Perception of Objects that Move in Depth, using Ecologically Valid Audio Cues

SONIA WILKIE and TONY STOCKMAN, Queen Mary University of London

Keywords: Auditory-visual looming, Psychoacoustics, Ecological Validity, Emotion, Depth cues, Human Factors,

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, and is dependent on the individuals capacity to accurately interpret depth and movement cues. However, 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 (such as sine or triangle 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 world stimuli.

This article presents an experiment on human perception where observers respond to objects that move in depth (on an approaching trajectory) using sounds that contain ecologically valid complex audio cues. We measure the participant's responses to the stimuli, asking them to indicate the approaching object's perceived contact time (measuring their amount of over-/underestimation); to rate their emotional (valence and arousal) responses; and to rate the engagement quality of the stimuli. Our results show that humans expressed a greater underestimation of the contact time for looming scenes which contained complex audio cues, than for scenes with no audio cues. Scenes that were rated as having greater engagement quality also correlated with greater ratings of emotion. This study provides new information on human looming perception using ecologically valid audio cues, and uses novel measurements of emotion.

1. INTRODUCTION

One feature of computer-generated environments (hyper and virtual reality, film, and gaming) is interacting with objects that move in space, particularly objects that move in depth towards the viewer. Examples can be seen in 3-D presentation where objects appear to leap out of the screen towards the viewer; and in gaming where judgements are made to avoid or attack approaching objects.

The extent to which a user can perceptually immerse within a multidimensional world and interact with moving objects is reliant on many elements. These include the effect of 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. To accurately generate a dynamic and rich perception of looming objects, the design of such complex stimuli should be based on firm scientific foundations that encompass what we know about how people visually and aurally perceive events and interactions.

Our research investigates human responses to the presentation of an object moving in depth on an approaching trajectory (auditory-visual looming), using ecologically valid stimuli that contain multiple audio cues.

1.1. Previous Research

Initial research on auditory looming found that humans associate an approaching object with at least three audio cues, namely, an increase in the amplitude, frequency change (the doppler shift), and interaural temporal differences [Rosenblum et al. 1987]. Results from Rosenblum's 1987 study also suggest that some audio cues have a greater affect on perception (and the amount of over- / under-estimation of the objects 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.

Later studies found that looming audio cues (in the form of an increase in the amplitude) created a greater underestimation of the contact time than receding audio cues (presented as a decrease in amplitude) [Neuhoff 1998, 2001; Cappe et al. 2009]. One explanation for this discrepancy in the perceived contact time suggests 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 concluded that tonal sounds (in the form of pitched sine tones) enabled easier detection of looming audio cues, than (white) noise [Neuhoff 1998, 2001; Ghazanfar et al. 2002; Maier et al. 2004].

When amplitude increase is used as an audio cue, the magnitude of the change is perceived to be greater than it physically is [Neuhoff and Heckel 2004; Neuhoff 1998, 2016] suggesting that the object is approaching at a faster rate. 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, therefore posing greater potential danger.

Many of these studies [Rosenblum et al. 1993; Neuhoff 1998, 2001; Cappe et al. 2009; Ghazanfar et al. 2002; Maier et al. 2008; Maier and Ghazanfar 2007; Maier et al. 2004] depict the approaching object using the single variable of increasing amplitude. This approach is understandable, since amplitude change has been shown to be an effective audio cue, and researchers are often motivated to increase experimental robustness through the absolute control of variables. However, the use of ecologically valid stimuli, real world sounds, and full audio cues as proposed by Gaver [1993a, 1993b] assists in building a comprehensive understanding of human perception of an objects movement.

Studies that investigate looming perception using real world sounds include Bach et al. [2009] and Tajadura-Jiménez et al. [2010], whilst studies that investigate looming perception using a 3-Dimensional virtual sound source with full spatial cues include Bach et al. [2009], Riskind et al. [2014], and Neuhoff et al. [2013, 2014]. The acoustic variables which comprise the full spatial cues include absolute delay, the Doppler shift, atmospheric filtering, gain attenuation due to atmospheric spreading, ground reflection attenuation, and HRTF's.

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 conducted on non-human primates (rhesus monkeys) [Maier e t al. 2004, 2008] and more recently extending the research to human observers [Romei et al. 2009; Cappe et al. 2009; Cappe et al. 2012; Harrison, 2012; Tyll et al. 2012; Cecere et al. 2014; Sutherland et al. 2014]. Representation of the approaching object involves the presentation of an expanding disc as the visual stimulus cue, and increasing the amplitude as a function of an auditory looming cue, with results indicating that the multisensory (auditory-visual) integration of looming information is occurring.

Whilst these studies have uncovered important information on the neural activity and mechanisms that underpin the cross-modal processing of auditory and visual information, the looming stimuli itself is somewhat abstract. If real world stimuli were to be used, it may alter (either by increasing or decreasing) any neural activity occurring between these two modalities, and will provide results on how people process information in real world scenarios.

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. A number of recent studies have begun to measure this factor [Bach et al. 2009; Tajadura- Jiménez et al. 2010] finding that approaching sounds were rated as more unpleasant (valence), and arousing, than receding sounds.

The Tajadura-Jiménez et. al. [2010] study had particularly interesting results, finding that when an unpleasant target image was paired with an approaching audio cue (increasing in amplitude only), the observers not only had faster response times to the negative target image, but also expressed greater arousal and unpleasantness, than when the negative target image was paired with a receding audio cue (decreasing in amplitude).

The results from these previous looming studies have provided important information on human perception and the audio cues that act as an indicator for approaching objects. However, the frequent use of single variables (often amplitude change) and simple sounds (often sine or triangle waves) invites the question how do humans perceive and respond to complex, ecologically valid looming sounds?

Information obtained would advance understanding of the audio cues involved in the motion detection of complex sounds, enable us to predict human perception and response to manipulation of the audio cues, and would also be useful for real world application.

1.2. Industry Application and Usage

In contrast to the simple audio parameters used in the scientific studies, the film and gaming industries require sound designers to use complex sounds, with the purpose of maximising the viewers experience, immersiveness, responsiveness to onscreen action, and overall perception of the virtual environment.

Many interactive games are adventure or sporting, in which case the player is faced with potential looming scenarios that require the gamer to make quick judgements on whether to attack or avoid approaching objects. The extent to which gamers progress through the game, and their continued survival, depends on the player's ability to quickly respond to approaching targets. As such, appropriate audio cues are crucial to successfully engage the player.

One of the features of 3D film presentation 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 multi-dimensional 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 experience from a passive one of motionless watching and listening, to an active one where viewers may physically move to avoid objects as an instinctive reaction to their perceived increasing proximity. Whilst the image representation of the objects movement in depth is often the focus of amazement, the generation of a rich perception of the event is dependent on the simultaneous presentation and integration of both sound and image, and the degree to which the sound accurately represents the objects movement.

Examination of people's responses to looming scenes that use complex audio cues will allow us to gain an understanding of how people respond in ecologically valid situations, and in what ways does greater sensory information cause their reactions to differ.

1.3. Feature Analysis Study

In this study we used 27 (film) looming scenes that were previously investigated with a feature analysis on the audio tracks, in order 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 features that were analysed included amplitude change, levels, slope, pan position, spectral centroid, spread, flux, roll-off, and image motion tracking of the object.

In summary, the analysis showed a number of changes in the features that were consistent among the variety of samples. This includes:

- Amplitude increases of an average of 45.05dB (SD = 15.32) on a linear / near-linear slope.
- An average spectral centroid *M* = 1957.8Hz (B6 16 cents) at the start of the sample, and an average peak at *M* = 3444.57Hz (A7 38 cents), an increase of 1486.77Hz (almost one octave).
- The pan position centrally placed, and close to the image position, however fluctuates more than the image position. This fluctuation emphasises the spatial movement without having to hard pan to a single channel.

In contrast to the previous auditory looming studies, the feature analysis of the film samples showed that the sounds designed for industry have:

- A greater range of variables used simultaneously to form complex looming stimuli (compared to the simple waves in the psychoacoustic studies, which often only increased the amplitude).
- A greater increase in the variable levels (ie 45.05dB amplitude increase in the film samples, versus the 10 - 30dB increase in the psychoacoustic studies).
- A rise in the spectral centroid, as opposed to a fall in the psychoacoustic studies.
- Less extreme variation in the spatial movement (the psychoacoustic studies tended to have the spatial movement hard panning from one channel to the other).

2. EXPERIMENT

In this study we examine human responses to the looming stimuli that use multiple complex audio cues from film scenes.

2.1. Aim

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

2.2. Hypothesis

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.

2.3. Method

2.3.1. 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:

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

2.3.2. 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.

2.3.3. Stimuli.

The stimuli consisted of 27 film scenes (listed in Appendix Table 1) 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 scenes ≤ 1000ms in duration. 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.

2.3.4. 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.

2.3.5. Dependent Variable Measurement.

Four dependent variable measurements were made, these are:

- **Perceived 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 1 Participant's response task to the stimuli.
Participant's pressed the keyboard space bar when they thought the object would reach them.

Using the below equation, 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 'Perceived Time-to-Impact'.

$$PTI = RT - P$$

where:

PTI = Perceived 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).

- **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.

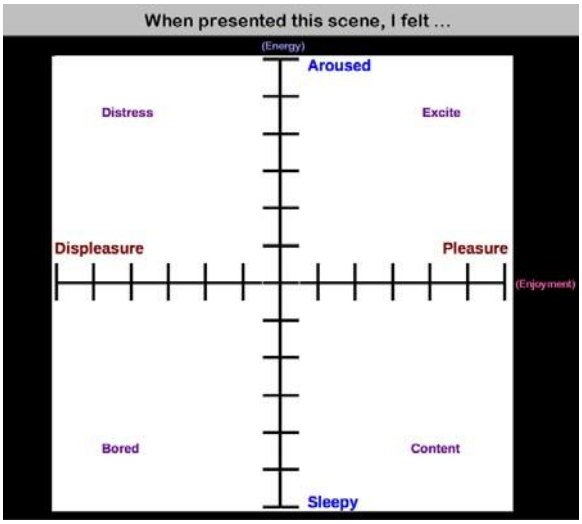

Figure 2. Valence / Arousal 2D Rating Scale .
Valence is measured on the X axis with 13-points ranging from displeasure to pleasure, whilst arousal is measured on the Y axis with 13-points ranging from sleepy to aroused. The minimum and maximum combinations of the valance / arousal sees the quadrants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

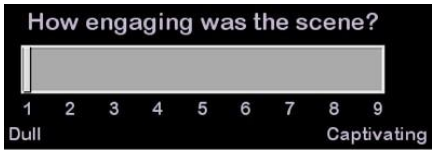

Figure 3 Engagement Rating Scale.
A 9-point visual analogue scale ranging from dull to captivating.

2.3.6. Procedure.

Participants sat at the computer workstation and were informed of the experiment 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.

Before commencing the experiment, the participants completed a practise test 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.

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

Each trial lasted for a total duration of 0.3 to 3.0 seconds (depending on the looming scene presented) and the participants 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' was displayed on the screen, and no sound output the headphones. The experiment lasted for approximately 25 minutes and participants were not given any information implying there might be 'correct', 'incorrect' or 'preferred' responses.

3. RESULTS

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

ANOVAS are sensitive to outliers therefore preliminary analyses were conducted on the data to check for outliers. Whilst outliers can provide interesting insights into human perception and action, as Ratcliff [p.510, 1993] noted in his investigation of reaction time outliers

> "*The processes that generate outliers can be fast guesses, guesses that are based on the subject's estimate of the usual time to respond, multiple runs of the process that is actually under study, the subject's inattention, or guesses based on the subject's failure to reach a decision.*".

Ratcliff [p.531, 1993] further recommended that "*…standard deviation cutoffs (depending on variability of subject means) be used to confirm more traditional analyses*". Therefore, any data points that were ± 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 perceived time-to-impact, 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 (using Tukey's HSD) with pairwise comparisons between the conditions were also conducted for each ANOVA. The Bonferroni adjustment was applied to correct for a possible increase in type 1 (false positive) errors associated with multiple comparisons). The descriptives are provided in the appendices, whilst in each analysis section we discuss the comparisons between the conditions and if the results support the hypothesis.

3.1. 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 over- estimation 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.

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

3.1.1. Presentation × Perceived Time-to-Impact

The perceived time-to-impact was averaged across all of the participants responses, for each scene sample presentation condition, and is plotted in Figure 4.

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 perceived time-to-impact was then averaged across all of the participants responses and scenes, for each presentation condition, and is plotted in Figure 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 under- estimation 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 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 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 $F(1.450, 36.257) = 5.725$, $p = 0.013$, $r = 0.39$.

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.

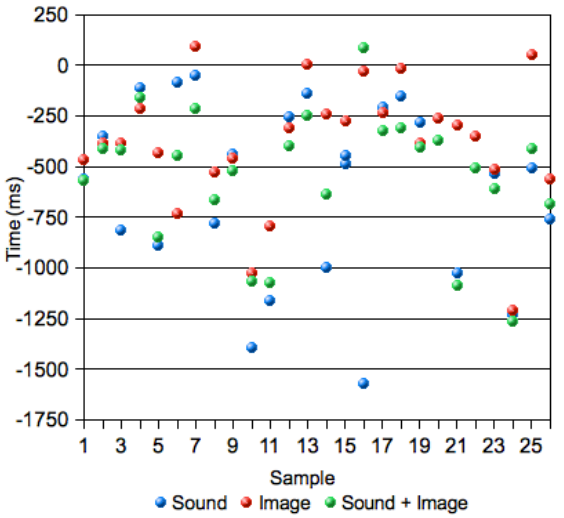

Figure 4 . Presentation x Perceived Time-to-Impact x Looming Scene Scatter Plot.

The perceived 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.

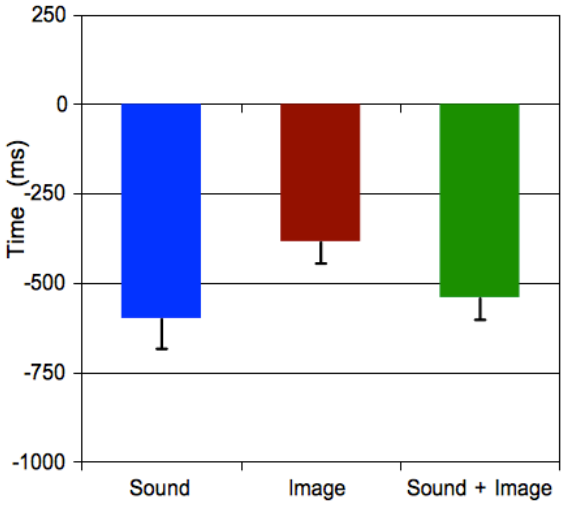

Figure 5 . Presentation × Perceived Time-to-Impact Bar Chart.

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 = −598.88ms, SE = 84.50), Image Only condition (M = −384.05ms, SE = 60.78), Sound + Image condition (M = −540.54ms, SE = 61.86).

3.1.2. Presentation x Emotion (Valence / Arousal)

The valance / arousal ratings x 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 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 rating.

The average was then calculated for each presentation condition, across all of the film looming scenes, and is plotted Figure 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 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 95% CI [1.008, 1.686] p =< 0.001; the Sound + Image condition versus the Sound Only condition 95% CI [0.700, 1.548] p =< 0.001. There was no significant difference between the two unimodal (Sound Only, and Image Only) conditions (see descriptives in Appendix Table 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 5).

The Sound + Image condition versus the Image Only condition 95% CI [1.570, 2.430] p =< 0.001; the Sound + Image condition versus the Sound Only condition 95% CI [0.700, 1.386] p =< 0.001; and the Sound Only condition versus the Image Only condition 95% CI [0.374, 1.540] 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.

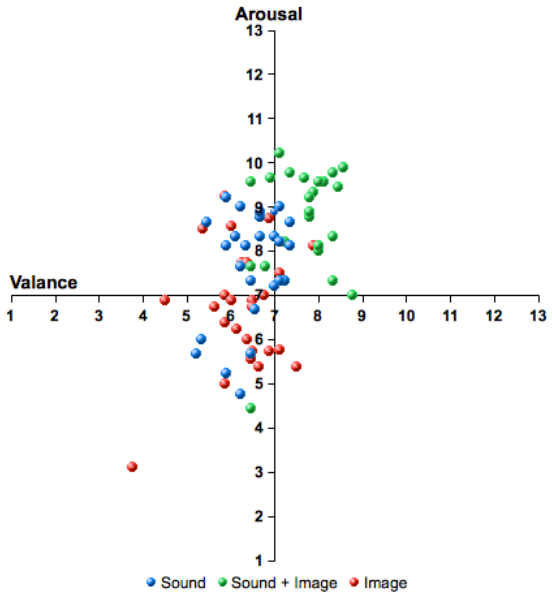

Figure 7. Presentation × Valence/Arousal Scatter Plot.
Participant's results were compiled and averaged, giving the valance / arousal ratings for each Presentation condition (Sound Only, Image Only, Sound + Image, by each of the 26 looming scenes.

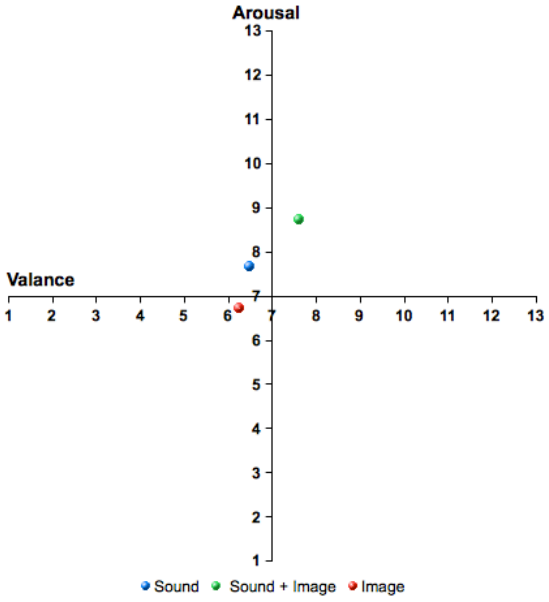

Figure 6. Presentation × Valence / Arousal (Averaged) Scatter Plot.
The valence / arousal results plotted in Figure 6 were then averaged across all of the looming scenes, for each presentation condition (Sound Only, Image Only, and Sound +Image). Sound: Valence M = 6.48, Arousal M = 7.68; Image: Valence M = 6.26, Arousal M = 6.72; Sound + Image: Valence M = 7.60, Arousal M = 8.72.

3.1.3. Presentation x Engagement

The engagement ratings were averaged across all of the participants responses and scene samples, for each presentation condition, and are plotted in Figure 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 6.

Mauchly's test indicated that the assumption of sphericity had been violated, $x^2(2) = 8.326$, $p = 0.016$, 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 95% CI [0.823, 1.61] $p =< 0.001$; the Sound + Image condition versus the Sound Only condition 95% CI [1.058, 1.794] $p =< 0.001$ (see pairwise comparisons listed in Appendix Table 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.

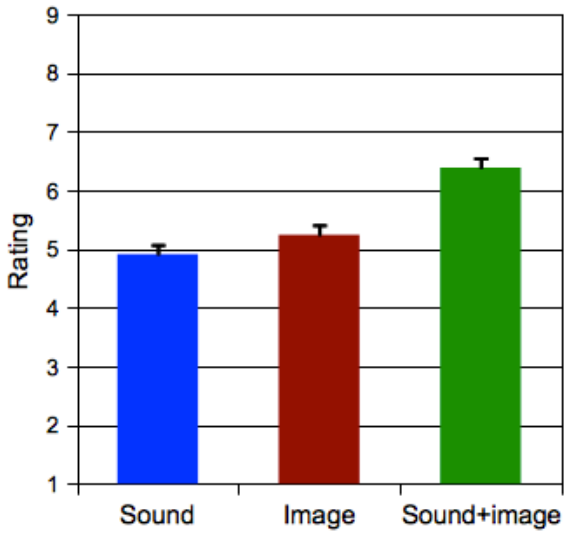

Figure 8. Presentation × Engagement Bar Chart. Presentation × Engagement Rating. *Results are plotted for each presentation condition (Sound Only, Image Only, Sound + Image) that were averaged across all of the participants ratings. Error bars indicate the standard error for each condition. The average 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).*

3.2. Correlations between the Dependent Variables

3.2.1. Engagement x Perceived 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 9.

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.

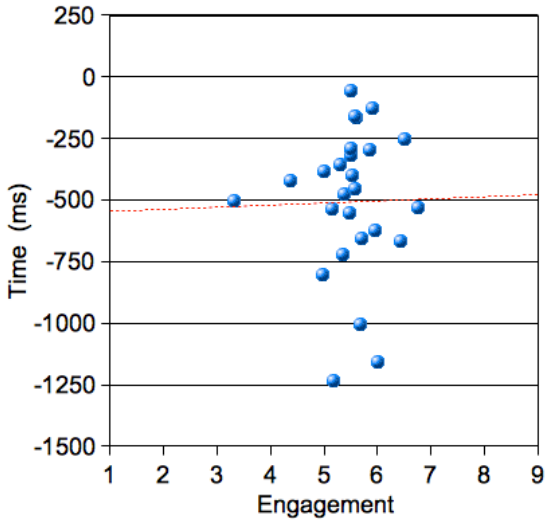

Figure 9. ***Perceived Time-to-Impact × Engagement Scatter Plot.*** *The average Time-to-Impact × Engagement rating is plotted for each scene. A linear regression analysis draws the line of best fit equation: $y = 8.49x - 554.61$, $r^2 = <0.01$. With a poor line of best fit, a negligible r^2, and a broad spread of data, we conclude that the results do not support hypothesis 2*

3.2.2. Valence / Arousal x Perceived Time-to-Impact

The perceived time-to-impact x valence and arousal ratings were averaged across all conditions, per scene sample (26 scenes) and are plotted on Figures 10 & 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.

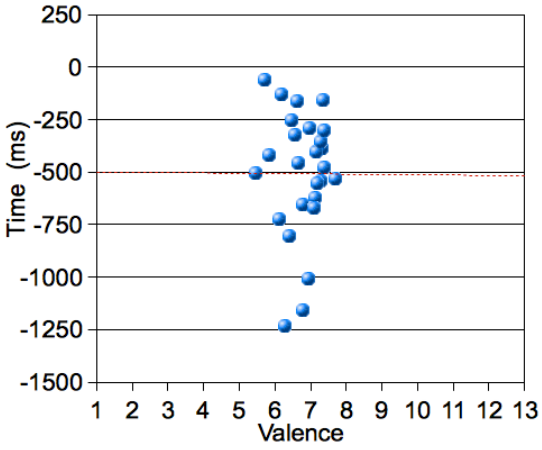

Figure 10. Perceived Time-to-Impact × Valence Scatter Plot. *The average Time-to-Impact × Valence rating is plotted for each scene. The linear regression line equation $y = -1.49x - 497.69$, $r^2 = <0.01$.*

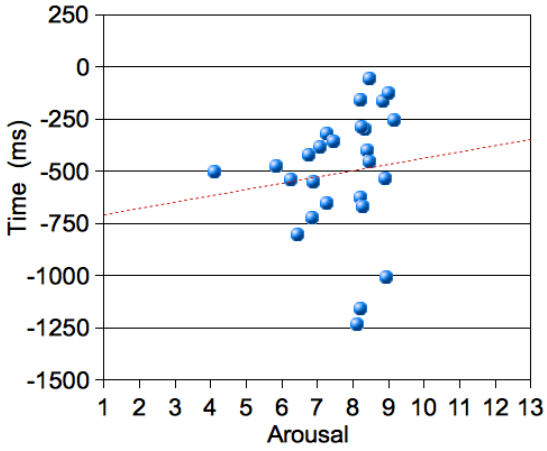

Figure 11. Perceived Time-to-Impact × Arousal Scatter Plot. *The average Time-to-Impact × Arousal rating is plotted for each scene. The linear regression line equation $y = 30.07x - 739.45$, $r^2 = 0.014$.*

3.2.3. Valence / Arousal x Engagement

The valence / arousal x engagement ratings were averaged across all conditions, per scene sample (26 scenes) and are plotted on Figures 12 & 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.

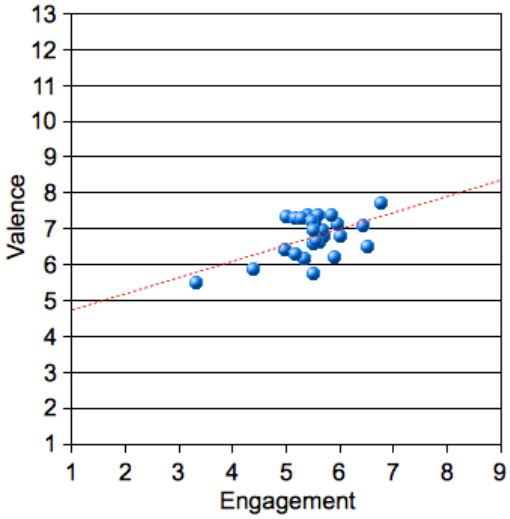

Figure 12. Valence × Engagement Scatter Plot.
The average Valence × Engagement rating is plotted for each scene. The linear regression equation of line is $y = 0.45x + 4.28$, $r^2 = 0.28$.

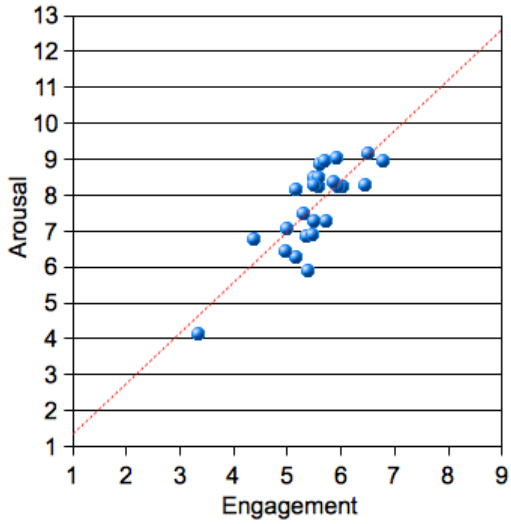

Figure 13. Arousal × Engagement Scatter Plot.
The average Arousal × Engagement rating is plotted for each scene. The linear regression equation of line is $y = 1.41x - 0.06$, $r^2 = 0.64$.

4. DISCUSSION

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). We mentioned in section 1.2 how relatively recent studies have begun to investigate multimodal auditory-visual looming in humans [Cappe et al. 2009; Tyll et al. 2012], with results indicating that multisensory (auditory-visual) integration of looming information is occurring. The results of the present study, in which complex audio looming cues caused observer's to underestimate the contact time of an approaching object, leading to faster reaction times in the sound + image condition, provides further evidence that the audio signal is biasing the response to the visual looming scene.

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. It is important to keep in mind that correlational studies with a small number of participants are underpowered, and further research with a greater number of participants will clarify the results.

And While the study by Tajadura-Jiménez et. al. [2010] found correlations between faster response times to targets with both the valence and arousal ratings, these results were obtained when comparing approaching versus receding objects, and objects with contrasting emotional associations (negative versus positive associations). Since our study focused on approaching objects only, and objects with the similar negative emotional association, the correlations in Tajadura-Jiménez' 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.

5. CONCLUSION

This study sought to determine if looming scenes with multiple complex audio cues would cause people to underestimate the contact time.

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 looming stimuli, which has been absent from the research corpus. This result suggests that further investigation of looming using complex audio cues and ecologically valid sounds is warranted, with results then being applicable to industry use in the design of sound effects in film, gaming and simulators.

Our study was also novel in measuring the emotional responses generated by looming information presented in different modalities. The measurement of emotion and engagement allowed us to investigate correlations between the stimuli's engagement quality, the amount of underestimation in the contact time, and the emotions experienced.

As such, the measurement of the emotional responses of participants and rating the sample engagement quality have been valuable tools, and leads to a better understanding of real world stimuli, emotional impact, and perceptions and actions generated as a result, therefore we recommend its use in future looming studies.

REFERENCES

- 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.
- Camponogara, I., Komeilipoor, N., & Cesari, P. (2015). When distance matters: Perceptual bias and behavioral response for approaching sounds in peripersonal and extrapersonal space. *Neuroscience*, 304, 101-108.
- Cappe, C., Thelen, A., Romei, V., Thut, G., and Murray, M. M. (2012). Looming Signals Reveal Synergistic Principles of Multisensory Integration. *Journal of Neuroscience, 32*(4), 1171-
- Cappe, D., Thut, G., Romei, V., and Murray, M. M. (2009). Selective integration of auditory-visual looming cues by humans. *Neuropsychologia; Neuropsychologia*, 47, 1045–1052.
- Cecere, R., Romei, V., Bertini, C., & Ladavas, E. (2014). Crossmodal enhancement of visual orientation discrimination by looming sounds requires functional activation of primary visual areas: A case study. *Neuropsychologia, 56*, 350-358.
- Gaver, W. W. (1993a). How do we hear in the world? Explorations in ecological acoustics. *Ecological psychology*, 5(4), 285-313.
- Gaver, W. W. (1993b). What in the world do we hear?: An ecological approach to auditory event perception. *Ecological psychology*, 5(1), 1-29.
- 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.
- Harrison, N. (2012). Auditory Motion in Depth is Preferentially 'Captured' by Visual Looming Signals. Seeing and Perceiving, 25(1), 71-85.
- 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.
- Neuhoff, J. G. (1998). Perceptual bias for rising tones. *Nature*, 395, 6698, 123–123.
- Neuhoff, J. G. (2001). An adaptive bias in the perception of looming auditory motion. *Ecological psychology*, 13, 2, 87–110.
- Neuhoff, J. G. (2016). Looming Sounds are Perceived as Faster than Receding Sounds. *Cognitive Research: Principles and Implications*, 1, 1, 15.
- Neuhoff, J. G., Hamilton, G., Gittleson, A., and Mejia, A. (2013). Babies in traffic: infant vocalizations modulate responses to looming sounds. *Journal of Cognitive Neuroscience*, 174.
- Neuhoff, J. G., Hamilton, G. R., Gittleson, A. L., and Mejia, A. (2014). Babies in Traffic: Infant Vocalizations and Listener Sex Modulate Auditory Motion Perception. *Journal of Experimental Psychology-Human Perception and Performance, 40*(2), 775-783.
- 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*.
- Ratcliff, R. (1993). Methods for dealing with reaction time outliers. Psychological bulletin, 114(3), 510.
- Riskind, J. H., Kleiman, E. M., Seifritz, E., & Neuhoff, J. G. (2014). Influence of anxiety, depression and looming cognitive style on auditory looming perception. *Journal of Anxiety Disorders, 28*(1), 45-50.
- Romei, V., Murray, M. M., Cappe, C., and Thut, G. (2009). Preperceptual and Stimulus-Selective Enhancement of Low-Level Human Visual Cortex Excitability by Sounds. *Current Biology*, 19(21), 1799-1805.
- 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., 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.
- Sutherland, C. A. M., Thut, G., and Romei, V. (2014). Hearing brighter: Changing in-depth visual perception

through looming sounds. *Cognition*, 132(3), 312-323.

- Tajadura-Jiménez, A., Väljamäe, A., Asutay, E., & Västfjäll, D. (2010). Embodied auditory perception: the emotional impact of approaching and receding sound sources. *Emotion*, 10, 2, 216.
- 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.

Appendix

#	Title	Year	DVD 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	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	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)

Table 1. List of Film Scenes Analysed.

A list of the looming scenes that were used in this experiment, with the year, DVD chapter, time frame, and object type.

Condition	N	Mean	Std. Dev.	Std. Error	95% Confidence Interval for Mean		Min	Max
					Lower	Upper		
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 2. Descriptive Statistics: Presentation X Perceived Time-to-Impact.

The descriptives results are tabled for the Presentation X Perceived Time-to-Impact, averaged across all of the participants. The columns are labelled 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 for Mean	
				Lower	Upper
Sound X Sound + Image	-58.34	71.974	1.000	-243.02	126.35
Sound X Image	-214.83*	77.768	0.032*	-414.38	-15.28
Sound + Image X Sound	58.34	71.974	1.000	-126.35	243.02
Sound + Image X Image	-156.49*	41.285	0.003*	-262.43	-50.55
Image X Sound	214.83*	77.768	0.032*	15.28	414.38
Image X Sound + Image	156.49*	41.285	0.003*	50.55	262.43

Table 3. Pairwise Comparisons: Presentation X Perceived Time-to-Impact

*The pairwise comparisons of Presentation X Perceived 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		Min	Max
					Lower	Upper		
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 4. Descriptive Statistics: Presentation X Valence / Arousal

The descriptives results are tabled for the Presentation X Valence / Arousal, averaged across all of the participants. The columns are labelled 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 for Mean	
				Lower	Upper
VALENCE					
Sound X Sound + Image	-1.124*	0.165	0.000*	-1.548	-0.700
Sound X Image	0.223	0.172	0.622	-0.219	0.664
Sound + Image X Sound	1.124*	0.165	0.000*	0.700	1.548
Sound + Image X Image	1.347*	0.132	0.000*	1.008	1.686
Image X Sound	-0.223	0.172	0.622	-0.664	0.219
Image X Sound + Image	-1.347*	0.132	0.000*	-1.686	-1.008
AROUSAL					
Sound X Sound + Image	-1.043*	0.134	0.000*	-1.386	-0.700
Sound X Image	0.957*	0.227	0.001*	0.374	1.540
Sound + Image X Sound	1.043*	0.134	0.000*	0.700	1.386
Sound + Image X Image	2.000*	0.168	0.000*	1.570	2.430
Image X Sound	-0.957*	0.227	0.001*	-1.540	-0.370
Image X Sound + Image	-2.000*	0.168	0.000*	-2.430	-1.570

Table 5. Pairwise Comparisons: Presentation X Valence / Arousal

*The pairwise comparisons of Presentation X Valence / Arousal. The * indicates the conditions where the mean difference is significant at α = 0.05. A Bonferroni adjustment has been applied to arousal, no adjustment was needed for Valence.*

Condition	N	Mean	Std. Dev.	Std. Error	95% Confidence Interval for Mean		Min	Max
					Lower	Upper		
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 6. Descriptive Statistics: Presentation X Engagement

The descriptives results are tabled for the Presentation X Engagement, averaged across all of the participants. The columns are labelled 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 for Mean	
				Lower	Upper
Sound X Sound + Image	-1.474*	0.141	0.000*	-1.836	-1.113
Sound X Image	-0.329	0.214	0.411	-0.879	0.221
Sound + Image X Sound	1.474*	0.141	0.000*	1.113	1.836
Sound + Image X Image	1.145*	0.155	0.000*	0.747	1.543
Image X Sound	0.329	0.214	0.411	-0.221	0.879
Image X Sound + Image	-1.145*	0.155	0.000*	-1.543	-0.747

Table 7. Pairwise Comparisons: Presentation X Engagement

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