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

Keywords:
Psychoacoustics, Audio Cues, Auditory Looming, Depth, Ecological Validity

Abstract

Objects that move in depth on an approaching trajectory (looming) are often encountered in both the real and virtual worlds. Examples include navigating oncoming traffic and sporting and gaming activities where judgements are made to avoid or attack approaching objects. How people react to looming objects may impact their survival and progression in the real, virtual, and gaming worlds, and relies on a person's ability to precisely interpret movement and depth cues. Psychological studies investigating auditory looming often depict an object's movement using simple audio cues (primarily amplitude increase) which are applied to tones (often sine or triangle waves) which are not normally encountered in the natural world. Whilst these studies provide valuable information about human perception and responses, technological advances allow us to present complex auditory stimuli with a range of audio cues and real-world sound sources, and to collect measurements on human perception and responses to ecologically valid stimuli.

In this study we investigate human responses to audio cues for movement in depth, and how the cues affect people's responses to the approaching object. We present an experiment on human perception where observers respond to auditory looming stimuli with real-world sounds that contain multiple audio cues, and introduce the direct-to-reverberant sound energy ratio as an audio cue. We measure the participants 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 listener's responses to the audio cues differed, revealing a hierarchy across the individual audio cues, with the amplitude increase being the most dominant cue, followed by the direct-to-reverberant energy ratio. The results also demonstrated that conditions with multiple audio cues prompted earlier estimates of the contact times, greater arousal and engagement ratings, than single audio cues; and that real world sound sources prompted significantly greater engagement ratings than both the artificial sound sources.

Introduction

An approaching object may be associated with three audio cues, namely, amplitude increase, frequency change (the Doppler shift), and interaural temporal differences [Rosenblum et al. 1987]. The audio cues have also been found to differ in their affect on perception and the object's perceived contact time, with some audio cue's prompting earlier responses to an approaching object, than other cues. Rosenblum et al [1987] compared these three cues (amplitude increase, frequency change, and interaural temporal differences) to determine how these cue's affected perception and if there was a hierarchy amongst the cues. The results indicated that the Doppler shift prompted a response before the object reached the observer whilst the amplitude increase prompted the fastest response to the contact time when the object had passed. Understanding which audio cues have a greater affect on perception than other audio cues would expand our understanding of human auditory perception and how these cues can be exploited in practical applications (for example, the application of audio cues to silent electric vehicles).

Auditory looming research using amplitude increase has found that judgements surrounding the magnitude of the increase is perceived to be greater than it physically is [Bach et al., 2007; Neuhoff and Heckel, 2004; Neuhoff 1998, 2016], suggesting that the object is approaching at a faster rate than it physically is. The endpoint loudness level is also perceived to be greater than it physically is for sounds (sine wave and noise in certain conditions) which are increasing in amplitude level as compared to sounds with no intensity change [Teghtsoonian et al., 2005; Neuhoff 1998], with the perceived loudness level being conversely true for sounds decreasing in amplitude.

Furthermore, the magnitude of the amplitude increase is perceived to be even greater when the auditory stimuli is 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 posing greater danger, than softer sounds. Many studies have used the single audio cue variable of amplitude increase to investigate auditory looming, finding it an effective audio cue [Rosenblum et al. 1993; Neuhoff 1998, 2001; Cappe et al. 2009; Ghazanfar et al. 2002; Maier et al. 2004, 2008; Maier and Ghazanfar 2007]. This approach is understandable since amplitude increase has been demonstrated as an effective audio cue, and researchers are often motivated to increase experimental robustness by absolute control of variables. However, investigation into a wider range of audio cues, as proposed by Gaver [1993a, 1993b] would assist in building a comprehensive understanding of human auditory looming perception.

Research has recently begun to investigate looming perception using a wider range of audio cues, including 3-Dimensional virtual sound sources with full spatial cues [Bach et al., 2009; Riskind et al., 2014; Neuhoff et al., 2013, 2014] and complex sound effects created by film sound designers [Wilkie and Stockman, 2018]. The acoustic variables used in these sound source stimuli include absolute delay, the Doppler shift, atmospheric filtering, gain attenuation due to atmospheric spreading, ground reflection attenuation, and HRTF's.

Surface reflections in the form of reverberation and the direct-to-reverberant energy ratio has been demonstrated extensively as an audio cue for depth perception and in determining the distance of a stationary object [Zahorik 2002a, b, 2001; Mershon and King, 1975; Bronkhorst and Houtgast, 1999; Bronkhorst, 2002; and Shinn-Cunningham 2000]. However, research on surface reflections as an audio cue for distance perception of stationary objects has not extended to dynamic objects moving in depth. Changes to the spectral cues in a sound source have also been found to induce motion perception [Baumgartner et al., 2007]. Decreasing the spectral contrast prompted listeners to perceive the sound source as approaching, whilst increasing the spectral contrast prompted listeners to perceive the sound source as receding.

Another important consideration with auditory looming stimuli is the sound source used to represent the object itself. Psychoacoustic studies often generate auditory stimuli using artificial sound sources such as sine, triangle, and square waves, which are rarely encountered in the natural world. The benefits of using an artificial sound source in experiments include increasing the study's internal validity, and limiting 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, meaning that the study's results may be limited in their capacity to transfer to real world applications, or improve our understanding of how people perceive and react in the real world. A number of recent studies have begun to investigate looming perception using real world sounds [Bach et al., 2009; Tajadura-Jiménez et al., 2010; and Wilkie and Stockman, 2018] finding that humans expressed a greater underestimation of the contact time for looming scenes which consisted of complex sounds with multiple audio cues.

The results from previous auditory looming studies have provided important information about human perception and the audio variables that act as a cue for approaching objects. However, the frequent use of single variables (often amplitude increase) and artificial sound sources (often sine or triangle waves) invites the question how do humans perceive and respond to real-world sound sources with multiple audio cues? And furthermore, how do surface reflections and the direct-to-reverberant sound energy ratio act as an audio cue for movement in depth?

The information obtained in answering these questions would advance understanding about 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 be useful for real world application, as well as assisting in developing a comprehensive understanding of human perception of looming objects.

Another factor to consider is that many studies investigate looming perception via a measurement of perceived contact time [Rosenblum et al., 1987; Neuhoff and Heckel, 2004; Neuhoff 1998, 2016], or corresponding neural activity [Baumgartner et al., 2007; Cappe et al., 2009, 2012; Tyll et al., 2012]. Whilst these studies have provided important information about looming perception, the salient nature of looming stimuli suggests that the measurement of emotion (Valence / Arousal and Engagement) 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 demonstrate the importance of measuring emotion in looming studies.

In this study, we examine people's responses to auditory looming scenes that use multiple audio cues and real-world sound sources, to gain an understanding of how people respond to a range of audio cues, and multiple audio cues, in real-world ecologically valid situations, and in what ways does greater sensory information cause their reactions to differ.

Experiment

Aims

This study has four aims, firstly, to determine if the direct-to-reverberant 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-reverberant energy ratio) affect looming perception, and if there is a hierarchy amongst 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.

Hypotheses

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

1. the direct-to-reverberant energy ratio parameter will act as an audio cue for an approaching object.
2. listener's responses to the individual audio cues for movement (amplitude increase, inter-aural differences, and the direct-to-reverberant energy ratio) will differ, revealing a hierarchy amongst the individual audio cues similar to the research findings of Rosenblum et al [1987] whereby the amplitude increase produced the strongest cue for movement in depth.
3. listener's responses to the trials with multiple (2 and 3) audio cues for movement will differ to the trials with single audio cues, reflecting the research findings of Tajadura-Jiménez et al. [2010] and Wilkie and Stockman [2018] whereby multiple cues elicited an earlier response to the perceived contact time than single cues.
4. listener's responses to the multiple audio cue combinations will differ, with some combinations affecting perception to a greater extent than other combinations, therefore revealing a hierarchy within the cue combination, reflecting the research findings of Rosenblum et al [1987].

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

5. listener's responses to approaching objects that present real world sound sources (i.e. a car tyre traction on a road surface) will differ from approaching objects that present artificial sound stimuli (i.e. a square wave or a noise band) reflecting the research findings of Bach et al. [2009] and Tajadura-Jiménez et al. [2010] whereby real world sounds prompted an emotional (Valence / Arousal) preference than an artificial sound source.

Method

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-Reverberant Sound Energy Ratio.

There were four dependent variables:
- Perceived time-to-impact,
- Valence,
- Arousal,
- Engagement.

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.

Stimuli

Visual information stating the experimental procedure was presented on a computer screen whereby a graphical user interface (GUI) displayed the trial number and presented the onscreen questionnaire after each trial asking the participants to rate their emotion and engagement responses.

Binaural auditory stimuli consisted of the following sound sources:

1. Car traction (Real world condition) with a fundamental frequency of approximately 400 Hz is the audio recording of the rubber car tyres rolling across the surface of an asphalt road.
2. 400 Hz square wave (Artificial condition).
3. Noise band (Artificial condition).

The three sound sources were each presented as the following Audio Cue conditions:

- Control condition (Ctrl) whereby no audio cues were applied to the sample, being equivalent to a stationary sound source. The sound was presented as its stereo sound file.

with the following single audio cues applied to the sound files as variables:

- Amplitude Increase (Amp) whereby the sound increased in amplitude over time.
- Inter-aural Differences (IAD) whereby the sound is presented binaurally between the two channels.
- Direct-to-Reverberant Sound Energy Ratio (Ref) whereby the ratio between the direct and reverberant sounds changed over time.

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

- Amplitude Increase + Inter-aural differences (Amp + IAD),
- Amplitude Increase + Reverberant ratio (Amp + Ref),
- Inter-aural differences + Reverberant ratio (IAD + Ref),
- Amplitude Increase + Inter-aural differences + Reverberant 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 or known perceptual responses to the stimuli (such as the sweep-induced fading bias [Canévet & Scharf, 1990; Teghtsoonian et al., 2005; Susini et al., 2010]), or the end-level bias that a particular amplitude 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 levels.

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. Whilst the psychoacoustic laws for inter-aural differences stipulate that for a frontal midline trajectory (approaching on a 0° 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 presenting the object moving at 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-reverberant energy ratio audio cue presents 6 first-order reflections off the surfaces. The reflections alter the overall sound by presenting reverberation (with more reverberation presented 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-reverberant 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 more audible and the overall sound content is comprised of a greater proportion of reverberant energy. As the object nears the observer, reflections at higher frequencies 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 (totaling 36 trials, listed in Appendix Table 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.

Generation of the Audio Cues for Movement Using Slab3D

The audio cues were generated using the virtual acoustic environment Slab3D Slabscape [NASA, 2013, 2019; Miller et al., 2019]. Slab3D draws on two head-related transfer function (HRTF) database collections to physically model and present the binaural cues. The HRTF databases used include the 'Listen' (IRCAM) and CIPID (UC Davis) collections (Miller et al., 2014).

We set the size of the virtual 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 1. Six 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 a 'perfect reflector', and the floor (ZN) as 'concrete'. The software default settings controlled the reflections, such as the manipulation of the direct-sound to reflected-sound ratio, spectral content, spectral scattering, and their change over time. Detailed description of the software and its settings is presented in Begault, et al. (2010), NASA (2013, 2019), and Miller et al. (2014, 2019).

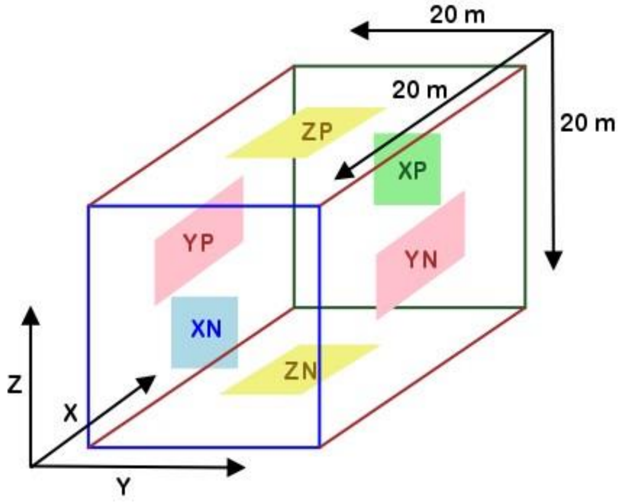

Figure 1. Slab3D Room Dimensions.

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), the listener position is located at (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 object's 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-impact.

Apparatus

Participants 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; present the trials in a randomised order, time the participants responses using the computer's internal clock, and collect the participant responses in a text file.

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

Dependent Variable Measurement

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

Perceived time-to-impact

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

Participants 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 1, the contact time (2000ms) 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'.

$$PTI = RT - CT \qquad \text{Equation 1}$$

where:

PTI = Perceived Time-to-impact, the amount of time (ms) which was under- / over-estimated. Underestimation is indicated in the negative value range, and overestimation is indicated in the positive value range.

RT = Participants Response Time, the time (in ms) when participants pressed the space bar when they thought the object reached them.

CT = Contact Time (2000ms),

Emotion: Valence and Arousal

To understand the participants 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 (see Figure ***2***). 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.

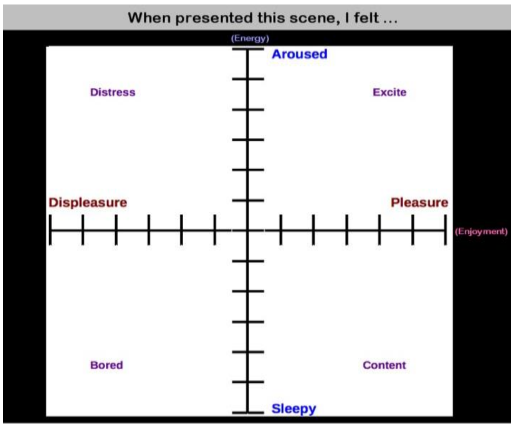

Figure 2. Valence / Arousal 2D Rating Scale.
Valence is measured on the X axis with 13-points ranging from displeasure to pleasure, whilst arousal is measured on the Y axis with 13-points ranging from sleepy to aroused. The minimum and maximum combinations of the valance / arousal sees the quadrants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

Engagement

To understand what the participants thought of the quality of the looming scene presented to them, 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 (see Figure 3).

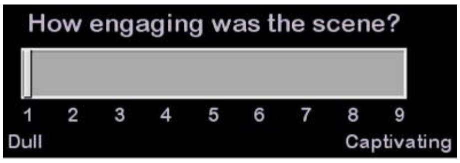

Figure 3. Engagement Rating Scale.
A 9-point visual analogue scale ranging from dull to captivating.

Procedure

Participants 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.

Before commencing the experiment, the participants completed a practice study using 6 looming scenes that were not presented in the experiment. These sessions were observed by the researcher in order to provide participants 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 the participants to listen to the sound 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 participants to rate their valence / arousal level and how engaging the scene was.

Each trial lasted for a total duration of 1700 milliseconds, however the participants 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 participants were not given any information implying there might be 'correct', 'incorrect' or 'preferred' responses.

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 with respect to 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.

The 15 participants 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 (Control, 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 (Control, IAD, Ref, IAD + Ref) were each averaged across the amplitude level (-18dB and -3dB).

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

Audio Cues

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

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

Audio Cues × Perceived time-to-impact

The results indicated that some of the data contained outliers. 12 outliers across 9 trial comparisons were removed leaving 36 trials per condition. The Perceived time-to-impact 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 4.

Looking at the plotted results, we see that the Control 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 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).

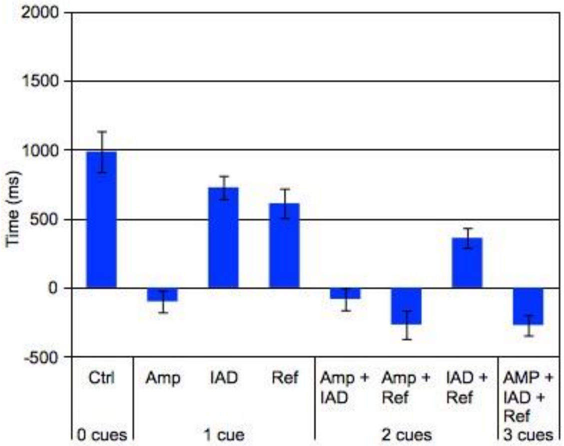

Figure 4. Audio Cue × Perceived time-to-impact Bar Chart.
The Perceived time-to-impact estimates for each audio cue condition (averaged across all of the participants ratings and sound sources) are plotted. Error bars indicate the standard error for each condition. The contact time occurs at 0ms, with any underestimation plotted in the negative range of the scale, and overestimation plotted in the positive range. Control condition: *M = 984.75 (S.E. = 147.87); Amp: M = -101.23 (S.E. = 78.11); IAD: M = 724.96 (S.E. = 84.25); Ref : M = 610.16 (S.E. = 106.43); Amp + IAD : M = -83.45 (S.E. = 80.82); Amp + Ref : M = -269.32 (S.E. = 102.88); IAD + Ref : M = 359.63 (S.E. = 71.73); Amp + IAD + Ref : M = -272.87 (S.E. = 74.19).*

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 the 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 been.

A one-way repeated measures ANOVA was conducted with the means and standard errors listed in Appendix Table 2. Mauchly's test of sphericity indicated that the assumption of sphericity had been violated, x2(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 reducing the Perceived time-to-impact $F(4.415,195.275) = 33.326, p =< 0.001, r = 0.683$, ($\alpha = 0.05$).

Post-hoc tests with pairwise comparisons were conducted and the descriptives are listed in Appendix Table 3. Pairwise comparisons of the Control condition to all other conditions revealed there was a significant difference (in the estimated Perceived time-to-impact) for all conditions with audio cues (except the IAD condition). This result supports hypothesis 1, that the direct-to-reverberant condition (Ref) acts as an audio cue for movement in depth, biasing people's responses to the perceived time-to-impact, and prompting earlier response times (than the Control 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 shorter 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 listener's responses to multiple cues differ from 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 earlier response times than the single audio cue conditions, supporting hypothesis 3. In regard to the exceptions (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, could arise from 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 listener's responses to multiple audio cues will differ from the trials with single audio cues) due to the existence of a hierarchy of cues, with some cues having greater affect than others.

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 per condition. The ratings were then averaged across all of the sound sources and participants responses for each audio cue condition, and are plotted in Figure 5.

Looking at the spread of the results, we see that the Control 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 in depth had greater arousal and valence ratings (with the exception of the IAD condition, which had M = 0.434 lower valence rating), than the Control 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 $x2(27) = 134.570$, $p =< 0.001$, 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 5. Whilst all of the conditions (except IAD) prompted greater valence ratings than the Control 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-reverberant variable (i.e. IAD vs IAD + Ref; Amp vs Amp + Ref; IAD vs Amp + IAD + Ref) reveals THAT the multiple cue condition prompts a significantly greater arousal rating, supporting hypothesis 3.

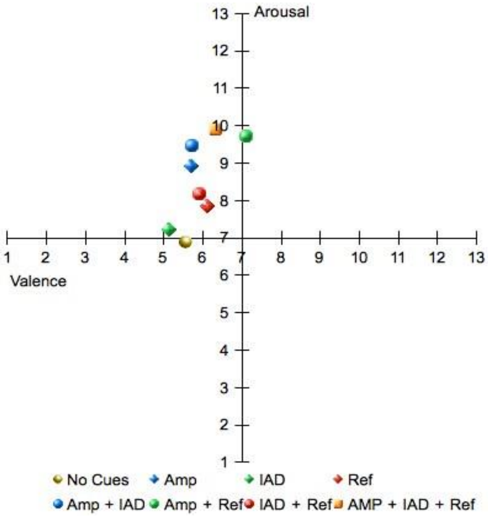

Figure 5. Audio Cue × Valence / Arousal Scatter Plot.
The valence / arousal ratings for each audio cue condition are plotted. Control condition: Valence: *M = 5.58 (S.E. = .29), Arousal: M = 6.88 (S.E. = .32); Amp: Valence: M = 5.71 (S.E. = .28), Arousal: M = 8.89 (S.E. = .40); IAD: Valence: M = 5.14 (S.E. = .28), Arousal: M = 7.19 (S.E. = .35); Ref : Valence: M = 6.13 (S.E. = .24), Arousal: M = 7.83 (S.E. = .30); Amp + IAD : Valence: M = 5.73 (S.E. = .36), Arousal: M = 9.44 (S.E. = .35); Amp + Ref : Valence: M = 7.11 (S.E. = .29), Arousal: M = 9.71 (S.E. = .29); IAD + Ref : Valence: M = 5.92 (S.E. = .24), Arousal: M = 8.14 (S.E. = .28); Amp + IAD + Ref : Valence: M = 6.33 (S.E. = .34), Arousal: M = 9.89 (S.E. = .30).*

For arousal, Mauchly's test indicated that the assumption of sphericity had also been violated x2(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, (α = 0.05).

Post-hoc tests on the arousal rating were conducted and the descriptives are listed in Appendix Table 5. When we compare the single audio cue conditions to the Control (no audio cues) condition, we see that all of the conditions (except IAD) prompted significantly greater arousal ratings, including the direct-to-reverberant 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.

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 participants responses, for each audio cue condition, and are plotted in Figure 6.

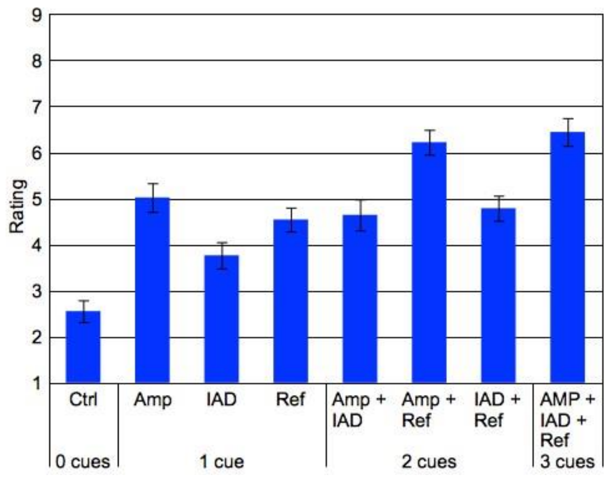

Figure 6. Audio Cue × Engagement Bar Chart.
The engagement rating for each audio cue condition (averaged across all of the participants ratings and sound sources) are plotted. Error bars indicate the standard error for each condition. *Control condition: M = 2.56 (S.E. = .24); Amp: M = 5.02 (S.E. = .31); IAD: M = 3.77 (S.E. = .29); Ref : M = 4.54 (S.E. = .26); Amp + IAD : M = 4.64 (S.E. = .34); Amp + Ref : M = 6.22 (S.E. = .27); IAD + Ref : M = 4.79 (S.E. = .27); Amp + IAD + Ref : M = 6.44 (S.E. = .31).*

Looking at the plotted results, we see that the Control 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 6. Mauchly's test indicated that the assumption of sphericity had been violated, x2(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, (α = 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 Control condition which contained no audio cues (see descriptives, and confidence intervals listed in Appendix Table 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 = ≤ 0.001) than all of the single audio cue conditions, supporting hypothesis 3.

Sound Source

To test hypothesis 5 (if listener's responses to an approaching object differs when presented with real world stimuli, as opposed to artificial stimuli) we investigate the affect 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 participants, totaling 120 trials per condition, 360 trials in total.

Sound Source × Perceived time-to-impact

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

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 10.

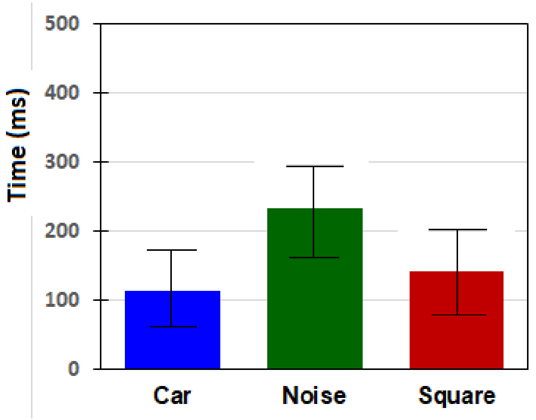

Figure 7. Sound Source × Perceived time-to-impact Bar Chart.
The Perceived time-to-impact for each sound source condition (averaged across all of the participants ratings and audio cues) are plotted. Error bars indicate the standard error for each condition. The contact time occurs at 0ms, with any underestimation plotted in the negative range of the scale, and overestimation plotted in the positive range. *Car: M = 114.73 (S.E. = 62.67); Noise: M = 234.30 (S.E. = 64.22); Square: M = 140.58 (S.E. = 60.86).*

A one-way repeated measures ANOVA was conducted with the means and standard errors listed in Appendix Table 10. Mauchly's test indicated that the assumption of sphericity had not been violated $x2(2) = 3.984, p = 0.136$, therefore the degrees of freedom did not need correction. The results indicate that the Perceived time-to-impact 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 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-impact.

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 8.

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.

One-way repeated measures ANOVA's were conducted with the means and standard errors listed in Appendix Table 12. For valence, Mauchly's test indicated that the assumption of sphericity had been violated $x2(2) = 25.745, p =< 0.001$, 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 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 13).

For arousal, Mauchly's test indicated that the assumption of sphericity had been violated $x2(2) = 6.366, p = 0.041$, 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.

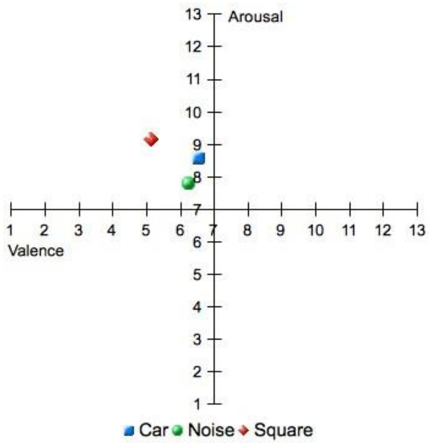

Figure 8. Sound Source × Valence / Arousal Scatter Plot.
The valence / arousal rating for each sound source condition (averaged across all of the participants' ratings and audio cues) are plotted. Car: *Valence: M = 6.55 (S.E. = .17), Arousal: M = 8.58 (S.E. = .23); Noise: Valence: M = 6.25 (S.E. = .14), Arousal: M = 7.78 (S.E. = .22); Square: Valence: M = 5.17 (S.E. = .21), Arousal: M = 9.13 (S.E. = .20).*

Sound Source × 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 9.

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 14. Mauchly's test indicated that the assumption of sphericity had been violated x2(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 15. This result supports hypothesis 5, that the sound source affects listeners engagement rating of an approaching object, with the real world car traction sound prompting a significantly greater engagement rating than the artificial square wave and noise band.

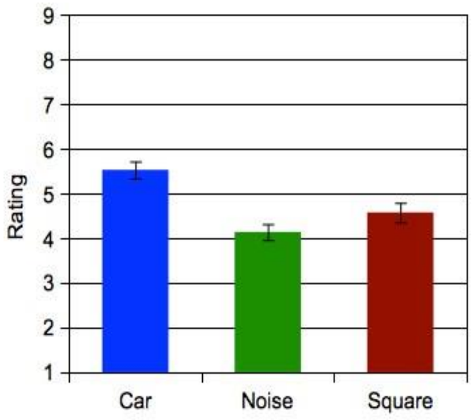

Figure 9 Sound Source × Engagement Bar Chart.
The engagement rating for each sound source condition (averaged across all of the participants ratings and audio cues) are plotted. Error bars indicate the standard error for each condition. *Car: M = 5.53 (S.E. = .19); Noise: M = 4.14 (S.E. = .18); Square: M = 4.58 (S.E. = .22).*

Discussion

In this study 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 Perceived time-to-impact and the addition of an occlusion period. When presented with the stimuli, participants 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-reverberant 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-impact and prompted an earlier response time; prompted a significantly greater arousal rating; and prompted a significantly greater engagement rating. Therefore, we conclude that the results support hypothesis 1, that the direct-to-reverberant ratio acted as an audio cue for movement in depth, influencing the perceived time-to-impact, arousal and engagement ratings.

The results also showed that for the single audio cues, listener's responses to the individual audio cues for movement in depth differed, revealing a hierarchy across the audio cues, supporting the research findings of Rosenblum et al. [1987]. 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-reverberant ratio (Ref) cue, and the inter-aural differences (IAD) cue; it prompted significantly greater arousal and engagement 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 a contrast effect may be occurring due to the angle of approach being of frontal midline. This hierarchy of audio cues 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, creating a greater contrast in the cue's signal and therefore increasing the magnitude of the audio cue information. Other factors which may also affect a cue's capacity to influence perception include the absolute sound level, room size which would increase or decrease the amount of reverberation, and the speed of the approaching object with a higher velocity producing a greater rate of change.

The amplitude's dominance as the strongest audio cue was also replicated when comparing the multiple cue conditions, whereby conditions containing the amplitude increase variable (Amp + Ref, Amp + IAD, Amp + IAD + Ref) prompted significantly earlier estimates of the perceived time-to-impact, 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-impact.

However, for measurements of engagement, the real world (car) sound source prompted significantly greater engagement ratings than both of the artificial sound sources. A possible reason for this result is that the sound of an approaching car (as opposed to a noise band or square wave) is often experienced on a daily basis for many people. The consequence of ignoring the cues of an approaching car can have an imminent and profound impact to one's survival prospects. Therefore achieving, and maintaining a high level of engagement is understandable.

Interestingly for measures of emotion, the artificial square wave, prompted significantly lower valence and significantly greater arousal ratings than the real world (car) recording. One possible explanation for this result may be that artificial tones often form the basis for alarms and warning signals. They are designed to gain attention and be uncomfortable so as to prompt people into action in order to make the stimuli stop. This result of the square wave prompting more negative valence and greater arousal ratings 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 (in particular the sound of an approaching car) differs to their responses to artificial sound sources.

From the results found in this study, we can see that the sound source stimuli affected human perception of an approaching object, and that the audio cues also impacted perception, with specific audio cues differing in the amount of under-/over-estimation of the approaching object. We therefore recommend that these factors are taken into consideration when investigating auditory looming perception, and suggest that further research into the parameters, in particular reverberation, and if the cue is more effective for natural or artificial sound sources, is needed in order to fully understand how sound influences human perception of approaching objects.

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.

Bach, D. R., Schächinger, H., Neuhoff, J. G., Esposito, F., Salle, F. D., Lehmann, C., and Seifritz, E. (2007). Rising sound intensity: an intrinsic warning cue activating the amygdala. *Cerebral Cortex*, 18(1), 145-150.

Baumgartner, R., Reed, D. K., Tóth, B., Best, V., Majdak, P., Colburn, H. S., and Shinn-Cunningham, B. (2017). Asymmetries in behavioral and neural responses to spectral cues demonstrate the generality of auditory looming bias. *Proceedings of the National Academy of Sciences*, 114(36), 9743-9748.

Begault, D., Wenzel, E. M., Godfroy, M., Miller, J. D., and Anderson, M. R. (2010). Applying spatial audio to human interfaces: 25 years of NASA experience. In *Audio Engineering Society Conference: 40th International Conference: Spatial Audio: Sense the Sound of Space. Audio Engineering Society*.

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.

Canévet, G., and Scharf, B. (1990). The loudness of sounds that increase and decrease continuously in level. *The Journal of the Acoustical Society of America*, *88*(5), 2136-2142.

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.

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.

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.

Miller, J. D., Godfroy-Cooper, M., and Szoboszlay, Z. P. (2019). Augmented-Reality Multimodal Cueing for Obstacle Awareness: Towards a New Topology for Threat-Level Presentation. [Online; accessed 12-September-2019].

Miller, J. D., Godfroy-Cooper, M., and Wenzel, E. M. (2014). Using published HRTFS with SLAB3D: Metric-based database selection and phenomena observed. Georgia Institute of Technology.

NASA (2013). Slab3D Home Page. http://slab3d.sonisphere.com. [Online; accessed 1-September-2013].

NASA (2019). Slab3D Research Page. https://human-factors.arc.nasa.gov/groups/ACD/projects/slab.php. [Online; accessed 8-November-2019].

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., and Neuhoff, J. G. (2014). Influence of anxiety, depression and looming cognitive style on auditory looming perception. *Journal of Anxiety Disorders, 28*(1), 45-50.

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*, 2*2*, 1467–1467.

Russell, J. A. (1980). A circumplex model of affect. *Journal of personality and social psychology 39,* 6, 1161– 1178.

Shinn-Cunningham, B. (2000). Distance cues for virtual auditory space. In *Proceedings of the IEEE-PCM*, volume 2000, pages 227–230.

Susini, P., Meunier, S., Trapeau, R., and Chatron, J. (2010). End level bias on direct loudness ratings of increasing sounds. *The Journal of the Acoustical Society of America*, 128(4), EL163-EL168.

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.

Teghtsoonian, R., Teghtsoonian, M., and Canévet, G. (2005). Sweep-induced acceleration in loudness change and the “bias for rising intensities”. *Perception & Psychophysics*, 67(4), 699-712.

Wilkie, S. and Stockman, T. (2018). Perception of Objects that Move in Depth, using Ecologically Valid Audio Cues. *The Journal of Applied Acoustics*, 134, 34-45.

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.

Appendix Tables

#	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 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	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 1 Descriptive Statistics: Audio Cues × Perceived time-to-impact

The descriptives results are tabled for the Perceived time-to-impact × 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
Control Condition (No Audio Cues) × Single Audio Cues:					
Ctrl × Amp	1085.941*	138.640	0.000*	624.849	1547.033
Ctrl × IAD	259.757	138.534	1.000	-200.980	720.494
Ctrl × Ref	374.554	114.053	0.056	-4.764	753.873
Control Condition (No Audio Cues) × Multiple Audio Cues:					
Ctrl × Amp + IAD	1068.161 *	161.213	0.000 *	531.998	1604.324
Ctrl × Amp + Ref	1254.032 *	165.680	0.000 *	703.011	1805.052
Ctrl × IAD + Ref	625.083 *	135.667	0.001 *	173.880	1076.287
Ctrl × Amp + IAD + Ref					
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 2 Pairwise Comparisons: Audio Cues × Perceived time-to-impact

The pairwise comparisons of Audio Cue condition × Engagement rating. The * indicates the conditions where the mean The pairwise comparisons of Audio Cue condition × Perceived time-to-impact. 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.

Condition #	Name	Valence N	Mean	Std. Error	95% Confidence Interval Lower	Upper	Arousal N	Mean	Std. Error	95% Confidence Interval Lower	Upper
1	Ctrl	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 3 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.

	Valence					Arousal				
				95% Confidence Interval					95% Confidence Interval	
Condition Pair	**Mean Diff.**	**Std. Error**	**Sig.**	**Lower**	**Upper**	**Mean Diff.**	**Std. Error**	**Sig.**	**Lower**	**Upper**
Control Condition (No Audio Cues) × Single Audio Cues:										
Ctrl × Amp	-.133	.253	1.000	-.976	.710	-2.011*	.498	.006 *	-3.669	-.354
Ctrl × IAD	.433	.166	.339	-.117	.984	-.311	.307	1.000	-1.331	.709
Ctrl × Ref	-.556	.246	.815	-1.375	.264	-.956 *	.285	.046 *	-1.902	-.009
Control Condition (No Audio Cues) × Multiple Audio Cues:										
Ctrl × Amp + IAD	-.156	.218	1.000	-.880	.569	-2.567 *	.434	.000 *	-4.011	-1.123
Ctrl × Amp + Ref	-1.533 *	.360	.003 *	-2.730	-.337	-2.833 *	.376	.000 *	-4.085	-1.582
Ctrl × IAD + Ref	-.344	.201	1.000	-1.013	.324	-1.267 *	.303	.004 *	-2.276	-.258
Ctrl × 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 4 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 α = 0.05. A Bonferroni adjustment was applied to correct for a possible increase in type 1 errors associated with multiple comparisons.

Condition				Std.	95% Confidence Interval	
#	**Name**	**N**	**Mean**	**Error**	**Lower**	**Upper**
1	Ctrl	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 5 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
Control Condition (No Audio Cues) × Single Audio Cues:					
Ctrl × Amp	-2.467*	.318	.000	-3.524	-1.409
Ctrl × IAD	-1.211*	.317	.012	-2.265	-.157
Ctrl × Ref	-1.989*	.306	.000	-3.006	-.972
Control (No Audio Cues) × Multiple Audio Cues:					
Ctrl × Amp + IAD	-2.089*	.314	.000	-3.132	-1.046
Ctrl × Amp + Ref	-3.667*	.339	.000	-4.793	-2.541
Ctrl × IAD + Ref	-2.233*	.325	.000	-3.313	-1.153
Ctrl × 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 Cue					
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 6 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 α = 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 Lower	Upper
Perceived time-to-impact:						
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 7 Descriptive Statistics: Amplitude Levels × Perceived time-to-impact / Valence / Arousal / Engagement

The descriptives results are tabled for the Amplitude Levels × Perceived time-to-impact / 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 × Perceived time-to-impact:					
-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 8 Pairwise Comparisons: Amplitude Levels × Engagement

The pairwise comparisons of Amplitude Level condition × Perceived time-to-impact / Valence / Arousal / 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.

Condition #	Name	N	Mean	Std. Error	95% Confidence Interval 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 9 Descriptive Statistics: Sound Source × Perceived time-to-impact

The descriptives results are tabled for the Sound Source × Perceived time-to-impact, 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 10 Pairwise Comparisons: Sound Source × Perceived time-to-impact

The pairwise comparisons of Sound Source condition × Perceived time-to-impact. 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.

Condition		Valence					Arousal				
#	Name	N	Mean	Std. Error	95% Confidence Interval Lower	Upper	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 11 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	Mean Diff.	Std. Error	Sig.	95% Confidence Interval Lower	Upper	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 12 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 α = 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 Lower	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 13 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.

Condition Pair	Mean Difference	Std. Error	Sig.	95% Confidence Interval Lower	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 14 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.