Applied Acoustics 134 (2018) 34–45

Contents lists available at ScienceDirect

Applied Acoustics

journal homepage: www.elsevier.com/locate/apacoust

Perception of objects that move in depth, using ecologically valid audio cues

Sonia Wilkie*, Tony Stockman

Queen Mary, University of London, UK

ARTICLE INFO

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-/under-estimation); 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 [21]. 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

* Corresponding author.
E-mail addresses: s.l.wilkie@qmul.ac.uk (S. Wilkie), t.stockman@qmul.ac.uk (T. Stockman).

https://doi.org/10.1016/j.apacoust.2017.12.032
Received 19 August 2017; Received in revised form 21 November 2017; Accepted 23 December 2017

amplitude) [12,13,3]. 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 [13]. It was also concluded that tonal sounds (in the form of pitched sine tones) enabled easier detection of looming audio cues, than (white) noise [12,13,7,11].

When amplitude increase is used as an audio cue, the magnitude of the change is perceived to be greater than it physically is [17,12,14] 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 [12,17] with louder sounds suggesting that the object is at a closer proximity, therefore posing greater potential danger.

Many of these studies [22,12,13,3,7,9,10,11] 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 [5,6] 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. [1] and Tajadura-Jiménez et al. [25], whilst studies that investigate looming perception using a 3-Dimensional virtual sound source with full spatial cues include Bach et al. [1], Riskind et al. [19], and Neuhoff et al. [15,16]. 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) [9,11] and more recently extending the research to human observers [20,3,2,8,26,4,24]. 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 [1,25] finding that approaching sounds were rated as more unpleasant (valence), and arousing, than receding sounds.

The Tajadura-Jiménez et al. [25] study had particularly interesting results, finding that when an unpleasant target image was paired with an approaching audio cue (amplitude increase), 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 (amplitude decrease).

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 with multiple audio cues?

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 45.05 dB (SD = 15.32) on a linear/near-linear slope.
- An average spectral centroid $M = 1957.8$ Hz (B6 16 cents) at the start of the sample, and an average peak at $M = 3444.57$ Hz (A7 38 cents), an increase of 1486.77 Hz (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: