Information Fusion 11 (2010) 45–50

Contents lists available at ScienceDirect

Information Fusion

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

Motion extrapolation of auditory–visual targets

Sophie Wuerger [a,*], Georg Meyer [a], Markus Hofbauer [b], Christoph Zetzsche [c], Kerstin Schill [c]

[a] School of Psychology, Eleanor Rathbone Building, Bedford Street South, University of Liverpool, Liverpool L69 7AZ, United Kingdom
[b] Neurologische Klinik, Ludwig-Maximilians-Universität, Marchioninistraße 23, 80377 München, Germany
[c] Kognitive Neuroinformatik, Universität Bremen FB3, Enrique–Schmidt–Straße 5, 28359 Bremen, Germany

ARTICLE INFO

Article history:
Received 30 January 2007
Received in revised form 18 October 2007
Accepted 1 April 2009
Available online 22 April 2009

Keywords:
Sensory integration
Auditory and visual motion
Motion extrapolation
Temporal localisation
Optimal integration

ABSTRACT

Many tasks involve the precise estimation of speed and position of moving objects, for instance to catch or avoid objects that cohabit in our environment. Many of these objects are characterised by signal representations in more than one modality, such as hearing and vision. The aim of this study was to investigate the extent to which the simultaneous presentation of auditory and visual signals enhances the estimation of motion speed and instantaneous position. Observers are asked to estimate the instant when a moving object arrives at a target spatial position by pressing a response button. This task requires observers to estimate the speed of the moving object and to calibrate the timing of their manual response such that it coincides with the true arrival time of the moving object. When both visual and auditory motion signals are available, the variability in estimating the arrival time of the moving object is significantly reduced compared to the variability in the unimodal conditions. This reduction in variability is consistent with optimal integration of the auditory and visual speed signals. The average bias in the estimated arrival times depends on the motion speed: for medium speeds (17 deg/s) observers' subjective arrival times are earlier than the true arrival times; for high speeds (47 deg/s) observers exhibit a (much smaller) bias in the other direction. This speed-dependency suggests that the bias is due to an error in estimating the motion speeds rather than an error in calibrating the timing of the motor response. Finally, in this temporal localization task, the bias and variability show similar patterns for motion defined by vision, audition or both.

1. Introduction

Significant progress has been made in understanding how signals from the auditory and visual modalities are combined for perceptual tasks such as spatial localization [2,4,7,14,30], the detection and extraction of motion [1,5,17,18,20,23,24], bimodal synchrony and grouping [13,15,19,21] or visual search [9]. Most studies on auditory–visual motion processing have focused on the detection of global motion embedded in noise [1,5,17] or motion biases introduced by one modality on the other modality (e.g. [23,24]). However, little is known about how humans integrate visual and auditory information in biologically relevant tasks where *motion speed* as well as instantaneous position have to be estimated from the auditory and visual modalities to initiate motor commands. To study motion extrapolation based on bimodal speed information we use a temporal localisation task, where the subject has to predict when the moving object arrives at a spatial target location.

When the target is defined visually, human observers can accurately point to the extrapolated final position of a moving target when feedback is given [6,22]. How this sensory signal is computed is still a matter of debate. Since the extrapolation of the final position of a target requires a correct estimation of an instantaneous position as well as the extraction of speed, it is clear that both spatial and temporal mechanisms must be involved in this task [29]. There is some evidence that the integration of auditory and visual motion signals occurs before speed is calculated, i.e. within spatial and temporal mechanisms [16]. This is consistent with the idea that the modality that is most reliable for a particular task will dominate the performance in this task [26,28]. More recently this hypothesis of 'modality appropriateness' has been formulated in terms of a quantitative framework [4,8,14,30] and numerous experimental studies have demonstrated its wide applicability to a variety of tasks (e.g. [7]).

In this study we are concerned with temporal localisation performance based on the integrated speed signals from the auditory and visual modality and we will address the following questions: (i) how the simultaneous availability of motion speed estimates from two modalities affects the performance in the temporal localisation task, and (ii) whether localization errors are similar in the auditory, visual and bimodal condition.

* Corresponding author. Tel.: +44 151 794 2173; fax: +44 151 794 2945.
E-mail address: s.m.wuerger@liverpool.ac.uk (S. Wuerger).

1566-2535/$ - see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.inffus.2009.04.005

2. Methods

2.1. Apparatus

The participants were seated in a chair in front of a 180-deg arc with a diameter of 2.0 m (Fig. 1; for details see also [10,18]) of 31 horizontally mounted LEDs and loudspeakers. The distance between the participant and the middle of the arc was 1.6 m. The LEDs and loudspeakers were switched on and off such that object movement was simulated. The LEDs and loudspeakers were controlled by a Tucker–Davis RP2 real-time signal processor (Tucker–Davis Technologies) which was connected to a Personal Computer. An additional LED, located in the centre of the arc, but slightly above the other LEDs, was on continuously and served as a fixation point.

2.2. Stimuli

The visual motion stimulus was generated by successive flashing of the LEDs while auditory motion signals were created by clicks, generated by a voltage steps, applied to a succession of loudspeakers. Motion could be defined either visually, auditorily or both. The motion sequence always started at the most distal LED/speaker in the array and moved towards the fixation point that also served as target position at the centre. The start point was chosen from a pseudorandom sequence.

The signals moved at 5 different speeds: 17, 23, 27, 34, and 47 deg/s. The different speed levels were randomly interleaved within each block to ensure that the observers had to extract the object's current position and speed to extrapolate the likely arrival time rather than rely on the duration of the motion signal or speed signal alone to estimate arrival time.

The motion signals were presented in a variable background of noise, generated by presenting clicks or flashes from transducers in random positions. 'Signal' is defined as the number of clicks or flashes at successive locations; 'Noise' is defined as the number of randomly generated clicks or flashes.

In a preliminary experiment we determined the signal-to-noise ratio (SNR) at which the motion can be reliably detected. The subjects had to discriminate random motion from coherent motion in a 2AFC setting where the noise level was adjusted using the QUEST procedure [27] to achieve 84% correct responses in either modality. We then equated the auditory and visual stimuli in terms of detectability so that both motion signals were comparable. Signal-to-noise ratio (SNR) was defined as the number of signal LEDs/loudspeakers divided by the number of 'noise' LEDs/loudspeakers over the whole duration of the motion stimulus. These motion thresholds (in terms of SNR) were determined at a motion speed of 30 deg/s and were 0.48 (or −6.74 dB) for visual motion and at 4.86 (13.73 dB) for auditory motion (for details see [18]). For the main experiment (motion extrapolation) the SNRs were fixed at these levels.

2.3. Procedure

The experiment was run in a darkened sound-proof room (IAC 1402A). The participants were seated in a chair with a headrest and instructed to look at the fixation point (Fig. 1). The start point was randomly chosen as either the left- or rightmost point in the array so that subjects could not fixate a known start point and track the signal from there. We confirmed that subjects maintained fixation by recording eye movements for two observers with an eye-tracker (ASL 5000 Series Model 501).

Observers were asked to estimate the instant at which the moving signal arrived at the target position, coincident with the fixation LED, by pressing a button. We measured the point in time when the button was pressed (=*estimated arrival time*). Feedback was given to the participant for: too early, too late, correct (object within 0.7 deg of the target position at the instance of response). The use of a spatial performance criterion and feedback to the subjects was based on ecological considerations: the task was to catch an object so that it had to be within a fixed range of the target point at the time of response to be caught and observers would expect to receive direct feedback for the task. To ensure that observers actu-

Fig. 1. Visual or auditory motion was generated by successively switching on/off 31 horizontally mounted LEDs or loudspeakers. The participant was seated in a chair in front of a 180-deg arc.