 1971; Ellensburg et al., 2009; Brenner et al., 1996; Calabro et al., 2011; Carlile et al., 2006; Colombo, 2000; Cornilleau-Pérès et al., 2002; Dill, 1974; Franconeri and Simons, 2003; Gabbiani et al., 2002; González et al., 2010; Gray and Regan, 1998, 1999, 2000; Hayes and Saiff, 1967; Hong and Regan, 1989; Kahan et al., 2011; King Jr et al., 1999; Khuu and Lee, 2010; Lee et al., 1983; Lin et al., 2009; Neppi-Mòdona et al., 2004; Parker and Alais, 2007; Preuss et al., 2006; Raviv and Joarder, 2000; Regan and Beverley, 1978; Regan and Vincent, 1995; Sahin and Gaudiano, 1998a,b; Savelsbergh et al., 1993; Schiff, 1965; Schiff et al., 1962; Terry et al., 2008; Vagnoni et al., 2012; Wang et al., 1993; Whiting et al., 1970; Yamawaki, 2011].

The implications of having a greater number of visual looming studies, has resulted in deeper investigation into human visual looming perception than auditory looming perception. A summary of the results, finds that approaching stimuli strongly captures attention whereas receding stimuli does not [Franconeri and Simons, 2003] suggesting that events which require rapid behavioural or motor responses for self preservation are more likely to receive attentional priority. Human perception (constructed from the visual information) of time-to-contact, object size, and rate of expansion, can be processed simultaneously, independently, and in parallel, when the object is placed in the observer's fovea. However, the capacity to process the visual information independently decreased as the object's position moved from the fovea into the periphery [Regan and Vincent, 1995]. Looming objects presented in the near field also elicited a different perception and response to those object's presented in the far field [Hong and Regan, 1989]. This is due to the perceived level of threat, with closer objects prompting a greater urgency for decision and action, than more distant objects.

When predictions were made on the perceived contact time, the conditions that were predicted most accurately were those that positioned the object at head or eye level and moved on a frontal midline trajectory that intercepted with the head, as opposed to conditions that placed the object level with other areas of the body or aligned at an eccentric angle to either the left or right side [Neppi-Mòdona et al., 2004]. This finding suggests that visual looming judgements require kinaesthetic information regarding the position of the object in relation to the position of head and potential contact point, in addition to the retinal information regarding the object's area expansion over time.

Since infants (aged 3 - 6 weeks) exhibited an avoidance response and became upset when presented with a looming shadow or object, it is suggested that the capture of human attention and reaction is not a learned response, but is innate [Ball and Tronick, 1971]. However, the object itself can bias an observers reaction. Objects that are perceived to be more threatening (for example snakes and spiders) can cause a greater underestimation in the perceived contact time, and have a greater reported fear rating, than non-threatening objects (i.e. butterflies and fluffy bunnies) [Vagnoni et al., 2012]. This finding suggests that the perceived contact time is not purely based on the innate visual looming cue (of an object's area expansion over time), and kinaesthetic information (regarding the position of the object in relation to the observers body), but is also dependent on emotional associations with the object, which may be learned.

This great volume of research investigating visual looming has also led to a rich diversity of studies, with research extending into non-human visual looming perception. The animals which also exhibited perception of, and response to, visual looming stimuli include primates [Schiff et al., 1962], birds [Wang et al., 1993; Schiff, 1965], amphibians [Schiff, 1965; King Jr et al., 1999], fish [Dill, 1974; Preuss et al., 2006], reptiles [Hayes and Saiff, 1967; Carlile et al., 2006], invertebrates [Schiff, 1965], and insects [Yamawaki, 2011; Terry et al., 2008]. By taking measurements of visual orientation, target fixation time, and response tasks, the results revealed that the perception of looming, and the reaction to an approaching object, is similar across this broad range of species, and that looming objects are highly salient to most animal visual systems.

Visual looming research has also extended to robotics, where it is used as a range sensor, and an obstacle avoidance system (in both robotics and vehicles) enabling autonomous moving robots to avoid colliding with obstacles [Raviv and Joarder, 2000; Sahin and Gaudiano, 1998a,b]. These studies use a number of looming measurement techniques, including:

- an increase in the object area size over time (as with human perception).
- a change in the object irradiance (brightness) over time (with the use of cameras).
- a change in the density of the objects texture over time.
- a change in the image blur over time.

These visual looming techniques particularly succeed in the detection of objects which have slanted surfaces (at $\geq 15°$ of line of sight) whereby sonar is unsuccessful due to the refraction of pulses [Sahin and Gaudiano, 1998a].

Further advantages of the visual looming detection systems include only requiring minimal equipment (only 1 camera is needed); it provides 3D (depth) information from a 2D image capture (by its measurements of area expansion etc); and the techniques can be used in cases where information (such as depth, range, object motion, and camera position) or processing capacity, are not available.

The research for this thesis is concentrating on auditory cues and their implications in auditory and auditory-visual looming. However, having this understanding of visual looming and the methods used in the unimodal studies provides us with experimental methods that may be applicable to our auditory looming studies (including the use of time-to-contact and emotional assessment as measurements of stimulus effect), in addition to information about how looming is perceived, and operates in a broader range of situations.

2.2.2.2 Auditory Looming

One of the functions that both the auditory and visual systems perform in scene analysis, is the identification of 'what' and 'where' [Bregman, 1994; Kraus and Nicol, 2005; Ungerleider and Pessoa, 2008]. After the initial research on looming perception in the visual modality, Rosenblum et al. investigated if a similar percept and response was also generated by presenting looming stimuli to the auditory modality [Rosenblum et al., 1987].

Initial research on auditory looming found that humans associate an approaching object with at least three audio cues, namely, an increase in the amplitude, a change in the fundamental frequency (the doppler shift), and inter-aural differences [Rosenblum et al., 1987]. Results from this study also suggest that some audio cues have a greater affect on perception, and the amount of over- / under-estimation of the object's perceived contact time, than other audio cues. For example, the change in amplitude elicited the fastest 'response to contact time' when the object passed, whilst the doppler shift prompted a response before the object had passed.

The original finding in visual looming studies that approaching stimuli created a greater underestimation of the contact time than receding stimuli, was replicated in the auditory modality [Neuhoff, 1998, 2001; Cappe et al., 2009] whereby looming audio cues (in the form of an increase in the amplitude) prompted a greater underestimation of the contact time than receding audio cues (presented as a decrease in the amplitude). This finding was evident in both the simulated condition (virtual environments which presented the computer generated audio cues) and those conditions where the audio cues were created by physical movement (i.e. a speaker swinging towards the listener) [Neuhoff, 2001].

The same explanation that was given in the visual looming studies for this discrepancy in the perceived contact time of the approaching object versus the receding object was applied to the auditory looming studies. That approaching objects present more danger, and that by underestimating the contact time observers are provided with more time to initiate the appropriate response (being fight or flight) therefore increasing self preservation [Neuhoff, 2001].

It was also demonstrated that tonal sounds in the form of pitched sine tones, enabled the detection of more looming audio cues, than (white) noise, which was evident in both humans [Neuhoff, 1998, 2001] and non-human primates (rhesus monkeys) [Ghazanfar et al., 2002; Maier et al., 2004].

Neuhoff's studies also showed that people overestimate the magnitude of intensity when presented with increasing stimuli. This implies that the increasing intensity of the approaching object is more dramatic than the extent of its physical approach. The results showed that for both the vowel sound and noise band stimuli, people overestimated the increase with the greatest overestimation being for the loudest vowel increase (60 - 90dB). This magnitude change is perceived to be even greater when presented at louder levels than at softer levels [Neuhoff, 1998; Neuhoff and Heckel, 2004] with louder levels implying that the object is at a closer proximity, than softer sounds which imply a farther distance. The female participants also expressed a greater over-estimation in the magnitude than the male participants [Neuhoff and Heckel, 2004] suggesting that females perceived a greater threat and that more safety time was needed to initiate the appropriate response.

In an evolutionary context for both the physical and virtual worlds, these overestimations of magnitude and underestimation of contact time provide an advantage to the observer, giving them more time to prepare (an increased safety margin) for the objects arrival, and to initiate the appropriate response (being fight or flight), therefore increasing their chance of survival.

2.2.3 Multimodal Auditory-Visual Looming

Real world looming scenarios such as approaching traffic often involve both auditory and visual information to assess a given situation. Studies have recently begun to investigate multimodal auditory-visual looming, with initial studies finding the response in non-human primates (rhesus monkeys) [Maier et al., 2004, 2008] and more recently has be replicated in human perception [Cappe et al., 2009; Tajadura-Jiménez et al., 2010; Tyll et al., 2012].

In Maiers' studies, representation of the approaching object often involved the presentation of an expanding disc as the visual stimulus cue, and the transmission of a 400Hz triangle wave for 1000ms duration as the auditory stimulus. An increase of the amplitude (55 - 75dB) functioned as the only auditory looming cue. The studies used preferential looking tests whereby longer durations were a measurement of attentional preference for certain conditions, although other physical responses were also reported when the animals were presented the auditory-visual looming stimuli, including the animals ducking, flinching, and jumping back to avoid the apparent approaching object, in addition to vocalising an alarm call.

The results indicated that the monkeys were able to associate an expanding visual object with a rising-intensity tone. It suggests that the non-human primates have an evolved capacity to integrate bimodal auditory and visual looming signals, which may also be applicable to humans.

Recent studies [Cappe et al., 2009, 2012; Tyll et al., 2012] have investigated this theory in human perception, with the results reflecting the findings in Maiers (Rhesus monkey) studies. The studies demonstrated that humans perceived the auditory-visual looming stimuli with quicker response times, than the visual (only) looming condition, however the auditory (only) looming condition prompted the earliest response time. This greater underestimation caused by the auditory only condition may be attributed to an unseen approaching object prompting people to err on the side of caution and respond earlier than necessary. The neural analysis conducted by Cappe et al. suggests that the bimodal auditory-visual stimuli was integrated and that it facilitated an earlier perception of an approaching object, as compared to the visual only condition.

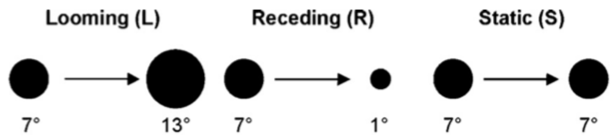

Figure 2.4: Visual Stimuli (Artificial)
From Cappe et al. [2009] the visual stimuli consisted of a white disc expanding on a black background (and vice-versa) expanding from $7^\circ - 13^\circ$ diameter over the 500ms duration.

These recent publications are the first studies to investigate human capacity to integrate auditory-visual looming stimuli. Whilst the focus of these studies are the neural integration of the multisensory looming information and the regions of the brain which are active during the looming events, more research on auditory-visual looming is needed, not only for the examination of the neural mechanisms that underpin motion perception and the responsiveness to possible danger (with judgements of fight or flight); but also to develop a greater understanding of the audio cues in complex sounds, which will enable the development of envelope algorithms for the generation of an objects movement in depth, and its implications in computer-generated environments. Further, a broader range of measurement techniques (other than time-to-contact and EEG) will allow a broader understanding of human perception and response.

The salient nature of looming stimuli suggests that the measurement of emotion would be a valuable tool to provide an insight on human experience in potentially threatening scenarios. Tajadura-Jiménez et al. [2010] recent study has begun to measure this factor, finding that people had a preference for ecologically valid sounds, over synthesised artificial tones. This preference may not be noticeably evident in simple time-to-contact measurements, however the measurement of emotion through valence and arousal ratings reveals this bias towards the stimuli. The study also revealed that approaching auditory-visual stimuli were rated as more unpleasant (lower valence) and arousing, than receding auditory-visual stimuli. This finding might be expected, given that the results reflect those of the previous auditory looming studies, however it only applied to the objects which had a negative and neutral association. When a target image was paired with an approaching negative sound source (a growling dog with an increase in the amplitude as the audio cue) the observers not only had faster response times to the target image, but also expressed greater arousal and unpleasantness, than when the target image was paired with the receding negative sound source (decreasing in amplitude). When the target image was paired with an approaching positive sound source (a giggling baby increasing in amplitude) the observers response times to the target image was not as fast as the negative source, and also expressed greater pleasantness and lower arousal. These responses to the positive versus negative sound sources may be expected, but interestingly the observers emotional responses to the receding positive source (a giggling baby decreasing in amplitude) expressed greater arousal and more unpleasantness than the approaching condition.

These results support and provide a broader insight to the biological preference and evolutionary explanation where receding positive objects and approaching negative objects present danger. By experiencing greater arousal and unpleasantness, in addition to underestimating the contact time, observers are prepared to initiate the appropriate response being fight or flight therefore increasing self preservation. The introduction of the emotion measurement in this study, has also introduced a new measurement technique that provides a broader understanding of human responses to looming stimuli that is not evident through traditional time-to-contact measurements alone.

2.2.4 Sound Sources

A survey of the literature reveals that many of the studies in the auditory and auditory-visual looming corpus use an artificial sound source to represent the object, namely the presentation of a sine, triangle, or square wave. However these sound sources are not regularly encountered in the natural world.

Examples of the sound sources used in the looming studies, include a 100-Hz tone (with five unspecified harmonics) [Lutfi and Wang, 1999]; 100 Hz synthesised vowel sound (with three formants at 450, 1450, and 2450 Hz) [Neuhoff, 1998, 2001]; 200 Hz sine wave [Ericson, 2007]; 400 Hz square wave [Bach et al., 2009; Cappe et al., 2009]; 400 Hz triangle wave [Maier et al., 2004, 2008; Maier and Ghazanfar, 2007; Cappe et al., 2009; Leo et al., 2011]; 500 Hz tone (with five unspecified harmonics) [Grassi, 2010]; 804 & 602 Hz to 765 & 572 Hz sine tones (crudely simulating an ambulance siren) [Rosenblum et al., 1987]; 1000 Hz sine wave [Neuhoff, 1998, 2001]; 1000 Hz square wave [Cappe et al., 2009]; 1000 Hz triangle wave [Ghazanfar et al., 2002; Neuhoff and Heckel, 2004; Cappe et al., 2009]; 3000 Hz sine wave [Tyll et al., 2012]; and a white noise band [Neuhoff, 1998, 2001; Ghazanfar et al., 2002; Maier et al., 2004, 2008; Maier and Ghazanfar, 2007].

A handful of studies have begun to use real world sound sources in there psychoacoustic looming experiments. The recording of an approaching vehicle has been used in three experiments [Schiff and Oldak, 1990; Rosenblum et al., 1993, 2000b]; with other real world sound sources being an approaching talking person [Schiff and Oldak, 1990]; an ambulance siren [Gordon and Rosenblum, 2005]; and one recent study investigating three real world sources, namely footsteps, a giggling baby, and a growling dog [Tajadura-Jiménez et al., 2010].

Looking across the research corpus, we can see that the studies have predominantly used artificial sound sources. Whilst the results from these studies using artificial sound sources have provided important information on human perception and response, can the conclusions drawn from these results, transfer to real world or hyper-real scenarios? Can these conclusions (which have been drawn from results based on artificial conditions) be used to predict and manipulate human perception and response in the real or hyper-real world? And can we use the conclusions (that have been drawn from results based on artificial conditions) to design audio cues for use as acoustic models in virtual environments and devices that have the capacity to precisely predict human responses and reactions to the stimuli?

We now have the technological ability to generate and manipulate complex auditory stimuli in precise detail. We are also able to collect measurements on human perception and response to complex and often real world stimuli with more than one measurement technique, including time-to-contact, EEG (Electroencephalograph), GSR (galvanic skin response), emotion (valence and arousal), and engagement ratings. This direction, is a direction in which experimental design should proceed. It would enable the collection of broader information on how humans perceive and respond in real world scenarios. Further, it would help to bridge the gap to the wealth of information from which conclusions have been drawn from artificial stimuli administered under experimental conditions that have a robust internal validity, but questionable and limited external validity.

Whilst a number of studies [Neuhoff, 1998, 2001; Ghazanfar et al., 2002; Maier et al., 2004] have concluded that tonal sounds (artificial pitched sine, triangle and square waves) enabled easier detection of looming audio cues, than (white) noise bands, more could be learnt about human perception by comparing responses to the artificial sound sources, with that of real world sound sources. One recent study (by Tajadura-Jiménez et al. [2010]) has made that comparison. The study compared human responses to a 1 kHz sine wave versus three real world sounds, namely a giggling baby (which has a positive emotional association), a growling dog (which has a negative emotional association), and footsteps (which have a neutral emotional association). The results indicated that humans preferred (with greater valence ratings) all three real world sound sources - positive, negative and neutral sound sources, to the artificial sine wave. This finding supports Gibsonian ecological realism and the importance of using rich real world stimuli for experimental design [Golonka and Wilson, 2012].

Further comparison of the three sound sources revealed that an emotional association with the object biased the perceived time-to-contact, whereby the object with the negative threatening association (the growling dog) prompted a greater underestimation in the perceived contact time, than the neutral and positive objects (the footsteps and giggling baby). This finding in the auditory modality parallels Vagnoni et al. [2012] finding in the visual modality, that people had greater underestimation of the (visual) contact time when presented images of animals that have a negative association, compared to their response time when presented animals that have a positive association.

2.2.5 Audio Cues

2.2.5.1 Single Cues

The first study investigating auditory looming [Rosenblum et al., 1987] demonstrated that three parameters of sound acted as audio cues for movement in depth, namely inter-aural differences, the doppler effect, and amplitude change. In addition to finding that all three parameters were associated with an approaching object, they found that the effect on perceived time-to-contact differed between the parameters, suggesting a hierarchy amongst the cues with some parameters prompting an earlier perceived contact time than others. The results showed that the change in amplitude elicited the earliest time-to-contact at the point in which the object had passed, whilst the Doppler effect prompted a response before the object had passed.

Amplitude Increase

An increase in the amplitude level is the parameter of sound that is used most often in psychoacoustic experiments as an audio cue to represent an approaching object. In many studies (for example Rosenblum et al. [1993]; Neuhoff [1998, 2001]; Neuhoff and Heckel [2004]; Cappe et al. [2009]; Ghazanfar et al. [2002]; Maier et al. [2004, 2008]; Maier and Ghazanfar [2007]) it has also been the only audio cue used in the experiment. Selection of this audio cue is understandable, given the amplitude increase parameter has been demonstrated to be a dominant audio cue (at velocities of < 10 meters per second (36 kph)) in the hierarchal studies [Rosenblum et al., 1987; Lutfi and Wang, 1999]. As an easy sound parameter to control and precisely replicate, researchers are often motivated to increase experimental robustness through the absolute control of variables, which has been achieved by the sole use of this parameter. However, this approach of investigating human perception of an object's movement that is depicted by the use of a single audio cue is questionable.

Zahorik et al. [2005] states that distance perception is likely to be much more complicated than reflecting a simple relationship with intensity at the ear, which could be encoded entirely within the auditory periphery. It is possible that this notion could be extended from the distance perception of a stationary object, to a moving dynamic object. Further, real world looming scenarios that are regularly encountered in the natural world are comprised of many parameters of sound generating multiple audio cues simultaneously, and need to be investigated in order to provide a greater understanding of human perception.

Nevertheless, when the amplitude increase parameter is used as an audio cue for a looming object, the magnitude of the change is perceived to be greater than it physically is [Neuhoff, 1998; Neuhoff and Heckel, 2004] suggesting that the object is approaching at a faster rate than it physically is. This change is perceived to be even greater when presented at louder levels, than at softer levels [Neuhoff, 1998; Neuhoff and Heckel, 2004] with louder sounds suggesting that the object is at a closer proximity to the observer, therefore poses grater potential danger, than softer levels which are perceived to be at a further distance. The results also showed that female participants expressed a greater over-estimation in the magnitude than male participants [Neuhoff and Heckel, 2004] suggesting that females perceived a potentially greater threat.

The magnitude of the amplitude increase has been consistent across the experiments. Many studies presented a 20dB increase in amplitude level [Rosenblum et al., 1987; Maier et al., 2004, 2008; Maier and Ghazanfar, 2007; Ghazanfar et al., 2002; Neuhoff and Heckel, 2004; Bach et al., 2009; Leo et al., 2011] or at a level ± 10dB (10dB [Cappe et al., 2009], 15dB [Neuhoff, 1998], 18dB [Tajadura-Jiménez et al., 2010], 30dB [Neuhoff and Heckel, 2004; Neuhoff, 2001]). These levels were presented over durations ranging from 250ms [Leo et al., 2011] to 7500ms [Rosenblum et al., 1987], however the 1000ms duration has been the duration presented most frequently [Maier et al., 2004, 2008; Maier and Ghazanfar, 2007; Lutfi and Wang, 1999; Ericson, 2007].

Although these amplitude increase levels are moderate and at a similar level, the impact of presenting them at different durations is not. Schiff and Oldak [1990] and Tajadura-Jiménez et al. [2010] found that longer durations (whereby the provision of more time to acquire information about the object, its velocity, trajectory, overall threat potential, and the amount of time needed to respond appropriately) did not improve time-to-contact accuracy. The 1.5 second duration which was the shortest duration presented in Schiff and Oldak [1990] study, was found to prompt the most accurate contact time, with accuracy decreasing up to the 6 second (and maximum) duration. One explanation for the shorter duration prompting the fastest and most accurate response, is that longer durations alter the rate of change of the amplitude increase. Shorter durations present a faster rate of change, and therefore presents the object moving at a faster velocity than longer durations. The presentation of a 20dB increase consistently between studies, but at different durations, means that the object in the Leo et al. [2011] study which was presented for a 250ms duration, was travelling at a velocity of approximately 1.44 kph, whereas in the Rosenblum et al. [1987] study which was presented for 7500ms duration, the object was travelling at an approximate velocity of 0.05 kph. This is quite a difference in the rate of change, that would have an impact on the perceived level of threat associated with the object, the participants response time, and any conclusions based on the results.

However, a surprising number of studies [Ericson, 2007; Grassi, 2010; Rosenblum et al., 1993, 2000b] do not state the magnitude of amplitude increase, even though it may have been the only audio cue used to represent an approaching object. As such, the conclusions drawn from their results are questionable, and may possibly simply reflect people's responses to unequal (possibly greater) levels and rates of change, rather than the specific audio cue.

Amplitude Envelope Slope

The slope of the amplitude envelope may bias viewers perception of the approaching object and the perceived contact time. According to the inverse square law, an approaching objects amplitude envelope will increase on a non-linear slope at approximately 6dB per halving of distance, with the greatest increase occurring at the closest proximities. A linear slope increases at a steeper rate earlier in the envelope than a non-linear (ISL) slope. This may prompt observers that perceive the velocity of the approaching object to be accelerating, causing people to estimate the contact time earlier than a non-linear (ISL) slope. Closer inspection of the stimuli used in the auditory(-visual) looming studies reveals that different slopes have been applied to the amplitude envelopes.

The majority of the looming studies that state the slope of the amplitude envelope, are increasing the amplitude on a linear slope. These studies include Rosenblum et al. [1987]; Neuhoff [1998]; Lutfi and Wang [1999]; Neuhoff [2001]; Bach et al. [2009]; Cappe et al. [2009]; Tajadura-Jiménez et al. [2010]; Tyll et al. [2012] and Gray [2011]. Whilst a linear slope makes reproduction of the experiment conditions easy, this is not how sound operates in the real world. Two studies, Ghazanfar et al. [2002] and Neuhoff and Heckel [2004], noted that they applied a non-linear exponential slope to the amplitude envelope. However, more studies applying a non-linear slope to the amplitude envelope need to be conducted, in order to fully understand human perception of looming stimuli.

Whilst it is a concern that only two studies, have used a non-linear envelop, what is of greater concern is the number of studies that do not mention the slope of the amplitude envelope in their publication whatsoever. These studies include Rosenblum et al. [1993]; Maier et al. [2004, 2008]; Maier and Ghazanfar [2007]; Leo et al. [2011]. Any conclusions drawn from these studies need to be treated with caution as they may simply be reflecting a bias caused by an unknown envelope slope.

Doppler Shift

The doppler shift, similar to an increase in the amplitude level, is one of the more prominent cues for movement in depth. The change in pitch is instantly identifiable when representing and characterising passing traffic. However, one study [Neuhoff and McBeath, 1996] investigating peoples understanding of the doppler shift, reveals that many people often erroneously believe that the pitch of an object (in this case, a train) approaching at a constant velocity, will rise as the object nears the observer. This commonly held belief was held by 261 out of the 292 (university psychology student) participants that were surveyed for the study. Whist the pitch and the rate of change are dependent on a number of factors, including the objects transmitted frequency, the objects velocity, the objects angle of trajectory, as well as environmental factors such as temperature and humidity, for an object approaching at a constant velocity, the pitch does not rise, but remains constant until the object is in close enough proximity of the final few wavelengths that the pitch drops to the actual source frequency, then below, as the object passes and moves farther away.

The doppler shift was one of the sound parameters investigated by Rosenblum et al. [1987] as an audio cue in the first study on auditory looming. In their study, Rosenblum et al. simulate an ambulance siren sound source that alternates between 2 tones. The doppler shift model presents the 2 siren tones starting at the frequencies of 804Hz and 602.9Hz which decreases to 764.6Hz and 572Hz as the proximity of the object nears, then passes the observer. In musical terms, these frequencies equate to the tone deviations of G5 $\pm$ 43 cents, and D5 $\pm$ 45 cents. The study revealed that the doppler shift model prompted people to underestimate the contact time of the approaching object by an average of $M =$ -557ms when the intercept time was actually 3000ms, and that this underestimation was greater than the other two audio cues presented in the study, namely amplitude increase and inter-aural differences.

Lutfi and Wang [1999] built upon the Rosenblum et al. study, investigating the doppler shift as an audio cue at two different velocities - the moderate velocity of 36 kph (10 m/s), and the fast velocity of 180 kph (50 m/s). Using a discrimination task to judge the object's displacement, velocity, and acceleration, the results indicate that people preferenced the doppler shift audio cue (over other audio cues) to judge the velocity and acceleration at 36 kph, and again preferenced the doppler shift audio cue (over other audio cues) to judge all three measurements (displacement, velocity and acceleration) at 180 kph.

Whilst the doppler shift has only been investigated as an audio cue for looming objects in a few studies, the results of the parameter's affect on perception and response time in these studies are encouraging.

Inter-Aural Differences

Inter-aural Differences was the third parameter of sound investigated by Rosenblum et al. [1987] as an audio cue for auditory looming. The magnitude of inter-aural differences presented is dependent on the object's offset angle and trajectory from the observer. If an object is approaching on a frontal midline trajectory, the inter-aural differences is theoretically non existent, however in reality human ears are not precisely symmetrical, so a small amount of inter-aural difference is present. This is particularly the case if binaural in-ear recordings were used to record the acoustic stimuli.

The Rosenblum et al. [1987] model presents the inter-aural temporal differences as a delay between the left and right channel ranging from 557ms (when the object was farthest in proximity) to 0ms (when the object was passing the observer), whilst the Lutfi and Wang [1999] model presented the inter-aural temporal differences at a range of 200ms to 0ms, and inter-aural level differences ranging from 0.5 dB to 0.0dB. The Gordon and Rosenblum study presented a siren passing the observer at an angle of 30°, however the magnitude of the inter-aural differences is not noted in the study. Further, the video camera recording the visual stimuli was placed in front of the human observer with the observers head located to the lower-right corner of the camera. The close position of the camera to the observers head created an obstacle that occluded the audio signal to the left ear. As binaural in ear microphones were used to record the stimuli, the signal level in the left channel was on average 1.03dB lower than that for the right ear. As a result, the inter-aural level differences between the two channels are greatly exaggerated, and the conclusions from this study regarding the audio cue's capacity to affect time-to-arrival estimation must be treated with caution.

In both the Rosenblum et al. [1987] and Lutfi and Wang [1999] studies, the inter-aural differences were concluded to be the least dominant audio cue out of the three cues presented, however in these studies, the inter-aural differences cue was not presented at its maximum magnitude of difference which is produced when the object is located at 90° from the observer and pans across [Howard, 2012]). It is reasonable to propose then, that in addition to Lutfi and Wang [1999] conclusion that velocity affects the capacity of an audio cue to inform or bias human perception of the approaching object, the angle of approach and magnitude of its change also need to be considered, in order to not overstate or generalise an audio cue's dominance, but rather place it in the context of the presentation.

Surface Reflections and the Direct-to-Reverberant Energy Ratio

Surface reflections and the direct-to-reverberant energy ratio have been investigated as audio cues in many range perception studies of stationary (non-moving) sources in both real and virtual environments [Griesinger, 2009; Von Békésy and Wever, 1960; Bronkhorst, 1995, 2002; Bronkhorst and Houtgast, 1999; Zahorik, 2002b; Mershon and King, 1975; Sheeline, 1983; Devore and Shinn-Cunningham, 2003; Valimaki et al., 2012]. These studies all find that the direct-to-reverberant energy ratio is a dominant audio cue for human perception of distance.

Griesinger [2009] notes that an object's sound source with no reflections causes the observer to perceive the object to be in close proximity, whilst the addition of reflections at 10 - 50ms after the direct sound signal will add a perceived distance to the object. In addition to creating distance between the observer and the object, the reflections also establishes a room impression with the walls defined by the reflective surfaces. Zahorik [2002b] found that the direct-to-reverberant energy ratio threshold for detection was 5 - 6dB, for the 0 - 20dB range of energy ratios examined

One looming study [Bach et al., 2009] has included the parameter of (ground) reflections in the sound stimuli presented in their looming study. However, its presence was simply due to the use of a live recording of the source, and no attempt was made to investigate the individual parameter or its effect as an audio cue. The parameter of surface reflections and the direct-to-reverberant energy ratio is yet to be explored independently as a dynamic audio cue for looming objects or its affect on human perception.

One explanation for the audio cue's notable omission from auditory looming research, is the extent to which the parameter can be precisely controlled and analysed in experimental conditions. It has only been with the development of acoustic modelling software (such as Catt-Acoustic [2006] or Slab3D [2013]) that the complexity of the parametric control, and the accurate generation of reflections has been possible. Further, the processing power required to generate the dynamic variations to a parametric model that is required for a moving object, has only recently become available.

As the acoustic modelling software and processing power has become readily available in recent years, the parameter cannot continue to be overlooked. We propose that the direct-to-reverberant energy ratio needs to be investigated as a dynamic audio cue in auditory looming, and introduce its use in this thesis.

2.2.5.2 Audio Cue Hierarchy

From the initial study on auditory looming, Rosenblum et al. [1987] suspected that the capacity for the parameters of sounds to act as audio cues would differ, with some cues biasing percept more than others. In their study, the authors concluded that the amplitude increase was the most dominant audio cue, prompting the fastest response to the contact time after the object had passed, and the inter-aural differences prompted the slowest response after the object had passed, whilst the doppler shift prompted a response before the object had reached the observer. The dominance of the amplitude increase variable, combined with its easy reproduction, has lead to many studies primarily using this parameter as an audio cue for approaching objects.

Lutfi and Wang [1999] built upon the Rosenblum et al. study, investigating the same audio cues at two different velocities - the moderate velocity of 36 kph (10 m/s), and the fast velocity of 180 kph (50 m/s). Using a discrimination task to judge the object's displacement, velocity, and acceleration, the results indicated that for a velocity of 36 kph, people preferenced the amplitude increase or the inter-aural differences to judge the object's displacement, whilst they preferenced the doppler shift audio cue to judge the velocity and acceleration. For the faster velocity of 180 kph, the observers again preferenced the doppler shift audio cue (over the other audio cues) to judge all three measurements. The authors concluded that motion perception (in this case, looming) is variable, and is not dependent on one single acoustic cue.

2.2.5.3 Multiple Cues

Reviewing the audio cues used in the previous auditory looming studies reveals that the majority studies [Rosenblum et al., 1993; Neuhoff, 1998, 2001; Neuhoff and Heckel, 2004; Cappe et al., 2009; Ghazanfar et al., 2002; Maier et al., 2008; Maier and Ghazanfar, 2007; Maier et al., 2004] present a single audio cue, usually the amplitude increase audio cue, to depict a moving object. This approach is understandable given that researchers are often motivated to increase experimental robustness through the absolute control of variables, and that the amplitude increase has proved itself to be an important audio cue. However, real world looming scenarios that are regularly encountered in the natural and virtual worlds are comprised of many parameters of sound generating multiple audio cues simultaneously, and need to be investigated in order to provide a greater understanding of human perception.

Neuhoff [2004] suggests that the main reason that early psychoacoustic studies were limited to using simple sound sources and few audio cues was due to the technology required to generate controlled complex and dynamic stimuli wasn't available. Whilst it is important to conduct experiments in controlled conditions, the use of simple tones to minimise external affects compromises the external validity of the results, so that information on the perception of sound and rarely heard artificial tones, may not be applicable to understanding the perception of complex sounds as heard in the real world. As such, more research needs to be undertaken so that information can be obtained on the perception of real world stimuli.

One such study [Bach et al., 2009] that noticed this lacuna between real world scenarios and the controlled experiments in looming research, sought to bridge the gap by comparing human responses to the amplitude increase (only) cue with that of a sound containing full motion cues from a live recording. The amplitude increase condition presented a 400 Hz square wave with a 20dB linear increase in amplitude. The full motion cues condition also presented the 400 Hz square wave, but due to the recording of a physically moving object, it was reported to contain the doppler shift, atmospheric filtering, gain attenuation, ground reflection attenuation, and head related transfer functions. The researchers concluded that responses to the amplitude increase condition were the same as the full motion cues condition, therefore they supported the design and conclusions of the controlled experiments. However the authors interpretation of the results is questionable. Whilst the perceived contact times to the amplitude increase condition was the same as the full motion cues, the skin conductance test showed greater response to full motion cues, than intensity only. Further, the question asking participants to rate the loudness change was flawed as the amplitude only condition was physically louder than the full motion cue condition. It is possible that this difference in amplitude levels may have further influenced the perceived magnitude of loudness change and contact times, whereby the louder amplitude condition suggested that the object was physically closer to the observer, therefore prompting a greater response to magnitude and faster response to contact time, or possibly even inducing an acoustic startle response.

Whilst Bach et al.'s study has shown initiative at approaching the complex topic of the perception of multiple audio cues, the interpretation of the results is questionable. Further, this study demonstrates the need to have a broad range of observer perception and response measurements, in order to gather an accurate and complete understanding of human looming perception.

2.2.6 Summary of Psychological Research

Our review of the psychoacoustic looming research has revealed a number of key points. There is limited published research on the perception of auditory-visual looming. Of the looming research that has been published, it is predominantly uni-modal studies investigating auditory looming or visual looming when it has been demonstrated that combined sensory information produces a more realistic representation of what is experienced by people in the real world.

The stimuli used in many of these psychoacoustic studies is also extremely controlled. The majority of the studies present the sound source as an artificial sine, square, or triangle wave often at 400 Hz or 1000 Hz, with limited exploration of the sound parameters that function as audio cues for movement in depth. In many cases, one single audio cue was presented, being the linear amplitude increase ranging from 10 - 30dB. There is a gap in our understanding of how humans perceive real world or hyper-real sound sources containing multiple audio cues, however the technology is now available to explore human perception of these complex sounds.

The salient nature of looming stimuli also suggests that the measurement of emotion would be a valuable tool to provide an insight on human experience in potentially threatening scenarios.

The results from these controlled psychoacoustic studies have provided important information on human perception of looming objects and the parameters of sound that act as audio cues for approaching objects. However, absolute experimental control through the use of artificial tones and the singular presentation of audio cues has resulted in a limited use and understanding of the audio cues in the experimental conditions. This does little to advance our understanding of the audio cues involved in complex sounds and limits the ecological validity of the results and real world applications. Further, it invites the question of how do humans perceive and respond to complex sounds with multiple audio cues? We will explore this question in later chapters with a feature analysis study examining the audio cues designed for film looming scenes, and a perceptual study investigating humans responses to the complex designed film looming stimuli. But first, we will continue to expand our knowledge of auditory-visual looming, with the third section in this background chapter exploring industry application of looming stimuli.

2.3 Industry Application

In this section, we examine the application of looming stimuli in Industry, to understand where, and how, people interact with looming stimuli in the real and virtual worlds. The industries investigated include the film, gaming, simulator, and auto technology industries, but looming cues and detection systems are also used in many other area's beyond the scope of this section, including robotics, obstacle detection, and obstacle avoidance systems.

2.3.1 Film

One of the features of 3D presentations that entices viewers to attend a 3D screening as opposed to a 2D screening, is the opportunity to see objects appear to leap out of the screen towards the viewer.

This presentation of objects moving through a multidimensional space assists in drawing the viewer into the created world and makes it appear more immersive, not only by presenting the third dimension of depth and bringing particular objects closer to the viewer, but also by transforming the medium from a passive experience of motionless watching and listening, to an active experience with viewers physically moving to avoid apparent objects as an instinctive reaction to the perceived proximity of the objects.

Stereoscopic and 3D presentation systems have been around since the late 1800's, however, it has only been in recent years that the technology has improved the experience with better image resolution and colour, and reduced the cost of production.

Systems and format technologies include Anaglyph stereography using the classic blue and red filtered glasses, Circular polarisation as used by RealD [Cowan and Officer, 2007] the Active shutter 3D system as used by Imax, and the Autostereoscopic lenticular system which doesn't require eyewear and used by Nintendo 3DS [Boev and Gotchev, 2011]

Although a great deal of research is currently being undertaken on the presentation of 3D images and the transformative eyewear, research also needs to be undertaken to develop post production tools and techniques to maximise the experience.

Studies on multimodal depth perception have demonstrated that the perceived depth of 2D images can be influenced, and extended, by the simultaneous presentation of sounds containing depth audio cues [Turner et al., 2011; Berry and Holliman, 2013].

The sound effects in film scenes are designed to articulate and emphasise the objects and events presented on and off screen. Believable (but not necessarily authentic) sound effects assist in drawing the viewer into the created world, adding a sense of presence and immersion [Serafin and Serafin, 2004].

2.3.2 Gaming

Interactivity in gaming was revolutionised in 2006 when Nintendo launched their new console, the Nintendo Wii. Previous to this console, gamers used hand held controllers and joysticks to direct the avatars' actions. Nintendo introduced the Wii remote which uses accelerometers, and the players physical movements and tilting to control the avatars' actions. This inception of physical movement to control onscreen actions led to the proliferation of active games featuring targets moving in depth towards the player. Examples include adventure, warfare, and sporting themes (tennis, boxing or baseball) where targets (such as ball's) that needed to be physically intercepted in order to survive and progress in the virtual world. The application of looming sound effects for these looming scenarios is critical to engaging players with the virtual world.

Figure 2.5: Screen Capture Of The Xbox 360 Kinect Game "Star Wars"
As the character of Luke Skywalker, the player has to deflect laser beams that are shot directly at them, with their lightsaber. An interactive and physical experience similar to hitting a baseball, there is however, no controller (lightsaber), and the players hand movements are tracked by the Kinect camera.

The most recent development is the Xbox Kinect, which has discarded the controller altogether and instead tracks the players movements via a 3D motion capture webcam. The tracking of the players entire body and physical movements has added another level to interactivity in gaming. Looming targets, no longer have to be intercepted (as with the wii remote) but can also be avoided through the players ducking to avoid the approaching object. Examples include ducking to avoid bullets in the case of warfare and adventure themed games (see Figure 2.5).

The players survival and progression through the game depends on their ability to quickly respond to the approaching target. They now have two options when faced with potential looming scenarios, either attack or avoid the approaching object. Their capacity to successfully interact with and respond to approaching objects, is reliant on quick decisions of fight or flight, and an individual's ability to accurately interpret depth and movement cues may affect their continued survival in the virtual world. As such appropriate audio cues are crucial to successfully engaging the player.

2.3.3 Simulators and Training Systems

The accurate modelling of auditory perception has a wide range of application areas, from aircraft simulators [Ploner-Bernard et al., 2005], modelling of perception-response cycles in car driver simulations [Gauduin and Boussard, 2009], to entertainment, broadcast and military applications [Begault et al., 1994].

In the context of serious games development in support of education and training, Michael Zyda wrote: "Spatial and immersive sound are key components for whatever training and educational systems researchers build with gaming. Developers must implement future engineering requirements and human-performance engineering to ensure that they can employ sound appropriately and effectively while minimizing cross-modal sensory conflicts" [Zyda, 2005].

2.3.4 Vehicle Technology

2.3.4.1 Electric Cars

Without the sound of the combustion engine purring away, electric cars naturally produce no engine sound. The only sound produced from an electric car is the broadband noise from the tyre traction moving across the road surface. As such, there are very little audio cues to inform bystanders about the presence of a nearby vehicle or its movement.

Pedestrians rely on the audio cues that engines produce to identify unseen vehicles, their proximity, and the velocity they are moving at, in order to avoid getting hit. Without the audio cues that inform people of a nearby moving vehicle, they remain unaware of a potential hazard. They have lost the advantage of time needed to prepare for the vehicle's arrival, and to initiate the appropriate response being fight or flight, therefore decreasing their chance of survival. In the real world, pedestrians and vehicles are regularly in close proximity. The introduction and proliferation of these silent electric vehicles potentially pose a serious risk to a large proportion of the population. As such, legislation was introduced in Japan and the USA (the pedestrian safety enhancement act) in 2010 which stipulates that electric and hybrid vehicles must generate an artificial proximity notification noise when travelling at 25 kph or less [Ashe, 2011].

Auto manufacturers have complied with this law by installing external speakers at the front of the car, which transmits the sound of a synthesised engine. This sound increases and decreases in spectral frequency content according to the speed and acceleration of the vehicle [Release, 2010; Goodwin, 2011b].

2.3.4.2 Driver Auditory Feedback

The advancement of noise reduction technology in the construction of vehicles has culminated in a much quieter cabin. However, this has resulted in drivers receiving limited auditory feedback regarding the movement and acceleration of the car.

When it comes to sport car enthusiasts, the auditory feedback from the engine sound particularly when accelerating is an important factor that makes the driving, and the watching of driving as a sport, more thrilling [Collantine, 2014].

BMW, in an attempt to increase driver enjoyment of its electric cars re-introduced auditory feedback for drivers in the M5 series sports car with the 'Active Sound Design Technology' [Goodwin, 2011a; Boeriu, 2011]. This system is a Digital Signal Processor that generates an eight cylinder engine sound effect. Its parameters are modified according to the vehicles speed, acceleration, deceleration, and torque, which is generated in real time with data obtained from the engine management system. When drivers want a more exhilarating experience, they switch from 'Sports' mode to 'Sports+' mode, which increases the frequency and spectral content. The sound is then transmitted through the cabins internal speakers, to provide feedback to the driver, and enhance their driving experience.

2.3.5 Summary of Industry Application

In this section, we conducted a review of various looming industries, namely the film, gaming, and auto technology industries, that use looming stimuli as a technique to convey information about the real or virtual environment and nearby objects, and as a tool to enhance user engagement and pleasure. This review enabled us to understand where and how people interact with looming stimuli in the real and virtual worlds.

2.4 Chapter Summary

In this chapter, we reviewed three key areas fundamental to possessing a thorough understanding of auditory-visual looming, and established a solid foundation from which we build our research.

In the first section of this chapter, we reminded ourselves of the laws of acoustics that describe the propagation of sound and how sound changes when it moves in depth, and the psychoacoustic factors that underpin human perception of, and response to, a moving sound source. This technical knowledge was provided as a scientific basis to understand how sound operates in an ideal scenario and perfect environment, which we draw upon in subsequent chapters.

In the second section of this chapter, we investigated Psychology's research corpus on looming. With limited published studies on auditory-visual multimodal research, we drew information from the unimodal looming studies of auditory looming and visual looming. We explored the design of experiments, the sound sources used to represent the looming object, the parameters of sound which were acting as audio cues for movement in depth, human perception and response to the stimuli, the lack of ecological validity in the design of the experiments, and questioned whether the conclusions drawn from these heavily controlled studies could be applied to real world scenarios.

In the third and final section of this chapter, we explored the application of looming stimuli in the film, gaming, and auto technology industries to see how looming stimuli has been applied as a tool to enhance user engagement and pleasure.

In the next chapter, we begin our series of experiments on auditory-visual looming, by first undertaking a feature analysis study on a sample of looming scenes designed for film. of the sound parameters which act as audio cues in film looming scenes. This analysis will provide information about how the sound parameters are designed, if the acoustic features are similar to those used in the psychoacoustic looming experiments, or if the acoustic features are similar to how they should operate according to the laws of acoustics. This feature analysis study is followed by a perceptual experiment on the same sample of film looming scenes, to acquire information about how humans perceive, and response to, looming scenes that have been designed for maximum effect and elicit a reaction from the viewers. We then take a closer inspection of the audio cues for looming and introduce the new cue of direct-to-reflections energy ratio, and lastly apply these audio cues to visual stimuli to investigate the audio cues affect on auditory-visual perception.

Chapter 3

A Feature Analysis Study of the Audio Cues in Film Looming Scenes

As we discussed in section 2.3.1, the sound effects in film scenes are designed to articulate and emphasise the imagery presented on and off screen. Believable sound effects assist in drawing the viewer into the created world, making it appear more immersive and gives it a sense of presence.

The presentation of objects moving in depth is naturally, a feature of 3D presentation. However it can also be effectively used in 2D presentation to draw the viewer into the scene. By presenting, or replicating the third dimension of depth, objects are brought closer to the viewer, transforming the medium from a passive experience of motionless watching and listening, to an active experience where viewers may, for example, physically move to avoid objects as an instinctive reaction to the perceived proximity of the objects, or motion towards them. Whilst many excellent text books explain the creative techniques for sound design (including Holman [2010]; Chion [1994]; Stevens and Raybould [2013]; Farnell [2010]; Ondaatje and Murch [2002]) they rarely present detailed DSP analyses of the acoustic features or compare them to how the features should behave according to the laws of physics.

Possible audio cues that may be used by the sound designers in film looming scenes include an increase in the amplitude level and manipulation to the slope of the magnitude of the increase, emphasis of the spatial (panning) movement, Head-related transfer functions (HRTFs), High Frequency (HF) scattering and blurred transients. It is not just the audio cues themselves, but how they change, and the rate of change, as the proximity of the object becomes closer.

Possible visual cues may include an expansion of the object's area, a change in the texture or luminance of the object as compared to the image background, how the visual cues change as the proximity of the object becomes closer, and the rate of change.

In this chapter we present a feature analysis study that was conducted on the audio track of the 27 film looming scenes to understand which features the sound designers and post production technicians use as audio cues to depict an approaching object, how the features change as the proximity of the object becomes closer to the observer. We compare the levels and magnitude of these features to those used in previous psychoacoustic looming studies, and also with how sound should behave according to the laws of physics. This will provide information about a broader range of sound parameters that can then be investigated more closely as audio cue variables in perceptual studies, to understand how humans perceive, and respond, to hyper-real stimuli which have been designed to maximise the impact and level of affect on viewers.

3.1 Aim

The aim of this study was to determine how the sound parameters that have been designed by sound designers and post production technicians in film looming scenes, change according to the proximity of the looming object, how the features compare to those used in the psychoacoustic studies, and how the features compare to how such sounds should behave according to the laws of physics.

3.2 Hypotheses

It was hypothesised that sound designers modify the following parameters of sound as audio cues to create a percept of an approaching object:

- Amplitude
- Spatialisation
- Spectral components

Due to the size of this chapter, detailed hypotheses specific to each feature analysed, are reported in the results section.

3.3 Method

The looming scenes from 27 films were analysed for key features with the full list of film scenes provided in Appendix Table A.

3.3.1 Scene Selection Criteria

Selection criteria for the looming scenes were applied at three stages of the initial review, being the film type, presentation, and sound clarity.

The criterion at the first stage was applied to the selection of the film by its genre, sourcing films from the Action, Science Fiction, Adventure, and Animation genres. The purpose of applying a criterion to the film genre, was to target films that most likely presented looming scenarios. As visual and audio special effects are often a feature of these genres (more so than romantic or drama) this criterion was applied to narrow the search time and therefore obtain initial results sooner.

The selection of recent films was also a criterion, to ensure that the technology (hardware and software) capable of the complex manipulation of the sound parameters was available (and perhaps not used to its full capabilities), and to ensure that the sound designers had access to a similar standard of technology. 23 secnes (85.19%) were produced with a time span of 10 years (M = 2005, SD = 5.67 years; min = 1983; max = 2010; $mode$ = 2005 & 2010 (5 scenes each)). It was also decided to take only one scene from each film, so as not to bias the overall results by having a disproportional number of scenes from the same sound designer.

The second stage criterion was applied to the image ensuring it was a looming scene - a frontal approaching object, that was not obscured by other objects in the scene.

The third stage criterion was applied to the sound in the looming scene. Attempts were made to select scenes with the 'cleanest' audio sample possible, to enable more accurate feature extraction and analysis. The criterion included selecting scenes in which the dominant (or preferably only) sound was generated by the looming object, and that no (or very little, if unavoidable) dialogue, music track, or sounds emitted by other objects were present in the audio sample. We also chose to select sounds that were constant, as opposed to sounds made up of discrete components (such as a horse galloping) so the results could be compared with those from the psychoacoustic looming studies, which all presented continuous sound samples.

Whilst these three criterion were used to aide and quicken the process of obtaining looming scenes, it is acknowledged that there are further limitations with the chosen stimuli. For example, genre by no means guarantees that the looming objects or scenes will possess a similar emotional association. And not all of the sound designers would be equally competent with their sound design skills and application of the potential audio cues. However, as this study is the first of its kind to investigate the application and design of looming scenes in film, it is a generalised study to determine if further research with deeper investigation is warranted. Replication of this study on a greater scale using a larger corpus of looming scenes, will provide enough data to control for parameters such as sound designer, object, emotional content, and the number of audio cues used.

3.3.2 Stimuli

Each sound source was a stereo track between 325 and 3007ms in duration (M = 1228ms, SD = 747ms; 13 scenes $\leq$1000ms). The sound files (.wav format) were a stereo mix down at a sampling rate of 44100 Hz with no automated spatialisation formatting such as Dolby digital or DTS, and were analysed using the MIRtoolbox v.1.3.4 [Lartillot and Toiviainen, 2007] for MATLAB.

3.4 Results

The following feature analyses were conducted on the film scenes' audio tracks:

3.4.1.1 Magnitude of the Amplitude Increase
3.4.1.2 Amplitude Envelope Slope
3.4.1.3 Amplitude Levels
3.4.1.4 Object Velocity (According to the Inverse Square Law)
3.4.2 Pan Position
3.4.3.1 Spectral Centroid
3.4.3.2 Spectral Spread

3.4.1 Amplitude

3.4.1.1 Magnitude of the Amplitude Increase

An increase in the amplitude has been the core, sometimes only, variable used in psychoacoustic looming studies to depict an approaching object. As discussed in section 2.2 these studies often exhibit similar choices for the presentation of the sound parameter, such as the magnitude of the amplitude increase, the slope of the increase, and the listening level. However, the ecological validity of these parameters and levels used is questionable. The analysis of the amplitude level and change in the film scenes will provide information on how this parameter functions in examples that use complex sounds, and whether the levels chosen in the previous looming experiments is suitable.

Aim

The aim of this analysis is to determine

- the magnitude of the amplitude increase used in film looming scenes, designed by post-production technicians and sound designers,

- if the magnitude of the amplitude increase is similar to that used in the psychoacoustic studies,
- if the magnitude of the amplitude increase is dependent on the scene duration.

Hypothesis

It was hypothesised that the amplitude will increase at a magnitude greater than the psychoacoustic looming studies, and relative to the time duration whereby shorter durations would allow less presentation time, therefore would have less amplitude increase than the longer durations.

Results

The amplitude increase analysis was conducted on 27 film looming scenes. The left and right channels amplitude levels were combined to give a total amplitude level and the envelope was measured from 46ms to the peak amplitude level in the sample. The scene specific plots that illustrate the amplitude envelope over time duration are provided in the Digital Appendix. As expected, all of the scenes increased the amplitude as the proximity of the object became closer. The amplitude levels for each scene are listed in Appendix Table A.2. We use the data from columns 'Total Increase' and 'Duration of Measurement', to explore the magnitude of the amplitude increase $\times$ scene duration which are plotted in Figure 3.1.

Looking at the plotted data, we see that the longest looming duration was 2961ms (The Day After Tomorrow) and the shortest duration was 279ms (Sin City). The spread of the data shows there were more looming scenes under 1500ms (21 scenes, 77.78% of total variation) than over 1500ms, and that the number of scenes under 1500ms were spread almost equally across that time span, with 6 scenes $\leq$ 500ms, 8 scenes in the 501 - 1000ms range, and 7 scenes in the 1001 - 1500ms range.

In regard to the magnitude of the amplitude increase, the average amplitude increase was calculated at M = 45.05dB (SD = 15.32), ranging from the minimum amplitude increase of min = 20.43dB (The Day After Tomorrow), to the maximum amplitude increase of max = 89.50dB (I Am Legend). This amplitude increase for the film looming scenes is greater than the levels used in the psychoacoustic looming experiments which ranged from 10-30dB (see section 2.2.2).

In only two of the film looming scenes (Alice in Wonderland at 28.25dB, and The Day After Tomorrow at 20.43dB) the measurement of the amplitude increase equalled the amount that used by the psychoacoustic studies. Further, these two scenes were expected to be at a soft level as they are comprised of broad band noise to depict air and snow, unlike the more focused tones (sine, square, and triangle waves) that was used in the experiments.

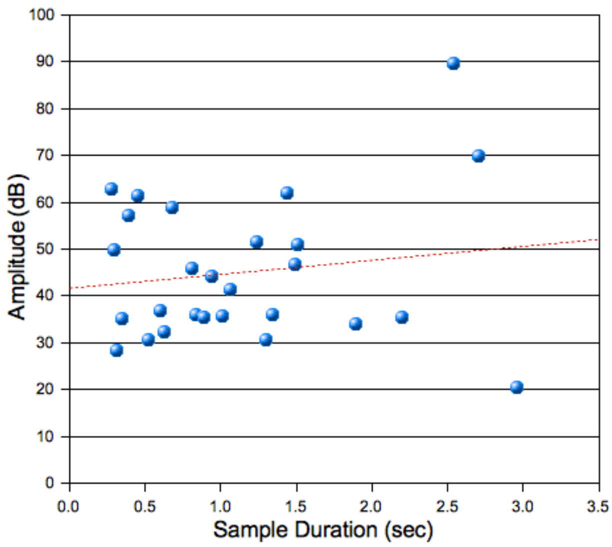

Figure 3.1: Magnitude of the Amplitude Increase × Scene Duration Scatter Plot
The amplitude increase × duration is plotted for each looming scene. A linear regression analysis draws the line of best fit equation: $y = 2.97x + 41.67$, $r^2 = 0.02$. With a poor line of best fit, a negligible r^2, and a broad spread of data, we conclude that the results do not support the hypothesis.

There is also a general trend for the scenes to cluster around the 30 - 60dB level, across all of the durations.

To test the hypothesis (that the magnitude of the amplitude increase would be related to the time duration), a linear regression analysis was conducted to see if there was a relationship between the magnitude of the amplitude increase and the scene duration. The line of best fit was calculated to be $y = 2.97x + 41.67$. With the broad spread of data, a poor line of best fit that increases by a small 2.97dB per second, a negligible coefficient of determination ($r^2 = 0.02$), we conclude that the results do not support the hypothesis, and that the magnitude of the amplitude increase is not related to the duration of the scene.

Magnitude of Amplitude Increase Discussion

Comparison of the film looming scenes amplitude envelope reveals that they all increased the amplitude over time, as is consistent with the (general) physics of an approaching object, and its application as a variable in the psychoacoustic looming studies. The magnitude of the amplitude increase used in the psychoacoustic looming experiments ranged from the 10dB increase [Rosenblum et al., 1987; Cappe et al., 2009] to the 30dB increase [Neuhoff, 2001; Neuhoff and Heckel, 2004]. The magnitude of the amplitude increase was greater for 25 of the film looming scenes, with an average increase of 45dB, and only two scenes increasing at the magnitude used by the psychoacoustic experiments.

It was hypothesised that the magnitude of the amplitude increase may be related to the duration of the looming scene, with longer scenes allowing a greater magnitude of increase, and that sound designers would exploit this opportunity. This hypothesis was rejected however, with no general change (either increasing or decreasing) in the magnitude over the scene durations. This conclusion was further supported by the linear regression analysis which indicated there was no relationship between the magnitude of the amplitude increase and the duration of the scene.

3.4.1.2 Amplitude Envelope Slope

The slope of the amplitude envelope may bias viewers perception of the approaching object and the perceived contact time. According to the inverse square law, an approaching object's amplitude envelope will increase on a non-linear slope at approximately 6dB per halving of distance, with the greatest increase occurring at the closest proximities. The majority of psychoacoustic looming studies however, applied a linear slope to the amplitude envelope. A linear slope presents a greater rate of change earlier in the envelope than a non-linear slope does, therefore the information presented to the observer suggests that the object is approaching at a faster rate, and may prompt people to estimate an earlier time-to-contact than for a non-linear increase. This may suggest to observers that the velocity of the approaching object is accelerating, prompting people to estimate the contact time earlier than a non-linear (ISL) increase.

Aim

The aims of this analysis is to determine

- if the slope of the amplitude envelope is manipulated as a cue to bias human perception of an approaching object,
- if so, is it increasing on a linear or non-linear slope,
- if the slope (m value) is dependent on the scene duration,
- and if it is at a rate similar to the psychoacoustic looming experiments.

Hypotheses

As hyper-real stimuli designed for the entertainment industry, it was hypothesised that the amplitude envelope would increase

1. on a linear slope, as a result of the application of a digital linear fade or crossfade modifying the change in amplitude level in audio editing software, or the sound designers physical movement of sliding a fader on a mixing desk.

2. inversely proportional to the time duration, whereby shorter durations limit the presentation time, thereby increasing the slope m value.

Results

The slope of the amplitude envelope was measured by applying linear, quadratic and cubic equations to determine the line of best fit. The measurement of the slope line for each scene was made from 46ms to the amplitude peak. The scene specific plots that illustrate the amplitude slope and line equations are provided in the Digital Appendix. The slope m-value for each scene listed in Appendix Table A.3.

Assessing the line of best fit for each of the sound sample's reveals that the amplitude envelope's for all scenes have a linear or near-linear slope. As it is not of a higher order relationship, further analyses of the polynomial equations were not conducted. This result supports hypothesis 1, that the film looming scenes apply a linear slope to the amplitude envelope, which may be due to the use of a digital fade or cross-fade, or the physical movement of sliding a fader on a mixing desk.

To test hypothesis 2, that the magnitude of the slope (m value) increases is inversely proportional to the scene duration, we use the envelope equation's m value from each scene which is listed in Appendix Table A.3. The m value was multiplied by 0.1 ($m \times 0.1$) to calculate the slope's increase (in dB) per 100ms and is also listed in Appendix Table A.3, and is plotted in Figure 3.2. The average increase across all of the looming scenes was calculated to be $M = 2.62$dB (standard deviation = 2.23dB, $min = 0.12$dB (Avatar), $max = 8.80$dB (Charlie and the chocolate Factory)).

A linear regression analysis was performed on the data and line equation was calculated to be $y = -0.89x + 3.63$ with the coefficient of determination $r^2 = 0.09$. Although the line of best fit is a poor fit and accounts for only small proportion (9%) of the data variation, the spread of the data and the regression line do illustrate a decreasing trend in the level of the m value with shorter durations having a greater m value, therefore greater increase in amplitude (dB) per 100ms, than the longer duration scenes. However we conclude that this result is not strong enough to support hypothesis 2.

Amplitude Envelope Slope Discussion

The slope of the amplitude envelope may bias human perception of an approaching object and the observers perceived contact time. A linear slope presents a greater rate of change earlier in the envelope than a non-linear slope does, therefore the information presented to the observer suggests that the object is approaching at a faster rate, and may prompt people to estimate an earlier time-to-contact than for a non-linear increase.

Whilst little is documented on the slope of the amplitude envelope in the psychoacoustic studies, with many studies failing to note the slope in their publication, this analysis

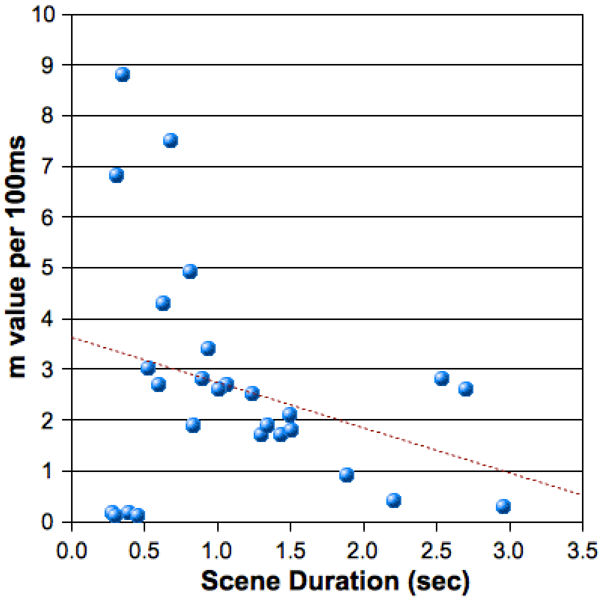

Figure 3.2: Slope m value per 100ms $\times$ Sample Duration Scatter Plot
The m value per 100ms is plotted for each looming scene. A linear regression analysis draws the line of best fit equation: $y = -0.89x + 3.63$, $r^2 = 0.09$.

of the film looming scenes showed that in this sample of film looming scenes, the sound designers applied a linear slope to the amplitude envelope. This technique may have been applied intentionally to bias viewers perception, or perhaps as a result of the audio workstation with digital linear faders modifying the change in amplitude level in audio editing software, or the sound designers physical movement of sliding a fader on a mixing desk.

3.4.1.3 Amplitude Levels

The amplitude level that the looming sound starts at (representing the farthest distance) and peaks at (representing the closest distance) may act as a cue to the approaching object's proximity or velocity.

Aim

The aim of this analysis is to determine if there is a relationship between the minimum and maximum amplitude levels.

Hypothesis

It was hypothesised that the film looming scenes which start at a greater amplitude level would also peak at at a greater level, than the film looming scenes which start at a lower level.

Results

The minimum and maximum amplitude levels for each of the film looming scenes were taken from the envelope analysis data in Appendix Table A.2 and are plotted in Figure 3.3.

The data was averaged for both the minimum and maximum levels, finding that the average minimum level M = -96.17dB (SD = 16.39; min = -143.7dB (I am legend); max = -69.02dB (Alice in Wonderland)); and the average maximum M = -51.12dB (SD = -12.35; min = -69.87dB (Despicable Me); max = -20.6dB (Sin City)).

Looking at the spread of the data we see that the distribution tends to cluster around the minimum level of -115 to -80dB, and the maximum level of -70 to -35 dB. There is a general upward trend with scenes that start at lower amplitude levels also peaking at a lower level, than scenes which start at a greater amplitude level.

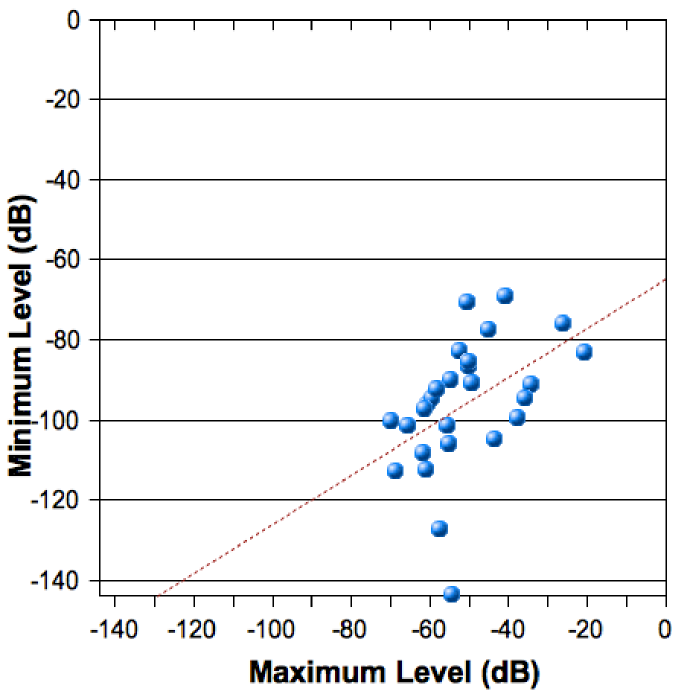

Figure 3.3: Amplitude Minimum × Maximum Levels Scatter Plot
The minimum and maximum amplitude level is plotted for each looming scene. A linear regression analysis draws the line of best fit equation: $y = 0.61x - 64.92, r^2 = 0.21$.

To test the hypothesis that film looming scenes which start at a greater amplitude level would also peak at a greater level, than the film looming scenes which start at a lower level, a linear regression analysis was performed on the data to see if there was a relationship between the level that samples started at (the minimum level) and peaked (the maximum level). The line of best fit was calculated to be $y = 0.61x - 64.92, r^2 = 0.21$. The equation indicates that for every 10dB increase in the maximum level, the minimum level also increased by 6.1dB. This equation accounts for a moderate 21% of the data variation, so we conclude that the results support the hypothesis.

Amplitude Levels Discussion

The use of carefully chosen amplitude levels in film looming scenes may assist in enhancing or exaggerating the distance and depth perception, of an approaching object.

The amplitude level that the looming starts at, and peaks at, may act as an audio cue as to the approaching object's proximity or velocity. As the psychoacoustic studies demonstrated, looming sources which were presented at greater amplitude levels, were perceived to be at a closer proximity and prompted the earlier estimation of contact times, than sources which were presented at lower amplitude levels. The feature analysis of the film looming scenes revealed that sounds which peaked at a greater amplitude level also started at a greater amplitude level.

3.4.1.4 Object Velocity (According to the Inverse Square Law)

As hyper-real stimuli constructed for the entertainment industry, the sound effects used in film looming scenes are designed to elicit an emotional, and sometimes physical, response. One way this may be achieved is by presenting the audio cues at exaggerated levels, prompting people to perceive the approaching object to be closer in proximity, or is approaching at a faster velocity, than how such sounds operates according to the laws of acoustics.

The psychoacoustic looming studies presented the sound sources moving at a broad range of velocities, ranging from 0.05 kph to 180 kph, and found that people preferred different audio cues for judgements of displacement, velocity, and acceleration, at different velocities.

In this analysis, we use the inverse square law to calculate the velocity of the approaching object, according to the magnitude of the amplitude increase. This will provide information about the prevalence of audio cues exaggerating the object's velocity, in our sample of looming scenes.

Aim

The aim of this analysis is to determine if the film scenes in our sample exaggerated the velocity (calculated from the magnitude of the amplitude increase) and if velocity presented is dependent on the duration of the scene.

Hypothesis

It was hypothesised that shorter durations would present the source moving at a faster velocity than longer scene durations.

Results

To calculate the sound source's velocity according to the magnitude of the amplitude increase, we used equation 3.1 the inverted Inverse Square Law by ratio equation [Nave, 2012b].

$$\left[\frac{D1}{D2}\right]^2 = \frac{I2}{I1} \tag{3.1}$$

where:
D1 = Object distance (meters) at the observer,
D2 = Object distance (meters) at the start of the sample (farthest distance),
I1 = Intensity (dB) emitted by the sound source at the observer,
I2 = Intensity (dB) emitted by the sound source at the start of the sample.

The velocity (m/s and kph) for each scene is reported in Appendix Table A.2, as well as the scene duration and distance travelled. The Velocity × Duration data points for each of the film scenes are plotted in Figure 3.4.

The average velocity across all of the scenes was calculated to be M = 36.33 kph (*SD* = 86.12). To test the hypothesis a linear regression analysis was performed on the data to see if there was a relationship between the velocity of the approaching object and the scene duration. The line of best fit was calculated to be $y = 22.87x + 10.29, r^2 = 0.04$, however as the coefficient of determination accounted for only 4% of the data variability we concluded that the line was a poor fit. Looking at the spread of the plotted data reveals that for the majority of scenes (n = 18) the source is moving at velocity of $\leq$ 10 kph. The remaining data points indicate that 5 scenes had a velocity between 10 - 50 kph (M = 32.99 kph, *SD* = 13.40), 2 scenes had a velocity between 50 - 100 kph

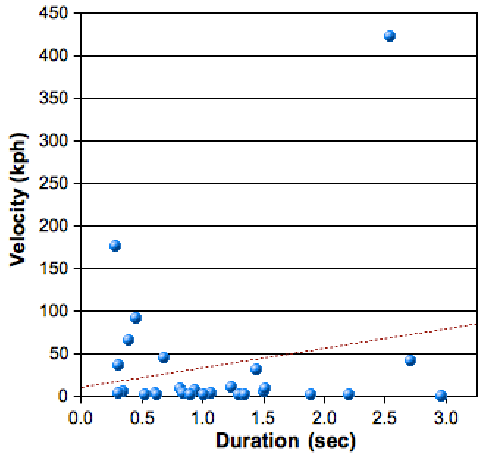

Figure 3.4: Velocity × Scene Duration Scatter Plot
The Velocity × Scene Duration plotted for each scene. **Linear Regression Line Equation:** $y = 22.87x + 10.29, r^2 = 0.04$.

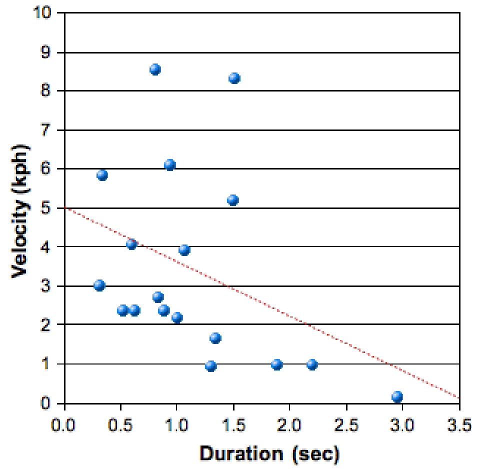

Figure 3.5: Velocity ($\leq$ 10 kph subset) × Scene Duration Scatter Plot
With 18 scenes less than 10 kph we take a closer inspection of the data. **Linear Regression Line Equation:** $y = -1.40x + 5.02, r^2 = 0.15$.

(M = 32.99 kph, SD = 13.40), and 2 scenes has a velocity $\geq$ 100 kph (M = 298.83 kph, SD = 175.27). This suggests that the film scenes in our sample have not overly exaggerated the velocity of the approaching object.

Closer inspection of the 18 scenes that had a velocity of $\leq$ 10 kph (plotted in Figure 3.5) reveals an average velocity of M = 3.41kph (SD = 2.48; min = 0.12 kph; max = 8.53 kph).

A regression analysis was again conducted on the data to test the hypothesis against the $\leq$ 10 kph subset to see if the relationship between the velocity and scene duration would be more accurately explained (with the variability accounting for a greater number of data points) by this densely populated subset.

The line of best fit was calculated to be $y = -1.40x + 5.02, r^2 = 0.15$ with the velocity decreasing by 1.40 kph for every 1 second increase in the duration of the scene, and the coefficient of determination accounting for 15% of the data variability. Whilst the coefficient is still only moderate, the line and spread of the data indicate a downward trend, with shorter durations ($\leq$ 1 second) having a velocity faster than those scenes 1 second duration. However, because the calculated decrease in velocity is only 1.4 kph per 1 second increase to the scene duration, we conclude that these results are not strong enough to support the hypothesis.

3.4.2 Pan Position

As discussed in section 3.3.1 Scene Selection Criteria, the scenes selected for this analysis were chosen as they presented objects approaching the viewer on a frontal trajectory. Any spatial movement and panning of the audio sound source, should also represent the object's direction of arrival. However, because the entertainment industry (film scenes) often use hyper-real stimuli, the scenes in our sample may have exaggerated audio cues to increase the inter-aural differences.

The sound designers may exaggerated the use of panning as a creative attempt to enhance or articulate the object's movement in the scene, alternatively the conservative use of panning may reflect a physical modelling of the object.

Aim

The aim of this analysis is to determine if spatialisation techniques and panning is applied to the audio stimuli, and if this spatial movement is consistent with proximity of the object.

Hypothesis

As a frontal approaching object, it was hypothesised that the audio spatial movement (panning) would reflect the direction of arrival, with the virtual audio source centrally placed with no hard pans to a particular (left or right) channel.

Method

The location of the virtual audio source was determined by applying the tangent panning law using equation 3.2 [Zölzer and Amatriain, 2002] which determines the virtual audio source position in degree's over time.

$$\tan\theta = \frac{AmpL - AmpR}{AmpL + AmpR}\,\tan\theta l \tag{3.2}$$

where:
$\tan\theta l = 45°$
$Amp\ L$ = Amplitude Left (channel)
$Amp\ R$ = Amplitude Right (channel)

The tan panning law was chosen over a simple linear interpolation of the two channels, as the linear interpolation would create a drop in the loudness - a “hole in the middle” of the stereo transmission [Zölzer and Amatriain, 2002]. Applying the tan panning law preserves the loudness of the audio source (object) across the azimuth. Tan was also chosen over the sin panning law, as the tan positions the audio at 45° which is it slightly wider, and will highlight any movement off-centre, than the more conservative 30° sin.

Results

The virtual sound source position was measured for each of the 27 film looming scenes with the sample specific plots that illustrate the position in degree's over time provided in the Digital Appendix. The range of the virtual sound source, from the leftmost position to the rightmost position is plotted for each scene in Figure 3.6.

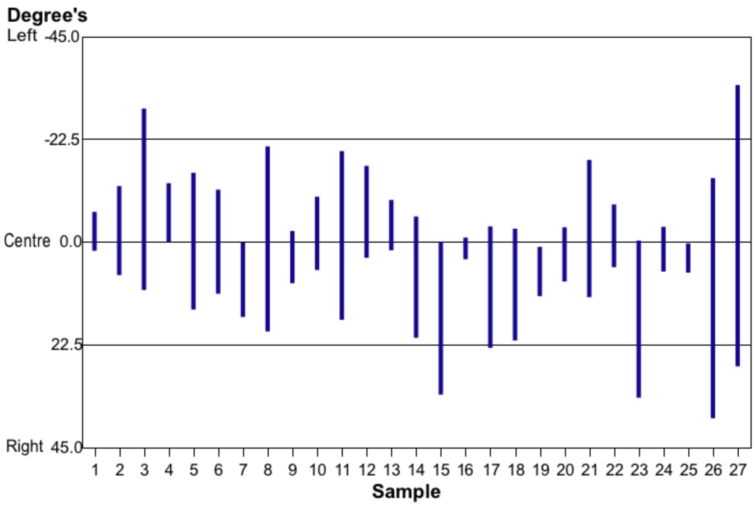

Figure 3.6: Audio Virtual Source Position × Sample
16 scenes (59.25% of total scenes; scenes 1, 3, 4, 7, 9, 12, 13, 14, 15, 17, 18, 19, 20, 23, 25, 26) tend to be weighted to one particular side, whereas 11 scenes (40.74% of total scenes; scenes 2, 5, 6, 8, 10, 11, 16, 21, 22, 24, 27) are more evenly weighted.

Looking at the spread of the results, overall, the scenes tended to remain somewhat central, with no scenes having a hard pan to either the right or left channel. Although there are no hard pans to a single channel, 16 scenes (59.25% of total scenes) tend to be weighted to one particular side, whereas 11 scenes (40.74% of total scenes) are more evenly weighted.

Looking at the average leftmost position across all of the scenes, the leftmost virtual audio source had a M = -9.58°, SD = 9.25, (min = -34.34°, max = 1.08°).

The average rightmost position across all of the scenes, was M = 13.86°, SD = 10.5, (min = -0.10°, max = 38.6°).

Pan Position Discussion

Overall, the spatial placement of the virtual source is as expected. It supports our hypothesis of a central position, which is representative of a central frontal approaching object. Although there were no hard pans to a single channel, a number of the scenes tended to be slightly weighted to one particular side, whereas other scenes were more evenly weighted. This may be a creative technique of the sound designer to draw a little spatial interest to the scene, or perhaps it may be a legacy from the position of the microphone when the audio sample was collected.

3.4.3 Spectral Content

3.4.3.1 Spectral Centroid

As discussed in section 2.1 Acoustics and Psychoacoustics of a Moving Object, frequency change over time is an important physics based cue for an approaching object, occurring both as a change in the fundamental frequency due to the doppler effect, and a change in the spectral spread due to environmental attenuation of certain frequencies according to the object's distance.

Frequency change, in the form of the doppler shift was also investigated in a number of the psychoacoustic looming studies (discussed in section 2.2.5) which concluded that the doppler shift was an important audio cue for perceiving movement in depth, judging an object's velocity or displacement, and is particularly preferenced for fast velocities ($\geq$ 180 kph). It was also demonstrated that many people erroneously believe that the doppler shift causes the pitch of an object approaching at a constant velocity to rise as the object nears the observer.

As hyper-real stimuli designed for the entertainment industry, the spectral centroid in our film looming scenes may be designed to model a physics based approach, or alternatively, it may be designed to model human expectation.

Conducting a feature analysis on the film looming scenes will reveal information about the spectral content such as the level of the spectral centroid, how the level changes as the object approaches the observer, and if this change is similar to physic's based or perceptual based models.

Aim

The aim of this analysis is to ascertain

- the level of the spectral centroid and how it changes as the proximity of the object approaches the observer,
- if there is a relationship between the minimum and maximum levels,

- the magnitude of the increase and duration.

Hypotheses

It was hypothesised that the

1. minimum spectral centroid frequency would affect the level of the maximum spectral centroid frequency, whereby scenes that have higher minimum frequencies also having higher maximum frequencies, than scenes with lower minimum frequencies.
2. duration of the measurement would affect the spectral centroid's magnitude of the change, with longer durations presenting a greater magnitude than shorter durations.

Results

Fourier analyses (FFT) were performed on each of the 27 film looming scenes, with the spectral centroid determined by the weighted mean of the frequencies in the spectrum. Scene specific plots that draw the spectral centroid over time are provided in the Digital Appendix, whilst the minimum level, maximum level, and range of spectral centroid are listed in Appendix Table A.4.

The spectral centroid was calculated for each of the film scenes as it changed over time. Time measurements were made between the minimum and maximum frequencies. The results indicate that more scenes increased the spectral centroid (n = 21, 77.78% of total scenes) as the proximity of the object became closer, than decreased (n = 6, 22.22% of total scenes). For the 21 scenes that increased the spectral centroid, it started on average at M = 1326.6 Hz (E6 +11 cents), and peaked on average at M = 4828.9 Hz (D8 +47 cents), an average increase of 3502.3 Hz (almost two octaves). For the 6 scenes that decreased the spectral centroid, it started on average at M = 4072.4Hz (C8 -48 cents), and decreased to average M = 1169.1Hz (D6 +8 cents), an average difference of 2903.3Hz (almost two octaves). It is interesting to note that the average minimum frequencies (notes E6 +11 cents and D6 +8 cents) for both the increase and decrease directions, and the average maximum frequencies (notes D8 +47 cents and C8 -48 cents) for both the increase and decrease directions, were only one musical tone apart.

The minimum and maximum spectral centroid frequencies for each film scene are plotted in Figure 3.7. Looking at the spread of the results, we see an upward trend, where the scene's that have a greater minimum frequency also having a greater maximum frequency.

To test hypothesis 1 a linear regression analysis was performed on the data to see if there was a relationship between the minimum and maximum frequency levels. The line

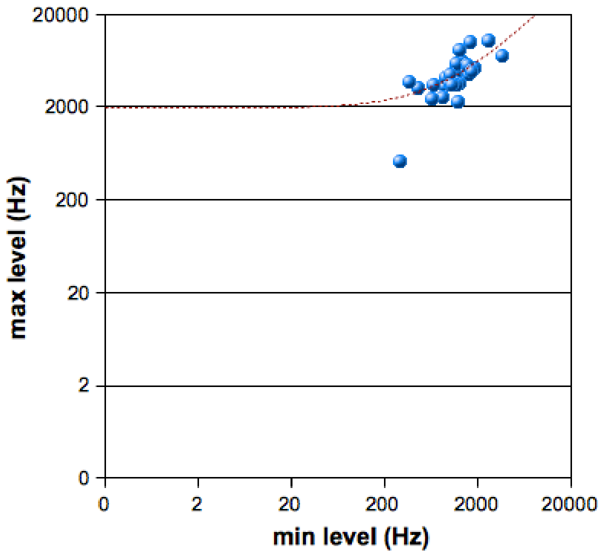

Figure 3.7: Spectral Centroid Minimum × Maximum Frequencies Scatter Plot
The minimum and maximum spectral centroid frequency is plotted for each looming scene. A linear regression analysis draws the line of best fit equation: $y = 2.11x + 1931.7$; $r^2 = 0.45$.

of best fit was calculated to be $y = 2.11x+1931.7$ with the spectral centroid's maximum frequency increasing by 2.11 Hz for every 1 Hz increase in the minimum level, and the coefficient of determination accounting for 45% of the data variability.

We conclude that this strong result support hypothesis 1, that the frequency of the minimum spectral centroid affects the level of the maximum spectral centroid frequency.

To investigate the relationship between the magnitude of the frequency change and the duration of the measurement, we plot the data points for each film looming scene in Figure 3.8.

Looking at the spread of the results, we see that shorter durations, in particular the durations ≤ 1 second have a much broader distribution of the magnitudes (ranging from 215 Hz to 8085 Hz) compared to the durations > 1 second (which range from 1701 Hz to 3577 Hz).

To test hypothesis 2 a linear regression analysis was performed on the data to see if there was a relationship between the magnitude of the frequency change and the duration of the measurement. The line of best fit was calculated to be $y = -1205.27x+4229.88$ with the magnitude of the spectral centroid frequency decreasing by 1205.27 Hz for every 1 second increase in duration, however the coefficient of determination only accounts for a moderate 15% of the data variability.

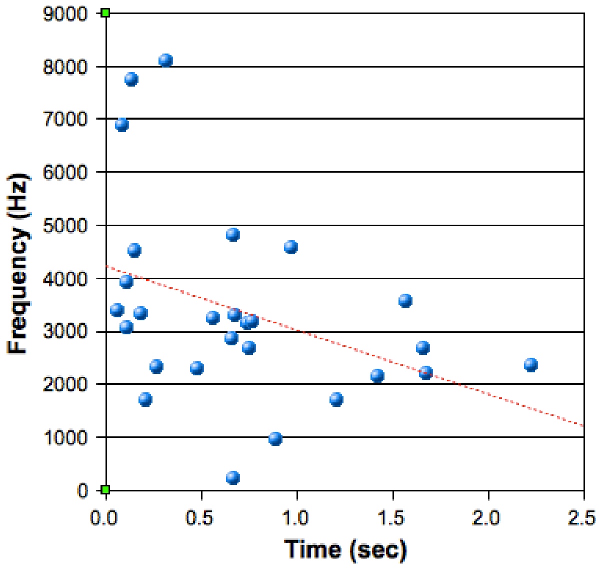

Figure 3.8: Spectral Centroid Magnitude of Frequency Change × Duration of Measurement Scatter Plot
The magnitude of the frequency change × duration is plotted for each looming scene. A linear regression analysis draws the line of best fit equation: $y = -1205.27x + 4229.88$; $r^2 = 0.15$.

Whilst the plotted results show a reduction in the distribution of the frequency magnitude, we conclude that the result from the regression analysis is not strong enough support hypothesis 2 regarding the relationship between the magnitude of the spectral centroid frequency and duration of the measurement.

Spectral Centroid Discussion

The majority of scenes increased the spectral centroid as the proximity of the object became closer. This is inconsistent with the doppler shift's decrease in frequency as an audio cue, which as we discussed in Section 2.1.2 the pitch of an object approaching on a frontal midline plane remains at a constant higher frequency until it reaches the observer where it drops to the actual transmitted frequency. However this result does reflect the mistaken common belief that the doppler shift actually causes a rise in the pitch as the object nears the observer [Neuhoff and McBeath, 1996]. Sound designers and post production technicians that hold this erroneous belief, or want to meet viewers expectations, may give one explanation for this rise in spectral centroid.

The minimum and maximum spectral centroid frequencies were moderately high, considering many of the approaching objects were large vehicles (such as cars, motorbikes, and spaceships) which could be expected to have lower spectral content. Regression analysis found a relationship between the minimum spectral centroid frequency, and the maximum spectral centroid frequency supporting hypothesis 1. There was an upward trend where the scene's that had a greater minimum frequency also had a greater maximum frequency.

Regression analysis of the magnitude of the spectral centroid change over time, suggested there was a small relationship for a decrease in the range of the magnitude as the duration increased, however it cannot be ignored that the shorter durations had a broader distribution of the data, than longer durations. As such, we conclude that the results were not strong enough to support the hypothesis.

3.4.3.2 Spectral Spread

As discussed in section 2.1.4 Environmental Attenuation, the distance of the object affects the frequencies which are audible.

For a sound source that is comprised of a wide range of frequencies (such as the looming film scenes in our sample) high frequencies are attenuated more than the lower frequencies. When an object is at a great distance from the observer, the spread of the spectrum will be narrower. As the proximity of the object becomes closer to the observer, the higher frequencies with shorter wavelengths become apparent, causing the magnitude of the spread to also increase, resulting in a broader frequency spectrum.

Aim

The aim of this analysis is to determine the magnitude of the spectral spread, and if it changes over the duration of the looming scene.

Hypotheses

It was hypothesised that the magnitude of the spectral spread would increase as the proximity of the object becomes closer to the observer.

Results

The scene specific plots that illustrate the spectral spread over time are provided in the Digital Appendix.

In all of the scenes the spectral spread decreased as the proximity of the object became closer to the observer. However this was not a smooth linear or near-linear transition.

Ten of the film looming scenes (37.04% of total trials) presented the maximum magnitude of spectral spread at the start of the scene (with an average spread of $M = 1327.3$ Hz, $SD = 653.5$) that decreased to the minimum spectral spread at the end of the scene (with an average of $M = 177.8$ Hz, $SD = 47.97$).

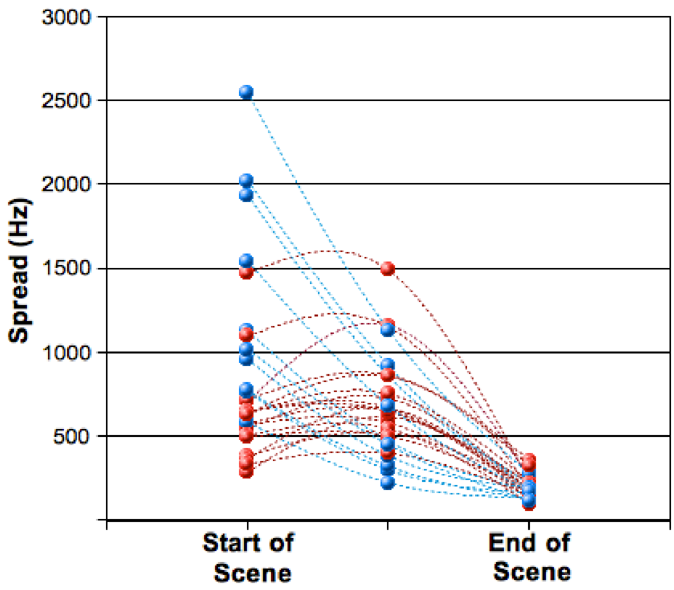

Figure 3.9: Spectral Spread Line Chart
The magnitude of the spectral spread is plotted. The blue data points represent the scenes which decrease in spread magnitude over the duration of the scene, whilst the red data points represent the scenes which first increase the magnitude of the spread, before decreasing at the end of the scene.

Seventeen looming scenes (67.96% of total trials) first increased the spread after the start of the sample, then decreased towards the end of the scene. They began with an average spectral spread of M = 624.07Hz (SD = 287.12), and increased to the maximum average spread of M = 752.70 Hz (SD = 280.96), then decreased to the minimum spectral spread at an average of M = 190.18 Hz (SD = 62.78).

Of these 17 scenes that first increased before decreasing, 8 of the scenes presented the maximum level just $\leq$ 0.025 seconds after the start of the signal (M = 0.013, SD = 0.007; and increased the spectral spread by M = 77.550Hz, SD = 98.121). One explanation for this result occurring so soon at the start of the sample, may be the signal-to-noise ratio in the first few windows, incorrectly measuring the broad spread of the noise. While for the 9 other scenes, the maximum spread occurred at 0.145 to 0.715 seconds after the start of the sample (M = 0.352 sec, SD = 0.172, however the spectral spread only increased by a small M = 174.033 Hz, SD = 146.147).

Spectral Spread Discussion

All of the scenes decreased the spectral spread as the proximity of the object neared the observer, disproving our hypothesis that the spectral spread would increase.

This was an unexpected result considering that it was somewhat consistent across all of the scenes in our sample. One explanation for this result is a signal-to-noise ratio impact on the analysis, whereby the signal level of the sound source was too low at the start of the sample, therefore the analysis erroneously indicated a greater spectral spread. As the object (sound source) becomes closer to the observer, the signal level becomes more accurately measurable and it provides a more reliable spectral spread. However the plotting of the spectral spread on the chart is now erroneously drawn as a starting with a great spectral spred, that decreases as the sounds signal becomes louder and more measurable.

The subcategorisation of the scenes into those in which the maximum spectral spread occurred at the start of the sample, versus those that increased the spectral spread, before decreasing was made to see if any early frequency changes to the envelope affected the sound, as a cue. However, the increase in spectral spread is small (from M = 77.55 Hz to M = 174.03 Hz) and occurred at a small time scale (from M = 0.013 to M = 0.352 sec). In the context that the spectral spread is decreasing overall (perhaps due to the signal-to-noise ratio) then any affect on the envelope and people's perception would be minimal to non-existent.

3.5 Discussion

This feature analysis study was conducted on the 27 looming scenes to understand which features might be acting as cues for approaching objects, how the features changed over time, and the magnitude of there change.

The audio features that were analysed include the magnitude of the amplitude increase, the amplitude levels, amplitude slope, audio pan position, spectral centroid, and spectral spread.

To summarise the results, the amplitude increased on a linear slope, at an average of 45.05dB (SD = 15.32) on a linear / near-linear slope, and there was no relationship between the magnitude of the amplitude increase and the duration of the sample. The magnitude of the amplitude increase in our sample of film looming scenes is greater than the magnitude of the amplitude increase used in psychoacoustic looming experiments (which ranged from a 10dB increase [Rosenblum et al., 1987; Cappe et al., 2009] to a 30dB increase [Neuhoff, 2001; Neuhoff and Heckel, 2004]). This greater magnitude of the amplitude increase presented in the film scenes may contribute to biasing viewers perception, emotion, and engagement levels.

The audio's virtual source position, showed that the position was mostly central, with none of the scenes having a hard pan to a single channel. The position was equally spread for 40.74% of the scenes, whereas 59.25% of the scenes tend to be slightly weighted to one particular side. Across all of the scenes, the leftmost position was M = -9.58°, whilst the rightmost position was M = 13.86°, a difference of 23.44°.

The frequency of the spectral centroid commenced at an average minimum frequency M = 1957.8 Hz and an average maximum frequency M = 3444.57 Hz. 21 film scenes

(77.78% of scenes) increased the spectral centroid as the proximity of the object heard the observer. It increased by an average $M = 3502.3$ Hz, almost two octaves, to $M =$ 4828.9Hz. This increase in the spectral centroid is inconsistent with the doppler shift as an audio cue. Regression analysis of the magnitude of the spectral centroid change per scene duration, showed there was a small relationship between the magnitude of the spectral centroid change, and it decreasing as the duration of the scene increased, however this was rejected due to the broader distribution of data points for magnitude, in the shorter durations.

The spectral centroid frequencies are somewhat high, considering many of the approaching objects are large vehicles (such as cars, motorbikes, and space-ships), which could be expected to have lower spectral content. We speculated that the sound effects may be reflecting environmental effects (geometric spreading, atmospheric absorption and ground reflection) that ensure that broader spectral content is received by the observer, than when the object is at a closer distance. However, further analysis of the spectral spread indicated that all of the scenes decreased the spectral spread as the proximity of the object became closer, disproving our hypothesis that the spectral spread would increase, consistent with environmental attenuation. This was an unexpected result considering that it was a consistent result across all of the scenes in our sample. One explanation for this result is a signal-to-noise ratio issue related to the analysis, whereby the signal level of the sound source was too low at the start of the sample, thereby erroneously indicating a greater spectral spread. As the object (sound source) approaches the observer, the signal level becomes more accurately measurable, therefore providing a more reliable spectral spread.

The results from this study of film looming scenes demonstrates that the sound effects have exaggerated key features which act as audio cues for objects moving in depth, than would be present in real world sounds generated according to the laws of physics, and the stimuli used in psychoacoustic looming studies investigating human perception and response to approaching objects. We propose that the use of sound effects which have been created by sound designers and post production technicians with the purpose of maximising the viewers experience and perception of the virtual environment warrant further investigation. A psychoacoustic study investigating human perception and response to the film looming stimuli would provide greater insight about how these cues influence the physiological and emotional responses to dynamic complex stimuli, with results being applicable to industry use in the design of sound effects in film, gaming and simulators.

Chapter 4

Responses to Designed Film Looming Stimuli

As we found from our feature analysis study in Chapter 3, the depiction of approaching objects in film often involves the presentation of hyper-real stimuli that use complex sounds with multiple audio cues. The sounds are designed to maximise the viewer experience and create an immersive environment that draws an observer into a believable virtual scenario.

Whilst parametric models exist that accurately generate and modify a sound sample according to various laws of physics, sound designers and post production technicians often prefer to draw on there own creative and listening skills when constructing a sound. A number of excellent textbooks thoroughly address (and teach) the technical and creative skills for designing sound (comprehensive examples include Farnell [2010]; Holman [2010]; Katz and Katz [2007]; Izhaki [2009]; Stevens and Raybould [2013]).

However, psychoacoustic looming studies measuring human perception of, and responses to, approaching objects often use artificial sound sources (sine, triangle, and square waves) and simple audio cues. Studies using real world or hyper-real stimuli, rarely make it to scientific peer-reviewed publication. This is perhaps due to the rigorous nature of critical review questioning the internal validity of the experimental design, and the assumption that results produced using artificial stimuli, are transferable to more complex stimuli. Nonetheless, the pursuit of absolute control in experiments has left a gap in the knowledge of how humans perceive and respond to real, and hyper-real looming situations. Further, it cannot conclusively be said that this gap can be filled by the information obtained from studies that use simple (auditory) stimuli, but have total experimental control.

In this chapter, we seek to address this gap (in the knowledge of how humans perceive and respond to real and hyper-real looming situations) in psychoacoustic experiments presenting hyper-real looming stimuli. We build upon our feature analysis study conducted in Chapter 3 by investigating human responses to the film looming stimuli, in order to understand how people respond to looming stimuli with multiple complex audio cues, that have been designed for maximum effect.

4.1 Aim

The aim of this study is to determine if a persons response to a looming object differs with the addition of sounds that use multiple audio cues, as opposed to looming scenes with no sound.

4.2 Hypotheses

It is hypothesised that:

1. the presentation of the sound stimuli that has multiple auditory looming cues applied to a complex sound source, will prompt people to:
   (a) underestimate the impact time of the approaching object, thereby eliciting a faster response time than the scenes with no sound;
   (b) express greater valence and arousal ratings, than the scenes with no sound;
   (c) express greater engagement ratings, than the scenes with no sound;
2. trials which prompt a greater underestimation of the impact time would also be rated with greater engagement, valence, and arousal levels, than the scenes with less underestimation of the impact time;
3. trials with a greater emotion (valence / arousal) ratings would also prompt a greater engagement ratings.

4.3 Method

4.3.1 Design

The study used a within-subjects design. There was one independent variable - Presentation, which was comprised of three levels:

- Image Only,
- Sound Only,
- Sound + Image.

There were four dependent variables:

- Time-to-Impact,
- Valence,
- Arousal,
- Engagement.

4.3.2 Participants

A sample of 15 participants naive to the aims and purpose of the study were recruited. They were Ph.D students and Postdoctoral researchers from Queen Mary, University of London aged between 20 and 36 years (M = 27.07 years, SD = 4.70), with more male participants than female participants (11 males, 4 females). The participants visual and auditory abilities were self reported in a questionnaire, and further physiological tests were not made. All participants reported normal hearing, with 6 participants correcting their vision with glasses or contact lenses. These participants wore their glasses during the experiment.

4.3.3 Stimuli

The stimuli consisted of 27 film scenes (listed in Appendix Table A.1 with the files included in the Digital Appendix) that presented object's moving towards the viewer, and were comprised of both audio and visual components. Each scene was between 313 and 3007ms in duration, with 13 samples $\leq$ 1000ms. The scenes were presented via computer with the visual stimulus displayed on the monitor, and the auditory stimulus was transmitted through a pair of headphones.

The 27 scenes were presented in each of the three presentation conditions -

- Image Only - which presented the looming image, with no sound stimuli.
- Sound Only - which presented the sound stimuli whilst a black screen was displayed on the computer.
- Sound + Image - which presented both the looming image and sound stimuli.

Each scene condition was presented once only (totalling 81 trials) and in a randomised order. The presentation of each trial was limited to once only as further presentations would have introduced memory and learning biases.

4.3.4 Apparatus

Participant's were located at a computer workstation with their head distanced approximately 40 cm from the computer monitor and eyes level with the centre of the monitor. A Mac Pro 1.1 with a NEC MultiSync EA221WM (LCD) monitor was used. The screen size was 22 inches with the resolution set to 1680×1050 pixels and the display was calibrated to a refresh rate of 60 Hz. The auditory stimulus was presented through Sennheiser HD515 headphones. The program MAX / MSP / Jitter version 4.6 was used to construct the software application that presented the auditory and / or visual stimuli; presented the trials in a randomised order, timed the participant's responses using the computer's internal clock, and collected the participant responses in a text file.

Using a computer dedicated to the experiments, the computer and monitor's brightness, frame rate, sound output level, and general equipment settings remained at the same set levels across the experiments.

4.3.5 Dependent Variable Measurement

Four dependent variable measurements were made, these are:

Time-to-Impact

Image motion tracking was performed on each scene to determine the approaching object's position and size over time. With the clock starting at frame 1, we timed the frame in which the object encompassed the greatest area on the screen, this is what we considered as the impact time and is called the 'peak'. Participant's responses to the stimuli (by pressing the keyboard space bar when they thought the object reached them) was also timed.

Figure 4.1: Participant's Response Task To Stimuli
Participant's pressed the keyboard space bar when they thought the object would reach them.

Using equation 4.1, the 'peak' time was subtracted from the response time, to give the amount of time that was underestimated or overestimated, and for the purpose of this study is called the 'Time-to-Impact'.

$$TI = RT - P \tag{4.1}$$

where:
TI = Time-to-Impact, the amount of time (ms) which was under- / over-estimated. Underestimation is indicted in the negative value range, and overestimation is indicated in the positive value range.
RT = Participant's response time (the time that participant's pressed the space bar when they thought the object reached them).
P = Peak time (the timed frame in which the object encompassed the greatest area, and measured from the image motion tracking).

Emotion: Valence and Arousal

To understand the participant's emotional response to the looming scenes, they were asked the question "When presented this scene, I felt..." and instructed to rate their emotion on a 2-dimensional 13-point valence / arousal scale. Valence was rated on the X axis and ranged from displeasure to pleasure, whilst arousal was rated on the Y axis, ranging from sleepy to aroused. To provide a reference for the combinations of the minimum and maximum valance / arousal, the quadrants were also labelled using the terms Distress, Excite, Content, and Bored, which were derived from Russell 1980.

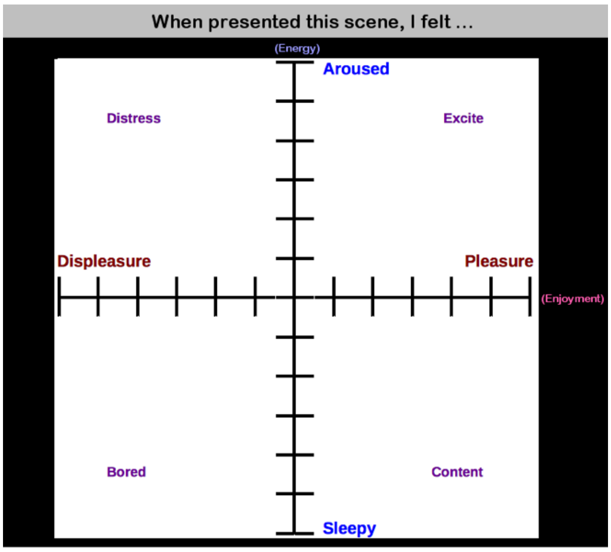

Figure 4.2: Valence / Arousal 2D Rating Scale
Valence is measured on the X axis with 13-points ranging from displeasure to pleasure, whilst arousal is measured on the Y axis with 13-points ranging from sleepy to aroused. The minimum and maximum combinations of the valance / arousal sees the quadrants labelled as distress, excite, content, and bored.

Engagement

To understand what the participant's thought of the quality of the looming scene they were presented, they were asked "How engaging was the scene?" and to rate their response on a 9-point visual analogue scale ranging from dull to captivating.

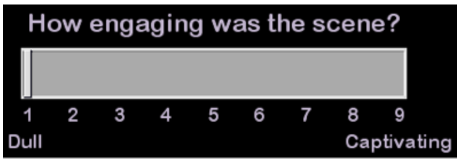

Figure 4.3: Engagement Rating Scale
A 9-point visual analogue scale ranging from dull to captivating.

4.3.6 Procedure

The participant's sat at the computer workstation and were informed of the experimental procedure. They were given an information sheet summarising both the procedure and the ethics approval, signed a consent form, and completed a background questionnaire asking questions on gender, age, and whether they have had corrections made to their vision or hearing (the documents are included in the Digital Appendix).

Before commencing the experiment, the participant's completed a practice study using 6 looming scenes (that were not additionally presented in the experiment). It was conducted as a supervised learning procedure to provide them with the opportunity to comprehend the experiment, the procedure, the micro time scale of the stimulus, and how to complete the task. Participant's were then instructed to start the experiment when ready.

The task required the participant's to watch and / or listen to the scene of an approaching object, and to press the keyboard space bar when they thought the object reached them. A pop-up questionnaire was then displayed on the computer screen, asking the participant's to rate their valence / arousal level and engagement.

Each trial lasted for a total duration of $0.3 - 3.0$ seconds (depending on the looming scene presented) and the participant's were not time restricted on how long they spent answering the questions. Once they had submitted their answers, a 4-second break was then given between each trial in which an image of 'visual white noise' (see Appendix Figure B.1) was displayed on the screen, and no sound was output through the headphones. The experiment lasted for approximately 25 minutes and participant's were not given any information implying there might be 'correct', 'incorrect' or 'preferred' responses.

4.4 Results

Results Analysis Method

To reduce repetition in the thesis the following method was used for each analysis and is explained here as a space saving measure.

As ANOVAS are sensitive to outliers preliminary analyses were conducted on the data to check for outliers. Any data points that were $\pm$ 3 standard deviations from the mean were removed, and are noted in each analysis.

One-way repeated measures ANOVAS were then conducted on the data to compare the audio cue or sound source condition by the time-to-contact, arousal, valence, and engagement ratings. The means and standard errors are noted in each analysis and provided in detail in the appendices.

The Mauchly's test of sphericity was also performed on the data for each of the ANOVAS to determine if the assumption of sphericity had been violated or not. It is noted in each analysis where the degrees of freedom needed to be corrected using either the Greenhouse-Geisser or Huynh-Feldt correction.

Post-hoc tests with pairwise comparisons between the conditions were also conducted for each ANOVA. The descriptives (with a Bonferroni adjustment to correct for a possible increase in type 1 (false positive) errors associated with multiple comparisons) are provided in the appendices, whilst in each analysis section we discuss the comparisons between the conditions and if the results support the hypothesis.

4.4.1 Presentation

Early exploration of the results showed a bias in the response of participant's to the Gattaca film scene, with an average overestimation of the contact time for the Image Only condition M = 1826.82ms. ANOVAS are sensitive to outliers, and as this overestimation was not caused by sound cues (since it was the Image Only condition), the scene (with each of its presentation conditions) were removed, and the data collected from the remaining twenty-six film scenes was used in the analyses.

A total of 390 trials were presented per presentation condition (Sound Only, Image Only, Sound + Image). The fifteen participant's each received seventy-eight trials, comprised of the twenty-six scenes presented once for each of the presentation conditions. The participant's responses were compiled and averaged for each trial.

4.4.1.1 Presentation × Time-to-Impact

The time-to-impact was averaged across all of the participants responses, for each scene sample presentation condition, and is plotted in Figure 4.4.

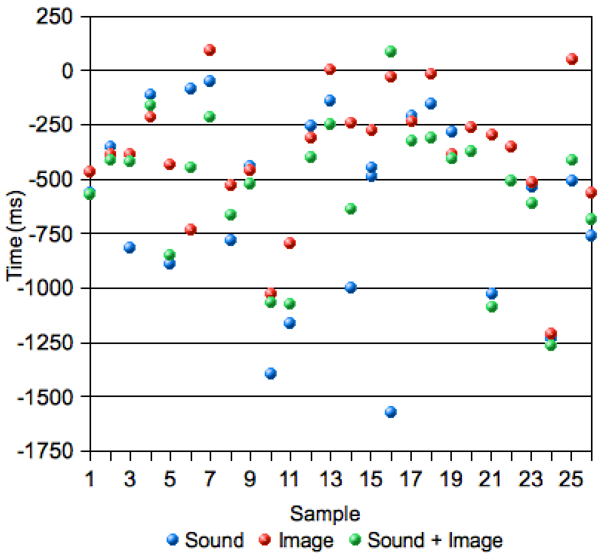

Figure 4.4: Presentation $\times$ Time-to-Impact $\times$ Looming Scene Scatter Plot The time-to-impact for each looming scene presentation condition was averaged across all of the participants, and is plotted. The contact time occurred at 0ms, with any underestimation shown in the negative range of the scale, and overestimation shown in the positive range.

Looking at the spread of the data, the majority of the trials prompted participant's to underestimate the contact time rather than overestimate it, with the conditions that contained sound (being the Sound Only condition, and the Sound + Image condition) having a greater underestimation, than the condition with no sound (the Image Only condition).

The time-to-impact was then averaged across all of the participants responses and scenes, for each presentation condition, and is plotted in Figure 4.5.

For all conditions the average time-to-impact value was before the 'peak', that is, when averaged across all of the looming scenes, each presentation condition (Sound Only, Image Only, and Sound + Image) prompted participant's to underestimate when they thought the object would contact.

The condition which generated the greatest 'time-to-impact' (therefore greatest underestimation of the impact time) was the Sound Only condition (M = −598.88ms, SE = 84.50); followed by the Sound + Image condition (M = −540.54ms, SE = 61.86); and the Image Only condition (M = −384.05ms, SE = 60.78).

To test hypothesis 1A (that the addition of the complex sound cues to a looming image will prompt people to underestimate the contact time of the approaching object, thereby eliciting a faster response time than the scenes with no sound), a one-way repeated measures ANOVA was conducted, and the descriptives are listed in Appendix Table B.2. Mauchly's test indicated that the assumption of sphericity had been violated, $x^2(2) = 11.436, p = 0.003$, therefore degrees of freedom were corrected using Greenhouse-Geisser

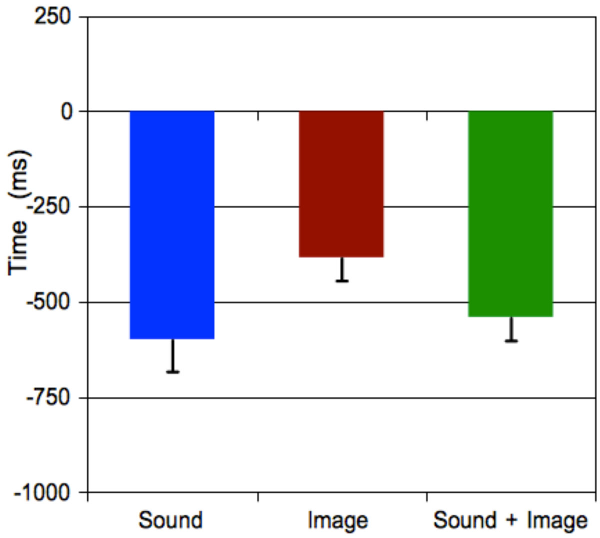

Figure 4.5: Presentation × Time-to-Impact Bar Chart
Presentation × time-to-impact. Results are plotted for each presentation condition (Sound Only, Image Only, Sound + Image) that were averaged across all of the participants ratings. Error bars indicate the standard error for each condition. The average time-to-impact for each condition are Sound Only condition ($M = -598.88$ms, $SE = 84.50$), Image Only condition ($M = -384.05$ms, $SE = 60.78$), Sound + Image condition ($M = -540.54$ms, $SE = 61.86$).

estimates of sphericity $\epsilon = 0.725$. The results indicate that the presentation condition had a significant and medium effect on the estimated time-to-impact $F(1.450, 36.257) = 5.725, p = 0.013, r = 0.39$.

Post-hoc tests with pairwise comparisons revealed a significant difference in the time-to-impact for the presentation conditions that had sound (Sound Only and Sound + Image) as compared to the condition (Image Only) that did not have sound. The Sound Only condition versus the Image Only condition $\mathrm{CI}_{.95} = -414.377$ (lower) -15.277 (upper), $p = 0.032$; the Sound + Image condition versus the Image Only condition $\mathrm{CI}_{.95} = -262.427$ (lower) -50.553 (upper), $p = 0.003$. The pairwise comparisons are listed in Appendix Table B.3.

The results indicate that the presentation condition had a significant and medium effect on the estimated time-to-impact, and that the conditions which presented sound (the Sound Only, and Sound + Image conditions) were underestimated to a significantly greater extent than the Image Only condition, therefore supporting hypothesis 1A.

4.4.1.2 Presentation × Emotion (Valence / Arousal)

The valance / arousal ratings × presentation condition (Sound Only, Image Only, Sound + Image) were averaged across all of the participants responses, for each of the 26 looming scenes and are plotted in Figure 4.6.

The spread of the data shows that the Sound + Image condition tends to have a greater number of trials with high valence / arousal ratings, whilst a comparison of the Sound Only and Image Only conditions suggest that they both have a similar spread of valence ratings, however the Sound Only condition tends to have more samples with greater arousal ratings.

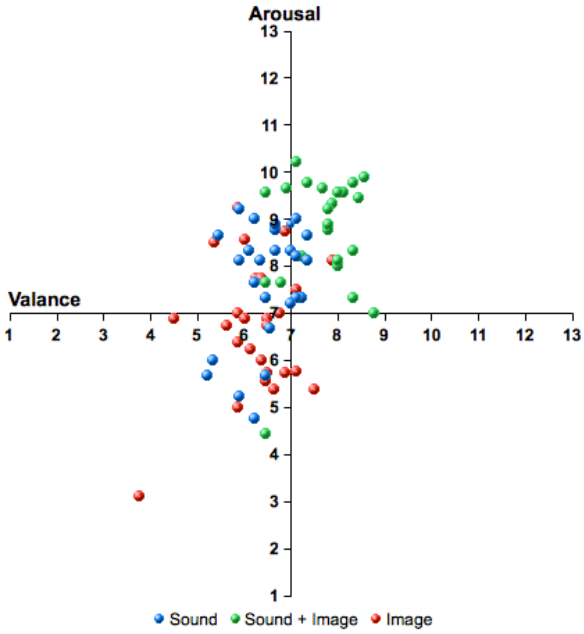

Figure 4.6: Presentation × Valence / Arousal Scatter Plot
Participant's results were compiled and averaged, giving the valance / arousal ratings for each Presentation condition (Sound Only, Image Only, Sound + Image, by each of the 26 looming scenes.

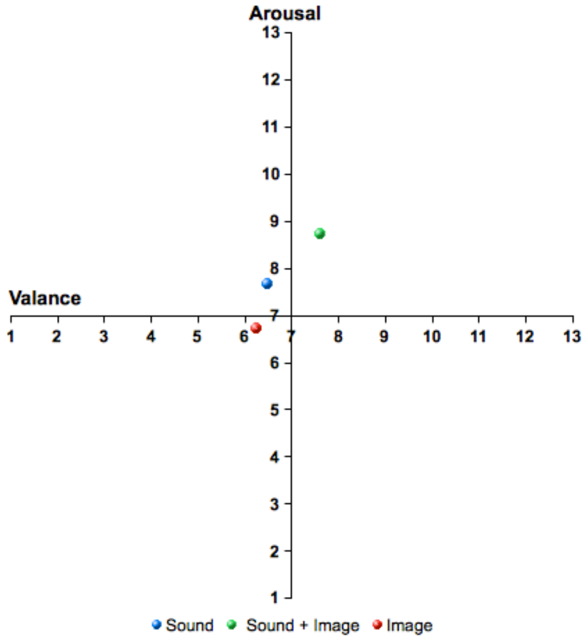

Figure 4.7: Presentation × Valence / Arousal (Averaged) Scatter Plot
The valence / arousal results plotted in Figure 4.6 were then averaged across all of the looming scenes, for each presentation condition (Sound Only, Image Only, and Sound + Image). **Sound:** Valence $M = 6.48$, Arousal $M = 7.68$; **Image:** Valence $M = 6.26$, Arousal $M = 6.72$; **Sound + Image:** Valence $M = 7.60$, Arousal $M = 8.72$.

The average was then calculated for each presentation condition, across all of the film looming scenes, and is plotted in Figure 4.7. The results indicate that the Sound + Image condition had the greatest valence and arousal ratings, followed by the Sound Only condition, and the Image Only condition.

To test hypothesis 1B (that the addition of the complex sound cues to a looming image will prompt people to have greater valence and arousal ratings, than the scenes with no sound) one-way repeated measures ANOVAS were conducted with the descriptives listed in Appendix Table B.4.

For valence, Mauchly's test indicated that the assumption of sphericity was not violated $x^2(2) = 2.269, p = 0.322$, therefore degrees of freedom did not need to be corrected. The results indicate that the presentation had a significant, and very large effect on the valence rating $F(2, 50) = 42.07, p =< 0.001, r = 0.84$.

Post-hoc tests revealed a significant difference in the valence rating for the Sound + Image presentation condition, compared to both of the uni-modal (Sound Only, and Image Only) conditions. The Sound + Image condition versus the Image Only condition $CI_{.95} = 1.008$ (lower) 1.686 (upper), $p =< 0.001$; the Sound + Image condition versus the Sound Only condition $CI_{.95} = 0.700$ (lower) 1.548 (upper), $p =< 0.001$. There was no significant difference between the two unimodal (Sound Only, and Image Only) conditions (see descriptives in Appendix Table B.5).

For arousal, Mauchly's test indicated that the assumption of sphericity had been violated $x^2(2) = 11.405, p = 0.003$, therefore degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity $\epsilon = 0.726$. The results indicate that the presentation condition had a significant and very large effect on the arousal rating $F(1.45, 36.278) = 61.529, p =< 0.001, r = 0.88$.

Post-hoc tests revealed a significant difference for all pairwise comparisons (see descriptives in Appendix Table B.5). The Sound + Image condition versus the Image Only condition $CI_{.95} = 1.570$ (lower) 2.430 (upper), $p =< 0.001$; the Sound + Image condition versus the Sound Only condition $CI_{.95} = 0.700$ (lower) 1.386 (upper), $p =< 0.001$; and the Sound Only condition versus the Image Only condition $CI_{.95} = 0.374$ (lower) 1.540 (upper), $p = 0.001$.

The results indicate that the presentation condition had a significant and very large effect on both the valence and arousal ratings. We see that the multimodal Sound + Image presentation condition had significantly greater ratings than both of the uni-modal conditions, however we also see that the Sound Only condition also had a significantly greater arousal rating than the Image Only condition. Therefore we conclude that the results support hypothesis 1B in regard to the arousal ratings, but does not support the hypothesis in regard to the valence ratings.

4.4.1.3 Presentation × Engagement

The engagement ratings were averaged across all of the participants responses and scene samples, for each presentation condition, and are plotted in Figure 4.8.

Looking at the plotted results, the Sound + Image condition had the greatest engagement rating, followed by the Image Only condition, and the Sound Only condition.

To test hypothesis 1C (that the addition of the complex sound cues to a looming image will prompt people to have greater engagement ratings, than the scenes with no sound), a one-way repeated measures ANOVA was conducted with the descriptives listed in Appendix Table B.6.

Mauchly's test indicated that the assumption of sphericity had been violated, $x^2(2) =$

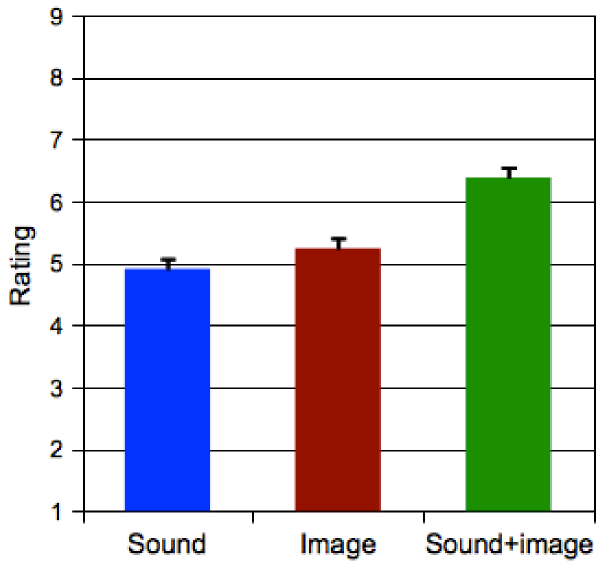

Figure 4.8: Presentation × Engagement Bar Chart
Presentation × Engagement Rating. Results are plotted for each presentation condition (Sound Only, Image Only, Sound + Image) that were averaged across all of the participants ratings. Error bars indicate the standard error for each condition. The average engagement rating for each condition are Sound Only condition ($M = 4.91$, $SE = 0.162$), Image Only condition ($M = 5.24$, $SE = 0.169$), Sound + Image condition ($M = 6.39$, $SE = 0.164$).

$8.326, p = 0.016$, therefore degrees of freedom were corrected using Huynh-Feldt estimates of sphericity $\epsilon = 0.815$. The results indicate that the presentation condition had a significant and very large effect on the engagement rating $F(1.629, 40.732) = 40.013, p =< 0.001, r = 0.84$.

Post-hoc tests with pairwise comparisons revealed a significant difference in the level of engagement for the multimodal (Sound + Image) presentation condition, compared to both of the uni-modal (Sound Only, and Image Only) conditions, the Sound + Image condition versus the Image Only condition $\mathrm{CI}_{.95} = 0.823$ (lower) 1.61 (upper), $p =< 0.001$; the Sound + Image condition versus the Sound Only condition $\mathrm{CI}_{.95} = 1.058$ (lower) 1.794 (upper), $p =< 0.001$ (see pairwise comparisons listed in Appendix Table B.7). There was no significant difference between the two unimodal (Sound Only, and Image Only) conditions.

The results indicate that the presentation condition had a significant and very large effect on the engagement ratings, however as the significant difference only occurred between the multimodal versus uni-modal conditions, and not between the Sound Only and Image Only conditions, the results do not support hypothesis 3 (that the addition of sound prompted greater engagement ratings), but more likely that multimodal presentation prompted greater engagement ratings.

4.4.2 Correlations between the Dependent Variables

4.4.2.1 Engagement × Time-to-Impact

To test hypothesis 2 (that trials which prompt a greater underestimation in the contact time would also be rated with greater engagement ratings, than the scenes with less underestimation of the contact time), a 2-tailed correlation analysis was conducted to see if there was a correlation between the amount of underestimation of the contact time and the engagement rating, with the alpha level for significance was set at 0.01.

The time-to-impact × engagement ratings were averaged across all conditions, per scene sample (26 scenes) and are plotted in Figure 4.9.

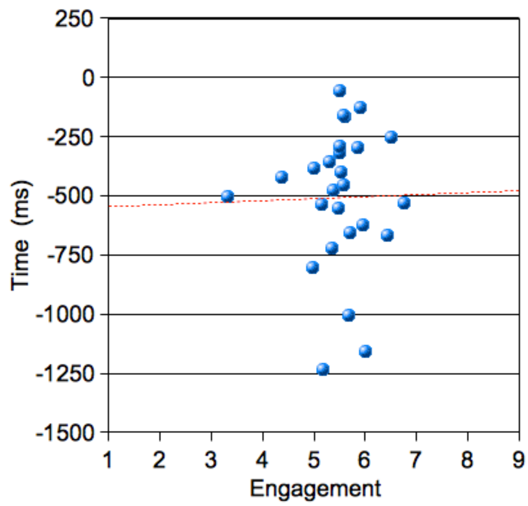

Figure 4.9: Time-to-Impact × Engagement Scatter Plot
The average Time-to-Impact × Engagement rating is plotted for each scene. A linear regression analysis draws the line of best fit equation: $y = 8.49x - 554.61, r^2 =< 0.01$. With a poor line of best fit, a negligible r^2, and a broad spread of data, we conclude that the results do not support hypothesis 2.

The relationship was investigated using Pearson's (product-moment) correlation coefficient. Preliminary analysis was performed to ensure no violations of the assumptions of normality, linearity and homoscedasticity.

With a low coefficient value, the results suggest there was no correlation between the amount of underestimation in the contact time and the engagement rating ($r = 0.018, n = 26, p = 0.929$), therefore hypothesis 2 is not supported and the underestimation in the impact time was not correlated with (either a higher or lower) engagement rating.

4.4.2.2 Valence / Arousal × Time-to-Impact

The time-to-impact × valence and arousal ratings were averaged across all conditions, per scene sample (26 scenes) and are plotted on Figures 4.10 & 4.11.

To test hypothesis 2 (that the trials which prompt a greater underestimation in the contact time would also be rated with greater valence and arousal ratings, than the scenes with less underestimation of the impact time) a 2-tailed correlation analysis was conducted using the same analysis method that was used in subsection 4.4.2. The results indicate there were no correlations between the time-to-impact and the valence rating ($r = -0.003, n = 26, p = 0.989$), and the time-to-impact and the arousal rating ($r = 0.119, n = 26, p = 0.563$), therefore hypothesis 2 is once again, not supported.

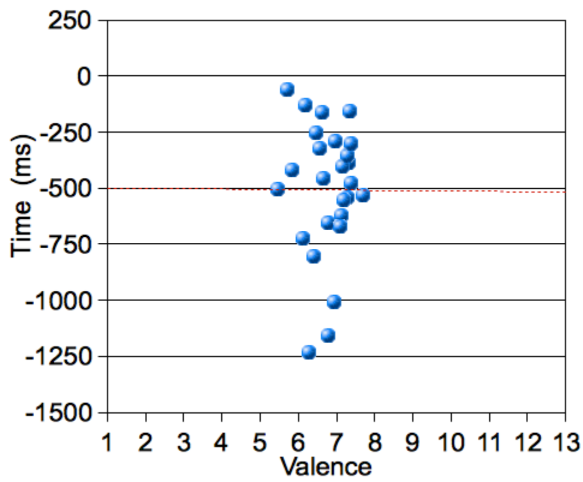

Figure 4.10: Time-to-Impact × Valence Scatter Plot
The average Time-to-Impact × Valence rating is plotted for each scene. The linear regression line equation $y = -1.49x - 497.69, r^2 =< 0.01$.

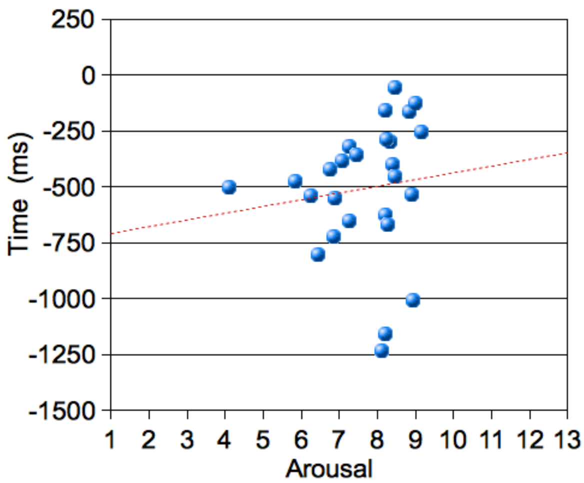

Figure 4.11: Time-to-Impact × Arousal Scatter Plot
The average Time-to-Impact × Arousal rating is plotted for each scene. The linear regression line equation $y = 30.07x - 739.45, r^2 = 0.014$.

4.4.2.3 Valence / Arousal × Engagement

The valence / arousal × engagement ratings were averaged across all conditions, per scene sample (26 scenes) and are plotted on Figures 4.12 & 4.13. The spread of the data shows an upward trend with samples that have greater valence and arousal ratings also having greater engagement ratings.

To test hypothesis 3 (that trials with a greater emotion (valence / arousal) ratings would also have greater engagement ratings), a 2-tailed correlation analysis was conducted using the same analysis method that was used in subsection 4.4.2. The results indicate there were large, positive correlations between valence and engagement ($r = 0.525, n = 26, p = 0.006$), and arousal and engagement ($r = 0.799, n = 26, p =< 0.001$), with greater valence and arousal ratings significantly correlated with greater engagement ratings. With such strong results (significant large positive correlations), we conclude that the results support hypothesis 3, that looming scenes which prompted greater valence and arousal ratings also prompted greater engagement ratings.

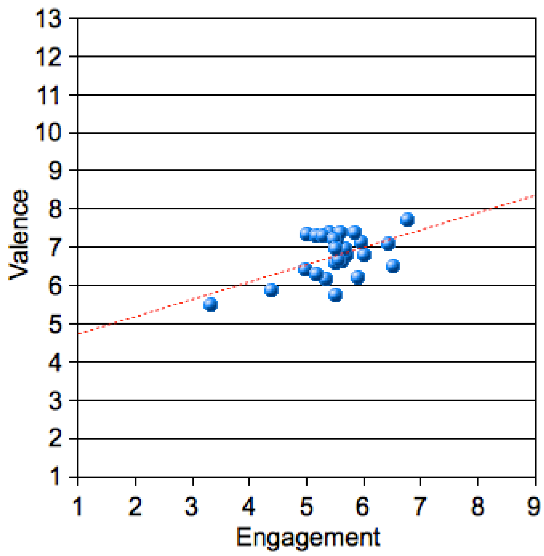

Figure 4.12: Valence × Engagement Scatter Plot
The average Valence × Engagement rating is plotted for each scene. The linear regression equation of line is $y = 0.45x + 4.28, r^2 = 0.28$.

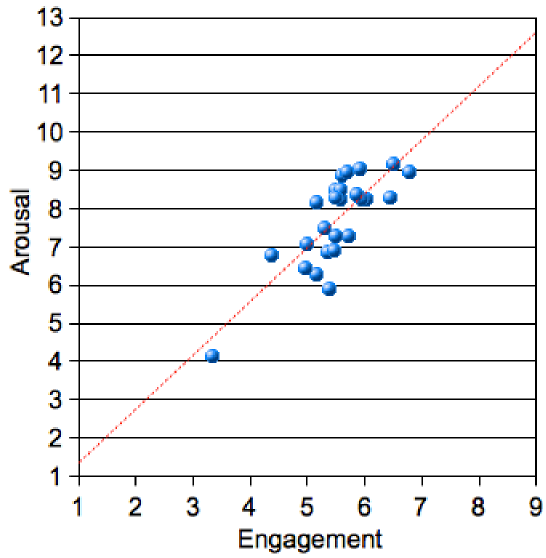

Figure 4.13: Arousal × Engagement Scatter Plot
The average Arousal × Engagement rating is plotted for each scene. The linear regression equation of line is $y = 1.41x - 0.06, r^2 = 0.64$.

4.5 Discussion

This study sought to obtain measurements of human perception and response to film looming scenes that are designed to present multiple audio cues over complex sound sources.

We recognise that this study has some limitations, due to the use of original audio tracks restricting the capacity to control and vary individual sound parameters. However, it did allow us to gain an insight into people's responses and reactions to ecologically valid real world and hyper-real looming stimuli, which has been absent from the psychoacoustic research corpus, but ubiquitous in everyday life.

Using a self reporting questionnaire to gather information about the participants visual and auditory abilities may also have had an impact on the results, whereby the participants auditory and visual acuity may not be as accurate as reported.

The results from this study have shown that the presentation of sound stimuli (which contained multiple auditory looming cues applied to complex sound sources) prompted observer's to underestimate the contact time of an approaching object, and by a significantly greater amount of time, than looming scenes with no sound stimuli. When the sound stimuli was added to visual looming scenes (the Sound + Image condition) the auditory stimuli continued to bias the observers perception, causing them to have a significantly greater underestimation of the contact time, than the scenes with no audio cues. Since both of the conditions that presented sound stimuli have a significantly greater underestimation of the time-to-impact, than the Image Only condition, we conclude that the results support hypothesis 1A (that the presentation of the sound stimuli that has multiple auditory looming cues applied to a complex sound source prompts people to underestimate the contact time of the approaching object, thereby eliciting a faster response time than the scenes with no sound).

Our study also sought to provide an insight concerning the emotional responses to looming stimuli, which has only recently been investigated by Bach et al. [2009]; Tajadura-Jiménez et al. [2010] who found that approaching stimuli elicited greater arousal ratings, than receding stimuli. Our study was novel, with our participant's comparing the modality of sensory information for looming stimuli (auditory, visual and auditory-visual). The results showed that sound stimuli had a significant effect on the arousal ratings, with significantly greater ratings for both conditions presenting sound stimuli, than the condition that did not (i.e. the Image Only condition), again supporting hypothesis 1B (that the presentation of the sound stimuli that has multiple auditory looming cues applied to a complex sound source will prompt people to have greater emotion ratings, than the scenes with no sound). However, hypothesis 1B was not supported in regard to the valence ratings. Whilst the Sound + Image condition had significantly greater valence ratings than the Sound Only and Image Only conditions, there was no difference between the Sound Only and Image Only conditions. Therefore we suggest, that the presentation of multimodal versus uni-modal stimuli had a greater affect on valence ratings, than the actual modality of the stimuli presented. This preference for multimodal stimuli over uni-modal stimuli was also evident in the Engagement ratings. Whilst the results showed that presentation had a significant and large effect on the engagement ratings, this difference only occurred between the multimodal versus uni-modal conditions, and not between the Sound Only and Image Only conditions. Therefore, the results do not support hypothesis 3 (that the addition of sound prompted greater engagement ratings), but more likely that multimodal presentation prompted greater engagement ratings than uni-modal presentation.

The measurement of participant's emotional responses, and the rating the scene's engagement quality have been valuable tools, leading to a better understanding of human responses to real world and hyper-real stimuli, the emotional impact of the stimuli, and the perceptions and actions generated as a result, therefore we recommend the use of the measurements in future looming studies. This will not only inform our understanding of human perception, but also provide detailed parameters for perception and response that is applicable to industry use in the design of sound effects in many virtual environments.

We also investigated if there were correlations between the emotion (valence and arousal) ratings, and the engagement ratings. The analyses indicated that there were significant large positive correlations between the valence and engagement ratings, and the arousal and engagement ratings, therefore we conclude that the results support hypothesis 3, that looming scenes which prompted greater valence and arousal ratings also prompted greater engagement ratings.

We investigated if there were correlations between the amount of (under-) estimation in the perceived impact time, and the emotion and engagement ratings given to the approaching object. We hypothesised that trials which prompted a greater underestimation in the impact time would also be rated with greater engagement, valence, and arousal levels, than the scenes with less underestimation of the contact time, however the results showed this was not the case. With low Pearsons correlation coefficient values, the results indicated there were no correlations between the amount of underestimation in the impact time and the engagement, valence, and arousal ratings, therefore hypothesis 2 is not supported.

Whilst the Tajadura et. al. [2010] study found correlations between faster response times to targets and the valence and arousal ratings, these results were obtained when comparing the approaching versus receding objects, and objects with contrasting emotional association (negative versus positive associations). Since our study focused on approaching objects only, and objects with the similar negative emotional association, the correlations in Tajadura's study perhaps do not extend to finer gradations in differences between the approaching audio cues, and objects of a similar emotive association.

Although the individual sound parameters that act as the audio cues for an approaching object could not be controlled and varied in this study, this investigation of the complex sounds in their original form (as created by the sound designers) has shown that the addition of sound with multiple audio cues, prompted people to have a greater underestimation of the contact time, than the looming scenes without the audio cues. This result indicates that further investigation is warranted, with future research exploring the complex stimulis' individual sound parameters as independent variables, and the perception generated as a result.

Chapter 5

The Effect of Audio Cues and Sound Source Stimuli on the Perception of Approaching Objects

We began our experimental investigations by undertaking a feature analysis study to understand which audio cues were being used in film looming scenes to emphasise approaching objects, which was then followed by a perceptual study of the film samples, in order to understand people's responses to designed complex audio cues.

In this chapter, we build upon the previous two studies by conducting a closer investigation of the audio cues for movement in depth, and how the cues affect people's responses to the approaching object.

Whilst the film looming scenes used complex sounds with multiple audio cues that were designed to maximise the viewer experience, as we discussed in the background chapter (Chapter 2.2) the majority of the previous psychoacoustic studies used artificial sound sources (such as sine, triangle, and square waves) that are rarely encountered in the natural world.

Artificial sound sources are often used in psychoacoustic experiments to increase the study's internal validity, and limit any bias that real world sounds may introduce. However, because such sounds are atypical of those encountered in the natural world, this leads to the criticism that the external validity of such experiments may be compromised, and as such, the studies results may be limited in their capacity to transfer into real world applications, or improve our understanding of how people perceive and react in the real world. Therefore, we seek to fill this gap in the auditory looming research corpus, between the psychoacoustic looming experiment which are predominantly well documented studies using artificial sounds, but rarely investigated using complex real world sounds), and the film industry's design of complex sounds which are exploit-ing audio features as cues for maximum effect, but have little documented scientific evidence to validate there use.

Another parameter which is often omitted in the body of psychoacoustic looming research is the audio cue created from surface reflections. Whilst surface reflections, in the form of reverberation and the direct-to-reverberant energy ratio has been extensively acknowledged as an audio cue for depth perception and determining the distance of a stationary object to an observer (for example, Zahorik [2002a,b, 2001]; Mershon and King [1975]; Bronkhorst and Houtgast [1999]; Bronkhorst [2002]; Shinn-Cunningham [2000]), the research on the audio cue for stationary objects has not been extended to dynamic objects that move in depth. We therefore investigate if this parameter of sound - the direct-to-reflections sound energy ratio is an audio cue for movement in depth.

Lastly, a large number of the psychoacoustic studies have investigated auditory looming using single cues, predominantly the amplitude increase based on claims of its dominance as the most salient audio cue. With the introduction of the direct-to-reflections energy ratio as an audio cue, we consider if this new cue is as effective as other audio cues, if there is a hierarchy amongst the individual audio cues, and further, if people's responses to an approaching object differs when presented multiple audio cues concurrently, as opposed to single audio cues.

5.1 Aims

This study therefore has four aims, firstly, to determine if the direct-to-reflections energy ratio acts as an audio cue for movement in depth, and if so, is it as affective as other well known and studied cues; Secondly, to establish the extent to which certain audio cues (amplitude increase, inter-aural differences, and the direct-to-reflections energy ratio) affect looming perception, and if there is a hierarchy amongst the cues with some cues prompting a greater response, than other cues; Thirdly, to determine if a listeners response to a looming object differs with the inclusion of sounds that use multiple audio cues, as opposed to single looming cues. And lastly, to determine if a listeners response to a looming object differs when the sound source is a real world sound, as opposed to an artificial sound.

5.2 Hypotheses

In regard to the audio cues, it is hypothesised that:

1. the direct-to-reflections energy ratio parameter, will act as an audio cue for an approaching object (as compared to the control condition with no audio cues).
2. listeners responses to the individual audio cues for movement (amplitude increase,

inter-aural differences, and the direct-to-reflections energy ratio) will differ, revealing a hierarchy amongst the individual audio cues.

3. listeners responses to the trials with multiple (2 and 3) audio cues for movement will differ to the trials with single audio cues.
4. listeners responses to the multiple audio cues combinations will differ, with some combinations affecting perception to a greater extent, than other combinations, therefore showing a hierarchy amongst the cue combination.

In regard to the sound source presented, it is hypothesised that:

5. listeners responses to approaching objects that present real world sound sources (i.e. a car tyre traction on a road surface) will differ to approaching objects that present artificial sound stimuli (i.e. a square wave or a noise band).

5.3 Method

5.3.1 Design

The study used a within-subjects design. There were two independent variables - sound source and audio cue, each comprising of three levels:

1. Sound Source:
    - Car traction (real world condition),
    - Square wave (artificial condition),
    - Noise band (artificial condition).
2. Audio Cue:
    - Amplitude Increase,
    - Inter-aural Differences,
    - Direct-to-Reflections Sound Energy Ratio.

There were four dependent variables:

- Time-to-Contact,
- Valence,
- Arousal,
- Engagement.

5.3.2 Participants

A sample of 15 participants naive to the aims and purpose of the study were recruited. They were Ph.D students and Postdoctoral researchers from Queen Mary, University of London aged between 22 and 34 years (M = 27.33 years, SD = 3.24), with more male participants than female participants (9 males, 6 females). The participants visual and auditory abilities were self reported in a questionnaire, and further physiological tests were not made. All participants reported normal hearing, with 4 participants correcting their vision with glasses or contact lenses. These participants wore their glasses during the experiment.

5.3.3 Stimuli

This study presented auditory stimuli only and no visual looming stimuli. However visual information was presented on a computer screen with a graphical user interface (GUI) displaying information about the experiment (trial number) and presenting the onscreen questionnaire asking participant's to rate their emotion and engagement responses. The sound files are included in the Digital Appendix.

The auditory stimuli was constructed using the following sound sources:

- Car traction (Real world condition) with a fundamental frequency of approximately 400 Hz,
- 400 Hz square wave (Artificial condition),
- Noise band (Artificial condition).

Each of the 3 sound sources were presented:

1. Sound only (Ctrl) the control condition with no audio cues applied to the sample,

with the following single audio cues applied as variables:

2. Amplitude Increase (Amp),
3. Inter-aural Differences (IAD),
4. Direct-to-Reflections Sound Energy Ratio (Ref),

and in combination as multiple (2 and 3) audio cue variables:

5. Amplitude Increase + Inter-aural differences (Amp + IAD),
6. Amplitude Increase + Reflections ratio (Amp + Ref),
7. Inter-aural differences + Reflections ratio (IAD + Ref),

8. Amplitude Increase + Inter-aural differences + Reflections ratio (Amp + IAD + Ref).

The amplitude increase audio cue increased non-linearly (according to the inverse square law) from -18dB to -3dB over 1700ms. The trials which did not include the amplitude increase variable, still needed to have a set amplitude level. To eliminate any response biases that a particular level may have (on the other audio cues which we were actually testing) we presented 2 amplitude levels, selecting the minimum -18dB and maximum -3dB levels. All trials that do not include the amplitude increase variable, are presented at both the -18dB and -3dB level.

The inter-aural differences audio cue is a binaural spatial rendering of the stimuli, where the auditory information is presented slightly differently to each channel. The trials which did not include the inter-aural differences were rendered on the monaural setting, presenting the same auditory information to both channels. Whilst the psychoacoustic laws for inter-aural differences stipulate that for a frontal midline trajectory (approaching on a $0\circ$ angle) there are theoretically no inter-aural differences. However, this rule is based on an idealised model where a person's head and the position of the ears are perfectly symmetrical. In reality, this is not the case. Small differences exist between the position of a person's ears, meaning that the sound is not presented equally to each ear, and therefore small amounts of inter-aural differences do occur. Whilst small inter-aural differences are present in the stimuli, they may however be too small to be distinguishable or affect perception. It is also acknowledged that greater angles would introduce greater amounts of inter-aural differences, causing the audio cue to be more salient. However, as this study is exploring the frontal midline trajectory and not other angles, we decided to err on the side of caution and include the audio cue for comparison with the psychoacoustic studies, than to exclude it.

The direct-to-reflections energy ratio audio cue presents 6 first-order reflections off the surfaces. The reflections alter the overall sound by presenting reverberation (with more reverberation when the object is at a farther distance) and a different overall spectral content (with more higher frequency reflections when the proximity of the object is closer to the observer).

The ratio (of the direct-to-reflections sound energy) changes as the proximity of the object becomes closer to the observer. When the object is at a farther distance, reflections at a lower frequency are transmitted, and the overall sound content is comprised of a greater proportion of reverberant energy. As the object nears the observer reflections at a higher frequency become apparent, whilst the proportion of the reverberant energy in the overall energy content, decreases as it becomes masked by the direct sound.

Each sample was 1700ms in duration, followed by 300ms of occlusion (silence). Each trial condition was presented once only (totalling 36 trials, listed in Appendix Table C.1) per observer, and in a randomised order. The presentation of each trial was limited to once only per observer as further presentations may have introduced learning, memory, and fatigue biases. The experiment was presented via a computer, with the GUI interface displayed on the computer monitor, and the auditory stimulus transmitted through a pair of headphones.

5.3.3.1 Generation of the Audio Cues for Movement Using Slab3D

The audio cues were generated using the physical modelling program Slab3D Slabscape (NASA [2013]). We set the size of the space at 20 cubic meters (the maximum dimensions possible) with the virtual observer positioned at one end of the space (XN) (a diagram of the space is illustrated in Figure 5.1). Reflections (6 first-order reflections) were produced from the left wall (YP), right wall (YN), and floor (ZN). We did not produce reflections from the roof (ZP), horizon wall (XP), or at the observer (XN) in order to maximise the available space for the approach, and limit interfering reflections from those surfaces. For the left and right walls (XP and XN) we set the surface material as ‘perfect reflector’, and the floor (ZN) as ‘concrete’. The software limited any further control of the reflections, such as the manipulation of the direct-sound to reflected-sound ratio, spectral content, spectral scattering, and their change over time.

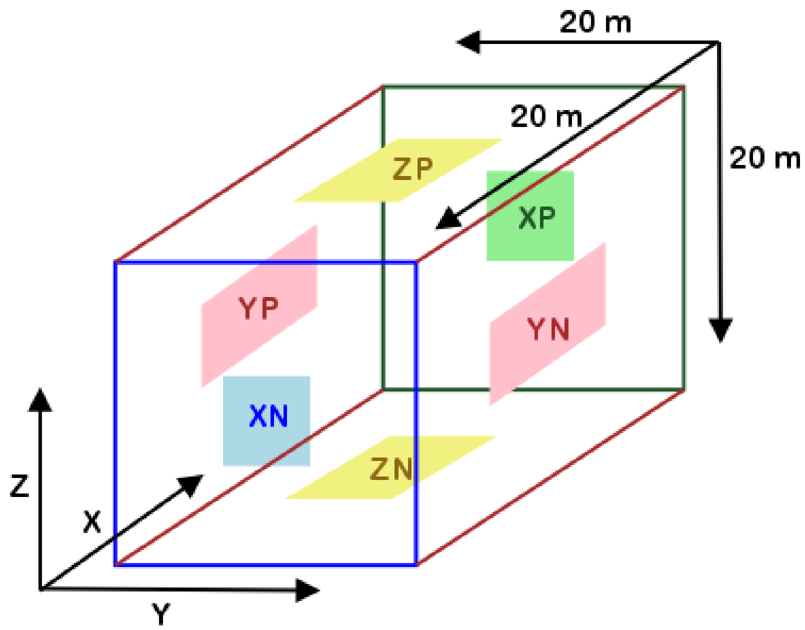

Figure 5.1: Slab3D Room Dimensions
The space dimensions are outlined with depth on the X axis, width on the Y axis, and height on the Z axis. The walls are illustrated with the left wall (YP), right wall (YN), roof (ZP) floor (ZN), horizon wall (XP), listener position is located (XN).

The room size, being a maximum size of 20 cubic meters limited the distance, velocity, and duration which the object could traverse. As such the virtual observer was positioned at one end of the space (XN) in order to maximise the distance, velocity, and duration which the object could traverse. The head of the virtual observer (yaw, pitch, roll) was set at zero so that the observer was facing towards the horizon (XP) and frontal towards the object, which approached on a midline trajectory at nose level.

We set the object's size at a diameter of 10 cm. Although this is a small size, any increase to the size of the object, increases the area it occupies in the limited space that was available, affecting the reflections produced. Therefore to minimise these biases we limited the size of the object to 10cm.

We set the objects velocity at 36 kph (10 meters per second) moving towards the observer on a frontal midline trajectory that intercepted with the observers head. Moving at that speed the object covers the 20 meters of the space, and intercepts with the observer at a time point of 2000ms. Using Audacity, we then edited the sample to 1700ms in duration, removing the final 300ms (a distance of 3 meters) to provide a period of occlusion (silence) for the listeners to predict the time-to-contact.

5.3.4 Apparatus

For this experiment, we used the same apparatus and computer workstation that was used in the Chapter 4 experiment. In a space and time saving measure, please refer to the Apparatus Section 4.3.4 for other methodological details.

5.3.5 Dependent Variable Measurement

For this experiment we had four dependent variable measurements, being time-to-contact, engagement, valence and arousal (emotion).

The emotion and engagement rating scales were the same as those used in Chapter 4. In a space and time saving measure, please refer to the Dependent Variable Measurement in Section 4.3.5, relevant subsections Emotion: Valence and Arousal and Engagement.

Time-To-Contact

Time-to-contact is a measurement technique that has been used extensively in visual and auditory looming studies as a measurement of the stimuli's effect on the perceived contact time (as discussed in Section 2.2).

The contact time was derived from the auditory stimuli which was generated using the Slab3D physical model with the contact time at 2000ms.

Participant's responses to the stimuli by pressing the keyboard space bar when they thought the object reached them was also timed, and for the purpose of this study is called the ‘Response Time’.

Using equation 5.1, the contact time was subtracted from the response time, to give the amount of time that was underestimated or overestimated, and for the purpose of this study is called the ‘time-to-contact.’.

$$RT - CT \quad = \quad TC \tag{5.1}$$

where:
RT = Participant’s Response Time, the time (in ms) when participant’s pressed the space bar when they thought the object reached them.
CT = Contact Time (2000ms),
TC = Time-to-contact, the amount of time (in ms) which was under- / over-estimated.

5.3.6 Procedure

Participant’s sat at the computer workstation and were informed of the experimental procedure. They were given an information sheet summarising both the procedure and the ethics approval, signed a consent form, and completed a background questionnaire asking questions on gender, age, and whether they have had corrections made to their vision or hearing (the documents are included in the Digital Appendix).

Before commencing the experiment, the participant’s completed a practice study using 6 looming scenes that were not additionally presented in the experiment. It was conducted as a supervised learning procedure to provide them with the opportunity to comprehend the experiment, the procedure, the micro time scale of the stimulus, and how to complete the task. Participant’s were then instructed to start the experiment when ready.

The task required the participant’s to listen to the sample of an approaching object. They were informed that the sound would be then occluded, but to imagine that the object was still moving towards them, and to press the keyboard space bar when they thought the object reached them. A pop-up questionnaire was then displayed on the computer screen, asking the participant’s to rate their valence / arousal level and how engaging the scene was.

Each trial lasted for a total duration of 1700 milliseconds however the participant’s were not time restricted as to how long they spent answering the questions. Once they had submitted their answers a 4 second break was then given between each trial. The experiment lasted for approximately 12 minutes and participant’s were not given any information implying there might be ‘correct’, ‘incorrect’ or ‘preferred’ responses.

5.4 Results

The 15 participant's were each presented the 4 audio cue conditions containing amplitude increase as a variable (Amp, Amp + IAD, Amp + Ref, Amp + IAD + Ref) 3 times (1 × Sound source (Car, Square, Noise)); and the 4 audio cue conditions which did not contain amplitude increase as a variable (Sound Only, IAD, Ref, IAD + Ref) 2 times each (1 × -18dB, and 1 × -3dB), × the 3 Sound Sources (Car, Square, Noise). To give an equal number of trials per audio cue, the data for the conditions not containing amplitude increase as a variable (Sound Only, IAD, Ref, IAD + Ref) were each averaged across the amplitude level (-18dB and -3dB).

This gives the 8 audio cue conditions (Sound only, Amp, IAD, Ref, Amp + IAD, Amp + Ref, IAD + Ref, Amp + IAD + Ref) × 3 sound source conditions (Car, Square, Noise), × 15 subjects, totalling 360 trials. The trials and conditions are listed in Appendix Table C.1.

5.4.1 Audio Cues

To test hypotheses 1 to 4 (does the direct-to-reflections energy ratio variable act as an audio cue?; do listeners responses to the individual audio cues differ?; do listeners responses to the trials with multiple audio cues differ to the trials with single audio cues?; and does listeners responses to the multiple audio cues differ?), we began by looking at the audio cues affect on the perceived time-to-contact, then emotion (valence and arousal), and lastly engagement rating.

Each analysis included eight within-subject variables for the audio cue condition (Sound Only (ctrl), Amp, IAD, Ref, Amp + IAD, Amp + Ref, IAD + Ref, Amp + IAD + Ref) and the alpha level for significance was set at 0.05.

5.4.1.1 Audio Cues × Time-to-Contact

The results indicated that some of the data contained outliers. 12 outliers across 9 trial comparisons were removed leaving 36 trials per condition. The time-to-contact was then averaged across all of the participants responses and sound sources for each audio (single and multiple) cue condition, and are plotted in Figure 5.2.

Looking at the plotted results, we see that the Sound Only condition (Ctrl - which contained no audio cues) had the greatest overestimation of the contact time (M = 984.716ms) as compared to all other conditions (see descriptives listed in Appendix Table C.2). This suggests that the application of audio cues (either single or multiple) caused people to alter (and lessen) their estimation of the contact time. The condition which contained all three audio cues (Amp + IAD + Ref) had the greatest overall underestimation of the contact time (M = -272.873ms).

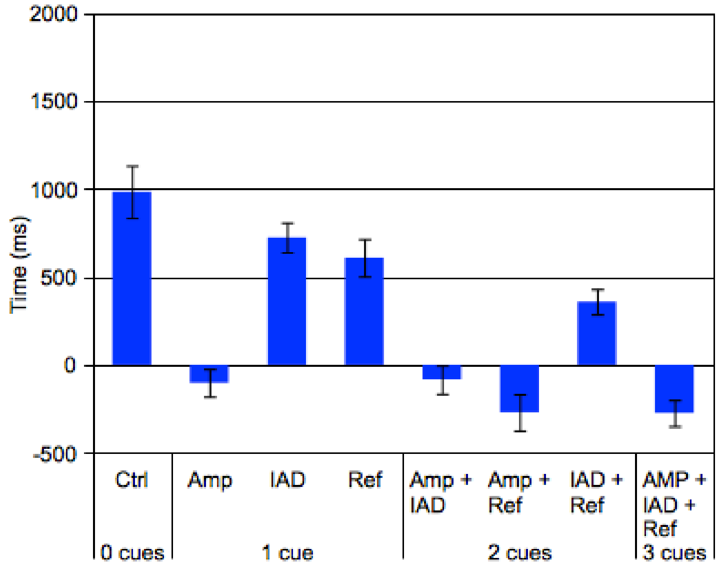

Figure 5.2: Audio Cue × Time-to-Contact Bar Chart
The time-to-contact estimates for each audio cue condition (averaged across all of the participants ratings and sound sources) are plotted. Error bars indicate the standard error for each condition. The contact time occurs at 0ms, with any underestimation plotted in the negative range of the scale, and overestimation plotted in the positive range. **Sound Only:** M = 984.716 ($S.E.$ = 147.867); **Amp:** M = -101.225 ($S.E.$ = 78.114); **IAD:** M = 724.960 ($S.E.$ = 84.245); **Ref :** M = 610.162 ($S.E.$ = 106.433); **Amp + IAD :** M = -83.445 ($S.E.$ = 80.818); **Amp + Ref :** M = -269.315 ($S.E.$ = 102.877); **IAD + Ref :** M = 359.633 ($S.E.$ = 71.731); **Amp + IAD + Ref :** M = -272.873 ($S.E.$ = 74.194).

When comparing the multiple audio cues to the single audio cues, the multiple audio cues had greater underestimation than the single audio cues, and in the case of IAD + Ref cue, had lesser overestimation than its related single cues. The conditions which contained amplitude increase as a cue in both single and multiple audio cues, all resulted in an underestimation of the contact time, prompting people to estimate the contact time to be earlier than it physically would have.

A one-way repeated measures ANOVA was conducted with the means and standard errors listed in Appendix Table C.2. Mauchly's test of sphericity indicated that the assumption of sphericity had been violated, $x^2(27) = 78.348, p =< 0.001$, therefore the degrees of freedom were corrected using the Greenhouse-Geisser estimates of sphericity $\epsilon = 0.631$. The results indicate that the audio cues had a significant, and strong positive effect on the estimated time-to-contact $F(4.415, 195.275) = 33.326, p =< 0.001, r = 0.683$, ($\alpha = 0.05$).

Post-hoc tests with pairwise comparisons were conducted and the descriptives are listed in Appendix Table C.3. Pairwise comparisons of the Sound only (ctrl) condition to all other conditions, revealed there was a significant difference (in the estimated time-to-contact) for all conditions with audio cues (except the IAD condition). This result supports hypothesis 1, that the direct-to-reflections condition (Ref) acts as an audio cue for movement in depth, biasing people's responses to the perceived time-to-contact, and prompting significantly ($p = 0.056$) earlier response times (than the Sound Only condition).

When we compare the single audio cue conditions (to determine if there is a hierarchy amongst the individual audio cues) we see that the amplitude only condition had significantly lesser estimated time than both of the other single audio cues (IAD and Ref), and that the Inter-aural differences condition had the greatest overestimation, supporting hypothesis 2. This pattern of results was again replicated in the multiple cue comparisons (supporting hypothesis 4), with the conditions containing the amplitude increase cue (Amp + Ref, and Amp + IAD) having a significantly greater underestimation.

When we compare the single versus multiple audio cue conditions (to determine if listeners responses to multiple cues differs to single cues) we see that in all condition comparisons (with the exception of Amp × Amp + IAD, and Amp × IAD + Ref), the multiple audio cue conditions prompted an earlier response times, than the single audio cue conditions, supporting hypothesis 3. In regard to the exception (the Amp × Amp + IAD, and the Amp × IAD + Ref pairwise comparisons), the single amplitude condition prompted an earlier estimation of the contact time (albeit a small -17.78ms earlier than the Amp + IAD condition, and a significantly greater -460.858ms earlier than the IAD + Ref condition). One explanation for this result, refer's to hypotheses 2 and 4 - the hierarchy of individual cues, and the strong capacity of the amplitude increase as an audio cue for movement in depth. The addition of the inter-aural differences (with the AMP+ IAD condition) had little impact (for movement in depth, as would be expected for a frontal mid-line plane); and the omission of an amplitude increase cue (in the IAD + Ref condition). We suggest this result supports hypothesis 3 (that listeners responses to multiple audio cues will differ to the trials with single audio cues), via the case of Hypotheses 2 and 4 (a hierarchy of cues, with some cues having greater affect than others).

5.4.1.2 Audio Cues × Emotion (Valence / Arousal)

The results showed that the valence data contained 1 outlier and the arousal data contained 10 outliers (across 8 trial comparisons). These were removed leaving 44 valence and 37 arousal trials and per condition. The ratings were then averaged across all of the sound sources and participant's responses for each audio cue condition, and are plotted in Figure 5.3.

Looking at the spread of the results, we see that the Sound Only (ctrl) condition (with no audio cues) had the lowest arousal rating and second lowest valence rating (with the IAD having the lowest valence and marginally greater arousal ratings); and the Amp + IAD + Ref (3 audio cues) condition had the greatest arousal rating and second greatest valence rating (whilst the Amp + Ref condition had the greatest valence rating). We can see that all of the conditions which presented one or more audio cues for movement

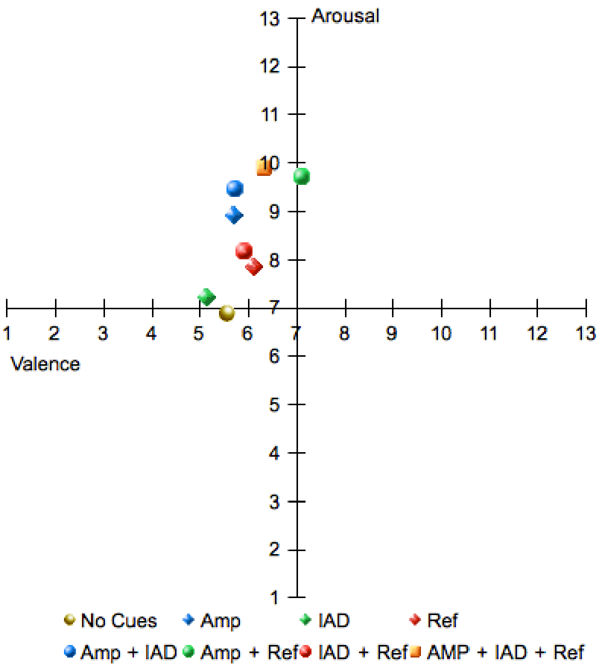

Figure 5.3: Audio Cue × Valence / Arousal Scatter Plot
The valence / arousal ratings for each audio cue condition are plotted. **Sound Only:** Valence: M = 5.578 ($S.E.$ = .293), Arousal: M = 6.878 ($S.E.$ = .324); **Amp:** Valence: M = 5.711 ($S.E.$ = .284), Arousal: M = 8.889 ($S.E.$ = .404); **IAD:** Valence: M = 5.144 ($S.E.$ = .281), Arousal: M = 7.189 ($S.E.$ = .351); **Ref :** Valence: M = 6.133 ($S.E.$ = .235), Arousal: M = 7.833 ($S.E.$ = .297); **Amp + IAD :** Valence: M = 5.733 ($S.E.$ = .364), Arousal: M = 9.444 ($S.E.$ = .350); **Amp + Ref :** Valence: M = 7.111 ($S.E.$ = .286), Arousal: M = 9.711 ($S.E.$ = .287); **IAD + Ref :** Valence: M = 5.922 ($S.E.$ = .236), Arousal: M = 8.144 ($S.E.$ = .277); **Amp + IAD + Ref :** Valence: M = 6.333 ($S.E.$ = .341), Arousal: M = 9.889 ($S.E.$ = .295).

in depth, had greater (average) arousal and valence ratings (with the exception of the IAD condition, which had $M = 0.434$ lower valence rating), than the Sound Only (ctrl) condition with no audio cues. There is also a general tendency for the conditions with multiple audio cues to have greater valence / arousal ratings than the single audio cue conditions. The four conditions that contained amplitude increase as an audio cue, all had greater arousal ratings, than the condition which did not contain the amplitude increase variable.

One-way repeated measures ANOVA's were conducted to compare the valence and arousal ratings by audio cue condition. For valence, Mauchly's test indicated that the assumption of sphericity had been violated $x^2(27) = 134.570, p =< 0.001$, therefore the degrees of freedom were corrected using the Greenhouse-Geisser estimates of sphericity $\epsilon = 0.583$. The results indicate that the application of audio cues had a significant, and moderate positive effect on the valence rating $F(4.084, 179.717) = 9.696, p =< 0.001, r = 0.367$, $(\alpha = 0.05)$.

Post-hoc tests on the valence rating were conducted and the descriptives are listed in Appendix Table C.5. Whilst all of the conditions (except IAD) prompted greater valence ratings that the Sound Only condition, only 1 of the conditions (Amp + Ref) reached the significance level, therefore hypothesis 1 cannot be supported in regard to the valence rating. Further pairwise comparisons between the single cue conditions, and again between the multiple cue conditions did not reveal any particular pattern of results for a hierarchy in the audio cues, therefore hypotheses 2 and 4 also cannot be supported in regard to the valence ratings. However when we compare the multiple cues to single cues, we can see that the addition of the amplitude increase variable (i.e. Ref vs Amp + Ref), and the direct-to-reflections variable (i.e. IAD vs IAD + Ref; Amp vs Amp + Ref; IAD vs Amp + IAD + Ref) reveals the multiple cue condition prompts a significantly greater arousal rating, supporting hypothesis 3.

For arousal, Mauchly's test indicated that the assumption of sphericity had also been violated $x^2(27) = 88.056, p =< 0.001$, therefore degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity $\epsilon = 0.678$. The results indicate that the arousal rating was significantly affected with a strong positive effect, by the application of audio cues $F(4.744, 208.755) = 19.665, p =< 0.001, r = 0.554$, $(\alpha = 0.05)$.

Post-hoc tests on the arousal rating were conducted and the descriptives are listed in Appendix Table C.5. When we compare the single audio cue conditions to the Sound Only (no audio cues) condition, we see that all of the conditions (except IAD) prompted significantly greater arousal ratings, including the direct-to-reflections ratio (Ref), supporting hypothesis 1, that the condition acts as an audio cue for movement in depth.

Comparing the single audio cue conditions (to determine if there is a hierarchy amongst the individual audio cues) we see that the amplitude increase condition prompted greater arousal ratings (that reached the significance level for the IAD pairwise comparison) supporting hypothesis 2, that some cues prompt a greater arousal rating than others. The capacity for the amplitude increase cue to increase arousal ratings, was replicated in the multiple cue conditions, and where the amplitude increase variable was added (i.e. Amp + IAD vs IAD + Ref; Amp + Ref vs IAD + Ref; IAD + Ref vs Amp + IAD + Ref) the arousal rating was significantly greater, supporting hypothesis 4. When we compare the multiple cues versus the single cues, generally the conditions with multiple cues prompted greater arousal ratings, than the single cue conditions, and again when the multiple cue condition contained amplitude increase as a variable (i.e. IAD vs Amp + IAD; Ref vs Amp + Ref; IAD vs Amp + IAD + Ref; Ref vs Amp + IAD + Ref) the multiple cue condition prompts a significantly greater arousal rating, supporting hypothesis 3.

5.4.1.3 Audio Cues × Engagement

Early exploration of the results showed that some of the data contained outliers. 7 outliers (across 7 trial comparisons) were removed, leaving 38 trials per condition. The engagement ratings were then averaged across all of the sound sources (car, noise, and square) and participant's responses, for each audio cue condition, and are plotted in Figure 5.4.

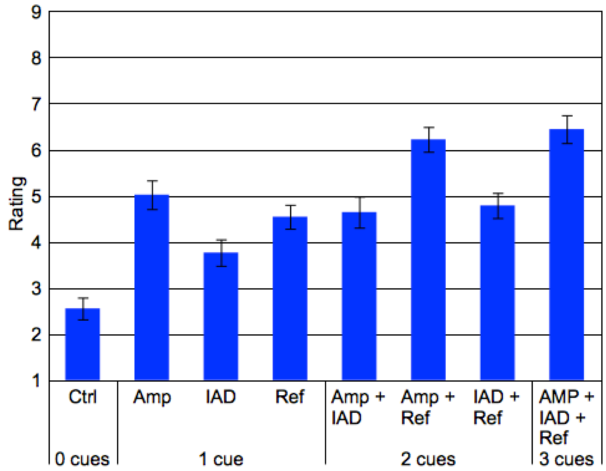

Figure 5.4: Audio Cue × Engagement Bar Chart
The engagement rating for each audio cue condition (averaged across all of the participants ratings and sound sources) are plotted. Error bars indicate the standard error for each condition. **Sound Only:** $M = 2.556$ ($S.E. = .236$); **Amp:** $M = 5.022$ ($S.E. = .311$); **IAD:** $M = 3.767$ ($S.E. = .287$); **Ref :** $M = 4.544$ ($S.E. = .259$); **Amp + IAD :** $M = 4.644$ ($S.E. = .337$); **Amp + Ref :** $M = 6.222$ ($S.E. = .269$); **IAD + Ref :** $M = 4.789$ ($S.E. = .273$); **Amp + IAD + Ref :** $M = 6.444$ ($S.E. = .301$).

Looking at the plotted results, we see that the Sound Only (ctrl) condition had the lowest engagement rating, and the condition with all 3 audio cues had the greatest rating. A one-way repeated measures ANOVA was conducted and the means, standard errors, and confidence intervals are listed in Appendix Table C.6. Mauchly's test indicated that the assumption of sphericity had been violated, $x^2(27) = 42.206, p = 0.032$, therefore degrees of freedom were corrected using Huynh-Feldt estimates of sphericity $\epsilon = 0.933$. The results indicate that the audio cue had a significant, and very strong positive effect on the engagement rating $F(6.533, 287.443) = 31.857, p =< 0.001, r = 0.717$, ($\alpha = 0.05$).

Post-hoc tests with pairwise comparisons revealed a significant difference in the engagement ratings for all audio cue combinations (both single and multiple) as compared to the Sound Only (ctrl) condition which contained no audio cues (see descriptives, and confidence intervals listed in Appendix Table C.7).

When we compare the single audio cue conditions (to determine if there is a hierarchy amongst the individual audio cues) we see that the amplitude only condition had greater engagement ratings than both of the other single audio cues (Ref, and IAD (significant at $p = 0.003$)) supporting hypothesis 2. Looking at the multiple cue conditions, we see that the condition containing all 3 audio cues had the greatest engagement rating, which was significantly greater ($p =< 0.001$) than the Amp + IAD and IAD + Ref conditions, supporting hypothesis 4; and was significantly greater ($p = \leq 0.001$) than all of the single audio cue conditions, supporting hypothesis 3.

5.4.1.4 Amplitude Level Presentation

Since the amplitude increase condition appears to be the strongest audio cue for movement in depth (on a frontal midline trajectory), we conduct a closer inspection of the cue comparing participant's responses to the condition increasing from -18 to -3dB, to two constant level presentations - the minimum -18dB level, and the maximum -3dB level.

Each of the three amplitude level presentation conditions (an increase from -3 to 18dB, the constant level of -3dB, and the constant level of -18dB) were presented 12 times, $\times$ fifteen participant's, totalling 180 trials per condition, 540 trials in total. The ratings were averaged across all of the sound sources (car, noise, and square) and participant's responses, for each amplitude level presentation condition, and are plotted in Figure 5.5.

We employed the same analysis methods as those detailed in Section 4.4. In a space saving measure please refer to this section for a broader explanation of the analysis methodology. The one-way repeated measures ANOVA's compared the time-to-contact, valence, arousal, and engagement ratings of the increasing amplitude condition with the constant level conditions. It contained three within-subject conditions for amplitude (increasing -18 to -3dB condition; and the two constant levels conditions being -18dB and -3dB), and the means, standard errors, and confidence intervals are listed in Appendix Table C.8.

Time-to-contact

Early exploration of the results showed that some of the data contained outliers, therefore 16 outliers (across 12 trials) were removed, leaving 164 trials per condition.

Mauchly's test indicated that the assumption of sphericity had not been violated, $x^2(2) = 1.382, p = 0.501$, therefore the degrees of freedom did not need correction. The results indicate that the audio cue condition had a significant, and very strong positive effect on the engagement rating $F(2, 326) = 136.401, p =< 0.001, r = 0.906$, ($\alpha = 0.05$). Post-hoc tests with pairwise comparisons showed a significant difference in the time-to-contact between the increasing amplitude condition (-18 to -3dB) and both of the constant level conditions (-18dB and -3dB).

Valence / Arousal

The valence data contained 4 outliers (across 4 trials), and the arousal data contained 6 outliers (across 6 trials) which were removed. The valence / arousal ratings were then averaged across all of the sound sources and participant's responses for each amplitude level condition, and are plotted in Figure 5.6.

Looking at the spread of the results, we see that (maximum) -3dB constant level condition, and the -18 to -3dB increasing level condition have greater arousal ratings than

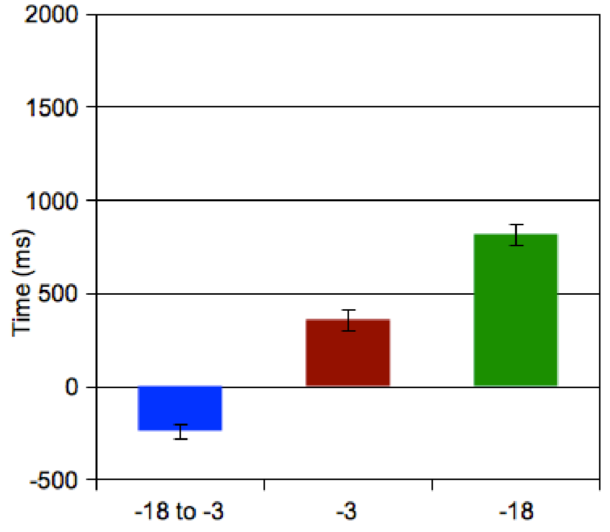

Figure 5.5: Amplitude Level × Time-to-Contact Bar Chart
The Time-to-contact rating for each Amplitude level presentation (averaged across all of the participants ratings and sound sources) are plotted. Error bars indicate the standard error for each condition. **-18 to -3dB:** M = -241.254 ($S.E.$ = 37.105); **-3:** M = 357.004 ($S.E.$ = 58.713); **-18:** M = 814.916 ($S.E.$ = 58.832).

the (minimum) -18dB level condition; whilst the louder -3dB constant level condition also had more negatively rated valence, than the -18dB condition, and the -18 to -3dB increasing level condition.

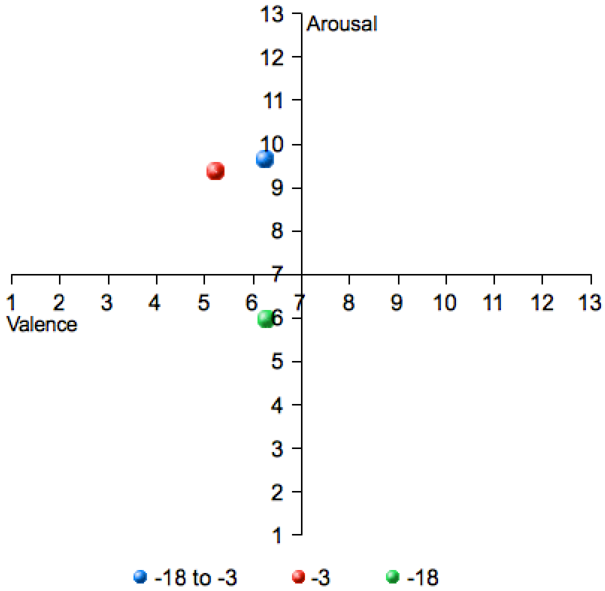

Figure 5.6: Amplitude Level × Valence / Arousal Scatter Plot
The valence / arousal rating for each amplitude level presentation (averaged across all of the participants ratings and sound sources) are plotted. **-18 to -3dB:** Valence: M = 6.250 ($S.E.$ = .163), Arousal: M = 9.649 ($S.E.$ = .156); **-3:** Valence: M = 5.239 ($S.E.$ = .179), Arousal: M = 9.356 ($S.E.$ = .154); **-18:** Valence: M = 6.278 ($S.E.$ = .139), Arousal: M = 5.948 ($S.E.$ = .222).

For valence, Mauchly's test indicated that the assumption of sphericity had been violated, $x^2(2) = 11.834, p = 0.003$, therefore degrees of freedom were corrected using Huynh-Feldt estimates of sphericity $\epsilon = 0.948$. The results indicate that the amplitude level presentation had a significant, and very strong positive effect on the valence rating $F(1.896, 331.843) = 23.926, p =< 0.001, r = 0.621$, ($\alpha = 0.05$).

Post-hoc tests with pairwise comparisons showed a significant difference in the valence rating between the loud constant level condition (-3dB) and the increasing amplitude condition (-18 to -3dB), and also between the loud (-3dB) and the soft (-18dB) constant level condition. The descriptives and confidence intervals listed in Appendix Table C.9.

For arousal, Mauchly's test indicated that the assumption of sphericity had also been violated, $x^2(2) = 10.738, p = 0.005$, therefore degrees of freedom were corrected using Huynh-Feldt estimates of sphericity $\epsilon = 0.953$. The results indicate that the amplitude level presentation had a significant, and very strong positive effect on the arousal rating $F(1.906, 329.735) = 160.268, p =< 0.001, r = 0.895$, ($\alpha = 0.05$). Post-hoc tests with pairwise comparisons revealed a significant difference in the arousal rating between the soft constant level condition (-18dB) and the increasing amplitude condition (-18 to -3dB), and also between the soft constant level condition (-18dB) and the loud constant level condition (-3dB). Descriptives and confidence intervals listed in Appendix Table C.9.

Engagement

Early exploration of the results showed there were no outliers, so the engagement ratings were averaged across all of the sound sources (car, noise, and square) and participant's responses, for each amplitude level presentation condition, and are plotted in Figure 5.6.

Looking at the plotted results, we see that -3dB constant level condition had a lower engagement rating than both the -18 to -3dB, and the -18dB conditions, which were similarly rated.

Mauchly's test indicated that the assumption of sphericity had been violated, $x^2(2) = 51.798, p = 0.000$, therefore degrees of freedom were corrected using Huynh-Feldt estimates of sphericity $\epsilon = 0.804$. The results indicate that the amplitude level presentation had a significant, and very strong positive effect on the engagement rating $F(1.609, 287.998) = 110.640, p =< 0.001, r = 0.880$, ($\alpha = 0.05$).

5.4.2 Sound Source

To test hypothesis 5 (if listeners responses to an approaching object differs when presented with real world stimuli, as opposed to artificial stimuli) we investigate the affect

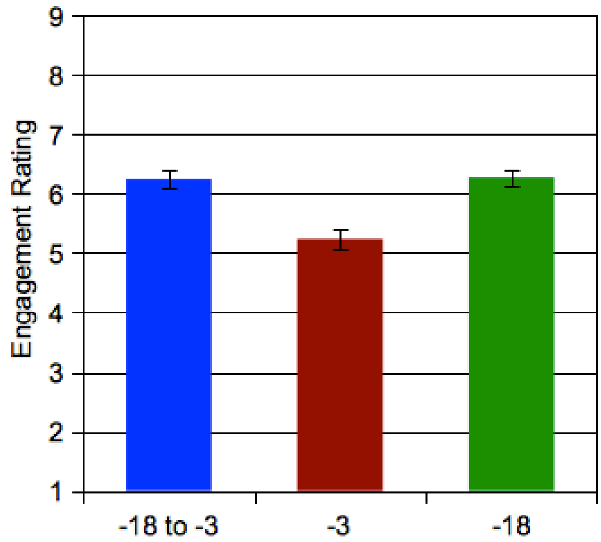

Figure 5.7: Amplitude Level × Engagement Bar Chart
The engagement rating for each amplitude level presentation (averaged across all of the participants ratings and sound sources) are plotted. Error bars indicate the standard error for each condition. **-18 to -3dB:** M = 5.583 ($S.E.$ = .162); **-3:** M = 4.322 ($S.E.$ = .164); **-18:** M = 3.506 ($S.E.$ = .144).

of sound source on human perception. There was 1 real world condition (consisting of a car traction sound) and 2 artificial sound source conditions (being the square wave, and the noise band presentations). Each of the 3 Sound Source conditions (Car traction, Noise Band, Square wave) were presented 8 times (conditions that were presented as both -18 and -3dB were averaged), × fifteen participant's, totalling 120 trials per condition, 360 trials in total. We employed the same analysis methods detailed in Section 4.4, in a space saving measure please refer to this section for a broader explanation of the analysis methodology.

5.4.2.1 Sound Source × Time-to-Contact

Early exploration of the results showed that some of the data contained outliers, therefore 10 outliers (across 7 trial comparisons) were removed, leaving 113 trials per condition. The time-to-contact was then averaged across all of the participants responses (and audio cues) for each sound source condition, and are plotted in Figure 5.8.

Looking at the spread of the data, we see that all three conditions prompted people to overestimate the contact time. The condition which generated the least amount of overestimation was the (real world) car condition (M = 114.728ms, SE = 62.672); followed by both the artificial conditions of the square wave (M = 140.582ms, SE = 60.852); and lastly the noise band condition (M = 234.297ms, SE = 64.224). Full descriptives are listed in Appendix Table C.10.

A one-way repeated measures ANOVA was conducted with the means and standard errors listed in Appendix Table C.10. Mauchly's test indicated that the assumption

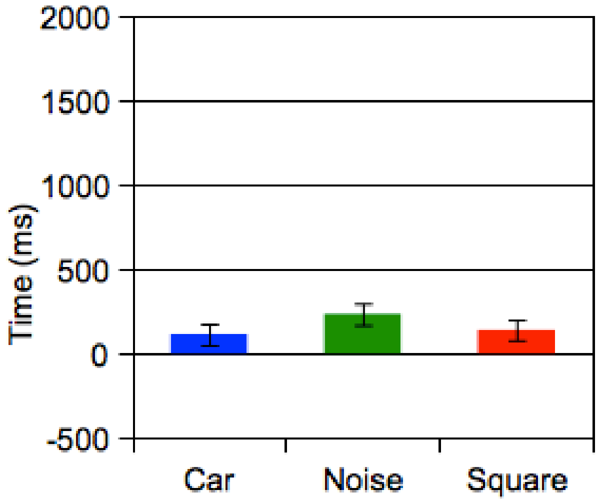

Figure 5.8: Sound Source × Time-to-Contact Bar Chart
The time-to-contact for each sound source condition (averaged across all of the participants ratings and audio cues) are plotted. Error bars indicate the standard error for each condition. The contact time occurs at 0ms, with any underestimation plotted in the negative range of the scale, and overestimation plotted in the positive range. **Car:** M = 114.728 ($S.E.$ = 62.672); **Noise:** M = 234.297 ($S.E.$ = 64.224); **Square:** M = 140.582 ($S.E.$ = 60.852).

of sphericity had not been violated $x^2(2) = 3.984, p = 0.136$, therefore the degree's of freedom did not need correction. The results indicate that the time-to-contact was not affected by the sound source condition $F(2, 224) = 2.051, p = 0.131, r = 0.063$, ($\alpha = 0.05$). Post-hoc tests with pairwise comparisons were conducted and the descriptives are listed in Appendix Table C.11. Please refer to the mean difference, significance levels, and confidence intervals listed in this table. The average difference in the time-to-contact for all pairwise comparisons did not meet the significance level and the greatest (average) difference between the conditions was 119.569ms (car × noise). As a result, hypothesis 5 cannot be supported when it comes to sound source affecting the perceived time-to-contact.

5.4.2.2 Sound Source × Emotion (Valence / Arousal)

The results showed that some of the valence data (but not arousal) contained outliers, therefore 3 valence outliers (across 3 trials) were removed, leaving 117 trials per condition for valence, and 120 trials per condition for arousal. The ratings were then averaged across all of the trials (and audio cues) for each sound source condition, and are plotted in Figure 5.9.

Looking at the spread of the results, we see that the (artificial) square wave had the greatest arousal rating and lowest valence ratings, whilst the (artificial) noise band and the (real world) car traction had similar valence ratings, although the car had a greater arousal rating.

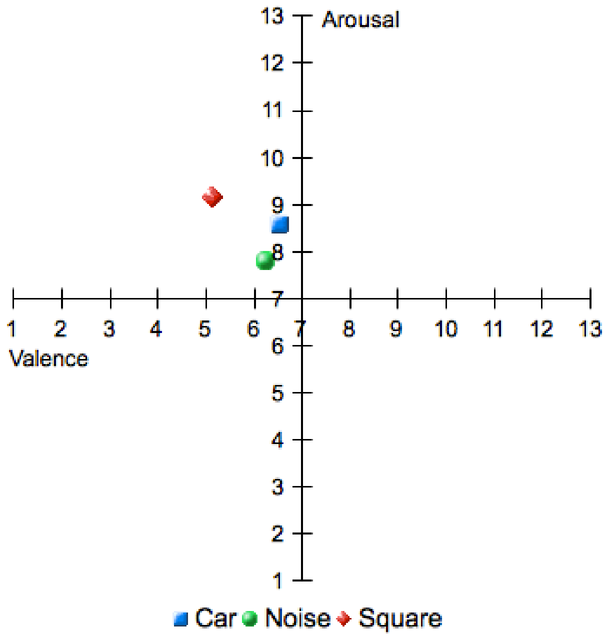

Figure 5.9: Sound Source × Valence / Arousal Scatter Plot
The valence / arousal rating for each sound source condition (averaged across all of the participants ratings and audio cues) are plotted. **Car:** Valence: $M = 6.551$ ($S.E. = .166$), Arousal: $M = 8.575$ ($S.E. = .230$); **Noise:** Valence: $M = 6.248$ ($S.E. = .136$), Arousal: $M = 7.783$ ($S.E. = .222$); **Square:** Valence: $M = 5.162$ ($S.E. = .208$), Arousal: $M = 9.133$ ($S.E. = .192$).

One-way repeated measures ANOVA's were conducted with the means and standard errors listed in Appendix Table C.12. For valence, Mauchly's test indicated that the assumption of sphericity had been violated $x^2(2) = 25.745, p =< 0.001$, therefore degrees of freedom were corrected using Huynn-Feldt estimates of sphericity $\epsilon = 0.844$. The results indicate that the sound source had a significant, and strong positive effect on the valence rating $F(1.687, 195.716) = 23.150, p =< 0.001, r = 0.596$, ($\alpha = 0.05$). are listed in Appendix Table C.13.

Post-hoc tests with pairwise comparisons revealed a significant difference in the valence rating for the (real world) car traction condition versus the (artificial) square wave CI$_{.95}$ = .762 (lower) 2.016 (upper), $p =< 0.001$; and between the two artificial conditions - the noise band versus the square wave CI$_{.95}$ = .613 (lower) 1.558 (upper), $p =< 0.001$. However there was no significant difference between the car traction and noise band (see descriptives listed in Appendix Table C.13).

For arousal, Mauchly's test indicated that the assumption of sphericity had been violated $x^2(2) = 6.366, p = 0.041$, therefore degrees of freedom were corrected using Huynn-Feldt estimates of sphericity $\epsilon = 0.965$. The results indicate that the sound source had a significant, and strong positive effect on the arousal rating $F(1.930, 229.692) = 15.050, p =< 0.001, r = 0.484$, ($\alpha = 0.05$). Post-hoc tests on the arousal rating revealed a significant difference for all pairwise comparisons, with the car versus the square wave conditions borderline significant at $p = 0.055$. We therefore conclude that these results support hypothesis 5, that the sound source affects listeners emotional (valence / arousal) responses to an approaching object.

5.4.2.3 Sound Source $\times$ Engagement

Early exploration of the results showed there were no outliers, so the engagement ratings were averaged across all of the participants responses (and audio cues) for each sound source condition, and are plotted in Figure 5.10.

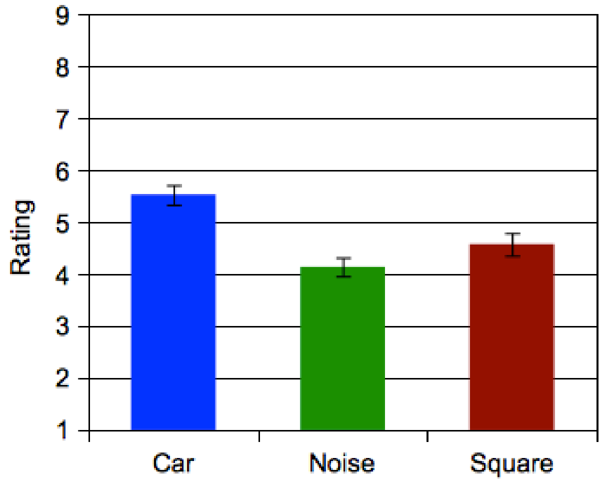

Figure 5.10: Sound Source $\times$ Engagement Bar Chart
The engagement rating for each sound source condition (averaged across all of the participants ratings and audio cues) are plotted. Error bars indicate the standard error for each condition. **Car:** M = 5.533 ($S.E.$ = .190); **Noise:** M = 4.142 ($S.E.$ = .179); **Square:** M = 4.579 ($S.E.$ = .218).

Looking at the plotted results, we see that the (real world) car condition had the greatest average engagement rating, followed by the 2 artificial conditions - the square wave and the noise band.

A one-way repeated measures ANOVA was conducted and the means and standard errors listed in Appendix Table C.14. Mauchly's test indicated that the assumption of sphericity had been violated $x^2(2) = 16.746$, p = < 0.001, therefore degrees of freedom were corrected using Huynn-Feldt estimates of sphericity $\epsilon = 0.895$. The results indicate that the sound source had a significant, and strong positive effect on the engagement rating $F(1.791, 213.120) = 22.893$, p = <0.000, r = 0.593, (α = 0.05).

Post-hoc tests with pairwise comparisons revealed a significant difference in the engagement rating for the (real world) car presentation condition, compared to both of the artificial conditions (square wave and noise band); the car condition versus the square wave condition $CI_{.95}$ = 0.358 (lower) .941 (upper), $p =< 0.001$; and the car condition versus the noise band condition $CI_{.95}$ = .941 (lower) 1.842 (upper), $p =< 0.001$. There was no significant difference between the two artificial (noise band and square wave) conditions. Please see the mean difference, confidence intervals, and significance levels listed in Appendix Table C.15. This result supports hypothesis 5, that the sound source affects listeners engagement rating of an approaching object, with the real world car traction prompting a significantly greater engagement rating, than the artificial square wave and noise band.

5.5 Discussion

In this chapter we conducted an experiment to have a closer inspection of individual audio cues for movement in depth, and the sound sources used to present the stimuli. In regard to the audio cues, the first observation we see is that a number of the conditions did not prompt an underestimation of the contact time. One explanation for this result, may have been the method for testing the time-to-contact and the addition of an occlusion period. When presented the stimuli, subjects may have waited for this occlusion period to start before considering when to predict the contact time. Further, as the occlusion period was only 300ms, it was perhaps too short to allow for any delays and individual discrepancies.

We introduced the direct-to-reflections sound energy ratio as an audio cue, and our first hypothesis was that the parameter would act as an audio cue for movement in depth. The results show that the presentation of the parameter biased the perceived time-to-contact and prompted an earlier response time; prompted a significantly greater arousal rating; and prompted a significantly greater engagement rating (p = $<$ 0.001). Therefore, we conclude that the results support hypothesis 1, that the direct-to-reflections ratio acted as an audio cue for movement in depth, influencing the perceived time-to-contact, arousal and engagement ratings.

The results also showed that for the single audio cues, listeners responses to the individual audio cues for movement in depth differed, revealing a hierarchy across the audio cues. Distribution of the results, shows that the amplitude increase cue (Amp) prompted the fastest response times, and the greatest arousal and engagement ratings, whilst the inter-aural differences cue prompted the slowest response time, and lowest arousal and engagement ratings. Further analysis showed that the amplitude increase cue (Amp) prompted a significantly earlier perceived contact time than both the direct-to-reflections ratio (Ref) cue ($p =< 0.001$), and the inter-aural differences (IAD) cue ($p =< 0.001$); it prompted significantly greater arousal ($p = 0.010$) and engagement ($p = 0.003$) ratings than the inter-aural differences (IAD) cue. We conclude that these significant results support hypothesis 2, that individual cues differ in there capacity to bias perception of an approaching object, with the amplitude increase being the most dominant cue, and the inter-aural differences being the least dominant cue for objects moving on a frontal midline trajectory. We also suggest that this hierarchy may change, and the capacity for the inter-aural differences to act as an audio cue may increase, as the object's angle of approach changes, increasing the magnitude of the difference between the two channels, and therefore increasing the magnitude of the audio cue information.

This pattern of results was also replicated when comparing the multiple cue conditions, with conditions containing the amplitude increase variable (Amp + Ref, Amp + IAD, Amp + IAD + Ref) prompting significantly earlier estimates of the time-to-contact, than the condition without the amplitude increase variable (IAD + Ref), supporting hypothesis 4.

We also saw that conditions with multiple audio cues generally prompted earlier estimates of the contact times, greater arousal and engagement ratings, than single audio cues. This result was significantly different for conditions which contained amplitude increase as one of the multiple audio cues, when compared to single cues that did not contain amplitude increase. Therefore, this result provides evidence in support of hypothesis 3, via the hierarchy of cues (hypotheses 2 and 4) with the multiple cues including amplitude increase having more affect than the associated single cues.

In this experiment, we also investigated if the sound source and the use of real world sound sources (in the form of a sound sample of an approaching car) as opposed to artificial sound sources (in the form of a square wave and a noise band) affect perception of the approaching object. Whilst the results showed that the real world (car traction) sound source prompted earlier estimates of the contact times than the artificial sound sources, it did not reach significance level, therefore does not support hypothesis 5 in regard to the estimated time-to-contact.

However, for measurements of engagement, the real world (car) sound source prompted significantly greater engagement ratings than both the artificial sound sources ($p =< 0.001$). And interestingly for measures of emotion, the artificial square wave, prompted significantly lower valence ($p =< 0.001$) and significantly greater arousal ratings ($p =< 0.055$) than the real world (car) recording. Whilst it may be expected that a square wave will prompt more negative valence and greater arousal ratings, this result may have implications for the use of artificial sound sources and the square wave in experimental conditions, the emotional responses to which may prompt results which are not automatically applicable to real world sounds. Therefore, in regard to the emotion and engagement ratings, we suggest that these results support hypothesis 5, that listeners responses to real world sound sources differ to their responses to artificial sound sources.

Chapter 6

Responses to Complex Auditory-Visual Looming

In the previous chapter, we investigated peoples responses to audio cues for movement in depth, at both the single and multiple cue level, as well as the sound source presented with the real world (car traction) sound source versus artificial sound sources (the square wave and noise band) which are regularly used in experimental conditions but rarely encountered in the natural world.

For our final experiment, we apply the audio cues and sound sources investigated in experiment 3 (Chapter 5) to visual stimuli (film sequences) to determine what, if any, impact they have on auditory-visual perception of the approaching object.

Whilst the audio cues are the foci of the study, we again consider the ecological validity as a factor. As we saw in the review of psychological research (Chapter 2.2), many of the psychoacoustic looming studies use an artificial sound source as the experimental stimuli. Whilst the results from such research has provided important information on human perception and responses, can the conclusions drawn from these results, which are based on artificial conditions, transfer to real world or hyper-real scenarios?

The experiment conducted for this chapter, investigates if human perception and response to artificial auditory-visual looming stimuli differs to real world auditory-visual looming stimuli. We present the real world stimuli of an approaching car (congruent with the visual stimuli of an approaching car presented with the sound of tyre traction over a road surface), and the two artificial auditory stimuli conditions (a square wave and a noise band which are presented with the visual stimulus that is most often used in the visual looming perception studies - a white disc expanding on a black background). We analyse if the observers responses to the artificial stimuli differs to the real world stimuli.

6.1 Aims

The aims of this study are, firstly, to determine if the audio cues (investigated in experiment 3) bias audiovisual perception of the approaching object (as compared to looming scenes with no audio cues), if so do certain audio cues affect visual perception more than others, and do multiple cues have a different affect to single cues?; secondly, do human responses differ when real world stimuli are presented, as compared to the artificial stimuli?.

6.2 Hypotheses

It is hypothesised that:

1. the addition of audio cues to a looming sound (in audiovisual presentation) will prompt people to
   (a) perceive the contact time (of the approaching object) to be sooner than the scenes which contain sound but no audio cues (the Sound (no cues) condition).
   (b) express greater valence, arousal, and engagement ratings than the scenes which contain sound but no audio cues (the Sound (no cues) condition).

2. observers responses to the looming audiovisual stimuli will differ according to the number of, and specific audio cue(s) presented. That observers responses to scenes with
   (a) single audio cues (amplitude increase, inter-aural differences, reflections ratio) will differ, suggesting a hierarchy amongst the individual audio cues.
   (b) multiple audio cues (Amp + IAD, Amp + Ref, IAD + Ref, Amp + IAD + Ref) will differ, suggesting a hierarchy amongst the combinations of, and number of audio cues.
   (c) multiple audio cues (two and three audio cues) will differ from the scenes with single audio cues.

3. observers responses to the looming audiovisual stimuli will differ, with the real world (car traction) sound source prompting people to
   (a) perceive the contact time (of the approaching object) to be sooner than the scenes which present an artificial sound source (noise band, square wave).
   (b) express greater valence, arousal, and engagement ratings than the scenes which have an artificial sound source (noise band, square wave).

4. observers responses to the congruent real world looming stimuli (e.g. car sound presented with the car moving image) will differ from the artificial congruent looming stimuli (the square wave with the expanding disc image, and the noise band with the expanding disc image).

6.3 Method

6.3.1 Design

The study used a within-subjects design. For the auditory stimulus, there were two independent variables - sound source and audio cue.

1. Sound Source:
    - Image Only (no sound),
    - Real World (car traction sample),
    - Artificial (square wave),
    - Artificial (noise band).
2. Audio Cue:
    - Image Only (no sound),
    - Amplitude Increase,
    - Inter-aural Differences,
    - Direct-to-Reflections Sound Energy Ratio.

The visual stimulus presented moving images and had one independent variable - visual presentation, which was comprised of two levels:

- Real World (the moving image of an approaching car),
- Artificial (the moving image of an expanding white disc on black background).

There were four dependent variables:

- Time-to-contact,
- Valence,
- Arousal,
- Engagement.

6.3.2 Participants

The same sample of 15 volunteers that were recruited in Experiment 3 Chapter 5 also participated in this study. They were Ph.D students and Postdoctoral researchers from Queen Mary, University of London aged between 22 and 34 years (M = 27.33 years, SD = 3.24), with more male participants than female participants (9 males, 6 females). All participants reported normal hearing, with 4 participants correcting their vision with glasses. A possible learning bias to the auditory stimuli may have been introduced in this experiment, due to the participants previous exposure to the auditory stimuli (presented once for each condition) in experiment 3. To reduce the impact of this possible bias, we will not compare responses to the auditory only stimuli, with auditory-visual stimuli.

6.3.3 Stimuli

The stimuli consisted of looming scenes (listed in Appendix Table D.1 with the files included in the Digital Appendix) that presented objects moving towards the viewer, and were comprised of both audio and visual components. Each scene was 1700ms in duration. The experiment was presented via a computer with the visual stimulus (and the experiment interface) displayed on the monitor, and the auditory stimulus was transmitted through a pair of headphones.

As our aim was to investigate how the audio cue variables (explored in experiment 3, Chapter 5) affect visual perception, we used the auditory stimuli generated for experiment 3 in this study also. An explanation of the auditory stimuli's construction is explained in Section 5.3.3.

The visual stimulus involved the presentation of film sequences showing an object (one being a car (Figure 6.1) and the other being a white disc on a black background (Figure 6.2)) thats area expanded over time, representing an object moving towards the viewer on a frontal midline (approaching) trajectory. The disc was chosen as it has been used extensively in visual looming studies [Schiff et al., 1962; Schiff, 1965; Gray and Regan, 1998, 1999; Regan and Beverley, 1978; Hong and Regan, 1989], whilst the car was chosen as it had also been the most used real-world object in visual looming studies [Schiff and Oldak, 1990; Caird and Hancock, 1994; Hancock and Manster, 1997; Horswill et al., 2005; Terry et al., 2008; Rodrigues et al., 2012] in addition to being regularly experienced in the real-world. Please refer to the discussion in Section 2.2.2.

The area of the disc and the rate of expansion was designed to match the area and expansion rate of the car, to eliminate any biases that object size and velocity may cause to the visual information between the two scenes. However it is acknowledged that the visual stimuli differed in equivalency for other aspects, such as colour, dimensionality, and emotional association, which may have affected the participants responses to the stimuli, however underscores the difference between the use of real-world and artificial stimuli used in experiments.

Figure 6.1: Visual Stimuli: Car

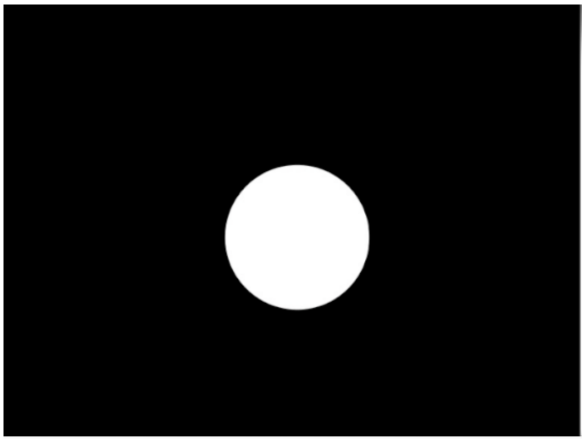

Figure 6.2: Visual Stimuli: Disc

In this study, we include two control conditions:

- The Image Only condition (whereby the visual stimuli were presented without sound),
- The Audiovisual Control condition (which presented the visual stimuli with the sound source stimuli, however did not contain any of the audio cue variables - i.e. no amplitude change, inter-aural differences or direct-to-reflections sound energy ratio).

The auditory stimuli were comprised of the following sound sources:

- Car traction (real world),
- Square wave (artificial),
- Noise band (artificial).

Each of the 3 sound sources had the following audio cue variables applied, as a single audio cue variable:

- Amplitude Increase (Amp),
- Inter-aural differences (IAD),
- Direct-to-Reflections Sound Energy Ratio (Ref).

and in combination as multiple audio cue variables:

- Amplitude Increase + Inter-aural differences (2 cues) (Amp + IAD),
- Amplitude Increase + Reflections ratio (2 cues) (Amp + Ref),

- Inter-aural differences + Reflections ratio (2 cues) (IAD + Ref),
- Amplitude Increase + Inter-aural differences + Reflections ratio (3 cues) (Amp + IAD + Ref).

The visual stimuli had two object presentation variables:

- Approaching car (real world) see Figure 6.1.
- Expanding white disc on black background (artificial) see Figure 6.2.

For the practice study we presented 6 trials that displayed an expanding black disc on a white background, and was not accompanied by an auditory stimulus.

There were 74 different trial conditions in total (listed in Appendix Table D.1) and each trial was presented once only in a randomised order. The presentation of each trial was limited to once only, as further presentations would have introduced learning, memory, and fatigue biases.

6.3.4 Apparatus

The apparatus used was the same as in the previous two perceptual experiments (Chapters 4 and 5). In a space and time saving measure, please refer to the apparatus Section 4.3.4 for other methodological details.

6.3.5 Dependent Variable Measurement

For this experiment we had four dependent variable measurements, being time-to-contact, engagement, valence and arousal (emotion). The emotion and engagement rating scales were the same as those used in the two previous perceptual experiments (Chapters 4 and 5). For a description of the measurement techniques, please refer to Section 4.3.5 subsections Emotion: Valence and Arousal and Engagement. The time-to-contact measurement was the same used in experiment 3, please refer to the description given in Section 5.3.5 Time-To-Contact.

6.3.6 Procedure

Participant's sat at the computer workstation and were informed of the experimental procedure. They were given an information sheet summarising both the procedure and the ethics approval, signed a consent form, and completed a background questionnaire asking questions on gender, age, and whether they have had corrections made to their vision or hearing (the documents are included in the Digital Appendix).

Before commencing the experiment, the participant's completed a practice study using 6 looming scenes that were not additionally presented in the experiment. It was conducted as a supervised learning procedure to provide them with the opportunity to comprehend the experiment, the procedure, the micro time scale of the stimulus, and how to complete the task. Participant's were then instructed to start the experiment when ready.

The task required the participant's to watch and listen to the scene of an approaching object. They were informed that the scene would be then blocked from both view and hearing, but to imagine that the object was still moving towards them, and to press the keyboard space bar when they thought the object reached them. A pop-up questionnaire was then displayed on the computer screen, asking the participant's to rate their valence / arousal level and how engaging the scene was.

Each trial lasted for a total duration of 1700ms and the participant's were not time restricted on the duration for answering the questions. Once they had submitted their answers a 4 second break was then given between each trial in which an image of 'visual white noise' (see Appendix Figure B.1) was displayed on the screen and no sound was output through the headphones. The experiment lasted for approximately 25 minutes and participant's were not given any information implying there might be 'correct', 'incorrect' or 'preferred' responses.

6.4 Results

A total of 1110 trials were presented. Each of the fifteen participant's received 74 trials, comprised of the three sound sources (car, square, and noise) presented twelve times for each visual presentation (car and disc). Two image only conditions were also presented (the car and disc were each presented once) whereby no sound was presented only the visual stimuli. The trials and conditions are listed in Appendix Table D.1.

The analyses performed in this chapter used the same analysis methods as those detailed in Section 4.4. In a space saving measure please refer to this section for a broader explanation of the analysis methodology.

6.4.1 Audio Cues

To test hypotheses 1 and 2, we began by looking at the audio cues affect on the perceived time-to-contact, then emotion (valence and arousal), and lastly engagement rating.

6.4.1.1 Audio Cues × Time-to-Contact

Early exploration of the results showed that some of the data contained outliers. ANOVAS are sensitive to outliers, therefore 7 outliers across 3 trial comparisons were removed, leaving 12 trials per audio cue (× visual presentation) condition. The time-to-contact was then averaged across all of the participants responses (and sound sources) for each audio cue condition per visual presentation, and is plotted in Figure 6.3.

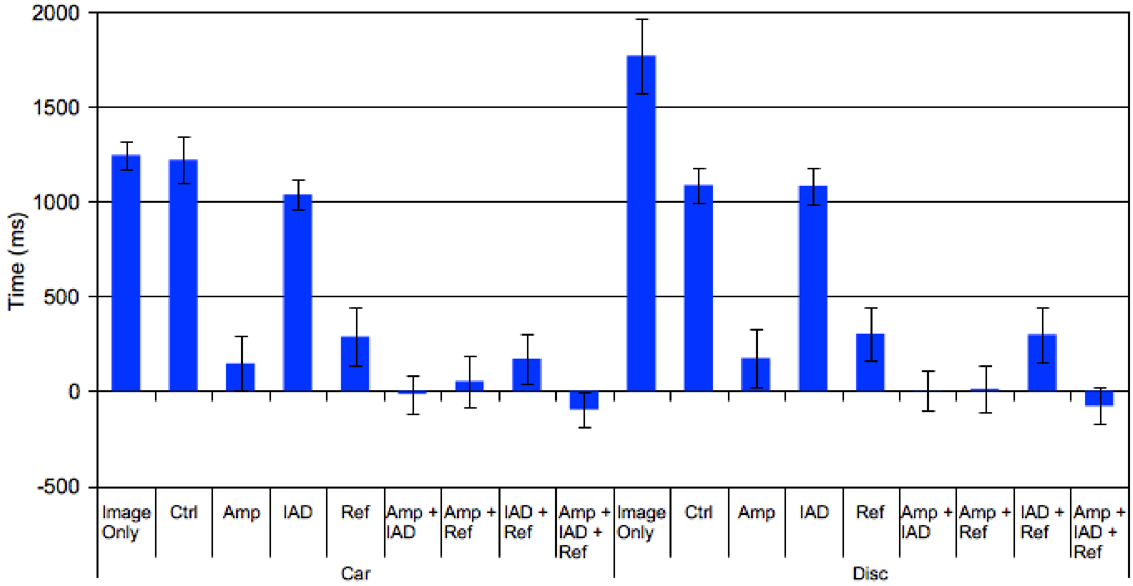

Figure 6.3: Audio Cue × Time-to-Contact Bar Chart
The time-to-contact estimates for each audio cue condition (averaged across all of the participants ratings and sound sources) are plotted. Error bars indicate the standard error for each condition. The contact time occurs at 0ms, with any underestimation plotted in the negative range of the scale, and overestimation plotted in the positive range. **Car Presentation - Image Only:** M = 1243.314 ($S.E.$ = 73.078); **Sound (no cues):** M = 1218.173 ($S.E.$ = 123.696); **Amp:** M = 145.326 ($S.E.$ = 145.057); **IAD:** M = 1035.353 ($S.E.$ = 76.119); **Ref:** M = 287.809 ($S.E.$ = 152.824); **Amp + IAD:** M = -15.885 ($S.E.$ = 102.085); **Amp + Ref:** M = 52.152 ($S.E.$ = 137.617); **IAD + Ref:** M = 170.414 ($S.E.$ = 131.153); **Amp + IAD + Ref:** M = -97.783 ($S.E.$ = 90.305); **Disc Presentation - Image Only:** M = 1767.856 ($S.E.$ = 195.996); **Sound (no cues):** M = 1084.738 ($S.E.$ = 92.904); **Amp:** M = 174.171 ($S.E.$ = 151.266); **IAD:** M = 1080.772 ($S.E.$ = 99.423); **Ref:** M = 301.549 ($S.E.$ = 140.099); **Amp + IAD:** M = 1.793 ($S.E.$ = 103.813); **Amp + Ref:** M = 11.253 ($S.E.$ = 119.458); **IAD + Ref:** M = 298.430 ($S.E.$ = 142.426); **Amp + IAD + Ref:** M = -79.058 ($S.E.$ = 96.061).

Looking at the results plotted in Figure 6.3 we see that the image only conditions have the greatest amount of overestimation in both the car and disc presentation, and that all other conditions containing sound and audio cues prompted less overestimation. This result shows that the application of audio cues (either single or multiple) caused people to alter (lessen) their estimation of the contact time. We see that the multiple cue conditions Amp + Ref and Amp + IAD, had the least amount of overestimation, and that the condition containing all 3 cues Amp + IAD + Ref actually prompted an underestimation of the perceived contact time (see descriptives listed in Appendix Table D.2). On closer inspection of the single audio cue conditions, we see that the Amp (only) condition prompted the least amount of overestimation and the IAD condition the most amount.

A one-way repeated measures ANOVA was conducted and the means and standard errors are listed in Appendix Table D.2. The results indicate that the audio cues had a significant, and very strong positive effect on the estimated time-to-contact $F(17, 187) = 47.234, p =< 0.001, r = 0.755$, ($\alpha = 0.05$). Post-hoc tests with pairwise comparisons were conducted and the descriptives are listed in Appendix Tables D.3, D.4, and D.5. In a space saving measure, please refer to the mean difference, significance level, and confidence intervals listed in these tables.

To test hypothesis 1A, we compare the Sound (no cues) condition to all of the conditions containing audio cues. The results show that the addition of audio cues to sound stimuli prompted people to perceive an earlier time-to-contact, than the condition with no cues (sound only). All of the conditions with audio cues (except IAD) were significantly earlier as compared to the Sound (no cues) condition (for both car and disc visual presentations), supporting hypothesis 1A. In a space saving measure, please refer to the mean difference, significance level, and confidence intervals listed in Appendix Tables D.3 and D.4.

When we compare the single audio cue conditions (to determine if there is a hierarchy amongst the individual audio cues, hypothesis 2A) we see that the amplitude increase condition, prompts the earliest perceived time-to-contact (which is significantly faster than the IAD condition for the car presentation), followed by the direct-to-reflections energy ratio, then the inter-aural differences, with the Amp and Ref conditions both eliciting significantly faster time-to-contact than the IAD condition. With such strong significant results, we conclude that the results support hypothesis 2A.

To determine if there is a hierarchy amongst the multiple cues (testing hypothesis 2B), we looked at the multiple cue pairwise comparisons (Amp + IAD, vs Amp + Ref, vs IAD + Ref, vs Amp + IAD + Ref). Only one of the pairwise comparisons reached the significance level, therefore the results are not strong enough to support hypothesis 2B; however, looking at the plotted results there does still appear to be a hierarchical ranking, with some audio cues affecting the perceived time-to-contact more than others. We see that the condition containing all three audio cues (Amp + IAD + Ref) not only had the earliest time-to-contact, it also prompted the participants to underestimate the contact time, for both the car and disc visual presentations. This was followed by the 2 multiple cue conditions which contained amplitude (Amp + IAD, and Amp + Ref), then lastly the IAD + ref condition. With the conditions containing amplitude all prompting earlier percepts of the contact time, it appears that the amplitude increase is a dominant audio cue (across both visual presentations), for objects moving on a frontal midline trajectory (and at this velocity).

When we compare the single versus multiple audio cue conditions (to determine if multiple audio cues prompted people to have a faster estimation than single audio cue conditions - hypothesis 2C) we see that the addition of a second audio cue, prompted earlier time-to-contact response times than when the stimuli was presented with just the single audio cue condition (i.e. Amp × Amp + IAD; Amp × Amp + Ref; IAD × Amp + IAD; IAD × Ref + IAD; Ref × Amp + Ref; Ref × IAD + Ref). Whilst these pairwise comparisons did not meet the significance level, we nevertheless see this pattern of results evident in both the car and disc visual presentations.

Looking at the other pairwise comparisons, we note that the condition with three cues (Amp + IAD + Ref) prompted significantly earlier time-to-contact response times than the IAD condition for both car and disc presentations; that the multiple cue Amp + IAD condition prompted significantly earlier time-to-contact response times than both the conditions which presented the cues individually; and that the Amp + Ref condition also prompted a significantly earlier time-to-contact response time than the IAD condition. Whilst the hypothesis can not be supported across all of the multiple versus single cue conditions, we propose that the significant results support the hypothesis for certain cue combinations, which are prompting earlier time-to-contact response times than other single cue conditions, primarily the inter-aural difference, which is perhaps due to the small differences in this cue.

There was one multiple cue condition (IAD + Ref condition) which did not prompt earlier times than the single condition (Amp only). Although it was only a small difference (see Amp × IAD + Ref pairwise comparison) with the amplitude increase -25.088ms faster for the car presentation, and -124.259ms for the disc presentation, one explanation for this result, refer's back to the hierarchy of individual cues, and the strong capacity of the amplitude increase as an audio cue for movement in depth, whereas the degree of inter-aural differences for an oncoming object, are small in comparison.

When comparing the multiple cue conditions which contained amplitude (as one of the multiple cues) (Amp + IAD, Amp + Ref, and Amp + IAD + Ref), against the single cues not containing amplitude (IAD and Ref), the multiple cues containing the amplitude increase cue all resulted in a significantly greater difference, with the multiple cues prompting earlier time-to-contact response times, than the single cues not containing the amplitude increase.

And lastly looking at the effect of audio cues on visual perception in relation to the ecological validity of the object, (e.g. car versus disc Appendix Table D.5), we see that for most conditions (with the exception of the Sound (no cues) pairwise comparison, and the Amp + Ref comparison), the time-to-contact results were earlier for the car presentation, than for the disc presentation, however none of the differences reached the significance level. This result suggests that the audio cues work for both real world and artificial scenarios, and although the time-to-contact was earlier for real world presentation, it did not reach the significance level.

6.4.1.2 Audio Cues × Emotion (Valence / Arousal)

Early exploration of the results showed that the Valence data contained 2 outliers (across 2 trial comparisons), and the Arousal ratings contained 4 outliers (across 3 trial comparisons). These were removed, leaving the data for 13 valence and 12 arousal trials and per condition. The emotion ratings were then averaged across all of the sound sources and participant's responses for each audio cue condition, and are plotted in Figure 6.4.

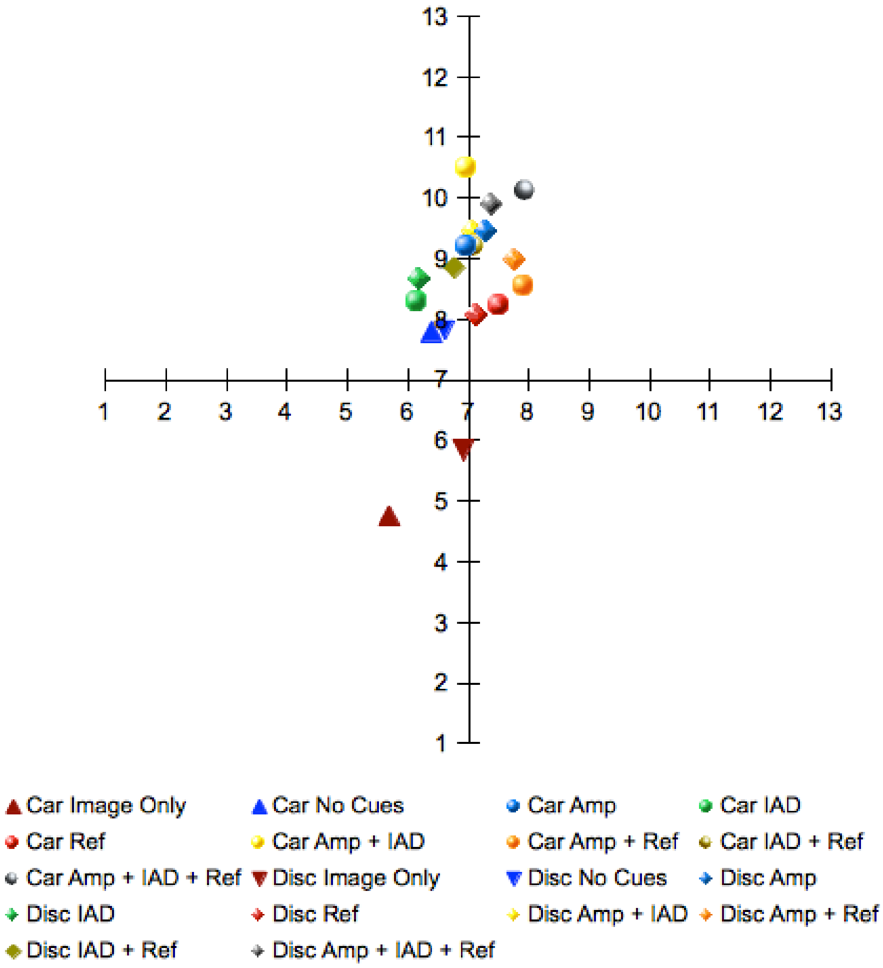

Figure 6.4: Audio Cue × Valence / Arousal Scatter Plot
The valence / arousal ratings for each audio cue condition (averaged across all of the participants ratings and sound sources) are plotted. (**Car: Image Only:** Valence M = 5.692, Arousal M = 4.750; **Sound (no cues):** Valence M = 6.395, Arousal M = 7.791; **Amp:** Valence M = 6.974, Arousal M = 9.195; **IAD:** Valence M = 6.154, Arousal M = 8.291; **Ref:** Valence M = 7.487, Arousal M = 8.223; **Amp + IAD:** Valence M = 6.948, Arousal M = 10.499; **Amp + Ref:** Valence M = 7.898, Arousal M = 8.555; **IAD + Ref:** Valence M = 7.102, Arousal M = 9.195; **Amp + IAD + Ref:** Valence M = 7.948, Arousal M = 10.138); (**Disc: Image Only:** Valence M = 6.923, Arousal M = 5.833; **Sound (no cues):** Valence M = 6.615, Arousal M = 7.820; **Amp:** Valence M = 7.308, Arousal M = 9.445; **IAD:** Valence M = 6.190, Arousal M = 8.654; **Ref:** Valence M = 7.141, Arousal M = 8.067; **Amp + IAD:** Valence M = 7.078, Arousal M = 9.444; **Amp + Ref:** Valence M = 7.770, Arousal M = 8.973; **IAD + Ref:** Valence M = 6.770, Arousal M = 8.848; **Amp + IAD + Ref:** Valence M = 7.386, Arousal M = 9.889).

One-way repeated measures ANOVAS were conducted with the means and standard errors listed in Appendix Table D.6. The results indicate that the application of audio cues had a significant, but week positive effect on the valence rating $F(17, 204) = 4.095, p =< 0.001, r = 0.171$, ($\alpha = 0.05$); and a significant, and very strong positive effect on the arousal rating $F(17, 187) = 22.610, p =< 0.001, r = 0.794$, ($\alpha = 0.05$).

Post-hoc tests with pairwise comparisons were conducted and the descriptives are listed in Appendix Tables D.7, D.8, and D.9. In a space saving measure, please refer to the mean difference, significance level, and confidence intervals listed in these tables.

Looking at the results, we see that the conditions with sound had greater valence / arousal ratings than the image only condition. That the conditions with audio cues had greater valence / arousal ratings than the Sound (no cues) condition (with exception of the valence rating for IAD which was rated on average .24 lower than the Sound (no cues) Condition - one explanation for this negligible difference could be due to the small level of cue in the IAD condition) and image only condition. We also see that the condition with all three cues (Amp + IAD + Ref) prompted significantly greater arousal rating, than the Sound (no cues) condition (for the disc visual presentation). As the significance level was not met for all pairwise comparisons, the hypothesis cannot be supported across all audio cues conditions, however the significant results for the Amp + IAD + Ref condition, supports hypothesis 1B, whereby the addition of multiple audio cues prompts greater arousal ratings.

When we compare the single audio cue conditions (to determine if there is a hierarchy amongst the individual audio cues, hypothesis 2A) we see that the Amp condition prompted the greatest arousal rating, followed by the IAD, then lastly the Ref. This order was evident in both the car and disc presentations, however the only pairwise comparison that met the level of significance was the Amp × Ref comparison for the disc visual presentation. We acknowledge that most of the comparisons did not reach the significance level, therefore the results are not strong enough to support hypothesis 2A outright across all conditions. However, the hierarchical order was evident in both car and disc presentations, it cannot be disqualified. Further, the fact that the amplitude increase prompted significantly greater arousal ratings than the reflections ratio condition, demonstrates that there is a difference in the amount of arousal which the audio cues prompt. As this difference did reach the significance level, we propose that in regard to the amplitude increase condition, the results support hypothesis 2A, that the amplitude increase is the strongest individual cue, and will prompt greater arousal ratings, than other cues.

To determine if there is a hierarchy amongst the multiple cues (hypothesis 2B) we looked at the multiple cue pairwise comparisons. We see that the order of the conditions differ between the car and disc presentations, and none of the pairwise comparisons (except the disc IAD + Ref × Amp + IAD + Ref condition) meet the level of significance. Whilst most of the comparisons did not reach the significance level, looking at the plotted data the order was somewhat consistent across both the car and disc presentation, therefore while we acknowledge that the results are not strong enough to support hypothesis 2B, the hierarchy cannot be completely ruled out, therefore we propose that the results are inconclusive for valence and arousal.

When we compare the single versus multiple audio cue conditions (to determine if multiple audio cues prompted people to have greater valence and arousal ratings than single audio cues conditions - hypothesis 2C), we see that for most pairwise comparisons, the addition of a second audio cue prompted greater valence and arousal ratings than the single audio cue condition (i.e. Amp × Amp + IAD; Amp × Amp + Ref; IAD × Amp + IAD; IAD × Ref + IAD; Ref × Amp + Ref; Ref × IAD + Ref). This is again repeated when the third audio cue was added (i.e. Amp × Amp + IAD + Ref; IAD × Amp + IAD + Ref; Ref × Amp + IAD + Ref), which for the IAD × Amp + IAD + Ref and Ref × Amp + IAD + Ref comparisons, the increase in valence and arousal ratings reached the significance level. As the results are not significant across all pairwise comparisons, hypothesis 2C cannot be supported outright across all of the multiple cue versus single cue conditions, however, we propose that the significant results for certain cue combinations, primarily the three audio cue combination (Amp + IAD + Ref), support hypothesis 2C, and that a hierarchy between the combinations of sound cues exists, that becomes evident with a maximal number of audio cues, prompting greater valence and arousal ratings than single cue conditions.

6.4.1.3 Audio Cues × Engagement

Early exploration of the results showed that some of the data contained outliers. 4 outliers across 3 trial comparisons were removed, leaving 12 trials per condition. The engagement ratings were then averaged across all of the sound sources (car, noise, and square), and all participant's responses, for each audio cue × visual presentation condition, and are plotted in Figure 6.5.

Looking at the plotted data, we see that all of the conditions with sound had greater engagement ratings than the image only (no sound) condition. This suggests that the participants found the multimodal presentations more engaging.The conditions with audio cues all had greater engagement ratings than the Sound (no cues) condition (sound but no audio cues), and that the condition containing all three audio cues (Amp + IAD + Ref) had the greatest engagement rating. These results were reflected in both the car and disc visual presentations.

A one-way repeated measures ANOVA was conducted and the means and standard errors listed in Appendix Table D.10. The results indicate that the audio cue condition had a significant, and strong positive effect on the engagement rating $F(17, 187) = 13.588, p =< 0.001, r = 0.456$, ($\alpha = 0.05$). Post-hoc tests with pairwise comparisons were conducted and the descriptives are listed in Appendix Tables D.11, D.12, and D.13. In a space saving measure, please refer to the mean difference, significance level,

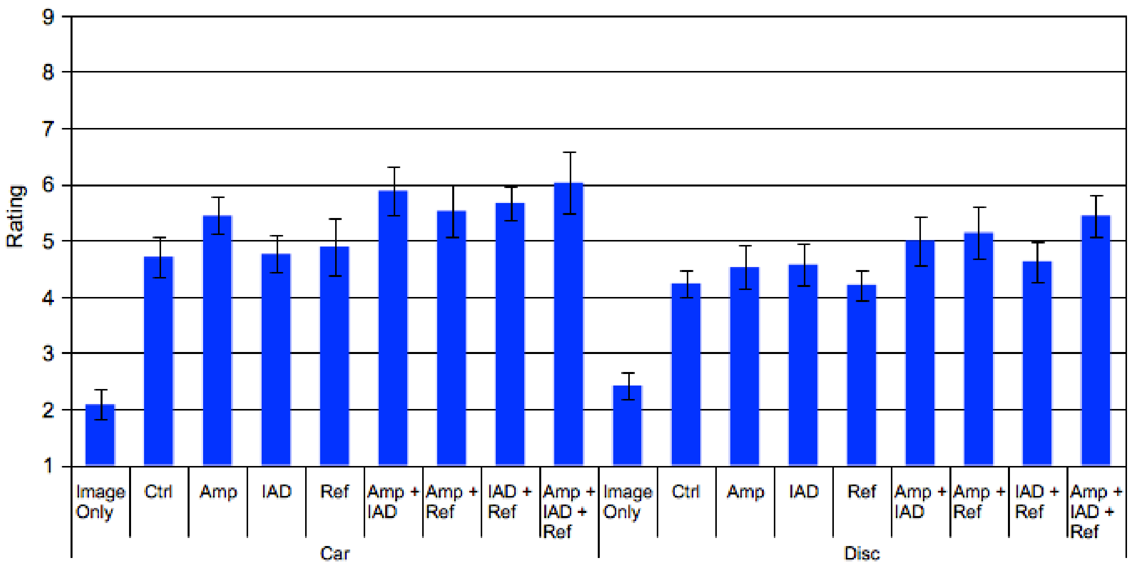

Figure 6.5: Audio Cue × Engagement Rating Bar Chart
The engagement rating for each audio cue condition (averaged across all of the participants ratings and sound sources) are plotted. Error bars indicate the standard error for each condition. (**Car:** **Image Only:** M = 2.467 ($S.E.$ = .389); **Sound (no cues):** M = 4.700 ($S.E.$ = .369); **Amp:** M = 5.799 ($S.E.$ = .350); **IAD:** M = 5.011 ($S.E.$ = .318); **Ref:** M = 4.910 ($S.E.$ = .451); **Amp + IAD:** M = 6.043 ($S.E.$ = .475); **Amp + Ref:** M = 5.756 ($S.E.$ = .416); **IAD + Ref:** M = 5.810 ($S.E.$ = .306); **Amp + IAD + Ref:** M = 6.221 ($S.E.$ = .513)); (**Disc:** **Image Only:** M = 2.733 ($S.E.$ = .358); **Sound (no cues):** M = 4.488 ($S.E.$ = .258); **Amp:** M = 4.979 ($S.E.$ = .392); **IAD:** M = 4.834 ($S.E.$ = .380); **Ref:** M = 4.512 ($S.E.$ = .302); **Amp + IAD:** M = 5.511 ($S.E.$ = .447); **Amp + Ref:** M = 5.534 ($S.E.$ = .427); **IAD + Ref:** M = 5.011 ($S.E.$ = .365); **Amp + IAD + Ref:** M = 5.823 ($S.E.$ = .371)).

and confidence intervals listed in these tables.

To test hypothesis 1B, we compare the Sound (no cues) condition, to all of the conditions containing audio cues (in both car and disc visual presentations). The results show that all of the audio cue conditions prompted significantly greater engagement ratings than the condition with no audio cues, therefore supporting hypothesis 1B.

When we compare the single audio cue conditions (to determine if there is a hierarchy amongst the individual audio cues, hypothesis 2A) we see that the amplitude increase condition, prompted the greatest engagement rating for both the car and disc presentations, followed by the Ref then the IAD for the disc presentation, and the IAD then Ref for the car visual presentation. None of the comparisons met the significance level, therefore the results are not strong enough to support hypothesis 2A; however, we do see that the amplitude increase condition consistently prompted the greatest engagement rating for single audio cues.

To determine if there is a hierarchy amongst the multiple cues (testing hypothesis 2B), we looked at the multiple cue pairwise comparison. We see that the condition containing all three audio cues (Amp + IAD + Ref) had the greatest engagement rating, and the IAD + Ref condition had the lowest (of the multiple cue) rating, which was evident in both the car and disc visual presentations. The ranked order of the middle two conditions differed between the car and disc presentations, with the Amp + Ref having the second greatest engagement ratings, followed by the Amp + IAD for the car presentation, and the Amp + IAD having the second greatest engagement rating, followed by the Amp + Ref, for the disc visual presentation. Again the results are not strong enough to support hypothesis 2B, however we acknowledge that the condition containing all three audio cues prompted the greatest engagement rating, and the IAD + Ref condition prompted the lowest engagement rating across both the car and disc presentations (for multiple cues).

When we compare the single versus multiple audio cue conditions (to determine if multiple audio cues prompted people to have greater engagement ratings than single audio cues conditions - hypothesis 2C) we see that the addition of a second audio cue, prompted a greater rating than the single audio cue condition (i.e. Amp $\times$ Amp + IAD; Amp $\times$ Amp + Ref; IAD $\times$ Amp + IAD; IAD $\times$ Ref + IAD; Ref $\times$ Amp + Ref; Ref $\times$ IAD + Ref), however in most cases it was not a significant difference. So whilst the results are not strong enough to support hypothesis 2C, we acknowledge that the multiple cues tended to have greater engagement ratings than the single cue conditions, which was replicated in both the car and disc visual presentations.

And lastly looking at the effect of audio cues on visual perception in relation to the ecological validity of the object, (e.g. car versus disc Appendix Table D.13), we see that for most conditions (with the exception of the image only pairwise comparison), the engagement ratings were greater for the car presentation, than for the disc presentation, however the differences were small with none of the differences reaching the level of significance. This suggests that the audio cues work with both ecologically valid and non-valid image scenarios, and although the real world situation had greater overall engagement ratings, it was not significant.

6.4.2 Sound Source

This analysis investigates the effect of the sound source on human perception of the approaching object, comparing responses to a real world sound source and visual stimuli (of an approaching car), versus artificial sound sources (a square wave and a noise band) and visual stimuli (an expanding white disc on a black background). The artificial sounds are often used as the stimuli in auditory looming experiments, whilst real world sounds are only rarely used in looming experiments, therefore a comparison of the responses to the artificial versus real world stimuli will allow us to consider if the conclusions drawn from artificial stimuli transfer to real world scenarios.

6.4.2.1 Sound Source × Time-to-Contact

Early exploration of the results showed that some of the data contained outliers. ANOVAS are sensitive to outliers, therefore 2 outliers across 2 trial comparisons were removed. The time-to-contact was then averaged across all of the participants responses (and audio cues) for each sound source × visual presentation condition, and are plotted in Figure 6.6.

Looking at the plotted results, we see that for all conditions, the (averaged) time-to-contact was overestimated. That the uni-modal image only presentations prompted a greater overestimation than the multimodal conditions, suggesting that the additional (multimodal) information prompted people to have earlier response times. Of the conditions with sound, we see that the conditions which presented the real world car sound source had the least amount of overestimation, followed by the square wave, then noise band. We can also see that the results were consistent across the visual presentations, with this pattern of results occurring in both visual stimuli (car and disc) presentations.

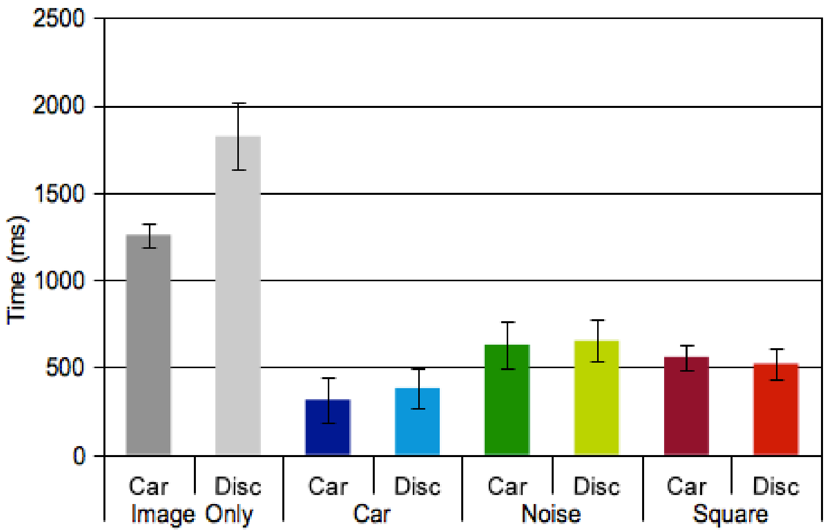

Figure 6.6: Sound Source × Time-to-Contact Bar Chart
The time-to-contact for each sound source × visual presentation condition (averaged across all of the participants ratings and audio cues) are plotted. Error bars indicate the standard error for each condition. The contact time occurs at 0ms, with overestimation plotted in the positive range of the scale. (**Image Only** - **Car:** $M = 1257.528$ ($S.E. = 68.709$), **Disc:** $M = 1822.595$ ($S.E. = 188.417$)); (**Car Sound Source** - **Car:** $M = 314.975$ ($S.E. = 129.602$), **Disc:** $M = 381.115$ ($S.E. = 115.661$)); (**Noise Band** - **Car:** $M = 629.291$ ($S.E. = 129.781$), **Disc:** $M = 652.810$ ($S.E. = 121.881$)); (**Square Wave** - **Car:** $M = 558.445$ ($S.E. = 69.709$), **Disc:** $M = 520.195$ ($S.E. = 90.051$));

A one-way repeated measures ANOVA was conducted with the means and standard errors listed in Appendix Table D.14. Mauchly's test indicated that the assumption of sphericity had been violated $x^2(27) = 103.499, p =< 0.001$, therefore the degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity $\epsilon = 0.283$. The results indicate that the sound source had a significant and strong positive effect on time-to-contact $F(1.983, 23.800) = 34.606, p =< 0.001, r = 0.691$, ($\alpha = 0.05$).

Post-hoc tests with pairwise comparisons were conducted and the descriptives are listed in Appendix Table D.15. In a space saving measure, please refer to the mean difference, significance level, and confidence intervals listed in this table.

To determine if observers response times to the looming audiovisual stimuli differs when real world sounds are presented, as opposed to artificial sounds (hypothesis 3A) we looked at the pairwise comparisons between the conditions which present the same visual stimuli but different sound sources (i.e. car-car × noise-car; car-car × square-car; noise-car × square-car; car-disc × noise-disc; car-disc × square-disc; noise-disc × square-disc).

The pairwise comparison of the car-car × noise-car conditions, revealed a significant difference in the perceived time-to-contact, with the car-car condition prompting a significantly earlier time-to-contact response time than the noise-car condition $CI_{.95} = -574.044$ (lower) -50.588 (upper), $p = 0.012$, however there was no significant difference between the car-car × square-car condition. This pattern of results was also evident in the expanding disc visual presentation. As the significance level was not met for all pairwise comparisons, the hypothesis cannot be supported across all conditions, however the significant results for the real world (car) × artificial (noise) condition for both visual presentations, supports hypothesis 3A, whereby the presentation of a real world sound source in an audiovisual presentation prompts earlier time-to-contact response times than the presentation of a noise band sound source.

To test hypothesis 4, do observers response times to congruent real world looming stimuli (the car sound presented with the car moving image) differ from artificial congruent looming stimuli (the square wave with the expanding disc image, and the noise band with the expanding disc image), we looked at the pairwise comparisons between the conditions (car-car × noise-disc; and car-car × square-disc).

The pairwise comparison of the car-car × noise-disc conditions, revealed a significant difference in the perceived time-to-contact, with the car-car condition prompting a significantly earlier time-to-contact response time than the noise-disc condition $CI_{.95} = -590.504$ (lower) -85.166 (upper), $p = 0.005$. However there was no significant difference between the car-car × square-disc condition. As the significance level was not met for both pairwise comparisons, the hypothesis cannot be supported across all conditions; however, the significant results for the real world (car-car) × artificial (noise-disc) condition supports hypothesis 4, whereby observers responses to the congruent real world looming stimuli (of the car sound presented with the car visual stimuli) differed to an artificial looming stimuli (of the noise band with the expanding disc visual stimuli).

6.4.2.2 Sound Source × Emotion (Valence / Arousal)

Early exploration of the results indicated that the valence data did not contain outliers, however some of the arousal data did. Therefore 3 arousal outliers (across 2 trial comparisons) were removed and the ratings were then averaged across all of the participants (and audio cues) for each sound source $\times$ visual presentation condition, with the results plotted in Figure 6.7.

Looking at the spread of the data, we see that the image only (car, disc) conditions had the lowest arousal ratings, with the multimodal audiovisual presentations prompting greater arousal ratings. The noise sound source conditions had a similar valence / arousal rating no matter which visual stimuli (real world or artificial) was presented; That the congruent real world car-car presentation had high valence and arousal ratings, with greater valence and (marginally greater) arousal ratings than the car-disc and square-car conditions; whilst the square-disc had greater arousal ratings than the square-car which had the lowest valence ratings overall. It is interesting to note that the valence ratings for car (and disc) image only conditions were greater than the square-car (and associated square-disc), with the results suggesting that the addition of the square wave to the moving image (no matter which visual stimuli was presented) lowered the valence rating, and that people preferred the visual stimuli without sound, rather than being accompanied by a square-wave.

One-way repeated measures ANOVAS were conducted to compare the valence and arousal ratings by sound source condition. For valence, Mauchly's test indicated that the assumption of sphericity had been violated $x^2(27) = 137.580, p =< 0.000$, therefore degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity $\epsilon = 0.310$. The results indicate that the sound source had a significant, and weak positive effect on the valence rating $F(2.167, 28.168) = 4.906, p =< 0.013, r = 0.207$, $(\alpha = 0.05)$.

For arousal, Mauchly's test indicated that the assumption of sphericity had been violated $x^2(27) = 131.282, p =< 0.001$, therefore degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity $\epsilon = 0.315$. The results indicate that the sound source had a significant, and very strong positive effect on the arousal rating $F(2.207, 24.279) = 40.147, p =< 0.001, r = 0.722$, $(\alpha = 0.05)$.

Post-hoc tests with pairwise comparisons were conducted and the descriptives are listed in Appendix Table D.17. In a space saving measure, please refer to the mean difference, significance level, and confidence intervals listed in this table.

To determine if observers valence / arousal ratings to the looming audiovisual stimuli differs when real world sounds are presented, as opposed to artificial sounds (hypothesis 3B) we looked at the pairwise comparisons between the conditions which present the same visual stimuli but different sound sources (i.e. car-car $\times$ noise-car; car-car $\times$ square-car; noise-car $\times$ square-car; car-disc $\times$ noise-disc; car-disc $\times$ square-disc; noise-disc $\times$ square-disc).

For valence, the pairwise comparisons revealed a significant difference in ratings, with the square-car condition prompting a significantly lower ratings than the car-car condition $\text{CI}_{.95} = 1.15$ (lower) 5.18 (upper), $p =< 0.001$, and the noise-car condition $\text{CI}_{.95}$

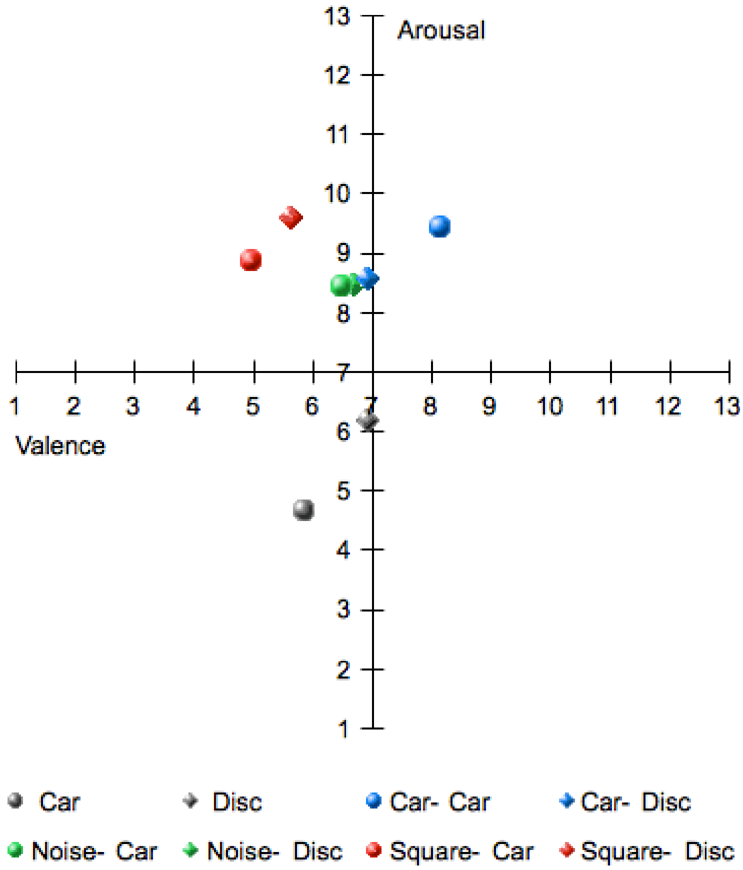

Figure 6.7: Sound Source × Valence / Arousal Scatter Plot
The valence / arousal rating for each sound source × visual presentation condition (averaged across all of the participants ratings and audio cues) are plotted. (**Image Only** - **Car:** Valence M = 5.857, Arousal M = 4.667; **Disc:** Valence M = 6.929, Arousal M = 6.167); (**Car Sound Source** - **Car:** Valence M = 8.136, Arousal M = 9.433; **Disc:** Valence M = 6.929, Arousal M = 8.550); (**Noise Band** - **Car:** Valence M = 6.489, Arousal M = 8.452; **Disc** Valence M = 6.721, Arousal M = 8.437); (**Square Wave** - **Car:** Valence M = 4.972, Arousal M = 8.882; **Disc:** Valence M = 5.661, Arousal M = 9.569);

$= -0.15$ (lower) 2.88 (upper), $p = 0.02$. However there was no significant difference between the car-car × noise-car condition. As the significance level was not met for all pairwise comparisons, the hypothesis cannot be supported across all conditions, however the significant results for the real world sound (car) × artificial (square wave) condition supports hypothesis 3B, whereby the presentation of a square wave artificial sound source in an audiovisual presentation prompts significantly lower (negative) valence ratings than the presentation of a noise band sound source or a real world car sound source.

For arousal, the pairwise comparison for the car-car × noise-car condition, revealed a significant difference in the arousal rating, with the car-car condition prompting a significantly greater ratings than the noise-car condition $CI_{.95} = 0.082$ (lower) 1.881 (upper), $p = 0.027$, however there was no significant difference between the car-car × square-car condition. As the significance level was not met for all pairwise comparisons the hypothesis cannot be supported across all conditions, however the significant results for the real world (car-car) $\times$ artificial (noise-car) condition for both visual presentations, supports hypothesis 3B, whereby the presentation of a real world car sound source in an audiovisual presentation prompts a greater arousal rating than the presentation of an artificial noise band sound source.

To test hypothesis 4, do observers valence / arousal ratings for congruent real world looming stimuli (the car sound presented with the car moving image) differ from artificial congruent looming stimuli (the square wave with the expanding disc moving image, and the noise band with the expanding disc moving image), we looked at the pairwise comparisons between the conditions (car-car $\times$ noise-disc; and car-car $\times$ square-disc).

For valence, the pairwise comparisons revealed a significant difference in ratings, with the square-disc condition prompting a significantly lower rating than the car-car condition $CI_{.95}$ = 1.69 (lower) 3.26 (upper), p =< 0.001, however the difference in the car-car $\times$ noise-disc condition did not meet the significance level. As the significance level was not met for both pairwise comparisons, the hypothesis cannot be supported across all conditions, however the significant results for the real world (car-car) $\times$ artificial (square-disc) condition supports hypothesis 4, whereby observers responses to the artificial looming stimuli of a square wave with an expanding disc differed, with significantly lower valence ratings than the presentation of congruent real world looming stimuli of a car sound source presented with the car visual stimuli.

For arousal, the square-disc condition prompted the greatest arousal rating, narrowly followed by the car-car condition, and lastly the noise-disc condition. However, none of the pairwise comparisons met the level of significance, therefore the results are not strong enough to support hypothesis 4.

6.4.2.3 Sound Source × Engagement

Early exploration of the results showed there were 4 outliers (across 4 trial comparisons). These were removed and the engagement ratings were then averaged across all of the participants responses (and audio cues) for each sound source $\times$ visual presentation condition, and are plotted in Figure 6.8.

Looking at the plotted results, we see that the conditions with sound all had greater engagement ratings, than the image only conditions, suggesting that people found multimodal presentation more engaging than a unimodal presentation. Of the multimodal audiovisual conditions, the conditions which presented the real world car visual presentation prompted greater engagement ratings that the expanding disc visual presentation. We also see that the congruent real world car-car condition had the greatest engagement rating, followed narrowly by the square-car and square-disc conditions, with the noise-disc condition prompting the lowest engagement rating out of the audiovisual conditions.

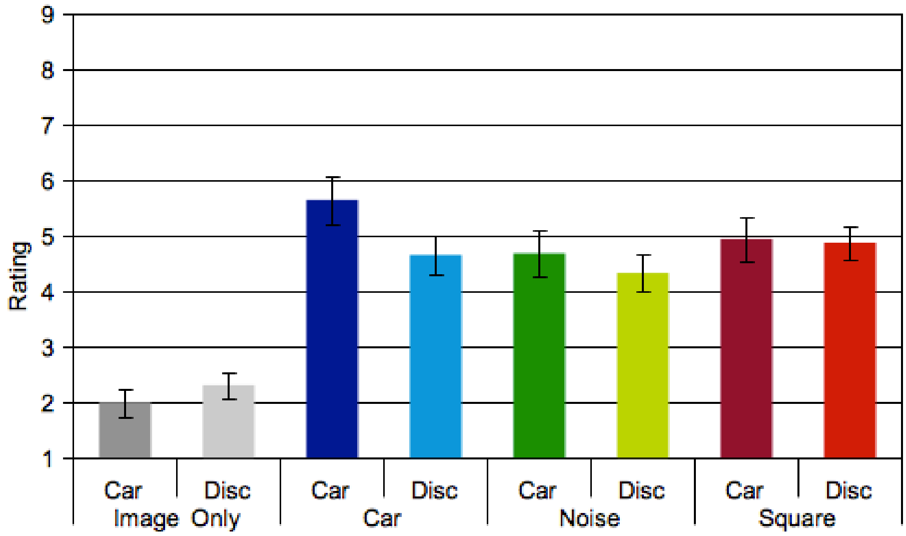

Figure 6.8: Sound Source × Engagement Bar Chart
The engagement rating × sound source for each visual presentation condition (averaged across all of the participants ratings and audio cues) are plotted. Error bars indicate the standard error for each condition. (**Image Only** - **Car:** $M = 2.000$ ($S.E. = .253$), **Disc:** $M = 2.308$ ($S.E. = .237$)); (**Car Sound Source** - **Car:** $M = 5.643$ ($S.E. = .424$), **Disc:** $M = 4.655$ ($S.E. = .349$)); (**Noise Band** - **Car:** $M = 4.685$ ($S.E. = .402$), **Disc:** $M = 4.335$ ($S.E. = .341$)); (**Square Wave** - **Car:** $M = 4.942$ ($S.E. = .390$), **Disc:** $M = 4.872$ ($S.E. = .311$));

A one-way repeated measures ANOVA was conducted with the means and standard errors listed in Appendix Table D.18. Mauchly's test indicated that the assumption of sphericity had been violated $x^2(27) = 128.674, p =< 0.001$, therefore degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity $\epsilon = 0.435$. The results indicate that the sound source had a significant, and strong positive effect on the engagement rating $F(3.044, 36.528) = 20.177, p =< 0.000, r = 0.561$, ($\alpha = 0.05$).

Post-hoc tests with pairwise comparisons were conducted and the descriptives are listed in Appendix Table D.19. In a space saving measure, please refer to the mean difference, significance level, and confidence intervals listed in this table.

To determine if observers engagement ratings for the looming audiovisual stimuli differs when real world sound sources are presented, as opposed to artificial sound sources (hypothesis 3B) we looked at the pairwise comparisons between the conditions which present the same visual stimuli but different sound sources (i.e. car-car × noise-car; car-car × square-car; noise-car × square-car; car-disc × noise-disc; car-disc × square-disc; noise-disc × square-disc).

Whilst none of the pairwise comparisons reached the level of significance, we see that the car-car condition had the greatest engagement rating, followed by the car-square condition, and lastly the car-noise.

For the car moving image presentation, we see that the car-car condition prompted the greatest engagement rating, followed by the square-car condition, and lastly the noise-car condition, whilst for the expanding disc presentation, the square-disc condition prompted the greatest engagement rating, followed by the car-disc condition, and lastly the noise-disc presentation. However, none of the pairwise comparisons met the level of significance, therefore the engagement results are not strong enough to support hypothesis 3B.

To test hypothesis 4, do observers engagement ratings for congruent real world looming stimuli (the car sound presented with the car visual stimuli) differ from artificial congruent looming stimuli (the square wave with the expanding disc image, and the noise band with the expanding disc image), we looked at the pairwise comparisons between the conditions car-car $\times$ noise-disc, and car-car $\times$ square-disc.

The real world car-car condition prompted the greatest engagement rating, followed by the square-disc condition, and lastly the noise-disc condition. However, none of the pairwise comparisons met the level of significance, therefore the engagement results are not strong enough to support hypothesis 4.

6.5 Discussion

In this chapter, we investigated human responses to audiovisual looming scenes to see whether variations to the audio cues and sound source bias perception of, and response to, the approaching object.

For our first hypothesis, we proposed that the addition of audio cues to a looming sound (in audiovisual presentation) would prompt people to perceive the objects time-to-contact to be earlier than the scenes which present sound but no audio cues (1A); and express greater valence, arousal, and engagement ratings than the scenes which present sound but no audio cues (1B). The results showed that the conditions with sound prompted earlier time-to contact, and greater valence, arousal, and engagement ratings than the image only condition, indicating that multi-modal presentations were preferred to the uni-modal presentations). The addition of audio cues to the sound (in audiovisual presentation) of a looming object prompted earlier time-to-contact response times than the sound (no cues) condition, which were significantly earlier for all of the audio cue conditions (except IAD) in both of the car and disc visual presentations, supporting hypothesis 1A.

Regarding the arousal ratings, the significance level was not met for all pairwise comparisons therefore the hypothesis cannot be supported across all audio cues conditions, however the condition with all three cues (Amp + IAD + Ref) was significantly greater than the Sound (no cues) condition. This significant result supports hypothesis 1B, whereby the addition of all three audio cues prompts greater arousal ratings. For the engagement ratings we see that all of the audio cue conditions prompted significantly greater engagement ratings than the Sound (no cues) condition, supporting hypothesis 1B, that the addition of audio cues prompted greater engagement ratings.

The next series of analyses we conducted were to determine if certain audio cues affected the perceived time-to-contact, emotion, and engagement ratings, more than other cues. Our second hypothesis proposed that observers responses to the looming audiovisual stimuli would differ according to the number of, and specific audio cue(s) presented. That observers responses to scenes with single audio cues (amplitude increase, inter-aural differences, and the reflections ratio) will differ, suggesting a hierarchy amongst the individual audio cues (2A). We see that the amplitude increase condition prompted the earliest time-to-contact out of the three conditions, and that the amplitude increase and reflections ratio conditions both prompted significantly earlier time-to-contact response times than the inter-aural differences condition. With such strong results, we conclude that the results support hypothesis 2A, that (across the single audio cue conditions) the amplitude increase and reflections ratio audio cues prompt an earlier time-to-contact than the inter-aural differences.

Whilst hypothesis 2A was not supported by the valence results, the arousal ratings again revealed that the amplitude increase condition prompted the greatest arousal rating, followed by the inter-aural differences, then lastly the reflections ratio. This order was present in both the car and disc audiovisual presentations, however the only pairwise comparison that met the level of significance was the Amp $\times$ Ref condition for the disc presentation. Since most of the comparisons did not reach significance, we acknowledge that the results are not strong enough to support hypothesis 2A across all conditions, however, as the hierarchical order was present in both car and disc presentations, it cannot be completely disqualified. We acknowledge that most of the comparisons did not reach the significance level, therefore the results are not strong enough to support hypothesis 2A outright across all conditions. However, as the order was evident in both car and disc presentations, it cannot be disqualified. Furthermore, as the amplitude increase prompted significantly greater arousal ratings than the reflections ratio condition, it demonstrates that there is a difference in the amount of arousal which the audio cues prompt. As this pairwise comparison reached the significance level, we propose that in regard to the amplitude increase condition, the results support hypothesis 2A, that the amplitude increase is the strongest individual cue, and prompts greater arousal ratings, than other audio cues.

Looking at the engagement ratings, we again see that the amplitude increase condition prompted the greatest engagement rating for both the car and disc audiovisual presentations, which was followed by the Ref then the IAD for the disc presentation, and the IAD then Ref for the car visual presentation. None of the comparisons met the significance level, therefore the engagement results are not strong enough to support hypothesis 2A, however we do see that the amplitude increase condition consistently prompted the greatest engagement rating for single audio cues.

We hypothesised that observers responses to scenes with multiple audio cues (Amp +

IAD, Amp + Ref, IAD + Ref, Amp + IAD + Ref) would differ, suggesting a hierarchy amongst the combinations of, and number of audio cues (2B). Again, the results are not strong enough to support the hypothesis outright across all conditions. In regard to time-to-contact, only one of the pairwise comparisons reached the significance level, therefore the time-to-contact results are not strong enough to support hypothesis 2B. However, looking at the plotted results, there does appear to be an ordering of the different conditions, with some audio cues affecting the perceived time-to-contact more than others, which is evident in both car and disc visual presentations. The condition containing all three audio cues (Amp + IAD + Ref) prompted the earliest time-to-contact response times (for both the car and disc presentations). This was followed by the 2 multiple cue conditions which contained amplitude (Amp + IAD, and Amp + Ref), then lastly the IAD + ref condition. As this order is evident in both the car and disc presentations, a hierarchy cannot be completely ruled out. Once again the conditions containing the amplitude increase as one of the multiple audio cues, all prompted earlier time-to-contact response times than the multiple cue conditions without the amplitude increase variable. It appears that the amplitude increase is a dominant audio cue (in both real world and artificial presentations), for looming objects moving on a frontal midline trajectory.

A similar pattern of results was apparent in the engagement ratings. Once again the condition containing all three audio cues prompted the greatest engagement rating, and the condition that did not contain the amplitude increase as one of the multiple cue variables (i.e. the IAD + Ref condition) prompted the lowest engagement rating (for the multiple cue conditions) across both the car and disc visual presentations. However, the pairwise comparisons did not meet the significance level, therefore the engagement results are not strong enough to support hypothesis 2B.

When comparing the multiple cue versus single cue conditions (2C), we hypothesised that observers responses to scenes with multiple audio cues (Amp + IAD, Amp + Ref, IAD + Ref, Amp + IAD + Ref) will differ, to the scenes with single audio cues (Amp, IAD, Ref). The results showed that the time-to-contact responses differed between the multiple and single cue conditions, with the multiple cue conditions prompting earlier contact time estimates than all of the single cue conditions, except for the amplitude increase condition which prompted an earlier contact time estimate than the IAD + Ref condition. This result shows that the amplitude increase is a dominant cue, and its absence (in the IAD + Ref condition) impacted on the perceived contact time. When additional cues are added to the single cue (i.e. Amp $\times$ Amp + Ref), the multiple cue conditions prompted earlier response times, which again prompted the earliest time-to-contact when a third audio cue was added. The Amp + IAD + Ref condition not only prompted the earliest time-to-contact (than all other cues), but was significantly earlier than the IAD condition (for both car and disc presentations). We also see that the multiple cue Amp + IAD and Amp + Ref conditions also prompted significantly earlier time-to-contact than the IAD condition. Whilst hypothesis 2C can not be supported across all of the multiple cue versus single cue comparisons, we propose that the significant results support the hypothesis for certain cue combinations, that certain stronger cue combinations (ie. the Amp + IAD + Ref multiple cue condition) prompt an earlier time-to-contact than other single cue conditions.

With regard to the valence, arousal, and engagement ratings, the results indicate that once again, the condition presenting all three audio cues (Amp + IAD + Ref) prompted the greatest ratings. We see that for most pairwise comparisons, the addition of a second audio cue prompted greater valence and arousal ratings than the single audio cue condition (i.e. Amp × Amp + IAD; Amp × Amp + Ref; IAD × Amp + IAD; IAD × Ref + IAD; Ref × Amp + Ref; Ref × IAD + Ref). This increase in the rating was again repeated when the third audio cue was added (i.e. Amp × Amp + IAD + Ref; IAD × Amp + IAD + Ref; Ref × Amp + IAD + Ref). which for the IAD × Amp + IAD + Ref and Ref × Amp + IAD + Ref comparisons, the increase in valence and arousal ratings reached the significance level, and the Ref × Amp + IAD + Ref comparison showed a significant increase in engagement rating, which was replicated in both the car and disc visual presentations. As the results are not significant across all pairwise comparisons, hypothesis 2C cannot be outrightly supported across all of the multiple cue versus single cue conditions for the valence, arousal, and engagement ratings, however, we propose that the significant results for certain cue combinations, primarily the three audio cue combination (Amp + IAD + Ref), support hypothesis 2C, and that a hierarchy between the combinations of sound cues exist, that becomes evident with a maximal number of audio cues, prompting greater ratings than single cue conditions.

For our third hypothesis, we proposed that observers responses to the sound source presented would differ, with the real world sound source (car traction) prompting people to perceive the time-to-contact (of the approaching object) to be sooner than the scenes which present an artificial sound source (noise band, square wave) (hypothesis 3A); and express greater valence, arousal, and engagement ratings than the scenes which present an artificial sound source (noise band, square wave) (hypothesis 3B). The results revealed a significant difference in the perceived time-to-contact, with the car-car condition prompting a significantly earlier time-to-contact than the noise-car condition, however there was no significant difference between the car-car × square-car condition. This pattern of results was also evident in the expanding disc visual presentation. As the significance level was not met for all pairwise comparisons, the hypothesis cannot be supported across all conditions, however the significant results for the real world (car) × artificial (noise) condition for both visual presentations, supports hypothesis 3A, whereby the presentation of a real world sound source in an audiovisual presentation prompted an earlier time-to-contact than the presentation of a noise band sound source.

In regard to the valence ratings, it is interesting to note that the image only (car and disc) conditions prompted greater ratings than the square-car and associated square-disc conditions, suggesting that the addition of the square wave to the moving image (no matter which visual stimuli was presented) lowered the valence rating, and that people preferred the visual stimuli without sound, rather than being accompanied by a square wave.

The pairwise comparisons revealed a significant difference in the valence ratings, with the square-car condition prompting a significantly lower ratings than both the car-car condition and the noise-car conditions As the significance level was not met for all pairwise comparisons, the hypothesis cannot be supported across all conditions, however the significant results for the real world sound (car) $\times$ artificial (square wave) condition supports hypothesis 3B, whereby the presentation of a square wave artificial sound source in a audiovisual presentation prompts significantly lower (negative) valence ratings than the presentation of a noise band sound source or a real world car sound source.

The results also revealed a significant difference in the arousal rating, with the car-car condition prompting significantly greater ratings than the noise-car condition, however there was no significant difference between the car-car $\times$ square-car condition. As the significance level was not met for all pairwise comparisons the hypothesis cannot be supported across all conditions, however the significant results for the real world (car-car) $\times$ artificial (noise-car) condition for both visual presentations, supports hypothesis 3B, whereby the presentation of a real world car sound source in an audiovisual presentation prompts a greater arousal rating than the presentation of an artificial noise band sound source.

In regard to the engagement ratings, we see that the conditions with sound all had greater engagement ratings than the image only conditions, suggesting that people found multimodal presentations more engaging than the unimodal presentations. Of the multimodal audiovisual conditions, the conditions which presented the real world car visual presentation prompted greater engagement ratings that the expanding disc visual presentation. We also see that the congruent real world car-car condition had the greatest engagement rating, followed narrowly by the square-car and square-disc conditions, with the noise-disc condition prompting the lowest engagement rating out of the audiovisual conditions. However as none of the pairwise comparisons met the level of significance, therefore the engagement results are not strong enough to support hypothesis 3B.

For our fourth hypothesis, we proposed that observers responses to congruent real world stimuli (ie. the car sound source presented with the moving image of an approaching car) would differ from the presentation of artificial congruent information (i.e. the square wave presented with the expanding disc visual stimuli, and noise band presented with the expanding disc visual stimuli). The results revealed a significant difference in the perceived time-to-contact with the car-car condition prompting a significantly earlier time-to-contact than the noise-disc condition. However there was no significant difference between the car-car $\times$ square-disc condition. As the significance level was not met for both pairwise comparisons, the hypothesis cannot be supported across all conditions, however the significant results for the real world (car-car) × artificial (noise-disc) condition supports hypothesis 4, whereby observers responses to the congruent real world car looming stimuli differed to the artificial noise band with the expanding disc looming stimuli.

Analysis of the valence ratings revealed that the square-disc condition prompted a significantly lower rating than the car-car condition, however the difference in the car-car × noise-disc condition did not meet the significance level. As the significance level was not met for both pairwise comparisons, the hypothesis cannot be supported across all conditions, however the significant results for the real world (car-car) × artificial (square-disc) condition supports hypothesis 4, whereby observers responses to the artificial looming stimuli of a square wave with an expanding disc differed, with significantly lower valence ratings than the presentation of congruent real world looming stimuli of a car sound source presented with the car visual stimuli.

Analysis of the arousal ratings showed that the square-disc condition prompted the greatest arousal rating, narrowly followed by the car-car condition, and lastly the noise-disc condition. However, none of the pairwise comparisons met the level of significance, therefore the results are not strong enough to support hypothesis 4.

The results also revealed that the real world car-car condition prompted the greatest engagement rating, followed by the square-disc condition, and lastly the noise-disc condition, however, none of the pairwise comparisons met the level of significance, therefore the engagement results are not strong enough to support hypothesis 4.

Chapter 7

Conclusions and Future Perspectives

In this thesis, we investigated the presentation of an object moving in depth on an approaching trajectory (looming), focusing on which parameters of sound acted as audio cues for movement in depth. The research investigated the audio cues that had been explored in previous research, namely amplitude increase and inter-aural differences, and introduced a third new audio cue of the 'direct-to-reflections sound energy ratio'. We examined the effect of audio cues on human perception when those cues are presented in combination as multiple audio cues compared to their effect when presented as single audio cues. The question of whether the effectiveness of audio cues differs when they are presented with real-world stimuli compared to their effectiveness when presented with artificial stimuli is also examined. To conclude this thesis we will summarise the research findings and then propose ideas and applications for future research directions

7.1 Research Summary

In chapter 2 - Background we reviewed the three key areas fundamental to understanding and conducting robust research on Auditory Looming. We began by reminding ourselves of the laws of acoustics which describe the propagation of sound and how sound changes when objects move in depth, and the psychoacoustic factors that underpin human auditory perception of an approaching object. We then reviewed psychological studies on auditory looming, highlighting key results and conclusions from previous experiments, whilst examining the experimental design, auditory stimuli, and parameters that may have affected human perception and the overall outcome or wider application of the results. We also explored the application of auditory looming in various industries and highlighted the gaps that exist from the limited publishing of experimental investigations.

We began our experimental studies in chapter 3 - A Feature Analysis Study of the Audio Cues in Film Looming Scenes. Conducting a feature analysis study on the audio cues from 27 film looming scenes enabled us to understand which features might be acting as cues for approaching objects, how the features changed over time, and the degree of their change. The audio features that were analysed included amplitude change, amplitude levels, amplitude slope, audio pan position, spectral centroid and spectral spread, in addition to image motion tracking of the object, comparison of the audio position to the image position, and feature contact time.

To summarise some of the results: the amplitude increased on a linear slope at an average of M = 45.05dB (SD = 15.32) and there was no correlation between the amount of increase and the duration of the sample. The amount of increase in these film samples is greater than the amount of increase used in psychoacoustic auditory looming experiments (which ranged from a 10dB increase (Rosenblum, et al. 1987; Cappe, et. al, 2009) to a 30dB increase (Neuhoff, 2001; Neuhoff & Heckel, 2004). This greater increase in the amplitude level used in the film scenes may contribute to biasing viewers' perception, engagement and surprise levels.

The average spectral centroid frequency was moderately high (M = 1957.8Hz to M = 3444.57Hz) considering that many of the approaching objects are large vehicles, such as cars, motorbikes and spaceships, which could be expected to have lower spectral content. The majority of samples (88.89%) increased the spectral centroid (M =1486.77Hz - almost one octave) as the proximity of the object became closer, which is inconsistent with the doppler shift as an audio cue. Regression analysis of the amount of spectral centroid change per sample duration, showed there was a small correlation for a decrease in the spectral centroid frequency, as the duration of the sample increased, however this is rejected due to the amount of decrease being too small to be perceptually noticeable.

We speculated that the sound effects may be modelling environmental effects (geometric spreading, atmospheric absorption and ground reflection) that ensure a broader spectral content is received by the observer when the object is at a close proximity. However, further analysis of the spectral spread indicated that there was a decrease in spectral spread as the object came closer, disproving our hypothesis. This was an unexpected result considering that it was evident across all of the samples. One explanation for this result is that it is perhaps a signal-to-noise ratio issue related to the analysis, whereby the signal level of the objects sound source is too low at the start of the sample, thereby erroneously indicating a greater spectral spread. As the object (sound source) becomes closer to the observer, the signal level becomes more accurately measurable, therefore providing a more reliable spectral spread.

Overall, the spatial position for both the audio and image objects tended to remain somewhat central, with only 7.41% of image samples having a hard pan to the right side, and none of the audio samples having hard pans to either side. The image and audio positions were generally similar, with 85.19% of samples overlapping in spatial position, although not necessarily at the same time. 14.81% of samples, however, had little (≤ 2 %) to no overlap in position.

This analysis of the hyper-real looming scenes demonstrates that the sound effects have exaggerated key features which act as audio cues for objects moving in depth, more than would be present in real world sounds generated according to the laws of physics, and more than the stimuli used in the psychoacoustic auditory looming studies.

In chapter 4 - Responses to Designed Film Looming Stimuli we describe our second investigation in which we built upon the feature analysis study by conducting a novel psychoacoustic experiment measuring human perception of, and response to, the film looming stimuli analysed in chapter 3. This experiment has provided new information on the human perception of, and response to, auditory(-visual) looming that presents multiple audio cues and complex sound sources, which have been designed for hyper-real scenarios and to generate emotional (valence and arousal) responses in observers. Whilst the study was limited in its capacity to control and vary individual sound parameters due to the use of original audio tracks, it allowed us to gain an insight into people's responses and reactions to ecologically valid real world and hyper-real looming stimuli, which has been absent from the research corpus, but ubiquitous in everyday life.

The results from this study demonstrated that the presentation of sound stimuli (which contained multiple auditory looming cues applied to complex sound sources) prompted observers to significantly underestimate the contact time of an approaching object, compared to looming scenes with no sound stimuli. When the sound stimuli was added to visual looming scenes (the Sound + Image condition) the auditory stimuli continued to bias the observers' perception, prompting them to significantly underestimate the contact time, compared to the scenes with no audio cues. As both of the conditions that presented sound stimuli prompted significantly greater underestimations of the time-to-impact than the Image Only condition, we conclude that the presentation of the sound stimuli that has multiple auditory looming cues applied to a complex sound source prompts people to underestimate the contact time of the approaching object, thereby eliciting a faster response time than the scenes with no sound.

Our research provided insights into the emotional responses to looming stimuli. Our study was novel in that we had our participants comparing the modalities of sensory information for looming stimuli (auditory, visual and auditory-visual). The results showed there was no significant difference in valence or engagement ratings. The findings did reveal that sound stimuli had a significant effect on the arousal ratings, with both of the conditions presenting sound stimuli prompting significantly greater arousal ratings than the condition that did not (the Image Only condition). This strong result added further support to our hypothesis that the presentation of the sound stimuli that have multiple auditory looming cues applied to a complex sound source prompts people to have greater arousal ratings than the scenes with no sound.

We investigated if there were correlations between the valence and arousal (emotion) ratings, and the engagement ratings. The analyses indicated that there were significant large positive correlations between the valence and engagement ratings, and the arousal and engagement ratings, indicating that looming scenes which prompted greater valence and arousal ratings also prompted greater engagement ratings.

The measurement of participants' emotional responses to looming stimuli, and their rating of the scene's engagement quality have been valuable tools, providing a better understanding of human responses to real world and hyper-real stimuli, the emotional impact of the stimuli, and the perceptions and actions generated as a result, that would not be gained from time-to-impact measurements alone. Therefore we recommend the use of these measurements in future looming studies. This not only improves our understanding of human perception, but also provides detailed parameters for perception and the associated responses to those parameters which are applicable to industry for use in the design of audio cues in many virtual environments.

For our third study, chapter 5 - The Effect of Audio Cues and Sound Source Stimuli on the Perception of Approaching Objects, we made a closer inspection of three of the audio cues for movement in depth. We introduced the new cue of 'direct-to-reflections sound energy ratio' which biased people's perceived time-to-contact, prompting significantly earlier response times, and significantly greater arousal and engagement ratings, than the conditions with no audio cues.

We compared the individual audio cues and found that individual cues differ in their capacity to bias perception of an approaching object. The amplitude increase variable was the most dominant cue and the inter-aural differences the weakest for objects moving on a frontal midline trajectory. The dominance of the amplitude increase variable was evident in both the single cue and multiple cue conditions. The conditions containing the amplitude increase variable (Amp + Ref, Amp + IAD, Amp + IAD + Ref) prompted significantly earlier estimates of the time-to-contact, and significantly greater arousal and engagement ratings than the conditions without the amplitude increase variable.

We also considered the complexity of the cues, comparing single versus multiple cues, finding that the presentation of multiple audio cues generally prompted earlier estimates of the time-to-contact, and greater arousal and engagement ratings, than the single audio cues. This result was significantly different for conditions that contained the amplitude increase as one of the multiple audio cues, when compared to single cues that did not contain amplitude increase.

And lastly, we also investigated the sound sources presented, comparing responses to artificial sound sources to real world sound sources. Whilst the results showed that the real world (car traction) sound source prompted earlier estimates of the time-to-contact than the artificial sounds, it did not reach the significance level, therefore it did not support our hypothesis in regard to the estimated time-to-contact. However, for measurements of engagement the real world (car) sound source prompted significantly greater engagement ratings than both of the artificial sound sources. Interestingly for measures of emotion, the artificial square wave prompted significantly lower valence, and significantly greater arousal ratings than the real world (car) recording. Whilst it may be expected that a square wave will prompt more negative valence and greater arousal ratings, this result has implications for the use of the square wave in experimental conditions. It is often argued that a result to a looming artificial sound source will be transferable to real world situations, therefore justifying the use of artificial sounds and the limitation of external validity. However, the results from our study demonstrate that this is not the case, and that emotional responses to artificial stimuli prompted results which are not automatically applicable to real world sounds. Therefore, in regard to the emotion and engagement ratings, the results strongly demonstrated that listeners' responses to real world sounds differ to their responses to artificial sounds. This result reinforces the need for experiments to use ecologically valid parameters, so that the research findings can be effectively applied to real world situations. It also reinforces the need for looming experiments to not just measure time-to-contact, or use this sole measurement as an indication of an audio cue's or sound source's effectiveness, and that additional measurements on emotion and engagement are critical to obtaining a full, accurate understanding of human perception and action.

In our fourth and final study, chapter 6 - Responses to Complex Auditory-Visual Looming, we took the audio cues investigated in chapter 6, and applied them to visual stimuli in order to measure how the audio cues bias human responses to the multimodal auditory-visual presentation.

We compared the individual audio cues and found that the individual cues differ in their capacity to bias auditory-visual perception of an approaching object. For the single cue presentations, the amplitude increase variable again prompted the earliest time-to-contact out of the three audio cue variables. Also, the amplitude increase and reflections ratio both prompted significantly earlier time-to-contact response times than the inter-aural differences. The amplitude increase variable also prompted significantly greater arousal ratings than the reflections ratio and inter-aural differences. Whilst the pairwise comparisons for the multiple cue conditions did not reach the level of significance, we again saw that the conditions containing the amplitude increase variable as one of the multiple audio cues all prompted earlier time-to-contact response times and greater engagement ratings, than the multiple cue conditions without the amplitude increase variable. It appears that the amplitude increase is the most dominant audio cue, in both real world and artificial presentations, for looming objects moving on a frontal midline trajectory.

We investigated the complexity of the cues comparing the presentation of single cues versus multiple cues. We found that the presentation of multiple audio cues, in general had a greater effect on the auditory-visual perception of the approaching object, with the conditions containing all three audio cues (Amp + IAD + Ref) prompting significantly earlier estimates of the time-to-contact, and significantly greater arousal and engagement ratings, than the single cue conditions.

We investigated the sound source presented, comparing responses to artificial sound sources with real world sound sources. The results showed that the real world (car-car) condition prompted significantly earlier estimates of the time-to-contact, than the the artificial noise band (noise-car) condition. This pattern of results was also evident in both the real world (car) visual presentation and the artificial (expanding disc) visual presentation. These strong results support our hypothesis that the presentation of a real world sound source in an audiovisual presentation prompted an earlier time-to-contact than the presentation of a noise band sound source. In regard to the valence ratings, it is interesting to note that the image only (both car and disc) presentations prompted greater valence ratings than the square-car and square-disc conditions, suggesting that the addition of the square wave to the moving image, no matter which visual stimulus was presented, lowered the valence rating, and that people preferred the visual stimuli without sound, rather than visual stimuli accompanied by a square wave. Both of the artificial square wave presentations prompted significantly lower valence ratings than the real world car-car condition. This strong result demonstrates that the presentation of an artificial square wave sound source in an audiovisual presentation prompts significantly lower valence ratings than the presentation of a real world car sound source. The abstract square wave conditions also prompted the greatest (square-disc) and third greatest (square-car) arousal ratings, with the real world (car-car) condition separating the two. Interpretation of this result may suggest that there is no difference between the abstract and real world sound sources, however there was a significant difference between the real world (car-car) and the artificial noise band (noise-car) conditions, indicating that individual artificial sound sources differ in their capacity to bias auditory-visual perception of an approaching object.

The principal contributions from this research and thesis are:

- The provision of new information on the human perception of, and responses to, auditory(-visual) looming that use multiple audio cues and complex sound sources.
- The provision of new information on human perception of, and responses to, ecologically valid stimuli using complex real-world, and hyper-real stimuli. This bridges the gap to the results and conclusions drawn from auditory looming studies that use artificial non-real world stimuli whilst providing a foundation for future experiments to incorporate real world stimuli and parameters into their design.
- The introduction of the sound parameter of 'direct-to-reflections sound energy ratio' as an audio cue for auditory(-visual) looming, and the measurement of human responses, both physical and emotional, to this cue.
- The measurement of human emotional responses (valence, arousal and engage-

ment) to auditory(-visual) looming, thereby increasing the limited information collected on this aspect of human perception and action.

- The development of experimental design and implementation of measurement techniques to evaluate human responses to complex stimuli, therefore strengthening the foundations for more ecologically valid experiments with greater external validity, bridging the gap between experimental looming research in laboratory conditions and real world applications.

7.2 Directions For Future Research

In the process of undertaking the research for this thesis, new research questions and ideas for application of the research arose. While these new questions and ideas were exciting, they had to be put aside in order to allow the experimental groundwork presented in this thesis to be undertaken, in order to provide a solid foundation on which to base these future projects.

Auditory-visual Looming Research

For our research, we presented and analysed objects that had a similar threatening emotive association. Whilst the decision was made to limit any bias that a particular object may have on observers' responses, and since our focus was on the audio cues, it did limit the range of the valence ratings. It would be interesting for future studies to expand the range of the objects, including objects that have positive emotive associations. Whilst a recent study [Tajadura-Jiménez et al., 2010] has conducted an initial investigation using a broader range of objects with positive and negative associations, the presentation of the visual stimuli is a still a target image, not an actual moving image, that is, a film sequence. Obtaining results from a study presenting positively and negatively associated moving objects would provide insight into human responses to a broader range of looming objects. Information could be obtained about how the audio cues may differ in their effect on the perceived time-to-contact and any under / over-estimation, valence, arousal and engagement ratings in these situations.

Another exciting avenue of research would be investigating auditory-visual looming in the context of computer games. The design of sound effects is critical to engaging players with the virtual world. The player's capacity to successfully interact with and respond to approaching objects, especially the sporting or warfare games reliant on quick decisions of fight or flight, may affect their continued survival in the virtual worlds, and is dependent on the individuals capacity to accurately interpret depth and movement cues. This is particularly important for interactive gaming systems such as the Xbox Kinnect and Playstation 4, whereby players no longer use controllers to manipulate the avatars' actions, but rather it is their physical actions that are manipulating the avatars actions.

Investigating the audio cues and sound source parameters used in our experiments in a gaming environment provides information on how the audio cues affect human perception and action in another virtual environment and industry that actively uses looming stimuli. Comparisons can then be made between the design of the audio cues used, and presented, in the gaming stimuli versus those of the film stimuli, giving greater insight into the hyper-real stimuli used, and regularly encountered, in everyday life. Further, many of the current looming studies capture human responses through the striking of a key on the keyboard, which in the pursuit of ecological validity, can be rather limiting. Using an interactive gaming console, enables the measurement of a human's physical reaction to an approaching object, by attempting to intercept the target object, or by side stepping to avoid the target object. This provides results based on physical human reactions, increasing the external validity and application of the findings, and demonstrating how humans respond in real world situations.

Further Statistical Analyses on the Collected Data

Further statistical analyses are to be performed on the data collected from the experiments we've already conducted.

In the analyses conducted thus far, we averaged the data across all of the participants. However, categorising the data according to various subsamples, such as gender (with male versus female participants), age, visual and auditory acuity (those participants who correct their vision with glasses or contact lenses versus those who do not, and those with greater visual or auditory acuity versus those which lesser acuity), may reveal that responses (the perceived time-to-contact, emotion, and engagement ratings) to the looming stimuli differ according to these subsamples.

Correlational analyses are also to be conducted on the results from Experiment 1 (the feature analysis study) with Experiment 2 (the perceptual study on the film looming stimuli), to investigate any relationships between the individual looming scenes (i.e. the type of object, audio cues, and duration of the scene) and the perceptual response to that particular looming scene. These analyses may reveal relationships between the time-to-contact, emotion and engagement ratings, with the particular looming scene, and its emotive association or audio cues presented. These can be further categorised into subsamples according to the type of object presented, for example, the responses to vehicles versus the responses to animals.

In regards to the audio cues, further correlational analyses are to be conducted on the data investigating any relationships between the film looming scenes with particular audio cue features (such as the scenes with a greater magnitude of amplitude increase, the slope of the increase, a broader spectral spread, the slope and an increase / decrease in the spectral centroid, spatial panning) and perceptual responses to the particular looming scenes. This may reveal any effect that the magnitude of the audio cue, and the slope of the change, has on human perception of an approaching object, and further how these cues can be manipulated for use in software applications, film, and gaming.

Subcategorisation of the data according to the emotional association of the presented object, and the resulting time-contact, emotion, and engagement ratings, would also yield information as to how the cues operate in threatening and non-threatening scenarios.

Further Experiments

To overcome a number of methodological limitations that were encountered in our studies, further experiments to be conducted include investigating stimuli of different durations to determine if the presentation of longer looming scenes, provides more information about the object's approach, velocity, and perceived threat, which in turn enables the observer to more accurately judge the contact time.

Presenting the object as moving on different trajectories and angles of approach will also provide more information about the audio cue's effectiveness, and hierarchy of the cues. For example, the inter-aural differences audio cue had the least effect on human perception in our studies which presented objects moving on a frontal midline approach. However, if the object approached from a more oblique angle, the inter-aural differences would be greater (reaching a maximal difference when the object is passing parallel to the observer), therefore the cue may be stronger, and more reliable for human perception.

Investigating different magnitudes and slopes of increase would also provide further insight to the cues effectiveness. Having a greater magnitude may suggest that the object is approaching at a faster rate, and prompt observers perceive the contact time to be sooner, and express greater valence and arousal, than lesser magnitudes.

Expanding the number and combinations of audio cues would also provide greater information about human looming perception. For example, other audio cues could include the doppler shift, head-related transfer functions (HRTF's), spectral components such as the spectral centroid, spectral spread, and spectral scattering of high frequencies, in addition to temporal (rhythmic) sounds which would allow for the repeated articulation and decay of these audio cues, rather than a continual sound source as used in our experiments.

A number of experimental limitations arose in our studies due to the film stimuli limitations.

The feature analysis and psychoacoustic studies of the film looming scenes allowed the gathering of information about human looming perception and response, to hyperreal scenes. Whilst novel, and new studies, they were nonetheless preliminary studies to gather general information about the stimuli used in the film industry. Further experiments could build upon our studies, using a much larger database of film looming scenes, whereby analyses could be conducted according to sound designer, year, object type, emotional association with the object and overall looming scene, duration, and of course audio cues used, would provide more specific and robust information on human perception to the looming cues.

In regards to the software, experimental limitations that also arose due to intrinsic software limitations, which could be developed for future experiments include, modelling other architectural spaces and investigating how the resulting reflections affect on looming perception, as compared to other spaces with different angles of reflections. Further control of the reflections, such as the manipulation of the direct-sound to reflected-sound ratio, spectral content, spectral scattering, and their change over time would also provide valuable information about the effectiveness of this audio cue.

Control of the visual looming cues, such as brightness, luminescence, colour, the object area expansion over time, the presentation of a 2-dimensional object versus 3-dimensional object, and the presentation of a computer generated object versus a real-world object, would also provide information about the impact of visual cues on auditory-visual looming perception.

Comparing congruent visual and auditory information (for example the rate of change, area, and magnitude of change) versus incongruent visual and auditory information (i.e. the presentation of auditory cues increasing at a greater magnitude and rate of change, than the visual cues) would yield information about which modality (auditory, visual, or a combination of auditory and visual) are dominating human judgement and action in looming scenarios.

Investigations into human adaptation to looming scenes would provide information on learning biases to the stimuli (the audio cues, stimuli duration, and sound source) and if repeated exposure to the stimlui prompts people to develop greater accuracy of the perceived contact time (and less underestimation), or if they will continue to err on the side of caution (and self preservation) and continue to underestimate the time-to-contact.

One interesting line of research to pursue, is that of emotive association to the approaching object, its relation to the audio cues presentation, and the observers response. Whilst our studies presented a number of different approaching objects, we averaged human responses across the object type.

A series of experiments controlling for the type of object (positive versus negative, with further subcategorisations of threat, for example approaching vehicles, dangerous animals, weapons), and how the emotive association with these objects and the level of threat, affects human perception and response to the approaching object. Further investigations about how the audio cues function when presented with objects that have these different emotive associations, and at different magnitudes, may also yield information about the cue's importance and position in the hierarchy of cues, and if certain cues have greater impact with certain cues or emotive association. Further building upon this theme, would be to also investigate any relations between the size of the object, area expansion over time, and velocity, with the audio cues, and the resulting affect on human perception.

Software Application

The results from this thesis has provided information on how the parameters of sound act as audio cues for movement in depth, and the effect that changes to these audio cues have on people's perception of, and response to, the approaching object. This information is instantiated in a software toolkit with the aim of building a software tool that presents a looming scene from a film or game. This software tool can be used to generate or reprocess the parameters of sound, according to experimental results, so that the sound is more representative of the visual stimuli, or to where the biasing and exaggeration of perception may lead to more immersive experiences. It will also include a library of stimuli envelope algorithms that determine various sound parameters, such as amplitude, spectral components, spatialisation and reverb, for looming scenarios. These algorithms can then be used to manipulate the viewers experience and response to the approaching object, depending on the users intentions of the experience, emotion, and the hyperreality they wish to generate. Further user evaluation studies will be run on the final software package to determine qualitative factors such as user perception and the effectiveness of the envelope algorithms.

It is intended that the software will have applications not only in commercial use, such as in films, gaming, flight or driving simulators, but also for psychoacoustic experiments and neural studies that will allow the identification of areas of the brain which are activated when presented with complex audiovisual (looming) cues.

Appendices

Appendix A

Chapter 3 Experiment 1

#	Title	Year	Chapter	Time (min : sec)	Object
1	The Matrix	1999	1	1:22 - 1:25	Flash light
2	Return of the Jedi	1983	3	0:20 - 0:24	Vehicle (Spaceship)
3	Revenge of the Sith	2005	31	3:08 - 3:09	Vehicle (Spaceship)
4	X-men	2006	15	0:35 - 0:36	Weapon (blade)
5	The Day After Tomorrow	2004	12	2:29 - 2:33	Vehicle (Helicopter)
6	King Arthur	2004	7	10:46 - 10:48	Weapon (Arrow)
7	Sherlock Holmes	2009	22	4:36 - 4:38	Bird
8	Van Helsing	2004	17	1:52 - 1:54	Trapese
9	I Am Legend	2007	17	0:00 - 0:03	Vehicle (Car)
10	Troy	2007	27	2:22 - 2:24	Weapon (Fire ball)
11	Beowulf	2007	2	4:03 - 4:05	Weapon (Axe)
12	The Bourne Identity	2002	12	2:10 - 2:12	Vehicle (Motorbike)
13	Charlie & the Chocolate Factory	2005	15	1:24 - 1:26	Mosquito
14	Mr and Mrs Smith	2005	20	0:40 - 0:44	Vehicle (Car)
15	Sin City	2005	18	1:06 - 1:07	Weapon (Blade)
16	28 Days Later	2002	11	0:01 - 0:04	Vehicle (Car)
17	Gattaca	1997	21	2:39 - 2:40	Vehicle (Car)
18	Alice in Wonderland	2010	15	0:19 - 0:20	Golfball
19	Avatar	2009	22	1:42 - 1:45	Weapon (Bomb)
20	Clash of the Titans	2010	13	4:11 - 4:13	Fire
21	Despicable Me	2010	18	2:23 - 2:24	Vehicle (Spaceship)
22	Kill Bill vol2	2004	6	0:03 - 0:06	Vehicle (Car)
23	Mission Impossible 3	2006	4	1:06 - 1:08	Vehicle (Helicopter)
24	Yogi Bear	2010	1	1:25 - 1:27	Trapese
25	Final Destination	2009	15	0:06 - 0:07	Golfball
26	Salt	2010	9	3:13 - 3:14	Vehicle (Motorbike)
27	Saving private ryan	1998	19	3:17 - 3:21	Vehicle (Aeroplane)

Table A.1: List of Film Scenes Analysed

A list of the looming scenes that were used in the feature analysis study, with the year, chapter, time, and object type.

#	Title	Decibel Level			Duration of Measurement (sec)	Velocity		Distance Travelled (meters)
		Start	Peak	Total Incr.		m/s	kph	
1	The Matrix	-90.63	-49.31	41.32	1.069	1.09	3.92	1.16
2	Return of the Jedi	-112.6	-61.08	51.52	1.243	3.03	10.92	3.77
3	Revenge of the Sith	-112.8	-68.78	44.02	0.940	1.69	6.08	1.59
4	X-men	-91.1	-34.04	57.06	0.395	18.06	65.02	7.13
5	The Day After Tomorrow	-70.86	-50.43	20.43	2.961	0.03	0.12	0.10
6	King Arthur	-96.15	-60.46	35.69	1.011	0.60	2.17	0.61
7	Sherlock Holmes	-101.3	-55.61	45.69	0.813	2.37	8.53	1.93
8	Van Helsing	-82.79	-52.35	30.44	1.301	0.26	0.92	0.33
9	I Am Legend	-143.7	-54.2	89.50	2.543	117.44	422.77	298.59
10	Troy	-90.03	-54.67	35.36	2.206	0.27	0.96	0.59
11	Beowulf	-99.64	-37.67	61.97	1.440	8.72	31.38	12.55
12	The Bourne Identity	-86.75	-50.06	36.69	0.604	1.13	4.07	0.68
13	Charlie & the Chocolate Factory	-94.49	-59.5	34.99	0.348	1.61	5.81	0.56
14	Mr and Mrs Smith	-97.4	-61.51	35.89	0.836	0.75	2.68	0.62
15	Sin City	-83.23	-20.6	62.63	0.279	48.58	174.90	13.54
16	28 Days Later	-108.4	-61.71	46.69	1.498	1.44	5.19	2.16
17	Gattaca	-94.65	-35.92	58.73	0.685	12.62	45.42	8.64
18	Alice in Wonderland	-69.02	-40.77	28.25	0.313	0.83	2.99	0.26
19	Avatar	-75.89	-26.24	49.65	0.302	10.06	36.23	3.04
20	Clash of the Titans	-101.5	-65.69	35.81	1.347	0.46	1.65	0.62
21	Despicable Me	-100.5	-69.87	30.63	0.522	0.65	2.34	0.34
22	Kill Bill vol2	-92.24	-58.27	33.97	1.893	0.26	0.95	0.50
23	Mission Impossible 3	-85.57	-50.26	35.31	0.894	0.65	2.35	0.58
24	Yogi Bear	-105.9	-55.05	50.85	1.510	2.31	8.32	3.49
25	Final Destination	-77.44	-45.2	32.24	0.627	0.65	2.35	0.41
26	Salt	-104.7	-43.45	61.25	0.453	25.51	91.84	11.55
27	Saving private ryan	-127.4	-57.63	69.77	2.706	11.38	40.98	30.80

Table A.2: Amplitude Levels Per Scene

The minimum and maximum amplitude levels measured for each scene (left and right channels combined); the total amplitude increase; the duration of the measurement; the distance travelled in meters per second (calculated using the Inverse square law, according to amplitude increase and a peak contact at a distance of 1cm (from ear), and distance / velocity in meters per second, kilometres per hour.

#	Title	Duration of Measurement (sec)	Linear Slope m value	dB Incr. per 100ms
1	The Matrix	1.069	27	2.7
2	Return of the Jedi	1.243	25	2.5
3	Revenge of the Sith	0.940	34	3.4
4	X-men	0.395	162.18	16.2
5	The Day After Tomorrow	2.961	2.8	0.28
6	King Arthur	1.011	26	2.6
7	Sherlock Holmes	0.813	49	4.9
8	Van Helsing	1.301	17	1.7
9	I Am Legend	2.543	28	2.8
10	Troy	2.206	4.1	0.41
11	Beowulf	1.440	17	1.7
12	The Bourne Identity	0.604	27	2.7
13	Charlie & the Chocolate Factory	0.348	88	8.8
14	Mr and Mrs Smith	0.836	19	1.9
15	Sin City	0.279	181.57	18.16
16	28 Days Later	1.498	21	2.1
17	Gattaca	0.685	75	7.5
18	Alice in Wonderland	0.313	68	6.8
19	Avatar	0.302	124.84	12.48
20	Clash of the Titans	1.347	19	1.9
21	Despicable Me	0.522	30	3.0
22	Kill Bill vol2	1.893	9.2	0.92
23	Mission Impossible 3	0.894	28	2.8
24	Yogi Bear	1.510	18	1.8
25	Final Destination	0.627	43	4.3
26	Salt	0.453	116.75	11.68
27	Saving private ryan	2.706	26	2.6

Table A.3: Amplitude Envelope Slope Per Scene

The duration of the looming scene; the slopes linear equation m value; and the linear slope increase per 100ms ($m \times 0.1$).

#	Film Scene	Minimum Frequency	Maximum Frequency	Magnitude of Change	Duration of Change (ms)	Spectral Spread Magnitude of Change
1	The Matrix	1373.0	6177.0	4804.0	0.670	944.7
2	Return of the Jedi	477.0	3157.0	2680.0	1.665	446.9
3	Revenge of the Sith	896.2	3223.0	2326.8	0.265	1255.1
4	X-men	1280.0	5808.0	4528.0	0.150	1717.4
5	The Day After Tomorrow	1445.0	5022.0	3577.0	1.570	434.2
6	King Arthur	295	509.3	214.7	0.670	1161.9
7	Sherlock Holmes	1931.0	5225.0	3294.0	0.680	893.7
8	Van Helsing	934.9	4096.0	3161.1	0.745	1844.1
9	I Am Legend	1151.0	4474.0	3323.0	0.190	505.2
10	Troy	1196.0	5779.0	4583.0	0.970	546.7
11	Beowulf	1316.0	8203.0	6887.0	0.090	597.5
12	The Bourne Identity	1607.0	4474.0	2867.0	0.665	479.2
13	Charlie & the Chocolate Factory	1020.0	4410.0	3390.0	0.060	644.9
14	Mr and Mrs Smith	1234.0	2211.0	977.0	0.895	447.4
15	Sin City	2641.0	10370.0	7729.0	0.140	521.0
16	28 Days Later	855.6	2557.0	1701.4	1.210	517.8
17	Gattaca	3789.0	6981.0	3192.0	0.765	2267.3
18	Alice in Wonderland	1600.0	5515.0	3915.0	0.115	429.8
19	Avatar	1704.0	9789.0	8085.0	0.315	316.4
20	Clash of the Titans	376.5	3626.0	3249.5	0.565	546.2
21	Despicable Me	665.7	2382.0	1716.3	0.210	447.1
22	Kill Bill vol2	1324.0	3521.0	2197.0	1.680	765.6
23	Mission Impossible 3	685.6	3354.0	2668.4	0.755	250.5
24	Yogi Bear	1193.0	3352.0	2159.0	1.420	331.0
25	Final Destination	1758.0	4836.0	3078.0	0.110	529.9
26	Salt	1051.0	3360.0	2309.0	0.481	1367.5
27	Saving private ryan	1075.0	3430.0	2355.0	2.230	898.8

Table A.4: Spectral Components Per Scene

The Spectral Centroid's minimum and maximum of the frequencies for each scene, magnitude of change, and duration of measurement are listed. Also listed is the Spectral Spread's magnitude of change.

Appendix B

Chapter 4 Experiment 2

#	Title	Presentation Conditions			Object
1	The Matrix	Sound Only	Image Only	Sound + Image	Flash light
2	Return of the Jedi	Sound Only	Image Only	Sound + Image	Vehicle (Spaceship)
3	Revenge of the Sith	Sound Only	Image Only	Sound + Image	Vehicle (Spaceship)
4	X-men	Sound Only	Image Only	Sound + Image	Weapon (blade)
5	The Day After Tomorrow	Sound Only	Image Only	Sound + Image	Vehicle (Helicopter)
6	King Arthur	Sound Only	Image Only	Sound + Image	Weapon (Arrow)
7	Sherlock Holmes	Sound Only	Image Only	Sound + Image	Bird
8	Van Helsing	Sound Only	Image Only	Sound + Image	Trapese
9	I Am Legend	Sound Only	Image Only	Sound + Image	Vehicle (Car)
10	Troy	Sound Only	Image Only	Sound + Image	Weapon (Fire ball)
11	Beowulf	Sound Only	Image Only	Sound + Image	Weapon (Axe)
12	The Bourne Identity	Sound Only	Image Only	Sound + Image	Vehicle (Motorbike)
13	Charlie & the Chocolate Factory	Sound Only	Image Only	Sound + Image	Mosquito
14	Mr and Mrs Smith	Sound Only	Image Only	Sound + Image	Vehicle (Car)
15	Sin City	Sound Only	Image Only	Sound + Image	Weapon (Blade)
16	28 Days Later	Sound Only	Image Only	Sound + Image	Vehicle (Car)
17	Gattaca	Sound Only	Image Only	Sound + Image	Vehicle (Car)
18	Alice in Wonderland	Sound Only	Image Only	Sound + Image	Golfball
19	Avatar	Sound Only	Image Only	Sound + Image	Weapon (Bomb)
20	Clash of the Titans	Sound Only	Image Only	Sound + Image	Fire
21	Despicable Me	Sound Only	Image Only	Sound + Image	Vehicle (Spaceship)
22	Kill Bill vol2	Sound Only	Image Only	Sound + Image	Vehicle (Car)
23	Mission Impossible 3	Sound Only	Image Only	Sound + Image	Vehicle (Helicopter)
24	Yogi Bear	Sound Only	Image Only	Sound + Image	Trapese
25	Final Destination	Sound Only	Image Only	Sound + Image	Golfball
26	Salt	Sound Only	Image Only	Sound + Image	Vehicle (Motorbike)
27	Saving private Ryan	Sound Only	Image Only	Sound + Image	Vehicle (Aeroplane)

Table B.1: List of Experiment Conditions
A list of the looming scenes that were used in this experiment (and the previous feature analysis study), with the presentation conditions, and object type.

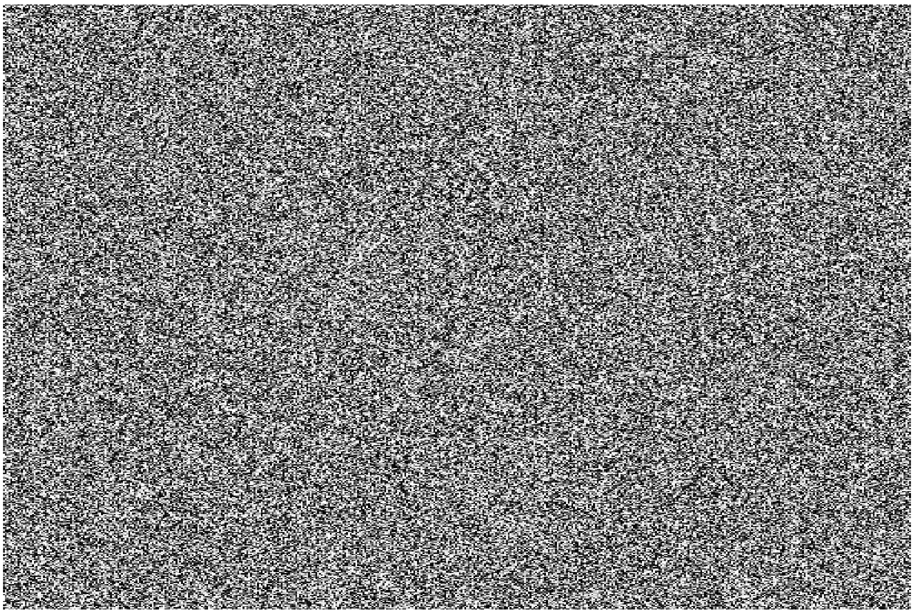

Figure B.1: Experiment 'Visual White Noise' Image Displayed Between Trials

This graphic displaying visual white noise was presented during the 4 second break between each trial in the experiment. It was chosen, over a simple black screen (which was displayed during the sound only trials) as an indicator of the break between trials.

Condition	N	Mean	Std. Dev.	Std. Error	95% Confidence Interval for mean Lower	Upper	Min	Max
Sound	26	-598.88	430.85	84.50	-772.90	-424.85	-1569.49	-50.50
Image	26	-384.05	309.89	60.78	-509.22	-258.88	-1211.69	91.41
Sound + Image	26	-540.54	315.43	61.86	-667.94	-413.14	-1266.22	85.34

Table B.2: Descriptive Statistics: Presentation × Time-to-Impact

The descriptives results are tabled for the Presentation × Time-to-Impact, averaged across all of the participants. The columns are labeled as condition number; condition name; number of trials; mean; standard error; and 95% confidence intervals for the mean.

Condition Pair	Mean Difference	Std. Error	Sig.	95% Confidence Interval Lower	Upper
Sound × Sound + Image	-58.34	71.974	1.000	-243.02	126.35
Sound × Image	-214.83*	77.768	0.032*	-414.38	-15.28
Sound + Image × Sound	58.34	71.974	1.000	-126.35	243.02
Sound + Image × Image	-156.49*	41.285	0.003*	-262.43	-50.55
Image × Sound	214.83*	77.768	0.032*	15.28	414.38
Image × Sound + Image	156.49*	41.285	0.003*	50.55	262.43

Table B.3: Pairwise Comparisons: Presentation × Time-to-Impact

The pairwise comparisons of Presentation × Time-to-Impact. The * indicates the conditions where the mean difference is significant at $\alpha = 0.05$. A Bonferroni adjustment was applied to correct for a possible increase in type 1 errors associated with multiple comparisons.

Condition	N	Mean	Std. Dev.	Std. Error	95% Confidence Interval for mean Lower	Upper	Min	Max
VALENCE								
Sound	26	6.48	0.62	0.12	6.23	6.73	3.35	9.85
Image	26	6.26	0.86	0.17	5.91	6.60	3.23	9.46
Sound + Image	26	7.60	0.73	0.14	7.31	7.90	3.88	11.00
AROUSAL								
Sound	26	7.68	1.28	0.25	7.16	8.19	4.50	10.69
Image	26	6.72	1.36	0.27	6.17	7.27	3.12	10.73
Sound + Image	26	8.72	1.24	0.24	8.22	9.22	5.40	11.73

Table B.4: Descriptive Statistics: Presentation × Valence / Arousal
The descriptives results are tabled for the Presentation × Valence / Arousal, averaged across all of the participants. The columns are labeled as condition number; condition name; number of trials; mean; standard error; and 95% confidence intervals for the mean.

Condition Pair	Mean Difference	Std. Error	Sig.	95% Confidence Interval Lower	Upper
VALENCE					
Sound × Sound + Image	-1.124*	0.165	0.000*	-1.548	-0.700
Sound × Image	0.223	0.172	0.622	-0.219	0.664
Sound + Image × Sound	1.124*	0.165	0.000*	0.700	1.548
Sound + Image × Image	1.347*	0.132	0.000*	1.008	1.686
Image × Sound	-0.223	0.172	0.622	-0.664	0.219
Image × Sound + Image	-1.347*	0.132	0.000*	-1.686	-1.008
AROUSAL					
Sound × Sound + Image	-1.043*	0.134	0.000*	-1.386	-0.700
Sound × Image	0.957*	0.227	0.001*	0.374	1.540
Sound + Image × Sound	1.043*	0.134	0.000*	0.700	1.386
Sound + Image × Image	2.000*	0.168	0.000*	1.570	2.430
Image × Sound	-0.957*	0.227	0.001*	-1.540	-0.370
Image × Sound + Image	-2.000*	0.168	0.000*	-2.430	-1.570

Table B.5: Pairwise Comparisons: Presentation × Valence / Arousal
The pairwise comparisons of Presentation × Valence / Arousal. The * indicates the conditions where the mean difference is significant at $\alpha = 0.05$. A Bonferroni adjustment has been applied to arousal, no adjustment was needed for Valence. The * indicates the conditions where the mean difference is $\alpha = 0.05$.

Condition	N	Mean	Std. Dev.	Std. Error	95% Confidence Interval for mean Lower	Upper	Min	Max
Sound	26	4.91	0.83	0.162	4.576	5.245	3.17	6.33
Image	26	5.24	0.86	0.169	4.892	5.587	3.00	6.40
Sound + Image	26	6.39	0.84	0.164	6.047	6.723	3.33	8.00

Table B.6: Descriptive Statistics: Presentation × Engagement
The descriptives results are tabled for the Presentation × Engagement, averaged across all of the participants. The columns are labeled as condition number; condition name; number of trials; mean; standard error; and 95% confidence intervals for the mean.

Condition Pair	Mean Difference	Std. Error	Sig.	95% Confidence Interval	
				Lower	Upper
Sound × Sound + Image	-1.474*	0.141	0.000*	-1.836	-1.113
Sound × Image	-0.329	0.214	0.411	-0.879	0.221
Sound + Image × Sound	1.474*	0.141	0.000*	1.113	1.836
Sound + Image × Image	1.145*	0.155	0.000*	0.747	1.543
Image × Sound	0.329	0.214	0.411	-0.221	0.879
Image × Sound + Image	-1.145*	0.155	0.000*	-1.543	-0.747

Table B.7: Pairwise Comparisons: Presentation × Engagement

The pairwise comparisons of Presentation × Engagement. The * indicates the conditions where the mean difference is significant at $\alpha = 0.05$. A Bonferroni adjustment was applied to correct for a possible increase in type 1 errors associated with multiple comparisons.

Appendix C

Chapter 5 Experiment 3

#	Sound Source	Audio Cue	Abbreviation	# Audio Cue Variables	Amplitude Level
1	Car Recording	None - Control	Ctrl	0	-3
2	Car Recording	None - Control	Ctrl	0	-18
3	Square Wave	None - Control	Ctrl	0	-3
4	Square Wave	None - Control	Ctrl	0	-18
5	Noise Band	None - Control	Ctrl	0	-3
6	Noise Band	None - Control	Ctrl	0	-18
7	Car Recording	Amplitude Increase	Amp	1	-18 to -3
8	Car Recording	Inter-aural Differences (binaural)	IAD	1	-3
9	Car Recording	Inter-aural Differences (binaural)	IAD	1	-18
10	Car Recording	Reflections	Ref	1	-3
11	Car Recording	Reflections	Ref	1	-18
12	Square Wave	Amplitude Increase	Amp	1	-18 to -3
13	Square Wave	Inter-aural Differences (binaural)	IAD	1	-3
14	Square Wave	Inter-aural Differences (binaural)	IAD	1	-18
15	Square Wave	Reflections	Ref	1	-3
16	Square Wave	Reflections	Ref	1	-18
17	Noise Band	Amplitude Increase	Amp	1	-18 to -3
18	Noise Band	Inter-aural Differences (binaural)	IAD	1	-3
19	Noise Band	Inter-aural Differences (binaural)	IAD	1	-18
20	Noise Band	Reflections	Ref	1	-3
21	Noise Band	Reflections	Ref	1	-18
22	Car Recording	Amplitude Increase + Inter-aural Differences	Amp + IAD	2	-18 to -3
23	Car Recording	Amplitude Increase + Reflections	Amp + Ref	2	-18 to -3
24	Car Recording	Inter-aural Differences + Reflections	IAD + Ref	2	-3
25	Car Recording	Inter-aural Differences + Reflections	IAD + Ref	2	-18
26	Square Wave	Amplitude Increase + Inter-aural Differences	Amp + IAD	2	-18 to -3
27	Square Wave	Amplitude Increase + Reflections	Amp + IAD	2	-18 to -3
28	Square Wave	Inter-aural Differences + Reflections	IAD + Ref	2	-3
29	Square Wave	Inter-aural Differences + Reflections	IAD + Ref	2	-18
30	Noise Band	Amplitude Increase + Inter-aural Differences	Amp + IAD	2	-18 to -3
31	Noise Band	Amplitude Increase + Reflections	Amp + IAD	2	-18 to -3
32	Noise Band	Inter-aural Differences + Reflections	IAD + Ref	2	-3
33	Noise Band	Inter-aural Differences + Reflections	IAD + Ref	2	-18
34	Car Recording	Amplitude Increase + Inter-aural Differences + Reflections	Amp + IAD + Ref	3	-18 to -3
35	Square Wave	Amplitude Increase + Inter-aural Differences + Reflections	Amp + IAD + Ref	3	-18 to -3
36	Noise Band	Amplitude Increase + Inter-aural Differences + Reflections	Amp + IAD + Ref	3	-18 to -3

Table C.1: List of Experiment Conditions

List of the trials and conditions that were used in the experiment. Listed are the trial number, sound source, audio cue, Number of audio cues (control vs trial; single versus multiple), and amplitude level.

Condition #	Name	N	Mean	Std. Error	95% Confidence Interval Lower	Upper
1	Sound only (Ctrl)	36	984.716	147.867	686.711	1282.722
Single Audio Cues:						
2	Amp	36	-101.225	78.114	-258.653	56.204
3	IAD	36	724.960	84.245	555.175	894.744
4	Ref	36	610.162	106.433	395.661	824.663
Multiple Audio Cues:						
5	Amp + IAD	36	-83.445	80.818	-246.322	79.433
6	Amp + Ref	36	-269.315	102.877	-476.650	-61.980
7	IAD + Ref	36	359.633	71.731	215.068	504.198
8	Amp + IAD + Ref	36	-272.873	74.194	-422.401	-123.345

Table C.2: Descriptive Statistics: Audio Cues × Time-to-Contact

The descriptives results are tabled for the time-to-contact × audio cue condition, averaged across all of the sound sources (and participants). The columns are labeled as condition number; condition name; number of trials; mean; standard error; and 95% confidence intervals for the mean.

Condition Pair	Mean Difference	Std. Error	Sig.	95% Confidence Interval Lower	Upper
Sound Only (No Audio Cues) × Single Audio Cues:					
Sound Only × Amp	1085.941*	138.640	0.000*	624.849	1547.033
Sound Only × IAD	259.757	138.534	1.000	-200.980	720.494
Sound Only × Ref	374.554	114.053	0.056	-4.764	753.873
Sound Only (No Audio Cues) × Multiple Audio Cues:					
Sound Only × Amp + IAD	1068.161 *	161.213	0.000 *	531.998	1604.324
Sound Only × Amp + Ref	1254.032 *	165.680	0.000 *	703.011	1805.052
Sound Only × IAD + Ref	625.083 *	135.667	0.001 *	173.880	1076.287
Sound Only × Amp + IAD + Ref	1257.589 *	168.915	0.000 *	695.811	1819.368
Single × Single Audio Cues:					
Amp × IAD	-826.184 *	101.416	0.000 *	-1163.476	-488.892
Amp × Ref	-711.387 *	120.738	0.000 *	-1112.938	-309.836
IAD × Ref	114.798	110.057	1.000	-251.233	480.828
Single × Multiple Audio Cues:					
Amp × Amp + IAD	-17.780	96.209	1.000	-337.753	302.193
Amp × Amp + Ref	168.090	99.040	1.000	-161.298	497.479
Amp × IAD + Ref	-460.858*	92.654	0.000	-769.008	-152.707
IAD × Amp + IAD	808.404*	104.050	.000	462.353	1154.455
IAD × Amp + Ref	994.275*	92.472	.000	686.729	1301.820
IAD × IAD + Ref	365.326	111.858	.059	-6.693	737.346
Ref × Amp + IAD	693.607*	119.585	.000	295.891	1091.323
Ref × Amp + Ref	879.477*	132.452	.000	438.965	1319.989
Ref × IAD + Ref	250.529	97.305	.377	-73.090	574.14
Amp × Amp + IAD + Ref	171.648	99.359	1.000	-158.801	502.097
IAD × Amp + IAD + Ref	997.832*	111.706	.000	626.318	1369.347
Ref × Amp + IAD + Ref	883.035*	128.961	.000	454.134	1311.935
Multiple × Multiple Audio Cues:					
Amp + IAD × Amp + Ref	185.870	131.349	1.000	-250.970	622.711
Amp + IAD × IAD + Ref	-443.078*	107.581	.005	-800.873	-85.283
Amp + Ref × IAD + Ref	-628.948*	124.485	.000	-1042.963	-214.934
Amp + IAD × Amp + IAD + Ref	189.428	89.720	1.000	-108.963	487.819
Amp + Ref × Amp + IAD + Ref	3.558	119.086	1.000	-392.499	399.614
IAD + Ref × Amp + IAD + Ref	632.506*	95.940	.000	313.428	951.584

Table C.3: Pairwise Comparisons: Audio Cues × Time-to-Contact

The pairwise comparisons of Audio Cue condition × Time-to-contact. The * indicates the conditions where the mean difference is significant at $\alpha = 0.05$. A Bonferroni adjustment was applied to correct for a possible increase in type 1 errors associated with multiple comparisons.

Condition #	Condition Name	VALENCE N	VALENCE Mean	VALENCE Std. Error	VALENCE 95% Confidence Interval Lower	VALENCE 95% Confidence Interval Upper	AROUSAL N	AROUSAL Mean	AROUSAL Std. Error	AROUSAL 95% Confidence Interval Lower	AROUSAL 95% Confidence Interval Upper
1	Sound Only	44	5.578	.293	4.987	6.168	37	6.878	.324	6.224	7.531
Single Audio Cues:											
2	Amp	44	5.711	.284	5.140	6.283	37	8.889	.404	8.076	9.702
3	IAD	44	5.144	.281	4.578	5.711	37	7.189	.351	6.481	7.897
4	Ref	44	6.133	.235	5.659	6.608	37	7.833	.297	7.235	8.432
Multiple Audio Cues:											
5	Amp + IAD	44	5.733	.364	4.999	6.468	37	9.444	.350	8.738	10.151
6	Amp + Ref	44	7.111	.286	6.534	7.688	37	9.711	.287	9.132	10.290
7	IAD + Ref	44	5.922	.236	5.447	6.397	37	8.144	.277	7.586	8.702
8	Amp + IAD + Ref	44	6.333	.341	5.646	7.020	37	9.889	.295	9.294	10.484

Table C.4: Descriptive Statistics: Audio Cues × Valence / Arousal

The descriptives results are tabled for the Audio Cues × Valence / Arousal, averaged across all of the sound sources (and participants). The columns are labeled as condition number; condition name; number of trials; mean; standard error; and 95% confidence intervals for the mean.

Condition Pair	VALENCE Mean Diff.	Std. Error	Sig.	95% Conf. Interval Lower	Upper	AROUSAL Mean Diff.	Std. Error	Sig.	95% Conf. Interval Lower	Upper
Sound Only (No Audio Cues) × Single Audio Cues:										
Sound Only × Amp	-.133	.253	1.000	-.976	.710	-2.011*	.498	.006 *	-3.669	-.354
Sound Only × IAD	.433	.166	.339	-.117	.984	-.311	.307	1.000	-1.331	.709
Sound Only × Ref	-.556	.246	.815	-1.375	.264	-.956 *	.285	.046 *	-1.902	-.009
Sound Only (No Audio Cues) × Multiple Audio Cues:										
Sound Only × Amp + IAD	-.156	.218	1.000	-.880	.569	-2.567 *	.434	.000 *	-4.011	-1.123
Sound Only × Amp + Ref	-1.533 *	.360	.003 *	-2.730	-.337	-2.833 *	.376	.000 *	-4.085	-1.582
Sound Only × IAD + Ref	-.344	.201	1.000	-1.013	.324	-1.267 *	.303	.004 *	-2.276	-.258
Sound Only × Amp + IAD + Ref	-.756	.301	.442	-1.756	.245	-3.011 *	.457	. .000 *	-4.531	-1.491
Single × Single Audio Cues:										
Amp × IAD	.567	.216	.339	-.153	1.287	1.700 *	.438	.010 *	.244	3.156
Amp × Ref	-.422	.204	1.000	-1.101	.257	1.056	.341	.096	-.079	2.191
IAD × Ref	-.989 *	.182	.000 *	-1.595	-.383	-.644	.259	.467	-1.506	.217
Multiple × Multiple Audio Cues:										
Amp + IAD × Amp + Ref	-1.378 *	.341	.006 *	-2.513	-.243	-.267	.314	1.000	-1.310	.777
Amp + IAD × IAD + Ref	-.189	.281	1.000	-1.124	.747	1.300 *	.366	.026 *	.084	2.516
Amp + Ref × IAD + Ref	1.189 *	.239	.000 *	.396	1.982	1.567 *	.309	.000 *	.538	2.595
Amp + IAD × Amp + IAD + Ref	-.600	.349	1.000	-1.760	.560	-.444	.360	1.000	-1.643	.754
Amp + Ref × Amp + IAD + Ref	.778	.365	1.000	-.436	1.992	-.178	.356	1.000	-1.361	1.005
IAD + Ref × Amp + IAD + Ref	-.411	.312	1.000	-1.449	.626	-1.744 *	.392	.002 *	-3.047	-.442
Single × Multiple Audio Cues:										
Amp × Amp + IAD	-.022	.325	1.000	-1.103	1.058	-.556	.378	1.000	-1.813	. 702
Amp × Amp + Ref	-1.400 *	.285	.000 *	-2.348	-.452	-.822	.407	1.000	-2.177	.533
Amp × IAD + Ref	-.211	.158	1.000	-.736	.313	.744	.419	1.000	-.649	2.138
IAD × Amp + IAD	-.589	.215	.249	-1.304	.127	-2.256 *	.401	.000 *	-3.590	-.921
IAD × Amp + Ref	-1.967 *	.296	.000 *	-2.950	-.983	-2.522 *	.373	.000 *	-3.763	-1.282
IAD × IAD + Ref	-.778 *	.162	.001 *	-1.318	-.237	-.956	.325	.145	-2.036	.125
Ref × Amp + IAD	.400	.279	1.000	-.528	1.328	-1.611 *	.339	.001 *	-2.740	-.483
Ref × Amp + Ref	-.978 *	.230	.003 *	-1.742	-.214	-1.878 *	.316	.000 *	-2.930	-.826
Ref × IAD + Ref	.211	.137	1.000	-.245	.667	-.311	.192	1.000	-.950	.328
Amp × Amp + IAD + Ref	-.622	.302	1.000	-1.627	.382	-1.000	.432	.712	-2.438	.438
IAD × Amp + IAD + Ref	-1.189 *	.314	.013 *	-2.232	-.146	-2.700 *	.380	.000 *	-3.963	-1.437
Ref × Amp + IAD + Ref	-.200	.316	1.000	-1.251	.851	-2.056*	.409	.000 *	-3.415	-.697

Table C.5: Pairwise Comparisons: Audio Cues × Valence / Arousal

The pairwise comparisons of Audio Cue condition × Valence / Arousal rating. The * indicates the conditions where the mean difference is significant at $\alpha = 0.05$. A Bonferroni adjustment was applied to correct for a possible increase in type 1 errors associated with multiple comparisons.

Condition #	Name	N	Mean	Std. Error	95% Confidence Interval for mean Lower	Upper
1	Sound Only	38	2.556	.236	2.080	3.031
Single Audio Cues:						
2	Amp	38	5.022	.311	4.396	5.648
3	IAD	38	3.767	.287	3.188	4.345
4	Ref	38	4.544	.259	4.023	5.066
Multiple Audio Cues:						
5	Amp + IAD	38	4.644	.337	3.966	5.323
6	Amp + Ref	38	6.222	.269	5.679	6.765
7	IAD + Ref	38	4.789	.273	4.239	5.338
8	Amp + IAD + Ref	38	6.444	.301	5.838	7.051

Table C.6: Descriptive Statistics: Audio Cue × Engagement

The descriptives results are tabled for the Audio Cues × Engagement, averaged across all of the sound sources (and participants). The columns are labeled as condition number; condition name; number of trials; mean; standard error; and 95% confidence intervals for the mean.

Condition Pair	Mean Difference	Std. Error	Sig.	95% Confidence Interval Lower	Upper
Sound Only (No Audio Cues) × Single Audio Cues:					
Sound Only × Amp	-2.467*	.318	.000	-3.524	-1.409
Sound Only × IAD	-1.211*	.317	.012	-2.265	-.157
Sound Only × Ref	-1.989*	.306	.000	-3.006	-.972
Sound Only (No Audio Cues) × Multiple Audio Cues:					
Sound Only × Amp + IAD	-2.089*	.314	.000	-3.132	-1.046
Sound Only × Amp + Ref	-3.667*	.339	.000	-4.793	-2.541
Sound Only × IAD + Ref	-2.233*	.325	.000	-3.313	-1.153
Sound Only × Amp + IAD + Ref	-3.889*	.334	.000	-4.999	-2.779
Single × Single Audio Cues:					
Amp × IAD	1.256*	.291	.003	.287	2.224
Amp × Ref	.478	.314	1.000	-.567	1.523
IAD × Ref	-.778	.299	.351	-1.772	.216
Multiple × Multiple Audio Cues:					
Amp + IAD × Amp + Ref	-1.578*	.373	.003	-2.817	-.338
Amp + IAD × IAD + Ref	-.144	.337	1.000	-1.266	.977
Amp + Reflections × IAD + Ref	1.433*	.327	.002	.346	2.521
Amp + IAD × Amp + IAD + Ref	-1.800*	.365	.000	-3.013	-.587
Amp + Ref × Amp + IAD + Ref	-.222	.330	1.000	-1.320	.876
IAD + Ref × Amp + IAD + Ref	-1.656*	.254	.000	-2.501	-.810
Single × Multiple Audio Cues:					
Amp × Amp + IAD	.378	.290	1.000	-.587	1.342
Amp × Amp + Ref	-1.200*	.292	.005	-2.172	-.228
Amp × IAD + Ref	.233	.314	1.000	-.813	1.279
IAD × Amp + IAD	-.878	.273	.069	-1.787	.032
IAD × Amp + Ref	-2.456*	.367	.000	-3.678	-1.233
IAD × IAD + Ref	-1.022*	.276	.017	-1.941	-.104
Ref × Amp + IAD	-.100	.336	1.000	-1.217	1.017
Ref × Amp + Ref	-1.678*	.303	.000	-2.686	-.669
Ref × IAD + Ref	-.244	.192	1.000	-.882	.393
Amp × Amp + IAD + Ref	-1.422*	.315	.001	-2.471	-.373
IAD × Amp + IAD + Ref	-2.678*	.317	.000	-3.732	-1.623
Ref × Amp + IAD + Ref	-1.900*	.300	.000	-2.898	-.902

Table C.7: Pairwise Comparisons: Audio Cues × Engagement

The pairwise comparisons of Audio Cue condition × Engagement rating. The * indicates the conditions where the mean difference is significant at $\alpha = 0.05$. A Bonferroni adjustment was applied to correct for a possible increase in type 1 errors associated with multiple comparisons.

Condition #	Condition Name	N	Mean	Std. Error	95% Confidence Interval for mean Lower	Upper
Time-to-contact:						
1	-18 to -3	164	-241.254	37.105	-314.522	-167.986
2	-3	164	357.004	58.713	241.068	472.941
3	-18	164	814.916	58.832	698.745	931.087
Valence:						
1	-18 to -3	176	6.250	.163	5.927	6.573
2	-3	176	5.239	.179	4.886	5.591
3	-18	176	6.278	.139	6.003	6.554
Arousal:						
1	-18 to -3	174	9.649	.156	9.341	9.958
2	-3	174	9.356	.154	9.053	9.660
3	-18	174	5.948	.222	5.511	6.386
Engagement:						
1	-18 to -3	176	5.583	.162	5.264	5.903
2	-3	176	4.322	.164	3.998	4.646
3	-18	176	3.506	.144	3.221	3.790

Table C.8: Descriptive Statistics: Amplitude Levels × Time-to-Contact / Valence / Arousal / Engagement

The descriptives results are tabled for the Amplitude Levels × Time-to-Contact / Valence / Arousal / Engagement, averaged across all of the sound sources, audio cues, and participants. The columns are labeled as condition number; condition name; number of trials; mean; standard error; and 95% confidence intervals for the mean.

Condition Pair	Mean Difference	Std. Error	Sig.	95% Confidence Interval Lower	Upper
Amp × Time-to-contact:					
-18 to -3 × -3	-598.258 *	61.604	.000 *	-747.275	-449.241
-18 to -3 × -18	-1056.170 *	63.980	.000 *	-1210.934	-901.406
-3 × -18	-457.911 *	66.714	.000 *	-619.288	-296.535
Amp × Valence:					
-18 to -3 × -3	1.011 *	.165	.000 *	.613	1.410
-18 to -3 × -18	-.028	.156	1.000	-.404	.347
-3 × -18	-1.040 *	.191	.000 *	-1.503	-.577
Amp × Arousal:					
-18 to -3 × -3	.293	.207	.475	-.207	.793
-18 to -3 × -18	3.701 *	.255	.000 *	3.085	4.317
-3 × -18	3.408 *	.225	.000 *	2.864	3.952
Amp × Engagement:					
-18 to -3 × -3	1.011 *	.165	.000 *	.613	1.410
-18 to -3 × -18	-.028	.156	1.000	-.404	.347
-3 × -18	-1.040 *	.191	.000 *	-1.503	-.577

Table C.9: Pairwise Comparisons: Amplitude Levels × Engagement

The pairwise comparisons of Amplitude Level condition × Time-to-contact / Valence / Arousal / Engagement. The * indicates the conditions where the mean difference is significant at $\alpha = 0.05$. A Bonferroni adjustment was applied to correct for a possible increase in type 1 errors associated with multiple comparisons.

Condition #	Condition Name	N	Mean	Std. Error	95% Confidence Interval for mean Lower	Upper
1	Car	113	114.728	62.672	-9.448	238.904
2	Noise	113	234.297	64.224	107.046	361.548
3	Square	113	140.582	60.852	20.011	261.153

Table C.10: Descriptive Statistics: Sound Source × Time-to-Contact
The descriptives results are tabled for the Sound Source × Time-to-Contact, averaged across all of the audio cues (and participants). The columns are labeled as condition number; condition name; number of trials; mean; standard error; and 95% confidence intervals for the mean.

Condition Pair	Mean Difference	Std. Error	Sig.	95% Confidence Interval Lower	Upper
Car × Noise	119.569	60.674	.154	-267.036	27.898
Car × Square	-25.853	57.790	1.000	-166.311	114.604
Noise × Square	93.716	67.501	.503	-70.343	257.774

Table C.11: Pairwise Comparisons: Sound Source × Time-to-Contact
The pairwise comparisons of Sound Source condition × Time-to-contact. The * indicates the conditions where the mean difference is significant at $\alpha = 0.05$. A Bonferroni adjustment was applied to correct for a possible increase in type 1 errors associated with multiple comparisons.

Condition #	Condition Name	VALENCE N	Mean	Std. Error	95% Confidence Interval Lower	Upper	AROUSAL N	Mean	Std. Error	95% Confidence Interval Lower	Upper
1	Car	117	6.551	.166	6.223	6.879	120	8.575	.230	8.120	9.030
2	Noise	117	6.248	.136	5.979	6.516	120	7.783	.222	7.343	8.224
3	Square	117	5.162	.208	4.750	5.575	120	9.133	.192	8.752	9.514

Table C.12: Descriptive Statistics: Sound Source × Valence / Arousal
The descriptives results are tabled for the Sound Source × Valence / Arousal, averaged across all of the audio cues (and participants). The columns are labeled as condition number; condition name; number of trials; mean; standard error; and 95% confidence intervals for the mean.

Condition Pair	VALENCE Mean Diff.	Std. Error	Sig.	95% Confidence Interval Lower	Upper	AROUSAL Mean Diff.	Std. Error	Sig.	95% Confidence Interval Lower	Upper
Car × Noise	.303	.184	.306	-.144	.750	.792 *	.232	.003 *	.228	1.355
Car × Square	1.389 *	.258	.000 *	.762	2.016	-.558 *	.233	.055 *	-1.125	.008
Noise × Square	1.085 *	.194	.000 *	.613	1.558	-1.350 *	.274	.000 *	-2.016	-.684

Table C.13: Pairwise Comparisons: Sound Source × Valence / Arousal
The pairwise comparisons of Sound Source condition × Valence / Arousal rating. The * indicates the conditions where the mean difference is significant at $\alpha = 0.05$. A Bonferroni adjustment was applied to correct for a possible increase in type 1 errors associated with multiple comparisons.

Condition #	Condition Name	N	Mean	Std. Error	95% Confidence Interval for mean Lower	95% Confidence Interval for mean Upper
1	Car	120	5.533	.190	5.158	5.909
2	Noise	120	4.142	.179	3.786	4.497
3	Square	120	4.579	.218	4.147	5.011

Table C.14: Descriptive Statistics: Sound Source × Engagement

The descriptives results are tabled for the Sound Source × Engagement, averaged across all of the audio cues (and participants). The columns are labeled as condition number; condition name; number of trials; mean; standard error; and 95% confidence intervals for the mean.

Condition Pair	Mean Difference	Std. Error	Sig.	95% Confidence Interval Lower	95% Confidence Interval Upper
Car × Noise	1.392 *	.185	.000 *	.941	1.842
Car × Square	.954 *	.245	.000 *	.358	1.550
Noise × Square	-.438	.195	.081	-.912	.037

Table C.15: Pairwise Comparisons: Sound Source × Engagement

The pairwise comparisons of Sound Source condition × Engagement rating. The * indicates the conditions where the mean difference is significant at $\alpha = 0.05$. A Bonferroni adjustment was applied to correct for a possible increase in type 1 errors associated with multiple comparisons.

Appendix D

Chapter 6 Experiment 4

#	Image Variables: Object Presented	Sound Variables: Sound Source	Audio Cue	# Audio Cues	Amp. Level	Audiovisual Consistency
1	Car	No sound - Control				
2	Disc	No sound - Control				
3	Car	Car	None - Control	0	-18	Congruent - Real world
4	Disc	Car	None - Control	0	-18	Incongruent
5	Car	Car	None - Control	0	-3	Congruent - Real world
6	Disc	Car	None - Control	0	-3	Incongruent
7	Car	Noise	None - Control	0	-18	
8	Disc	Noise	None - Control	0	-18	
9	Car	Noise	None - Control	0	-3	
10	Disc	Noise	None - Control	0	-3	
11	Car	Square	None - Control	0	-18	Incongruent
12	Disc	Square	None - Control	0	-18	Congruent - Abstract
13	Car	Square	None - Control	0	-3	Incongruent
14	Disc	Square	None - Control	0	-3	Congruent - Abstract
15	Car	Car	Amplitude	1	-18 to -3	Congruent - Real world
16	Disc	Car	Amplitude	1	-18 to -3	Incongruent
17	Car	Noise	Amplitude	1	-18 to -3	
18	Disc	Noise	Amplitude	1	-18 to -3	
19	Car	Square	Amplitude	1	-18 to -3	Incongruent
20	Disc	Square	Amplitude	1	-18 to -3	Congruent - Abstract
21	Car	Car	IAD	1	-18	Congruent - Real world
22	Disc	Car	IAD	1	-18	Incongruent
23	Car	Car	IAD	1	-3	Congruent - Real world
24	Disc	Car	IAD	1	-3	Incongruent
25	Car	Noise	IAD	1	-18	
26	Disc	Noise	IAD	1	-18	
27	Car	Noise	IAD	1	-3	
28	Disc	Noise	IAD	1	-3	
29	Car	Square	IAD	1	-18	Incongruent
30	Disc	Square	IAD	1	-18	Congruent - Abstract
31	Car	Square	IAD	1	-3	Incongruent
32	Disc	Square	IAD	1	-3	Congruent - Abstract
33	Car	Car	Reflections	1	-18	Congruent - Real world
34	Disc	Car	Reflections	1	-18	Incongruent
35	Car	Car	Reflections	1	-3	Congruent - Real world
36	Disc	Car	Reflections	1	-3	Incongruent
37	Car	Noise	Reflections	1	-18	
38	Disc	Noise	Reflections	1	-18	
39	Car	Noise	Reflections	1	-3	
40	Disc	Noise	Reflections	1	-3	
41	Car	Square	Reflections	1	-18	Incongruent
42	Disc	Square	Reflections	1	-18	Congruent - Abstract
43	Car	Square	Reflections	1	-3	Incongruent
44	Disc	Square	Reflections	1	-3	Congruent - Abstract
45	Car	Car	Amp + IAD	2	-18 to -3	Congruent - Real world
46	Disc	Car	Amp + IAD	2	-18 to -3	Incongruent
47	Car	Noise	Amp + IAD	2	-18 to -3	? - ?
48	Disc	Noise	Amp + IAD	2	-18 to -3	
49	Car	Square	Amp + IAD	2	-18 to -3	Incongruent
50	Disc	Square	Amp + IAD	2	-18 to -3	Congruent - Abstract
51	Car	Car	Reflections + Amp	2	-18 to -3	Congruent - Real world
52	Disc	Car	Reflections + Amp	2	-18 to -3	Incongruent
53	Car	Noise	Reflections + Amp	2	-18 to -3	
54	Disc	Noise	Reflections + Amp	2	-18 to -3	
55	Car	Square	Reflections + Amp	2	-18 to -3	Incongruent
56	Disc	Square	Reflections + Amp	2	-18 to -3	Congruent - Abstract
57	Car	Car	Reflections + IAD	2	-18	Congruent - Real world
58	Disc	Car	Reflections + IAD	2	-18	Incongruent
59	Car	Car	Reflections + IAD	2	-3	Congruent - Real world
60	Disc	Car	Reflections + IAD	2	-3	Incongruent
61	Car	Noise	Reflections + IAD	2	-18	
62	Disc	Noise	Reflections + IAD	2	-18	
63	Car	Noise	Reflections + IAD	2	-3	
64	Disc	Noise	Reflections + IAD	2	-3	
65	Car	Square	Reflections + IAD	2	-18	Incongruent
66	Disc	Square	Reflections + IAD	2	-18	Congruent - Abstract
67	Car	Square	Reflections + IAD	2	-3	Incongruent
68	Disc	Square	Reflections + IAD	2	-3	Congruent - Abstract
69	Car	Car	Reflections + Amp + IAD	3	-18 to -3	Congruent - Real world
70	Disc	Car	Reflections + Amp + IAD	3	-18 to -3	Incongruent
71	Car	Noise	Reflections + Amp + IAD	3	-18 to -3	
72	Disc	Noise	Reflections + Amp + IAD	3	-18 to -3	
73	Car	Square	Reflections + Amp + IAD	3	-18 to -3	Incongruent
74	Disc	Square	Reflections + Amp + IAD	3	-18 to -3	Congruent - Abstract

Table D.1: List of Experiment Conditions

Listing the trial number, image object presented; sound source, audio cue variable, number of audio cues, and amplitude level.

Condition #	Image	Audio Cue	N	Mean	Std. Error	95% Confidence Interval Lower	Upper
Car:							
1	Car	-	12	1243.314	73.078	1082.470	1404.159
2	Car	Sound Only (no cues)	12	1218.173	123.696	945.920	1490.425
Single Audio Cues:							
3	Car	Amp	12	145.326	145.057	-173.942	464.594
4	Car	IAD	12	1035.353	76.119	867.816	1202.890
5	Car	Ref	12	287.809	152.824	-48.554	624.172
Multiple Audio Cues:							
6	Car	Amp + IAD	12	-15.885	102.085	-240.573	208.803
7	Car	Amp + Ref	12	52.152	137.617	-250.741	355.046
8	Car	IAD + Ref	12	170.414	131.153	-118.252	459.080
9	Car	Amp + IAD + Ref	12	-97.783	90.305	-296.543	100.976
Disc:							
10	Disc	-	12	1767.856	195.996	1336.471	2199.240
11	Disc	Sound Only (no cues)	12	1084.738	92.904	880.257	1289.218
Single Audio Cues:							
12	Disc	Amp	12	174.171	151.266	-158.764	507.105
13	Disc	IAD	12	1080.772	99.423	861.943	1299.600
14	Disc	Ref	12	301.549	140.099	-6.807	609.905
Multiple Audio Cues:							
15	Disc	Amp + IAD	12	1.793	103.813	-226.697	230.284
16	Disc	Amp + Ref	12	11.253	119.458	-251.671	274.178
17	Disc	IAD + Ref	12	298.430	142.426	-15.046	611.906
18	Disc	Amp + IAD + Ref	12	-79.058	96.061	-290.486	132.371

Table D.2: Descriptive Statistics: Audio Cues $\times$ Time-to-Contact

The descriptives results are tabled for the time-to-contact $\times$ audio cue condition, averaged across all of the sound sources (and participants). The columns are labeled as condition number; condition name; number of trials; mean; standard error; and 95% confidence intervals for the mean.

CAR Condition Pair	Mean Difference	Std. Error	Sig.	95% Confidence Interval Lower	Upper
Image Only × Image + Sound:					
Image Only × Sound Only	25.142	157.683	1.000	-784.237	834.520
Image Only × Amp	1097.988*	136.413	.001	397.788	1798.189
Image Only × IAD	207.961	120.938	1.000	-412.807	828.729
Image Only × Ref	955.505*	182.433	.043	19.084	1891.926
Image Only × Amp + IAD	1259.199*	119.972	.000	643.389	1875.009
Image Only × Amp + Ref	1191.162*	139.709	.001	474.041	1908.282
Image Only × IAD + Ref	1072.900*	147.915	.003	313.656	1832.144
Image Only × Amp + IAD + Ref	1341.097*	134.974	.000	648.279	2033.916
Image Only					
Sound Only × Single Audio Cues:					
Sound Only × Amp	1072.847*	139.447	.001	357.073	1788.621
Sound Only × IAD	182.819	82.905	1.000	-242.728	608.367
Sound Only × Ref	930.363*	94.840	.000	443.555	1417.172
Sound Only × Multiple Audio Cues:					
Sound Only × Amp + IAD	1234.058*	86.095	.000	792.136	1675.979
Sound Only × Amp + Ref	1166.020*	73.759	.000	787.417	1544.623
Sound Only × IAD + Ref	1047.758*	53.874	.000	771.225	1324.292
Sound Only × Amp + IAD + Ref	1315.956*	104.166	.000	781.274	1850.638
Single × Single Audio Cues:					
Amp × IAD	-890.027*	115.482	.001	-1482.791	-297.264
Amp × Ref	-142.483	114.785	1.000	-731.671	446.704
IAD × Ref	747.544*	95.655	.001	256.551	1238.537
Single × Multiple Audio Cues:					
Amp × Amp + IAD	161.211	62.600	1.000	-160.113	482.535
Amp × Amp + Ref	93.173	100.136	1.000	-420.823	607.169
Amp × IAD + Ref	-25.088	109.046	1.000	-584.819	534.642
IAD × Amp + IAD	1051.238*	68.263	.000	700.847	1401.629
IAD × Amp + Ref	983.201*	101.423	.000	462.602	1503.800
IAD × IAD + Ref	864.939*	92.251	.000	391.420	1338.459
Ref × Amp + IAD	303.694	85.854	.712	-136.994	744.382
Ref × Amp + Ref	235.657	103.131	1.000	-293.709	765.022
Ref × IAD + Ref	117.395	99.714	1.000	-394.434	629.224
Amp × Amp + IAD + Ref	243.109	145.762	1.000	-505.080	991.298
IAD × Amp + IAD + Ref	1133.137*	58.567	.000	832.517	1433.757
Ref × Amp + IAD + Ref	385.593	99.812	.404	-126.736	897.921
Multiple × Multiple Audio Cues:					
Amp + IAD × Amp + Ref	-68.037	75.938	1.000	-457.822	321.747
Amp + IAD × IAD + Ref	-186.299	68.364	1.000	-537.211	164.612
Amp + Ref × IAD + Ref	-118.262	44.949	1.000	-348.983	112.459
Amp + IAD × Amp + IAD + Ref	81.898	101.025	1.000	-436.658	600.454
Amp + Ref × Amp + IAD + Ref	149.936	125.184	1.000	-492.632	792.503
IAD + Ref × Amp + IAD + Ref	268.198	115.112	1.000	-322.667	859.062

Table D.3: Pairwise Comparisons: Car Audio Cues × Time-to-Contact

The pairwise comparisons of Time-to-contact × Car Audio Cue condition averaged across all of the sound sources (and participants). The * indicates the conditions where the mean difference is significant at $\alpha = 0.05$. A Bonferroni adjustment was applied to correct for a possible increase in type 1 errors associated with multiple comparisons.

DISC Condition Pair	Mean Difference	Std. Error	Sig.	95% Confidence Interval Lower	Upper
Image Only × Image + Sound:					
Image Only × Sound Only	683.118	209.293	1.000	-391.177	1757.413
Image Only × Amp	1593.685*	200.077	.001	566.695	2620.675
Image Only × IAD	687.084	257.449	1.000	-634.392	2008.560
Image Only × Ref	1466.307*	208.266	.003	397.286	2535.328
Image Only × Amp + IAD	1766.062*	174.940	.000	868.104	2664.021
Image Only × Amp + Ref	1756.602*	199.516	.000	732.492	2780.713
Image Only × IAD + Ref	1469.426*	204.195	.003	421.301	2517.550
Image Only × Amp + IAD + Ref	1846.913*	217.822	.001	728.839	2964.987
Sound Only × Single Audio Cues:					
Sound Only × Amp	910.567*	87.758	.000	460.107	1361.026
Sound Only × IAD	3.966	100.045	1.000	-509.563	517.495
Sound Only × Ref	783.188*	81.854	.000	363.037	1203.340
Sound Only × Multiple Audio Cues:					
Sound Only × Amp + IAD	1082.944*	55.352	.000	798.824	1367.064
Sound Only × Amp + Ref	1073.484*	68.735	.000	720.669	1426.300
Sound Only × IAD + Ref	786.308*	73.362	.000	409.745	1162.870
Sound Only × Amp + IAD + Ref	1163.795*	70.556	.000	801.635	1525.955
Single × Single Audio Cues:					
Amp × IAD	-906.601	183.803	.069	-1850.053	36.852
Amp × Ref	-127.378	70.229	1.000	-487.862	233.106
IAD × Ref	779.222*	151.603	.049	1.050	1557.395
Single × Multiple Audio Cues:					
Amp × Amp + IAD	172.377	53.210	1.000	-100.749	445.504
Amp × Amp + Ref	162.917	118.108	1.000	-443.327	769.162
Amp × IAD + Ref	-124.259	32.882	.467	-293.040	44.522
IAD × Amp + IAD	1078.978*	142.331	.002	348.400	1809.557
IAD × Amp + Ref	1069.518*	123.196	.000	437.159	1701.878
IAD × IAD + Ref	782.342	164.025	.089	-59.593	1624.276
Ref × Amp + IAD	299.756*	57.808	.046	3.031	596.481
Ref × Amp + Ref	290.296	102.335	1.000	-234.986	815.578
Ref × IAD + Ref	3.119	54.296	1.000	-275.582	281.820
Amp × Amp + IAD + Ref	253.228	139.955	1.000	-465.155	971.611
IAD × Amp + IAD + Ref	1159.829*	95.083	.000	671.772	1647.886
Ref × Amp + IAD + Ref	380.607	124.277	1.000	-257.305	1018.518
Multiple × Multiple Audio Cues:					
Amp + IAD × Amp + Ref	-9.460	86.627	1.000	-454.115	435.195
Amp + IAD × IAD + Ref	-296.637*	46.985	.009	-537.810	-55.463
Amp + Ref × IAD + Ref	-287.177	106.938	1.000	-836.085	261.732
Amp + IAD × Amp + IAD + Ref	80.851	102.856	1.000	-447.108	608.809
Amp + Ref × Amp + IAD + Ref	90.311	70.431	1.000	-271.208	451.830
IAD + Ref × Amp + IAD + Ref	377.487	120.568	1.000	-241.383	996.358

Table D.4: Pairwise Comparisons: Disc Audio Cues × Time-to-Contact

The pairwise comparisons of Time-to-contact × Disc Audio Cue condition averaged across all of the sound sources (and participants). The * indicates the conditions where the mean difference is significant at $\alpha = 0.05$. A Bonferroni adjustment was applied to correct for a possible increase in type 1 errors associated with multiple comparisons.

Condition Pair	Mean Difference	Std. Error	Sig.	95% Confidence Interval Lower	Upper
Car Audio cues × Disc Audio Cues:					
Image Only × Image Only	-524.542	147.435	.687	-1281.320	232.237
Sound Only × Sound Only	133.435	64.777	1.000	-199.065	465.935
Amp × Amp	-28.845	41.557	1.000	-242.156	184.466
IAD × IAD	-45.418	78.395	1.000	-447.818	356.981
Ref × Ref	-13.740	79.096	1.000	-419.739	392.259
Amp + IAD × Amp + IAD	-17.678	34.157	1.000	-193.007	157.650
Amp + Ref × Amp + Ref	40.899	53.718	1.000	-234.835	316.633
IAD + Ref × IAD + Ref	-128.016	84.243	1.000	-560.432	304.401
Amp + IAD + Ref × Amp + IAD + Ref	-18.726	66.293	1.000	-359.007	321.556

Table D.5: Pairwise Comparisons: Car Multiple Audio cues vs Disc Multiple Audio Cues × Time-to-Contact

The pairwise comparisons of Car Multiple Audio Cues Time-to-contact × Disc Audio Cues Time-to-contact averaged across all of the sound sources (and participants). The * indicates the conditions where the mean difference is significant at $\alpha = 0.05$. A Bonferroni adjustment was applied to correct for a possible increase in type 1 errors associated with multiple comparisons.

Condition textbf#	Image	Audio Cue	VALENCE N	Mean	Std. Error	95% Confidence Interval Lower	Upper	AROUSAL N	Mean	Std. Error	95% Confidence Interval Lower	Upper
Car:												
1	Car	-	13	5.692	.746	4.067	7.317	12	4.750	.617	3.392	6.108
2	Car	Sound Only	13	6.395	.212	5.933	6.856	12	7.791	.280	7.173	8.408
Single Audio Cues:												
3	Car	Amp	13	6.974	.328	6.259	7.689	12	9.195	.395	8.326	10.064
4	Car	IAD	13	6.154	.249	5.612	6.696	12	8.291	.299	7.633	8.949
5	Car	Ref	13	7.487	.414	6.586	8.388	12	8.223	.154	7.884	8.561
Multiple Audio Cues:												
6	Car	Amp + IAD	13	6.948	.326	6.237	7.659	12	10.499	.341	9.748	11.251
7	Car	Amp + Ref	13	7.898	.512	6.782	9.015	12	8.555	.211	8.091	9.019
8	Car	IAD + Ref	13	7.102	.392	6.248	7.957	12	9.195	.149	8.866	9.524
9	Car	Amp + IAD + Ref	13	7.948	.420	7.033	8.864	12	10.138	.337	9.396	10.880
Disc:												
10	Disc	-	13	6.923	.288	6.296	7.550	12	5.833	.474	4.790	6.877
11	Disc	Sound Only	13	6.615	.307	5.948	7.283	12	7.820	.255	7.258	8.382
Single Audio Cues:												
12	Disc	Amp	13	7.308	.388	6.463	8.152	12	9.445	.277	8.836	10.054
13	Disc	IAD	13	6.190	.352	5.423	6.957	12	8.654	.304	7.984	9.324
14	Disc	Ref	13	7.141	.403	6.263	8.018	12	8.067	.208	7.610	8.525
Multiple Audio Cues:												
15	Disc	Amp + IAD	13	7.078	.458	6.080	8.075	12	9.444	.460	8.432	10.456
16	Disc	Amp + Ref	13	7.770	.380	6.942	8.598	12	8.973	.452	7.979	9.967
17	Disc	IAD + Ref	13	6.770	.339	6.032	7.508	12	8.848	.247	8.304	9.391
18	Disc	Amp + IAD + Ref	13	7.386	.553	6.180	8.592	12	9.889	.190	9.472	10.307

Table D.6: Descriptive Statistics: Audio Cues × Valence / Arousal

Descriptives results for Valence / Arousal ratings × Audio cue, averaged across all of the sound sources (and participants). The columns are labeled as condition number; condition name; number of trials; mean; standard error; and 95% confidence intervals for the mean.

Car Condition Pair	VALENCE Mean Diff.	Std. Error	Sig.	95% Conf. Interval Lower	Upper	AROUSAL Mean Diff.	Std. Error	Sig.	95% Conf. Interval Lower	Upper
Image Only × Image + Sound:										
Image Only × Sound Only	-.702	.799	1.000	-4.67	3.268	-3.041*	.465	.006	-5.426	-.655
Image Only × Amp	-1.282	.969	1.000	-6.094	3.53	-4.445*	.626	.003	-7.658	-1.232
Image Only × IAD	-.462	.839	1.000	-4.630	3.707	-3.541	.707	.061	-7.168	.086
Image Only × Ref	-1.795	1.021	1.000	-6.865	3.276	-3.473*	.579	.014	-6.447	-.498
Image Only × Amp + IAD	-1.256	.839	1.000	-5.423	2.910	-5.749*	.915	.009	-10.447	-1.052
Image Only × Amp + Ref	-2.206	1.072	1.000	-7.530	3.117	-3.805*	.488	.001	-6.308	-1.302
Image Only × IAD + Ref	-1.410	.966	1.000	-6.207	3.387	-4.445*	.611	.002	-7.583	-1.307
Image Only × Amp + IAD + Ref	-2.256	.992	1.000	-7.184	2.671	-5.388*	.881	.012	-9.912	-.865
Sound Only × Single Audio Cues:										
Sound Only × Amp	-.579	.239	1.000	-1.766	.608	-1.404	.356	.348	-3.229	.421
Sound Only × IAD	.241	.141	1.000	-.460	.941	-.500	.318	1.000	-2.132	1.132
Sound Only × Ref	-1.092	.365	1.000	-2.906	.721	-.432	.196	1.000	-1.437	.573
Sound Only × Multiple Audio Cues:										
Sound Only × Amp + IAD	-.554	.261	1.000	-1.851	.743	-2.708	.531	.053	-5.436	.020
Sound Only × Amp + Ref	-1.504	.538	1.000	-4.175	1.167	-.764	.215	.694	-1.869	.340
Sound Only × IAD + Ref	-.708	.331	1.000	-2.353	.937	-1.404*	.206	.005	-2.464	-.345
Sound Only × Amp + IAD + Ref	-1.554	.379	.223	-3.434	.326	-2.347	.477	.069	-4.794	.099
Single × Single Audio Cues:										
Amp × IAD	.820	.196	.192	-.152	1.792	.904	.212	.203	-.184	1.992
Amp × Ref	-.513	.286	1.000	-1.936	.910	.973	.312	1.000	-.630	2.575
IAD × Ref	-1.333	.272	.056	-2.683	.017	.068	.222	1.000	-1.069	1.206
Multiple × Multiple Audio Cues:										
Amp + IAD × Amp + Ref	-.950	.449	1.000	-3.182	1.282	1.944	.437	.151	-.301	4.190
Amp + IAD × IAD + Ref	-.154	.230	1.000	-1.298	.991	1.304	.330	.344	-.388	2.997
Amp + Ref × IAD + Ref	.796	.289	1.000	-.641	2.234	-.640	.168	.438	-1.501	.221
Amp + IAD × Amp + IAD + Ref	-1.000	.254	.298	-2.259	.259	.361	.126	1.000	-.284	1.006
Amp + Ref × Amp + IAD + Ref	-.050	.238	1.000	-1.233	1.133	-1.583	.401	.348	-3.641	.474
IAD + Ref × Amp + IAD + Ref	-.846*	.102	.000	-1.355	-.338	-.943	.289	1.000	-2.428	.542
Single × Multiple Audio Cues:										
Amp × Amp + IAD	.025	.205	1.000	-.992	1.043	-1.304	.486	1.000	-3.797	1.189
Amp × Amp + Ref	-.925	.369	1.000	-2.756	.907	.640	.277	1.000	-.783	2.063
Amp × IAD + Ref	-.128	.155	1.000	-.898	.641	.000	.319	1.000	-1.636	1.636
IAD × Amp + IAD	-.795	.270	1.000	-2.137	.548	-2.208*	.325	.005	-3.877	-.540
IAD × Amp + Ref	-1.745	.446	.316	-3.960	.471	-.264	.272	1.000	-1.661	1.132
IAD × IAD + Ref	-.948	.241	.303	-2.146	.249	-.904	.207	.169	-1.964	.156
Ref × Amp + IAD	.538	.435	1.000	-1.621	2.698	-2.277*	.362	.009	-4.137	-.417
Ref × Amp + Ref	-.412	.330	1.000	-2.050	1.227	-.332	.149	1.000	-1.098	.433
Ref × IAD + Ref	.385	.275	1.000	-.979	1.749	-.973*	.071	.000	-1.335	-.610
Amp × Amp + IAD + Ref	-.975*	.191	.040	-1.923	-.026	-.943	.415	1.000	-3.071	1.184
IAD × Amp + IAD + Ref	-1.795*	.305	.011	-3.310	-.279	-1.848*	.240	.001	-3.077	-.618
Ref × Amp + IAD + Ref	-.462	.285	1.000	-1.875	.952	-1.916*	.319	.014	-3.555	-.276

Table D.7: Pairwise Comparisons: Car Audio Cues × Valence / Arousal

The pairwise comparisons of Valence / Arousal rating × Car Audio Cue condition averaged across all of the sound sources (and participants). The * indicates the conditions where the mean difference is significant at $\alpha = 0.05$. A Bonferroni adjustment was applied to correct for a possible increase in type 1 errors associated with multiple comparisons.

DISC Condition Pair	VALENCE Mean Diff.	Std. Error	Sig.	95% Conf. Interval Lower	Upper	AROUSAL Mean Diff.	Std. Error	Sig.	95% Conf. Interval Lower	Upper
Image Only × Image + Sound:										
Image Only × Sound Only	.308	.448	1.000	-1.919	2.534	-1.987*	.359	.027	-3.828	-.145
Image Only × Amp	-.385	.461	1.000	-2.673	1.904	-3.612*	.569	.008	-6.530	-.693
Image Only × IAD	.733	.475	1.000	-1.626	3.092	-2.821*	.304	.000	-4.382	-1.260
Image Only × Ref	-.218	.482	1.000	-2.612	2.177	-2.234	.535	.235	-4.978	.509
Image Only × Amp + IAD	-.155	.600	1.000	-3.134	2.825	-3.611*	.622	.018	-6.805	-.417
Image Only × Amp + Ref	-.847	.452	1.000	-3.094	1.400	-3.140	.738	.208	-6.930	.650
Image Only × IAD + Ref	.153	.453	1.000	-2.095	2.401	-3.014*	.405	.002	-5.091	-.937
Image Only × Amp + IAD + Ref	-.463	.666	1.000	-3.773	2.847	-4.056*	.483	.001	-6.537	-1.575
Sound Only × Single Audio Cues:										
Sound Only × Amp	-.692	.248	1.000	-1.925	.541	-1.625*	.264	.011	-2.981	-.269
Sound Only × IAD	.425	.090	.075	-.022	.873	-.834*	.159	.043	-1.653	-.016
Sound Only × Ref	-.525	.234	1.000	-1.688	.637	-.247	.239	1.000	-1.473	.978
Sound Only × Multiple Audio Cues:										
Sound Only × Amp + IAD	-.462	.192	1.000	-1.416	.492	-1.624	.427	.448	-3.816	.568
Sound Only × Amp + Ref	-1.155	.290	.279	-2.595	.286	-1.153	.477	1.000	-3.602	1.295
Sound Only × IAD + Ref	-.155	.104	1.000	-.674	.364	-1.027	.255	.303	-2.336	.281
Sound Only × Amp + IAD + Ref	-.771	.344	1.000	-2.479	.938	-2.069*	.295	.003	-3.584	-.554
Single × Single Audio Cues:										
Amp × IAD	1.118	.298	.424	-.363	2.598	.791	.368	1.000	-1.096	2.678
Amp × Ref	.167	.112	1.000	-.388	.722	1.378*	.129	.000	.715	2.040
IAD × Ref	-.951	.256	.453	-2.223	.321	.587	.355	1.000	-1.236	2.410
Multiple × Multiple Audio Cues:										
Amp + IAD × Amp + Ref	-.692	.371	1.000	-2.535	1.151	.471	.313	1.000	-1.137	2.079
Amp + IAD × IAD + Ref	.308	.216	1.000	-.765	1.380	.597	.587	1.000	-2.418	3.612
Amp + Ref × IAD + Ref	1.000	.227	.133	-.129	2.129	.126	.572	1.000	-2.809	3.061
Amp + IAD × Amp + IAD + Ref	-.308	.219	1.000	-1.395	.778	-.445	.532	1.000	-3.174	2.284
Amp + Ref × Amp + IAD + Ref	.384	.366	1.000	-1.434	2.202	-.916	.491	1.000	-3.436	1.604
IAD + Ref × Amp + IAD + Ref	-.616	.312	1.000	-2.165	.933	-1.042*	.138	.002	-1.748	-.336
Single × Multiple Audio Cues:										
Amp × Amp + IAD	.230	.257	1.000	-1.045	1.505	.001	.366	1.000	-1.876	1.878
Amp × Amp + Ref	-.462	.299	1.000	-1.948	1.024	.472	.291	1.000	-1.021	1.964
Amp × IAD + Ref	.538	.232	1.000	-.616	1.691	.598	.318	1.000	-1.036	2.231
IAD × Amp + IAD	-.888	.215	.211	-1.954	.178	-.790	.562	1.000	-3.674	2.094
IAD × Amp + Ref	-1.580*	.311	.042	-3.127	-.033	-.319	.613	1.000	-3.467	2.829
IAD × IAD + Ref	-.580*	.108	.025	-1.114	-.046	-.193	.192	1.000	-1.179	.792
Ref × Amp + IAD	.063	.230	1.000	-1.080	1.206	-1.377	.371	.529	-3.283	.529
Ref × Amp + Ref	-.629	.308	1.000	-2.160	.901	-.906	.282	1.000	-2.352	.540
Ref × IAD + Ref	.371	.212	1.000	-.681	1.423	-.780	.310	1.000	-2.373	.813
Amp × Amp + IAD + Ref	-.078	.254	1.000	-1.339	1.182	-.444	.264	1.000	-1.799	.911
IAD × Amp + IAD + Ref	-1.196	.360	.934	-2.985	.593	-1.235	.291	.211	-2.728	.258
Ref × Amp + IAD + Ref	-.245	.220	1.000	-1.339	.848	-1.822*	.252	.003	-3.117	-.526

Table D.8: Pairwise Comparisons: Disc Audio Cues × Valence / Arousal

The pairwise comparisons of Valence / Arousal rating × Disc Audio Cue condition averaged across all of the sound sources (and participants). The * indicates the conditions where the mean difference is significant at $\alpha = 0.05$. A Bonferroni adjustment was applied to correct for a possible increase in type 1 errors associated with multiple comparisons.

Car vs Disc Condition Pair	VALENCE Mean Diff.	Std. Error	Sig.	95% Conf. Interval Lower	Upper	AROUSAL Mean Diff.	Std. Error	Sig.	95% Conf. Interval Lower	Upper
Car Image Only × Disc Audio Cues:										
Image Only × Image Only	-1.231	.826	1.000	-5.331	2.870	-1.083	.570	1.000	-4.010	1.843
Sound Only × Sound Only	-.221	.280	1.000	-1.611	1.170	-.029	.096	1.000	-.520	.462
Amp × Amp	-.334	.193	1.000	-1.295	.627	-.250	.526	1.000	-2.951	2.451
IAD × IAD	-.036	.199	1.000	-1.023	.951	-.363	.191	1.000	-1.342	.616
Ref × Ref	.346	.246	1.000	-.875	1.567	.155	.207	1.000	-.908	1.218
Amp + IAD × Amp + IAD	-.129	.411	1.000	-2.170	1.912	1.055	.700	1.000	-2.536	4.646
Amp + Ref × Amp + Ref	.128	.408	1.000	-1.896	2.153	-.418	.528	1.000	-3.128	2.292
IAD + Ref × IAD + Ref	.332	.125	1.000	-.288	.952	.348	.146	1.000	-.401	1.096
Amp + IAD + Ref × Amp + IAD + Ref	.562	.237	1.000	-.615	1.740	.249	.202	1.000	- -.787	1.285

Table D.9: Pairwise Comparisons: Car Multiple Audio Cues × Disc Audio Cues × Valence / Arousal

The pairwise comparisons of Car Multiple Audio Cues Valence / Arousal rating × Disc Audio Cues ratings averaged across all of the sound sources (and participants). The * indicates the conditions where the mean difference is significant at $\alpha = 0.05$. A Bonferroni adjustment was applied to correct for a possible increase in type 1 errors associated with multiple comparisons.

Condition #	Image	Audio Cue	N	Mean	Std. Error	95% Confidence Interval Lower	Upper
Car:							
1	Car	-	12	2.467	.389	1.633	3.300
2	Car	Sound Only	12	4.700	.369	3.908	5.492
Single Audio Cues:							
3	Car	Amp	12	5.799	.350	5.047	6.550
4	Car	IAD	12	5.011	.318	4.328	5.693
5	Car	Ref	12	4.910	.451	3.943	5.877
Multiple Audio Cues:							
6	Car	Amp + IAD	12	6.043	.475	5.026	7.061
7	Car	Amp + Ref	12	5.756	.416	4.865	6.647
8	Car	IAD + Ref	12	5.810	.306	5.154	6.466
9	Car	Amp + IAD + Ref	12	6.221	.513	5.121	7.322
Disc:							
10	Disc	-	12	2.733	.358	1.965	3.501
11	Disc	Sound Only	12	4.488	.258	3.934	5.042
Single Audio Cues:							
12	Disc	Amp	12	4.979	.392	4.138	5.820
13	Disc	IAD	12	4.834	.380	4.019	5.649
14	Disc	Ref	12	4.512	.302	3.864	5.160
Multiple Audio Cues:							
15	Disc	Amp + IAD	12	5.511	.447	4.552	6.470
16	Disc	Amp + Ref	12	5.534	.427	4.619	6.449
17	Disc	IAD + Ref	12	5.011	.365	4.228	5.795
18	Disc	Amp + IAD + Ref	12	5.823	.371	5.027	6.618

Table D.10: Descriptive Statistics: Audio Cues × Engagement

Descriptives results for Engagement ratings × Audio cue, averaged across all of the sound sources (and participants). The columns are labeled as condition number; condition name; number of trials; mean; standard error; and 95% confidence intervals for the mean.

CAR Condition Pair	Mean Difference	Std. Error	Sig.	95% Confidence Interval Lower	Upper
Image Only × Image + Sound:					
Image Only × Sound Only	-2.625*	.467	.024	-5.021	-.229
Image Only × Amp	-3.359*	.461	.002	-5.728	-.990
Image Only × IAD	-2.680*	.446	.014	-4.971	-.389
Image Only × Ref	-2.804*	.465	.013	-5.189	-.419
Image Only × Amp + IAD	-3.804*	.419	.000	-5.956	-1.652
Image Only × Amp + Ref	-3.445*	.423	.001	-5.614	-1.276
Image Only × IAD + Ref	-3.582*	.304	.000	-5.140	-2.024
Image Only × Amp + IAD + Ref	-3.943*	.444	.000	-6.220	-1.666
Sound Only × Single Audio Cues:					
Sound Only × Amp	-.734	.222	1.000	-1.875	.406
Sound Only × IAD	-.055	.178	1.000	-.968	.858
Sound Only × Ref	-.179	.302	1.000	-1.727	1.369
Sound Only × Multiple Audio Cues:					
Sound Only × Amp + IAD	-1.179	.280	.225	-2.618	.260
Sound Only × Amp + Ref	-.820	.344	1.000	-2.586	.946
Sound Only × IAD + Ref	-.957	.220	.174	-2.083	.170
Sound Only × Amp + IAD + Ref	-1.318	.408	1.000	-3.415	.778
Single × Single Audio Cues:					
Amp × IAD	.679	.158	.197	-.134	1.492
Amp × Ref	.555	.371	1.000	-1.352	2.462
IAD × Ref	-.124	.366	1.000	-2.001	1.753
Single × Multiple Audio Cues:					
Amp × Amp + IAD	-.445	.374	1.000	-2.365	1.475
Amp × Amp + Ref	-.086	.391	1.000	-2.094	1.922
Amp × IAD + Ref	-.223	.266	1.000	-1.590	1.145
IAD × Amp + IAD	-1.124	.394	1.000	-3.147	.899
IAD × Amp + Ref	-.765	.359	1.000	-2.609	1.079
IAD × IAD + Ref	-.902	.280	1.000	-2.338	.534
Ref × Amp + IAD	-1.000	.195	.051	-2.003	.003
Ref × Amp + Ref	-.641	.214	1.000	-1.741	.460
Ref × IAD + Ref	-.778	.256	1.000	-2.089	.534
Amp × Amp + IAD + Ref	-.584	.489	1.000	-3.094	1.926
IAD × Amp + IAD + Ref	-1.263	.452	1.000	-3.586	1.059
Ref × Amp + IAD + Ref	-1.139*	.168	.005	-2.001	-.277
Multiple × Multiple Audio Cues:					
Amp + IAD × Amp + Ref	.359	.294	1.000	-1.151	1.870
Amp + IAD × IAD + Ref	.222	.167	1.000	-.633	1.078
Amp + Ref × IAD + Ref	-.137	.287	1.000	-1.611	1.338
Amp + IAD × Amp + IAD + Ref	-.139	.277	1.000	-1.559	1.281
Amp + Ref × Amp + IAD + Ref	-.498	.215	1.000	-1.602	.605
IAD + Ref × Amp + IAD + Ref	-.362	.323	1.000	-2.022	1.299

Table D.11: Pairwise Comparisons: Car Audio Cues × Engagement

The pairwise comparisons of Engagement × Car Audio Cue condition averaged across all of the sound sources (and participants). The * indicates the conditions where the mean difference is significant at $\alpha = 0.05$. A Bonferroni adjustment was applied to correct for a possible increase in type 1 errors associated with multiple comparisons.

DISC Condition Pair	Mean Difference	Std. Error	Sig.	95% Confidence Interval Lower	Upper
Image Only × Image + Sound:					
Image Only × Sound Only	-1.819	.411	.156	-3.928	.290
Image Only × Amp	-2.113	.466	.131	-4.506	.281
Image Only × IAD	-2.153	.458	.099	-4.504	.197
Image Only × Ref	-1.793	.378	.093	-3.733	.148
Image Only × Amp + IAD	-2.582	.511	.056	-5.205	.040
Image Only × Amp + Ref	-2.723	.583	.105	-5.717	.271
Image Only × IAD + Ref	-2.208	.508	.177	-4.814	.397
Image Only × Amp + IAD + Ref	-3.028*	.440	.004	-5.284	-.772
No Audio Cues (Sound Only) × Single Audio Cues:					
Sound Only × Amp	-.293	.301	1.000	-1.836	1.250
Sound Only × IAD	-.334	.309	1.000	-1.921	1.253
Sound Only × Ref	.027	.157	1.000	-.782	.835
Sound Only × Multiple Audio Cues:					
Sound Only × Amp + IAD	-.763	.344	1.000	-2.528	1.001
Sound Only × Amp + Ref	-.904	.304	1.000	-2.465	.657
Sound Only × IAD + Ref	-.389	.133	1.000	-1.074	.296
Sound Only × Amp + IAD + Ref	-1.209	.278	.178	-2.637	.219
Single × Single Audio Cues:					
Amp × IAD	-.041	.410	1.000	-2.147	2.065
Amp × Ref	.320	.182	1.000	-.614	1.254
IAD × Ref	.361	.313	1.000	-1.246	1.967
Single × Multiple Audio Cues:					
Amp × Amp + IAD	-.470	.112	.226	-1.044	.104
Amp × Amp + Ref	-.611	.232	1.000	-1.802	.580
Amp × IAD + Ref	-.096	.255	1.000	-1.403	1.211
IAD × Amp + IAD	-.429	.482	1.000	-2.905	2.046
IAD × Amp + Ref	-.570	.526	1.000	-3.272	2.132
IAD × IAD + Ref	-.055	.346	1.000	-1.832	1.722
Ref × Amp + IAD	-.790	.230	.844	-1.969	.389
Ref × Amp + Ref	-.931	.251	.530	-2.220	.358
Ref × IAD + Ref	-.416	.171	1.000	-1.292	.460
Amp × Amp + IAD + Ref	-.916	.180	.055	-1.842	.010
IAD × Amp + IAD + Ref	-.875	.447	1.000	-3.170	1.420
Ref × Amp + IAD + Ref	-1.236*	.163	.002	-2.071	-.401
Multiple × Multiple Audio Cues:					
Amp + IAD × Amp + Ref	-.141	.212	1.000	-1.230	.948
Amp + IAD × IAD + Ref	.374	.276	1.000	-1.044	1.793
Amp + Ref × IAD + Ref	.515	.233	1.000	-.680	1.710
Amp + IAD × Amp + IAD + Ref	-.446	.207	1.000	-1.511	.619
Amp + Ref × Amp + IAD + Ref	-.305	.185	1.000	-1.256	.646
IAD + Ref × Amp + IAD + Ref	-.820	.271	1.000	-2.211	.571

Table D.12: Pairwise Comparisons: Disc Audio Cues × Engagement

The pairwise comparisons of Engagment × Disc Audio Cue condition averaged across all of the sound sources (and participants). The * indicates the conditions where the mean difference is significant at $\alpha = 0.05$. A Bonferroni adjustment was applied to correct for a possible increase in type 1 errors associated with multiple comparisons.

Condition Pair	Mean Difference	Std. Error	Sig.	95% Confidence Interval Lower	Upper
Car Image Only × Disc Audio Cues:					
Image Only × Image Only	-.333	.355	1.000	-2.157	1.491
Sound Only × Sound Only	.473	.328	1.000	-1.211	2.156
Amp × Amp	.913	.529	1.000	-1.803	3.630
IAD × IAD	.193	.476	1.000	-2.250	2.637
Ref × Ref	.678	.428	1.000	-1.516	2.873
Amp + IAD × Amp + IAD	.888	.521	1.000	-1.788	3.565
Amp + Ref × Amp + Ref	.388	.389	1.000	-1.606	2.383
IAD + Ref × IAD + Ref	1.040	.256	.284	-.272	2.352
Amp + IAD + Ref × Amp + IAD + Ref	.582	.447	1.000	-1.714	2.877

Table D.13: Pairwise Comparisons: Car Multiple Audio cues × Disc Multiple Audio Cues × Engagement

The pairwise comparisons of Car Multiple Audio Cues Engagement rating × Disc Audio Cues ratings averaged across all of the sound sources (and participants). The * indicates the conditions where the mean difference is significant at $\alpha = 0.05$. A Bonferroni adjustment was applied to correct for a possible increase in type 1 errors associated with multiple comparisons.

Condition #	Image	Sound	N	Mean	Std. Error	95% Confidence Interval for mean Lower	Upper
1	Car	-	13	1257.528	68.709	1107.825	1407.231
2	Disc	-	13	1822.595	188.417	1412.070	2233.119
3	Car	Car	13	314.975	129.602	32.596	597.353
4	Disc	Car	13	381.115	115.661	129.111	633.120
5	Car	Noise	13	629.291	129.781	346.523	912.058
6	Disc	Noise	13	652.810	121.881	387.254	918.366
7	Car	Square	13	558.445	69.709	406.561	710.328
8	Disc	Square	13	520.195	90.051	323.990	716.401

Table D.14: Descriptive Statistics: In-/Congruent Presentation × Time-to-Contact

Descriptives results for Time-to-contact × Sound Source, averaged across all of the audio cues (and participants). The columns are labeled as condition number; condition name; number of trials; mean; standard error; and 95% confidence intervals for the mean.

Condition Pair	Mean Difference	Std. Error	Sig.	95% Confidence Interval Lower	Upper
Image only × Image Only:					
Image Only - Car × Image Only - Disc	-565.067	141.546	.050	-1130.185	.051
Image only (Car) × Sound + Image:					
Image Only - Car × Car-Car	942.553*	155.799	.002	320.527	1564.579
Image Only - Car × Car-Disc	876.412*	120.991	.000	393.359	1359.465
Image Only - Car × Noise-Car	628.237*	135.640	.016	86.699	1169.775
Image Only - Car × Noise-Disc	604.718*	126.052	.012	101.459	1107.976
Image Only - Car × Square-Car	699.083*	107.237	.001	270.943	1127.223
Image Only - Car × Square-Disc	737.332*	125.204	.002	237.458	1237.207
Image only (Disc) × Sound + Image:					
Image Only - Disc × Car-Car	1507.620*	228.201	.001	596.533	2418.707
Image Only - Disc × Car-Disc	1441.479*	173.061	.000	750.539	2132.420
Image Only - Disc × Noise-Car	1193.304*	186.319	.001	449.429	1937.179
Image Only - Disc × Noise-Disc	1169.785*	200.935	.002	367.556	1972.013
Image Only - Disc × Square-Car	1264.150*	188.420	.001	511.888	2016.412
Image Only - Disc × Square-Disc	1302.399*	196.090	.001	519.513	2085.285
Sound + Image: × Sound + Image:					
Car-Car × Noise-Car	-314.316*	65.054	.012	-574.044	-54.588
Car-Car × Square-Car	-243.470	82.924	.349	-574.542	87.602
Car-Disc × Noise-Disc	-271.695*	66.373	.042	-536.689	-6.701
Car-Disc × Square-Disc	-139.080	67.232	1.000	-407.504	129.344
Car-Car × Noise-Disc	-337.835*	63.286	.005	-590.504	-85.166
Car-Car × Square-Disc	-205.221	66.098	.255	-469.115	58.673
Car-Disc × Noise-Car	-248.175*	39.795	.001	-407.055	-89.295
Car-Disc × Square-Car	-177.329	72.839	.881	-468.137	113.478
Noise-Car × Square-Car	70.846	78.094	1.000	-240.943	382.636
Noise-Car × Square-Disc	109.095	72.483	1.000	-180.293	398.483
Noise-Disc × Square-Car	94.365	98.728	1.000	-299.804	488.535
Noise-Disc × Square-Disc	132.615	84.785	1.000	-205.888	471.117
Car-Car × Car-Disc	-66.141	59.682	1.000	-304.420	172.139
Noise-Car × Noise-Disc	-23.519	79.559	1.000	-341.159	294.120
Square-Car × Square-Disc	38.249	38.079	1.000	-113.778	190.277

Table D.15: Pairwise Comparisons: In-/Congruent Presentation × Time-to-Contact

The pairwise comparisons of In-/Congruent Presentation × Time-to-Contact. The * indicates the conditions where the mean difference is significant at $\alpha = 0.05$. A Bonferroni adjustment was applied to correct for a possible increase in type 1 errors associated with multiple comparisons

Condition #	Image	Sound	VALENCE N	Mean	Std. Error	95% Confidence Interval Lower	Upper	AROUSAL N	Mean	Std. Error	95% Confidence Interval Lower	Upper
1	Car	-	14	5.857	.710	4.324	7.391	12	4.667	.607	3.330	6.003
2	Disc	-	14	6.929	.267	6.353	7.504	12	6.167	.474	5.123	7.210
3	Car	Car	14	8.136	.557	6.934	9.339	12	9.433	.258	8.865	10.001
4	Disc	Car	14	6.929	.367	6.136	7.721	12	8.550	.283	7.926	9.174
5	Car	Noise	14	6.489	.489	5.432	7.546	12	8.452	.179	8.059	8.845
6	Disc	Noise	14	6.721	.363	5.937	7.504	12	8.437	.269	7.845	9.028
7	Car	Square	14	4.972	.288	4.350	5.594	12	8.882	.213	8.413	9.350
8	Disc	Square	14	5.661	.517	4.543	6.778	12	9.569	.173	9.188	9.950

Table D.16: Descriptive Statistics: In-/Congruent Presentation × Valence / Arousal

Descriptives results for Valence / Arousal × Sound Source, averaged across all of the audio cues (and participants). The columns are labeled as condition number; condition name; number of trials; mean; standard error; and 95% confidence intervals for the mean.

Condition Pair	VALENCE Mean Diff.	Std. Error	Sig.	95% Conf. Interval Lower	Upper	AROUSAL Mean Diff.	Std. Error	Sig.	95% Conf. Interval Lower	Upper
Image only × Image Only:										
Car × Disc	-1.07	0.77	1	-4.1	1.96	-1.500	.515	.395	-3.607	.607
Image only (Car) × Sound + Image:										
Car × Car-Car	-2.28	1.06	1	-6.41	1.85	-4.767*	.683	.001	-7.560	-1.973
Car × Car-Disc	-1.07	0.92	1	-4.68	2.54	-3.883*	.473	.000	-5.820	-1.947
Car × Noise-Car	-0.63	1.09	1	-4.91	3.65	-3.785*	.524	.000	-5.929	-1.641
Car × Noise-Disc	-0.86	0.95	1	-4.58	2.86	-3.770*	.545	.001	-6.001	-1.539
Car × Square-Car	0.89	0.84	1	-2.41	4.18	-4.215*	.549	.000	-6.461	-1.969
Car × Square-Disc	0.2	1.04	1	-3.88	4.27	-4.903*	.627	.000	-7.467	-2.338
Image only (Disc) × Sound + Image:										
Disc × Car	1.07	0.77	1	-1.96	4.1	1.500	.515	.395	-.607	3.607
Disc × Car-Car	-1.21	0.63	1	-3.67	1.26	-3.267*	.377	.000	-4.810	-1.724
Disc × Car-Disc	0	0.53	1	-2.07	2.07	-2.383*	.423	.004	-4.114	-.652
Disc × Noise-Car	0.44	0.59	1	-1.85	2.73	-2.285*	.377	.002	-3.829	-.741
Disc × Noise-Disc	0.21	0.5	1	-1.75	2.16	-2.270*	.525	.034	-4.419	-.121
Disc × Square-Car	1.956*	0.46	0.03	0.16	3.75	-2.715*	.380	.001	-4.268	-1.162
Disc × Square-Disc	1.27	0.63	1	-1.2	3.74	-3.403*	.384	.000	-4.972	-1.833
Sound + Image: × Sound + Image:										
Car-Car × Noise-Car	1.65	0.63	0.61	-0.83	4.12	.982*	.220	.027	.082	1.881
Car-Car × Square-Car	3.164*	0.51	0	1.15	5.18	.552	.167	.199	-.133	1.236
Car-Disc × Noise-Disc	0.21	0.18	1	-0.49	0.9	.113	.153	1.000	-.511	.738
Car-Disc × Square-Disc	1.27	0.32	0.05	0	2.54	-1.019	.260	.068	-2.084	.046
Car-Car × Noise-Disc	1.42	0.54	0.57	-0.68	3.51	.997	.379	.655	-.554	2.547
Car-Car × Square-Disc	2.476*	0.2	0	1.69	3.26	-.136	.144	1.000	-.723	.451
Car-Disc × Noise-Car	0.44	0.33	1	-0.86	1.74	.098	.163	1.000	-.569	.766
Car-Disc × Square-Car	1.956*	0.26	0	0.96	2.96	-.332	.271	1.000	-1.439	.776
Noise-Car × Square-Car	1.516*	0.35	0.02	0.15	2.88	-.430	.133	.228	-.976	.116
Noise-Car × Square-Disc	0.83	0.54	1	-1.29	2.95	-1.117*	.135	.000	-1.671	-.564
Noise-Disc × Square-Car	1.749*	0.23	0	0.84	2.66	-.445	.270	1.000	-1.548	.658
Noise-Disc × Square-Disc	1.06	0.43	0.83	-0.64	2.76	-1.133	.287	.065	-2.308	.043
Car-Car × Car-Disc	1.21	0.4	0.28	-0.36	2.78	.883	.365	.955	-.611	2.378
Noise-Car × Noise-Disc	-0.23	0.21	1	-1.03	0.57	.015	.201	1.000	-.809	.839
Square-Car × Square-Disc	-0.69	0.4	1	-2.24	0.87	-.688*	.166	.046	-1.367	-.008

Table D.17: Pairwise Comparisons: In-/Congruent Presentation × Valence / Arousal

The pairwise comparisons of In-/Congruent Presentation × Valence / Arousal. The * indicates the conditions where the mean difference is significant at $\alpha = 0.05$. A Bonferroni adjustment was applied to correct for a possible increase in type 1 errors associated with multiple comparisons

Condition #	Image	Sound	N	Mean	Std. Error	95% Confidence Interval for mean Lower	Upper
1	Car	-	13	2.000	.253	1.448	2.552
2	Disc	-	13	2.308	.237	1.791	2.824
3	Car	Car	13	5.643	.424	4.719	6.567
4	Disc	Car	13	4.655	.349	3.894	5.417
5	Car	Noise	13	4.685	.402	3.810	5.561
6	Disc	Noise	13	4.335	.341	3.592	5.077
7	Car	Square	13	4.942	.390	4.091	5.792
8	Disc	Square	13	4.872	.311	4.195	5.549

Table D.18: Descriptive Statistics: In-/Congruent Presentation × Engagement

Descriptives results for Sound Source and Image Presentation × Engagement rating, averaged across all of the audio cues (and participants). The columns are labeled as condition number; condition name; number of trials; mean; standard error; and 95% confidence intervals for the mean.

Condition Pair	Mean Difference	Std. Error	Sig.	95% Confidence Interval Lower	Upper
Image only × Image Only:					
Image Only - Car × Image Only - Disc	-.308	.328	1.000	-1.617	1.001
Image only (Car) × Sound + Image:					
Image Only - Car × Car-Car	-3.643*	.354	.000	-5.058	-2.228
Image Only - Car × Car-Disc	-2.655*	.302	.000	-3.860	-1.451
Image Only - Car × Noise-Car	-2.685*	.377	.000	-4.192	-1.179
Image Only - Car × Noise-Disc	-2.335*	.246	.000	-3.317	-1.353
Image Only - Car × Square-Car	-2.942*	.486	.002	-4.880	-1.003
Image Only - Car × Square-Disc	-2.872*	.375	.000	-4.369	-1.375
Image only (Disc) × Sound + Image:					
Image Only - Disc × Car-Car	-3.335*	.520	.001	-5.413	-1.258
Image Only - Disc × Car-Disc	-2.348*	.495	.013	-4.324	-.371
Image Only - Disc × Noise-Car	-2.378*	.532	.021	-4.500	-.255
Image Only - Disc × Noise-Disc	-2.027*	.475	.031	-3.924	-.130
Image Only - Disc × Square-Car	-2.634*	.497	.005	-4.619	-.648
Image Only - Disc × Square-Disc	-2.565*	.388	.001	-4.113	-1.016
Sound + Image: × Sound + Image:					
Car-Car × Noise-Car	.958	.342	.448	-.407	2.323
Car-Car × Square-Car	.702	.441	1.000	-1.058	2.461
Car-Disc × Noise-Disc	.321	.083	.066	-.013	.654
Car-Disc × Square-Disc	-.217	.373	1.000	-1.708	1.274
Car-Car × Noise-Disc	1.308	.494	.595	-.664	3.281
Car-Car × Square-Disc	.771	.477	1.000	-1.132	2.674
Car-Disc × Noise-Car	-.030	.399	1.000	-1.622	1.562
Car-Disc × Square-Car	-.286	.511	1.000	-2.328	1.756
Noise-Car × Square-Car	-.256	.239	1.000	-1.211	.698
Noise-Car × Square-Disc	-.187	.244	1.000	-1.160	.786
Noise-Disc × Square-Car	-.607	.540	1.000	-2.764	1.550
Noise-Disc × Square-Disc	-.538	.401	1.000	-2.139	1.064
Car-Car × Car-Disc	.988	.524	1.000	-1.105	3.080
Noise-Car × Noise-Disc	.351	.412	1.000	-1.296	1.997
Square-Car × Square-Disc	.069	.234	1.000	-.865	1.003

Table D.19: Pairwise Comparisons: In-/Congruent Presentation × Engagement.

The pairwise comparisons of In-/Congruent Presentation × Engagement ratings. The * indicates the conditions where the mean difference is significant at $\alpha = 0.05$. A Bonferroni adjustment was applied to correct for a possible increase in type 1 errors associated with multiple comparisons

Appendix E

Digital Appendix Contents

Chapter 3 Experiment 1

- Feature Analysis Plots
  - Amplitude Envelope Slope & Levels
  - Audio Virtual Source Position
  - Image Motion Tracking
  - Spectrogram, Spectral Centroid & Spectral Spread
- Film Looming Scenes

Chapter 4 Experiment 2

- Experiment Stimuli
- Experiment Documents
  - Participant Information Sheet
  - Participant Consent Form
  - Participant Questionnaire

Chapter 5 Experiment 3

- Experiment Stimuli
- Experiment 3 & 4 Documents
  - Participant Information Sheet
  - Participant Consent Form
  - Participant Questionnaire

Chapter 6 Experiment 4

- Experiment Stimuli

Bibliography

Ashe, S. (2011). Follow-up report: President obama signs pedestrian safety enhancement act into law.

Bach, D. R., Neuhoff, J. G., Perrig, W., and Seifritz, E. (2009). Looming sounds as warning signals: The function of motion cues. *International Journal of Psychophysiology*, 74(1):28–33.

Ball, W. and Tronick, E. (1971). Infant responses to impending collision: Optical and real. *Science; Science*.

Begault, D. R. et al. (1994). *3-D sound for virtual reality and multimedia*, volume 955. Ap Professional Boston etc.

Berry, J. and Holliman, N. (2013). Electronic imaging & signal processing using auditory depth cues to enhance stereoscopic visual media. *SPIE*.

Boeriu, H. (2011). Active sound design brings the m5 engine sound into the cabin.

Boev, A. and Gotchev, A. (2011). Comparative study of autostereoscopic displays for mobile devices. In *IS&T/SPIE Electronic Imaging*, pages 78810B–78810B. International Society for Optics and Photonics.

Bregman, A. S. (1994). *Auditory scene analysis: The perceptual organization of sound*. MIT press.

Brenner, E., Van den Berg, A., and Van Damme, W. (1996). Perceived motion in depth. *Vision Research*, 36(5):699–706.

Bronkhorst, A. W. (1995). Localization of real and virtual sound sources. *The Journal of the Acoustical Society of America*, 98:2542.

Bronkhorst, A. W. (2002). Modeling auditory distance perception in rooms. In *Proc. EAA Forum Acusticum Sevilla*.

Bronkhorst, A. W. and Houtgast, T. (1999). Auditory distance perception in rooms. *Nature*, 397(6719):517–520.

Brunswik, E. (1943). Organismic achievement and environmental probability. *Psychological review*, 50(3):255.

Brunswik, E. (1955). Representative design and probabilistic theory in a functional psychology. *Psychological review*, 62(3):193.

Brunswik, E. and Kamiya, J. (1953). Ecological cue-validity of 'proximity' and of other gestalt factors. *The American journal of psychology*, pages 20–32.

Caird, J. and Hancock, P. (1994). The perception of arrival time for different oncoming vehicles at an intersection. *Ecological Psychology*, 6(2):83–109.

Calabro, F. J., Beardsley, S. A., and Vaina, L. M. (2011). Different motion cues are used to estimate time-to-arrival for frontoparallel and looming trajectories. *Vision research.*

Cappe, C., Thelen, A., Romei, V., Thut, G., and Murray, M. M. (2012). Looming signals reveal synergistic principles of multisensory integration. *The Journal of Neuroscience*, 32(4):1171–1182.

Cappe, C., Thut, G., Romei, V., and Murray, M. M. (2009). Selective integration of auditory-visual looming cues by humans. *Neuropsychologia; Neuropsychologia*, 47:1045–1052.

Carlile, P. A., Peters, R. A., and Evans, C. S. (2006). Detection of a looming stimulus by the jacky dragon: selective sensitivity to characteristics of an aerial predator. *Animal behaviour*, 72(3):553–562.

Chion, M. (1994). *Audio-vision: sound on screen.* Columbia University Press.

Coleman, P. D. (1968). Dual role of frequency spectrum in determination of auditory distance. *The Journal of the Acoustical Society of America*, 44:631.

Collantine, K. (2014). What f1 spectators say about the engine noise debate. *F1 Fanatic.*

Colombo, C. (2000). Time to collision from first-order spherical image motion. *Robotics and Autonomous Systems*, 31(1):5–15.

Cornilleau-Pérès, V., Wexler, M., Droulez, J., Marin, E., Miège, C., and Bourdoncle, B. (2002). Visual perception of planar orientation: dominance of static depth cues over motion cues. *Vision research*, 42(11):1403–1412.

Cowan, M. and Officer, C. S. (2007). Reald 3d system.

Dalenbäck, B. (2006). Catt-acoustic.

Devore, S. and Shinn-Cunningham, B. (2003). Perceptual consequences of including reverberation in spatial auditory displays. In *Proc of International Conference on Auditory Displays*, pages 75–78.

Dill, L. M. (1974). The escape response of the zebra danio (¡ i¿ brachydanio rerio¡/i¿) i. the stimulus for escape. *Animal Behaviour*, 22(3):711–722.

Driver, J. and Spence, C. (2000). Multisensory perception: beyond modularity and convergence. *Current Biology*, 10(20):R731–R735.

Ellensburg, W., Blau, D., and Bronk, T. (2009). Visual looming effect in the landscape: Research, analysis, and case study. *AECOM.*

Ericson, M. A. (2007). Detection and discrimination of approaching and receding puretones. *ICAD.*

Farnell, A. (2010). *Designing sound.* MIT Press.

Franconeri, S. L. and Simons, D. J. (2003). Moving and looming stimuli capture attention. *Attention, Perception, & Psychophysics*, 65(7):999–1010.

Gabbiani, F., Krapp, H. G., Koch, C., and Laurent, G. (2002). Multiplicative computation in a visual neuron sensitive to looming. *Nature*, 420(6913):320–4.

Gauduin, B. and Boussard, P. (2009). High fidelity sound rendering for car driving simulators. In *Proceedings of Driving Simulation Conference DSC*, pages 283–294.

Ghazanfar, A. A., Neuhoff, J. G., and Logothetis, N. K. (2002). Auditory looming perception in rhesus monkeys. *Proceedings of the National Academy of Sciences*, 99(24):15755–15757.

Gibson, J. J. (2013). *The ecological approach to visual perception*. Psychology Press.

Gibson, J. J., Reed, E. S., and Jones, R. (1982). *Reasons for realism: Selected essays of James J. Gibson*. Lawrence Erlbaum Associates.

Golonka, S. and Wilson, A. D. (2012). Gibson's ecological approach-a model for the benefits of a theory driven psychology.

González, E. G., Allison, R. S., Ono, H., and Vinnikov, M. (2010). Cue conflict between disparity change and looming in the perception of motion in depth. *Vision research*, 50(2):136–143.

Goodwin, A. (2011a). Bmw m5 generates fake engine noise using stereo.

Goodwin, A. (2011b). Prius' artificial engine noise demonstrated, explained.

Gordon, M. S. and Rosenblum, L. D. (2005). Effects of intrastimulus modality change on audiovisual time-to-arrival judgments. *Perception & psychophysics*, 67(4):580–594.

Grassi, M. (2010). Sex difference in subjective duration of looming and receding sounds. *Perception*, 39(10):1424.

Gray, R. (2011). Looming auditory collision warnings for driving. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 53(1):63–74.

Gray, R. and Regan, D. (1998). Accuracy of estimating time to collision using binocular and monocular information. *Vision Research; Vision Research*, 38(4):499–512.

Gray, R. and Regan, D. (1999). Adapting to expansion increases perceived time-to-collision. *Vision research*, 39(21):3602–3607.

Gray, R. and Regan, D. (2000). Estimating the time to collision with a rotating nonspherical object. *Vision Research*, 40(1):49–63.

Griesinger, D. (2009). The importance of the direct to reverberant ratio in the perception of distance, localization, clarity, and envelopment. In *Audio Engineering Society Convention 126*, pages 7724–7736.

Hammond, K. R. (1998). Ecological validity: Then and now. *Unpublished manuscript available electronically at: http: 00brunswik. org0notes0essay2. html.*

Hancock, P. A. and Manster, M. (1997). Time-to-contact: More than tau alone. *Ecological Psychology*, 9(4):265–297.

Harris, C. M. (1966). Absorption of sound in air versus humidity and temperature. *The Journal of the Acoustical Society of America*, 40(1):148–159.

Hayes, W. N. and Saiff, E. I. (1967). Visual alarm reactions in turtles. *Animal Behaviour*, 15(1):102–106.

Holman, T. (2010). *Sound for film and television*. Taylor & Francis.

Hong, X. and Regan, D. (1989). Visual field defects for unidirectional and oscillatory motion in depth. *Vision Research*, 29(7):809–819.

Horswill, M. S., Helman, S., Ardiles, P., and Wann, J. P. (2005). Motorcycle accident risk could be inflated by a time to arrival illusion. *Optometry & Vision Science*, 82(8):740–746.

Howard, I. P. (2012). *Perceiving in Depth, Volume 1: Basic Mechanisms: Basic Mechanisms*, volume 3. Oxford University Press, USA.

Illingworth, V. (2004). *The Penguin dictionary of physics.* Penguin.

Ingård, U. (1953). A review of the influence of meteorological conditions on sound propagation. *The Journal of the Acoustical Society of America*, 25:405.

ISO, D. (1993). 9613–1: 1993. acoustics. attenuation of sound during propagation outdoors. part 1: Calculation of the absorption of sound by the atmosphere. *International Organization for Standardization, Geneva.*

Izhaki, R. (2009). *Mixing audio: concepts, practices and tools.* Elsevier Academic Press, Oxford.

Kahan, T. A., Colligan, S. M., and Wiedman, J. N. (2011). Are visual features of a looming or receding object processed in a capacity-free manner? *Consciousness and cognition*, 20(4):1761–1767.

Katz, B. and Katz, R. A. (2007). *Mastering audio: the art and the science.* Focal Press.

Khuu, S. K. and Lee, Terence CP anzahod Hayes, A. (2010). Object speed derived from the integration of motion in the image plane and motion-in-depth signaled by stereomotion and looming. *Vision research*, 50(9):904–913.

King Jr, J. G., Lettvin, J. Y., and Gruberg, E. R. (1999). Selective, unilateral, reversible loss of behavioral responses to looming stimuli after injection of tetrodotoxin or cadmium chloride into the frog optic nerve. *Brain research*, 841(1):20–26.

Kraus, N. and Nicol, T. (2005). Brainstem origins for cortical what and where pathways in the auditory system. *Trends in neurosciences*, 28(4):176–181.

Lartillot, O. and Toiviainen, P. (2007). A matlab toolbox for musical feature extraction from audio. In *Proceedings of the 10th International Conference on Digital Audio Effects, Bordeaux, France.*

Lee, D., Young, D., Reddish, P., Lough, S., and Clayton, T. (1983). Visual timing in hitting an accelerating ball. *The Quarterly Journal of Experimental Psychology*, 35(2):333–346.

Leo, F., Romei, V., Freeman, E., Ladavas, E., and Driver, J. (2011). Looming sounds enhance orientation sensitivity for visual stimuli on the same side as such sounds. *Experimental brain research*, 213(2-3):193–201.

Licklider, J. and Webster, J. (1950). The discriminability of interaural phase relations in two-component tones. *The Journal of the Acoustical Society of America*, 22:191.

Lin, J. Y., Murray, S. O., and Boynton, G. M. (2009). Capture of attention to threatening stimuli without perceptual awareness. *Current Biology*, 19(13):1118–1122.

Lutfi, R. A. and Wang, W. (1999). Correlational analysis of acoustic cues for the discrimination of auditory motion. *The Journal of the Acoustical Society of America*, 106(2):919–928.

Maier, J. X., Chandrasekaran, C., and Ghazanfar, A. A. (2008). Integration of bimodal looming signals through neuronal coherence in the temporal lobe. *Current Biology*, 18(13):963–968.

Maier, J. X. and Ghazanfar, A. A. (2007). Looming biases in monkey auditory cortex. *The Journal of neuroscience*, 27(15):4093–4100.

Maier, J. X., Neuhoff, J. G., Logothetis, N. K., and Ghazanfar, A. A. (2004). Multisensory integration of looming signals by rhesus monkeys. *Neuron*, 43(2):177–181.

Mershon, D. H. and King, L. E. (1975). Intensity and reverberation as factors in the auditory perception of egocentric distance. *Perception & Psychophysics*, 18(6):409–415.

NASA (2013). Slab3D Home Page. `http://slab3d.sonisphere.com`. [Online; accessed 1-September-2013].

Nave, C. (2012a). Hyperphysics Website, Department of Physics and Astronomy, Georgia State University. `http://hyperphysics.phy-astr.gsu.edu/hbase/sound/dopp.html#c4`. [Online; accessed 1-Feb-2011].

Nave, C. (2012b). Hyperphysics Website, Department of Physics and Astronomy, Georgia State University. `http://hyperphysics.phy-astr.gsu.edu/hbase/acoustic/isprob.html#c3`. [Online; accessed 1-Feb-2011].

Neppi-Mòdona, M., Auclair, D., Sirigu, A., and Duhamel, J.-R. (2004). Spatial coding of the predicted impact location of a looming object. *Current biology*, 14(13):1174–1180.

Neuhoff, J. G. (1998). Perceptual bias for rising tones. *Nature*, 395(6698):123–124.

Neuhoff, J. G. (2001). An adaptive bias in the perception of looming auditory motion. *Ecological Psychology*, 13(2):87–110.

Neuhoff, J. G. (2004). *Ecological psychoacoustics*. Elsevier Academic Press Amsterdam.

Neuhoff, J. G. and Heckel, T. (2004). Sex differences in perceiving auditory looming produced by acoustic intensity change. In *Proceedings of the 10th Meeting of the International Conference on Auditory Display*.

Neuhoff, J. G. and McBeath, M. K. (1996). The doppler illusion: the influence of dynamic intensity change on perceived pitch. *Journal of Experimental Psychology: Human Perception and Performance*, 22(4):970.

Norma, I. (1996). 9613–2: Acoustic attenuation of sound during propagation outdoors–general methods of calculation. *International Standard*.

Ondaatje, M. and Murch, W. (2002). *The conversations: Walter Murch and the art of editing film*. Bloomsbury Publishing.

Parker, A. and Alais, D. (2007). A bias for looming stimuli to predominate in binocular rivalry. *Vision research*, 47(20):2661–2674.

Peters, R., Smith, B. J., and Hollins, M. (2011). *Acoustics and noise control*. Pearson Education.

Ploner-Bernard, H., Sontacchi, A., Lichtenegger, G., Vössner, S., and Braunstingl, R. (2005). Sound-system design for a professional full-flight simulator. In *Proceedings of the 8th International Conference on Digital Audio Effects (DAFx05), Madrid, Spain, September 20*, volume 22, pages 36–41.

Preuss, T., Osei-Bonsu, P. E., Weiss, S. A., Wang, C., and Faber, D. S. (2006). Neural representation of object approach in a decision-making motor circuit. *The Journal of neuroscience*, 26(13):3454–3464.

Raviv, D. and Joarder, K. (2000). The visual looming navigation cue: a unified approach. *Computer Vision and Image Understanding*, 79(3):331–363.

Regan, D. and Beverley, K. (1978). Looming detectors in the human visual pathway. *Vision Research*, 18(4):415–421.

Regan, D. and Vincent, A. (1995). Visual processing of looming and time to contact throughout the visual field. *Vision Research*, 35(13):1845–1857.

Release, T. M. C. P. (2010). Tmc to sell approaching vehicle audible system for 'prius'.

Rodrigues, J., Pinto, M., Dommes, A., Cavallo, V., and Vienne, F. (2012). Simulated traffic and auditory information: The impact on street crossing in young and old adults. *Actes INRETS*, pages 95–104.

Rosenblum, L. D., Carello, C., and Pastore, R. E. (1987). Relative effectiveness of three stimulus variables for locating a moving sound source. *Perception*, 16(2):175–186.

Rosenblum, L. D., Gordon, M. S., and Jarquin, L. (2000a). Echolocating distance by moving and stationary listeners. *Ecological Psychology*, 12(3):181–206.

Rosenblum, L. D., Gordon, M. S., and Wuestefeld, A. P. (2000b). Effects of performance feedback and feedback withdrawal on auditory looming perception. *Ecological psychology*, 12(4):273–291.

Rosenblum, L. D., Wuestefeld, A. P., and Saldana, H. M. (1993). Auditory looming perception: Influences on anticipatory judgments. *Perception*, 22:1467–1467.

Russell, J. A. (1980). A circumplex model of affect. *Journal of personality and social psychology*, 39(6):1161–1178.

Rutherford, R. E. and Fancher, A. (2012). *Pioneers of psychology : a history.* New York W.W. Norton.

Sahin, E. and Gaudiano, P. (1998a). Mobile robot range sensing through visual looming. In *Intelligent Control (ISIC), 1998. Held jointly with IEEE International Symposium on Computational Intelligence in Robotics and Automation (CIRA), Intelligent Systems and Semiotics (ISAS), Proceedings*, pages 370–375. IEEE.

Sahin, E. and Gaudiano, P. (1998b). Visual looming as a range sensor for mobile robots. Technical report.

Savelsbergh, G., Whiting, H., Pijpers, J., and Van Santvoord, A. (1993). The visual guidance of catching. *Experimental Brain Research*, 93(1):148–156.

Schiff, W. (1965). Perception of impending collision: A study of visually directed avoidant behavior. *Psychological Monographs: General and Applied*, 79(11):1–26.

Schiff, W., Caviness, J. A., and Gibson, J. J. (1962). Persistent fear responses in rhesus monkeys to the optical stimulus of" looming". *Science*, 136(3520):982–983.

Schiff, W. and Oldak, R. (1990). Accuracy of judging time to arrival: effects of modality, trajectory, and gender. *Journal of Experimental Psychology: Human Perception and Performance*, 16(2):303.

Serafin, S. and Serafin, G. (2004). Sound design to enhance presence in photorealistic virtual reality. In *ICAD*.

Sheeline, C. W. (1983). *An investigation of the effects of direct and reverberant signal interaction on auditory distance perception*. PhD thesis, Stanford University.

Shinn-Cunningham, B. (2000). Distance cues for virtual auditory space. In *Proceedings of the IEEE-PCM*, volume 2000, pages 227–230.

Stevens, R. and Raybould, D. (2013). *The Game Audio Tutorial: A Practical Guide to Creating and Implementing Sound and Music for Interactive Games*. Taylor & Francis.

Tajadura-Jiménez, A., Väljamäe, A., Asutay, E., and Västfjäll, D. (2010). Embodied auditory perception: The emotional impact of approaching and receding sound sources. *Emotion*, 10(2):216–229.

Terry, H. R., Charlton, S. G., and Perrone, J. A. (2008). The role of looming and attention capture in drivers' braking responses. *Accident analysis and prevention*, 40(4):1375–1382.

Thomas, J. A., Moss, C. F., and Vater, M. (2004). *Echolocation in bats and dolphins*. University of Chicago Press.

Turner, A., Berry, J., and Holliman, N. (2011). Can the perception of depth in stereoscopic images be influenced by 3d sound? In *IS&T/SPIE Electronic Imaging*, pages 786307–786307. International Society for Optics and Photonics.

Tyll, S., Bonath, B., Schoenfeld, M. A., Heinze, H. J., Ohl, F. W., and Noesselt, T. (2012). Neural basis of multisensory looming signals. *NeuroImage*, 65:13–22.

Ungerleider, L. G. and Pessoa, L. (2008). What and where pathways. *Scholarpedia*, 3(11):5342.

Vagnoni, E., Lourenco, S. F., and Longo, M. R. (2012). Threat modulates perception of looming visual stimuli. *Current Biology*, 22(19):R826–R827.

Valimaki, V., Parker, J. D., Savioja, L., Smith, J. O., and Abel, J. S. (2012). Fifty years of artificial reverberation. *Audio, Speech, and Language Processing, IEEE Transactions on*, 20(5):1421–1448.

Von Békésy, G. and Wever, E. G. (1960). *Experiments in hearing*, volume 8. McGraw-Hill New York.

Wang, Y.-c., Jiang, S., Frost, B. J., et al. (1993). Visual processing in pigeon nucleus rotundus: luminance, color, motion, and looming subdivisions. *Visual neuroscience*, 10:21–21.

Whiting, H., Gill, E., and Stephenson, J. (1970). Critical time intervals for taking in flight information in a ball-catching task. *Ergonomics*, 13(2):265–272.

Wiener, F. M. and Keast, D. N. (1959). Experimental study of the propagation of sound over ground. *The Journal of the Acoustical Society of America*, 31(6):724–733.

Wilkie, S. and Stockman, T. (2012). The perception of auditory-visual looming in film. *Computer Music Modelling and Retrieval Symposium*.

Wilkie, S. and Stockman, T. (2013). *The Perception of Auditory-Visual Looming in Film*. Springer.

Wilkie, S., Stockman, T., and Reiss, J. D. (2012). Amplitude manipulation for perceived movement in depth. *Audio Engineering Society Convention 132*.

Yamawaki, Y. (2011). Defence behaviours of the praying mantis¡ i¿ tenodera aridifolia¡/i¿ in response to looming objects. *Journal of Insect Physiology*, 57(11):1510–1517.

Zahorik, P. (2001). Estimating sound source distance with and without vision. *Optometry & Vision Science*, 78(5):270–275.

Zahorik, P. (2002a). Assessing auditory distance perception using virtual acoustics. *The Journal of the Acoustical Society of America*, 111(4):1832–1846.

Zahorik, P. (2002b). Direct-to-reverberant energy ratio sensitivity. *The Journal of the Acoustical Society of America*, 112(5):2110–2117.

Zahorik, P., Brungart, D. S., and Bronkhorst, A. W. (2005). Auditory distance perception in humans: A summary of past and present research. *Acta Acustica united with Acustica*, 91(3):409–420.

Zölzer, U. and Amatriain, X. (2002). *DAFX: digital audio effects*, volume 1. Wiley Online Library.

Zwislocki, J. and Feldman, R. (1956). Just noticeable differences in dichotic phase. *The Journal of the Acoustical Society of America*, 28:860.

Zyda, M. (2005). From visual simulation to virtual reality to games. *Computer*, 38(9):25–32.