BRILL

Multisensory Research 29 (2016) 365–395

MULTISENSORY RESEARCH

brill.com/msr

Audiovisual Integration of Time-to-Contact Information for Approaching Objects

Patricia R. DeLucia [1,*], **Doug Preddy** [1] **and Daniel Oberfeld** [2]

[1] Department of Psychological Sciences, MS 2051, Texas Tech University, Lubbock, TX 79409-2051, USA

[2] Department of Psychology, Johannes Gutenberg-Universität, 55099 Mainz, Germany

Received 17 March 2015; accepted 20 October 2015

Abstract

Previous studies of time-to-collision (TTC) judgments of approaching objects focused on effectiveness of visual TTC information in the optical expansion pattern (e.g., visual tau, disparity). Fewer studies examined effectiveness of auditory TTC information in the pattern of increasing intensity (auditory tau), or measured integration of auditory and visual TTC information. Here, participants judged TTC of an approaching object presented in the visual or auditory modality, or both concurrently. TTC information provided by the modalities was jittered slightly against each other, so that auditory and visual TTC were not perfectly correlated. A psychophysical reverse correlation approach was used to estimate the influence of auditory and visual cues on TTC estimates. TTC estimates were shorter in the auditory than the visual condition. On average, TTC judgments in the audiovisual condition were not significantly different from judgments in the visual condition. However, multiple regression analyses showed that TTC estimates were based on both auditory and visual information. Although heuristic cues (final sound pressure level, final optical size) and more reliable information (relative rate of change in acoustic intensity, optical expansion) contributed to auditory and visual judgments, the effect of heuristics was greater in the auditory condition. Although auditory and visual information influenced judgments, concurrent presentation of both did not result in lower response variability compared to presentation of either one alone; there was no multimodal advantage. The relative weightings of heuristics and more reliable information differed between auditory and visual TTC judgments, and when both were available, visual information was weighted more heavily.

Keywords

Time-to-collision, tau, audition, visual perception, multisensory integration, looming, psychophysical reverse correlation, observer weights

* To whom correspondence should be addressed. E-mail: pat.delucia@ttu.edu

DOI:10.1163/22134808-00002520

1. Audiovisual Integration of TTC Information for Approaching Objects

The avoidance and creation of collisions is ubiquitous. For example, the ability to avoid rear-end collisions is essential while driving. The ability to create collisions is important in sports when hitting a ball with a bat. Understanding how people perceive collision has widespread practical importance (DeLucia, 2015), for example, when developing assistive technologies to help the visually impaired with navigation and automotive technologies to help drivers avoid rear-end collisions.

Avoiding and creating collisions involves judgments of whether a collision will occur (collision detection), and when an impending collision will occur (time-to-collision; TTC estimation). Subsequently, actions can be executed to avoid or create the collision. Most prior studies focused on judgments of time-to-collision which is the focus of this paper.

The dominant theory of TTC perception is the tau-hypothesis (Hecht and Savelsbergh, 2004). When an object approaches the eye, the resulting optical expansion pattern contains accurate information about the time remaining until the object would hit the eye (Hoyle, 1957; assuming certain conditions, such as constant velocity, are met; Lee, 1976). This information is the optical invariant, tau (τ), which is defined as the ratio of the object's optical size by its instantaneous rate of optical expansion. We refer to this as visual τ to distinguish it from auditory τ, discussed later. Notably, visual τ relies solely on information available in the optic array and does not require knowledge about the physical size, velocity, or distance of the object. τ-like variables provide both monocular and binocular TTC information (Gray and Regan, 1998) and also provide time-to-passage information (von Hofsten and Lee, 1985). Tau theory is attractive because it does not require the observer to rely on perceptual estimates of the object's distance and velocity. In addition, τ provides TTC information directly and does not require the use of cognitive processes or depth cues (for examples of counter evidence, see Brendel *et al.*, 2012; DeLucia, 2004, 2013).

Most studies of TTC perception require observers to make judgments about a silent approaching object that is presented in the visual domain. However, in many ordinary situations, approach events provide auditory and visual information, for example, a car approaching a pedestrian at a crosswalk. The relevance of auditory information when navigating through traffic is evident, for example, in the discussion of potential risks posed by 'quiet' electric cars (Ashmead *et al.*, 2012). More generally, humans use visual and auditory (and other) information during navigation, making it important to understand the manner in which auditory information is used and how it is combined with visual information. In fact, a multitude of auditory cues are related to motion, distance, and TTC of an object (Jenison, 1997; Kaczmarek, 2005; Lutfi and

Wang, 1999; Porschmann and Storig, 2009; Rosenblum *et al.*, 1987; Zakarauskas and Cynader, 1991), just as in the visual domain (Calabro *et al.*, 2011; DeLucia *et al.*, 2003; Gonzalez *et al.*, 2010; Gray and Regan, 1998, 1999; Khuu *et al.*, 2010; Yan *et al.*, 2011). In the present study, we controlled the change in acoustic intensity of the approaching object as was done in previous research on audiovisual TTC judgments (Zhou *et al.*, 2007). For an object on a straight, direct collision course with the observer, the interaural time delay and the interaural level difference remain constant during the approach, and thus do not provide information about TTC, and the Doppler frequency shift also remains constant (Jenison, 1997). In contrast, the dynamic changes of the sound intensity during the approach of the sound source provide accurate information about TTC: The ratio between the instantaneous acoustic intensity and the instantaneous rate of intensity change is a τ-like variable that specifies TTC (Shaw *et al.*, 1991). For a point source in an acoustic free field (i.e., without any acoustically reflecting surfaces), the acoustic intensity $I(r)$ as a function of distance of the sound source from the observer (r) obeys an inverse-square law (Hartmann, 2005), $I(r) \propto 1/r^2$, where '$\propto$' means 'is proportional to'. This implies a 6-dB intensity loss for each doubling of distance (Zahorik, 2002). For an object approaching the listener on a straight collision path and at constant velocity v, the actual time-to-contact at time t is given by $\mathrm{TTC}(t) = r(t)/\dot{r}(t) = r(t)/v$, where $r(t)$ is the instantaneous distance from the observer and $\dot{r}(t)$ is the first derivative of $r(t)$ with respect to time, that is, the velocity (v) of the object. The acoustic intensity at the observer's ear depends on the distance of the sound source from the observer as $I(t) = k/r^2(t)$, where k is a constant that reflects the (constant) acoustic power emitted by the source and the position of the receiver relative to the directivity pattern of the sound source. The derivative of $I(t)$ with respect to time is

$$\dot{I}(t) = 2k\frac{\dot{r}(t)}{r^3(t)}. \tag{1}$$

If one computes the τ-like ratio $I(t)/\dot{I}(t)$ (Hoyle, 1957; Lee, 1976), the higher-order terms of $r(t)$ cancel out, resulting in

$$I(t)/\dot{I}(t) = \frac{r(t)}{2\dot{r}(t)}. \tag{2}$$

Because the TTC of an object approaching at constant velocity is, by definition, $r(t)$ divided by $\dot{r}(t)$, $I(t)/\dot{I}(t) = \mathrm{TTC}/2$. Thus, $I(t)/\dot{I}(t)$ specifies TTC (Shaw *et al.*, 1991). The preceding derivation holds for the acoustic far field, that is, for distances between the sound source and listener that are more than 3.0 m (Blauert, 1996); this generally occurs in situations requiring TTC judgments in traffic.Whereas the described relation, strictly speaking, applies only to the acoustic free field, the presence of a reflecting ground surface (e.g., a car approaching a listener in an open space) causes additional reverberant energy to reach the listener's ear. However, the reflections will reach the observer from approximately the same angle as the direct sound if the distance from the sound source (e.g., car engine) is greater than about 10 m. Thus, it can be assumed that the reflections will only minimally change the inverse law relation between sound intensity and distance of the sound source. Only in the presence of many additional reflecting surfaces (e.g., in a narrow urban road), will acoustic reflections result in a smaller change of acoustic energy at the listener's ear when the distance to the sound source changes. On the other hand, at larger distances of 15.0 m or more (Blauert, 1996) sound absorption in the air will contribute to the distance-dependent change in sound intensity, in addition to the intensity changes that occur due to geometric spreading of sound energy (Jenison, 1997) and with a stronger effect at high than low frequencies. Thus, the potential reduction of the distance-dependent level change due to reverberation is partly compensated for by the additional contribution of absorption for real-world traffic scenarios. In short, equation (2) can be viewed as a reasonable approximation.

Previous research demonstrated that listeners are capable of estimating TTC on the basis of auditory information (Gordon and Rosenblum, 2005; Gordon *et al.*, 2013; Hellmann, 1996; Kaczmarek and Niewiarowicz, 2013; Rosenblum *et al.*, 1987, 1993; Schiff and Oldak, 1990), although Guski (1992) proposed that auditory information is primarily used for detecting an approaching object, rather than for actually estimating its TTC. Even infants younger than seven months are probably sensitive to auditory TTC. Freiberg *et al.* (2001) observed avoidance behavior (defensive leaning back) in response to a sound that gradually increased, but not decreased, in sound pressure level (SPL). Interestingly, the defensive response was stronger for fast changes in SPL (which occur at short TTCs) than for slow changes (which occur at long TTCs). Prior research indicated that blind or visually impaired individuals can make safe decisions about crossing a street, presumably on the basis of auditory information from oncoming traffic (in addition to residual vision if present); however, such decisions were slower and more dangerous compared to sighted individuals (Ashmead *et al.*, 2005; Emerson *et al.*, 2011).

A few studies have also compared the integration of auditory and visual information in the context of TTC estimation, or of motion perception in general. On a basic level, there is evidence for audiovisual integration in motion detection (Wuerger *et al.*, 2003), especially for 'looming' signals, that is, objects that are on a collision course with the observer (Cappe *et al.*, 2009; Conrad *et al.*, 2013; Maier *et al.*, 2004). With respect to TTC judgments, some studies compared performance when only visual, only auditory, or auditory and visual information were available. The results showed higher accuracy (small difference between estimates of TTC and true TTC values) in the visual-only compared to the auditory-only condition, and varying degrees of an advantage to having both (Hassan, 2012; Schiff and Oldak, 1990; Zhou *et al.*, 2007).

Schiff and Oldak (1990) used color sound films recorded in a field setting (a real vehicle approaching on a road). The sound of the vehicle was recorded via a microphone affixed to the top of the camera. The filmed approaching objects were presented for 4–6 s (with only visual, only auditory, or concurrent auditory and visual information) and then ended (before the vehicle reached the camera). The participant's task was to press a button when they thought that the object would reach or pass them had the object continued moving after it was no longer visible and audible. This type of task was used in the present experiment. In Schiff and Oldak's study, average TTC judgments were underestimates of the actual TTC in all modality conditions; participants' TTC estimates were less than the actual TTC. The underestimation was smallest in the visual-only condition, largest in the auditory-only condition, and intermediate in the audiovisual condition. Thus, the results of Schiff and Oldak are compatible with audiovisual integration, but show no multimodal advantage in the sense of higher accuracy, even though the real-world recordings contained not only the intensity change of the approaching sound source, but also contained potential spectral changes; the latter were due to the greater attenuation of high frequencies in the air that occur at large distances and to reflections from the ground surface. Notably, the study by Schiff and Oldak as well as other prior studies did not quantify how much weight participants assigned to each modality or how much weight was assigned to different cues within each modality. We introduce a novel method to do so.

Zhou *et al.* (2007) presented a simulation of an approaching car moving along a ground surface at a constant speed. The acoustic simulations were not described in detail but the method section explicitly mentions that intensity changes were presented. In a two-interval task, the participants decided which of two sequentially presented approaching cars had the shorter TTC at the moment of disappearance. The average accuracy in comparing the two TTCs was slightly higher in the audiovisual condition (smaller just-noticeable difference, JND) than in the unimodal conditions. However, the audiovisual accuracy fell between the two unimodal conditions when one participant who showed a higher JND in the visual than in the auditory condition was excluded (see Note 1). Thus, the data by Zhou *et al.* (2007) show, at most, a weak multimodal advantage.

In a study by Hassan (2012), participants stood at a crossing point on a real street. After observing and/or listening to traffic, participants were prompted to rate whether there was enough time to cross the road. In the visual-only condition, participants heard white noise through noise-cancelling headphones and had foam inserts in their ears. In the auditory-only condition, participants closed their eyes. In the audiovisual condition, they had auditory and visual information. The car's actual TTC at the time of the prompt was measured. Using a signal-detection model for the participant's response (Hassan and Massof, 2012), the TTC at which the participant perceived the TTC of the car to be sufficiently long to be able to cross the street was estimated, which can be viewed as the mean accepted gap duration (e.g., Baurès *et al.*, 2014). The three modality conditions were compared with respect to 'bias' which was defined as the difference between a participant's mean accepted gap duration and their measured street crossing time. The closer the bias value is to zero, the better the accuracy. On average, normally sighted participants achieved better accuracy when auditory and visual information was present concurrently than when either was present alone. However, only the difference between the audiovisual condition and the auditory-only condition was significant.

There are two noteworthy limitations of the previous studies. First, they provided only indirect information about the relative contribution of auditory and visual information when both were presented concurrently. Because auditory TTC and visual TTC were always identical, the relative importance of auditory and visual information could only be inferred in two rather indirect ways. One approach is based on the assumption that the modality that contributed more to the judgments was the one that resulted in mean TTC estimates most similar to those in the audiovisual condition. However, the present results demonstrate that participants used auditory information, although mean TTC estimates in the audiovisual condition were closer to the condition in which only visual information was provided. In the second approach, if auditory TTC and visual TTC are always identical in the audiovisual condition, the occurrence of multimodal integration is determined by comparing the reliability (inverse of variance) of the TTC estimates in the two unimodal and the audiovisual conditions. If reliability in the audiovisual condition is greater than in the unimodal conditions, this is taken as audiovisual integration (multimodal advantage); if reliability in the audiovisual condition approximates the sum of the reliabilities in the unimodal conditions, this is taken as optimal integration, compatible with a Bayesian framework (cf. Oruc *et al.*, 2003). However, this analysis cannot account for reliance on other cues such as final size (discussed subsequently), and typically assumes unbiased estimates and uncorrelated cues (but see Oruc *et al.*, 2003). For example, the classical 'inverse effectiveness' approach, in which the inverse variance of the estimates based on single cues is compared to the inverse variance based on the combined cues, assumes that the cues are uncorrelated, although a correction for correlation can be applied (Oruc *et al.*, 2003). TTC estimates are typically biased (Tresilian, 1995), and the cues are correlated (Oberfeld *et al.*, 2011). In fact, the present results show that subjects used both auditory and visual cues, although no multimodal advantage was observed. In summary, to measure the weights assigned to different cues, it is necessary to dissociate these cues experimentally, as pointed out by Rushton and Wann (1999).

The second limitation of prior studies of audiovisual integration in TTC judgments, is that it was not determined which of multiple cues presented *within* a given modality (Tresilian, 1994) contributed to TTC estimates in the audiovisual condition. For example, simple 'heuristic' cues such as the final optical size of an approaching object are known to have to have a strong influence on TTC estimates in the visual domain, even when reliable TTC information is also available (DeLucia, 1991). In other words, approaching objects with relatively large optical sizes appear as closer and as arriving earlier, compared to objects with smaller optical sizes, consistent with a size heuristic of 'larger is closer' (DeLucia, 2004). It becomes important to measure the influence of such approximate cues on audiovisual TTC estimates.

These limitations were addressed in the current study. We used an experimental design to directly measure the influence of different auditory and visual cues on TTC estimates. Participants judged the TTC of an approaching object that was presented in the visual modality, auditory modality, or in both modalities concurrently. When the auditory and visual stimuli were presented concurrently, the TTC information provided by the two modalities was jittered slightly against each other (cf. Rushton and Wann, 1999), so that the auditory and visual TTCs were not perfectly correlated as they were in previous studies. This allowed us to use a psychophysical reverse correlation approach (e.g., Ahumada and Lovell, 1971; Beard and Ahumada, 1998) to estimate the influence of auditory and visual cues on the TTC estimates, and to assess their relative importance. In psychophysical reverse correlation, also termed perceptual weight analysis (Berg, 1989), the trial-by-trial data are analyzed to estimate the influence of different sources of information or stimulus components on the response of the participant (for example applications see Murray, 2011). For example, this method has been used to determine the relative weight of three different acoustic cues for the discrimination of auditory motion (Lutfi and Wang, 1999).

We focused on two questions. First, is information from auditory and visual modalities integrated in TTC estimates? We examined whether TTC estimates were influenced by visual and auditory information when both were available and, if they were, we determined the relative contributions of each. Second, is there a multimodal advantage? We examined whether TTC estimates were more accurate (i.e., closer to the veridical value) or more precise (i.e., less variable) when visual and auditory information were concurrently available compared to either one alone.

We will see that although a multimodal advantage does not occur, both auditory and visual cues are used. However, they are not weighted equally in TTC judgments, and the reliance on cues that more reliably provide accurate information (τ) compared to other more heuristic cues differs between the two modalities.

2. Method

2.1. Participants

Twenty-four students from Texas Tech University participated for course credit (12 men, 12 women; ages 18–24 yrs, $M = 19.68$, SD $= 1.55$). All participants self-reported normal or corrected-to-normal visual acuity and normal hearing.

2.2. Apparatus and Displays

Displays were created with a Dell Optiplex 390 computer with an AMD Radeon HD 6350 graphics card and a Sound Blaster X-fi Titanium sound card. Visual displays were created in 800 × 600 resolution and presented with a monitor refresh rate of 75.0 Hz. Motion appeared smooth and without flicker.

The set-up is shown in Fig. 1. Displays were viewed with two eyes (binocular disparity information was not presented) on a monitor with a 43.18-cm diagonal. Auditory stimuli were generated digitally (sampling rate 48.0 kHz, 24 bits resolution) and presented through a mono speaker located on top of the monitor viewed by the participants. Using a single loudspeaker ensured that the sound source was perceived as being in front of the listener, and the sound intensity was well above the threshold of hearing even in the presence of the background noise in the room (about 34.0 dBA). The speaker was preferable to headphones which can result in sounds being perceived as localized within the head, sometimes even if sophisticated binaural simulations are used (Begault and Wenzel, 2001).

Visual and auditory displays simulated an object that approached the virtual eye/ear at a constant speed for one second and provided TTC information. The visual object consisted of an untextured colored square that expanded sym-

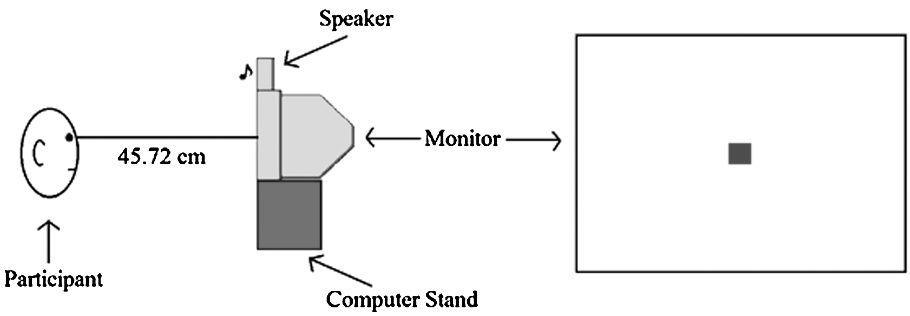

Figure 1. Diagram of experimental set-up and display.

metrically about the observer's line of sight. The auditory object consisted of a 1.0 kHz tone that increased in intensity while maintaining a constant frequency (see Note 2). The visual expansion and the increase in acoustic intensity corresponded to an object approaching the observer at constant speed along the line of sight. We used simple displays so that we could directly compare the influence of auditory tau, which is specified by increasing intensity (Shaw *et al.*, 1991), and visual tau, which is specified by increasing optial size (Lee, 1976), without potentially confounding effects of other depth cues which might occur with more complex scenes (DeLucia, 2004).

The stimuli were presented for 1 s and then became invisible/inaudible. The object's actual TTC at the time of its disappearance was 0.5 s, 1.0 s, 1.5 s, 2.0 s, 2.5 s, or 3.0 s. To create a variety of scenes so that participants would not exhibit stereotyped responses, we varied the object's starting distance within each level of TTC. In the 'near distance' condition the object was closer to the virtual eye when the scene started and when it ended, compared to 'far distance' condition. As shown in Table 1, within each level of TTC, the object moved faster in the far distance condition than in the near distance condition so that both scenes resulted in the same TTC. In Table 1, the optical properties of the approaching object are described in degrees of visual angle. The auditory properties of the object are described in terms of sound pressure level (SPL), not to be confused with sound intensity (Hartmann, 2005). A representative scene that would result in these properties is also shown in Table 1. In this virtual scene the visual object had a diameter of 0.5 m and the auditory object had a sound pressure level of 85.9 dB SPL at a distance of 2.0 m from the observer. We selected the range of levels so that all sounds were clearly audible during their entire duration (at least 20.0 dB above the background noise level), but still not uncomfortably loud at the end of the stimulus.

The minimal difference in sound pressure level between the beginning and the end of the sounds was 2.41 dB, which is above the intensity difference limen for brief 1.0 kHz pure tones at a level of approximately 50.0 dB SPL (Florentine, 1986). Pilot observations were conducted to ensure that differences in intensity changes among the auditory stimuli were discriminable (see also Oberfeld *et al.*, 2014). The change in optical size during the approach event was above threshold for detection of change in optical size from motion in depth (Hills, 1975). The presentation of the visual and auditory stimuli was controlled and synchronized with DirectRT (v. 2010).

There were three modality conditions. In the *auditory condition*, only the auditory object was presented. In the *visual condition*, only the visual object was presented. In both of these conditions, the object's TTC values at the end of the approach were 0.5, 1.0, 1.5, 2.0, or 2.5 s (the 3.0 s TTC was used only in the audiovisual condition); each was represented in the near and far distances.

Table 1.
Optical and auditory display parameters, and representative 3D scene

		Near distance						Far distance					
		TTC (s)						TTC (s)					
		0.5	1.0	1.5	2.0	2.5	3.0	0.5	1.0	1.5	2.0	2.5	3.0
Optical and auditory parameters													
Visual angle (deg)	First frame	2.24	1.79	1.53	1.40	1.32	1.29	0.92	0.71	0.59	0.50	0.44	0.40
	Last frame	6.54	3.49	2.48	2.06	1.82	1.70	2.67	1.38	0.97	0.74	0.61	0.53
Sound pressure level (dB SPL)	First frame	69.81	67.86	66.48	65.70	65.19	64.98	62.08	59.79	58.18	56.74	55.68	54.82
	Last frame	79.12	73.66	70.69	69.08	67.99	67.39	71.33	65.62	62.53	60.14	58.50	57.23
Representative 3D scene (assuming sphere diameter of 0.5 meters)													
Distance (m)	First frame	12.775	16.00	18.75	20.50	21.75	22.275	31.125	40.50	48.75	57.50	65.00	71.775
	Last frame	4.375	8.20	11.55	13.90	15.75	16.875	10.725	20.70	29.55	38.90	47.00	54.375
Velocity (km/h)		31.50	29.25	27.00	24.75	22.50	20.25	76.50	74.25	72.00	69.75	67.50	65.25

Each of these ten unique scenes was presented five times for a total of 50 trials in each of the auditory and visual conditions.

In the *audiovisual condition*, auditory and visual objects were presented simultaneously. In this condition, we varied the difference between the auditory and the visual TTC so that the two TTCs were no longer perfectly correlated. The difference between the auditory TTC (denoted TTC_a) and the visual TTC (denoted TTC_v) was 0.5 s, 0 s, or -0.5 s. To create scenes in which TTC_v was less than TTC_a, each of the visual objects with TTC values of 0.5, 1.0, 1.5, 2.0, and 2.5 s was paired with an appropriate condition from Table 1 such that the auditory TTC was 0.5 s longer (i.e., $TTC_a = TTC_v + 0.5$ s). To create scenes in which TTC_v was greater than TTC_a, each of the visual objects with TTC values of 1.0, 1.5, 2.0, and 2.5 s was paired with an appropriate condition such that the auditory TTC was 0.5 s shorter (i.e., $TTC_a = TTC_v - 0.5$ s). The visual TTC of 0.5 s could not be combined with a shorter auditory TTC because TTC cannot be zero. The pairings were used in both the near and far conditions, resulting in 28 unique scenes. Each was replicated five times for a total of 140 trials in the audiovisual condition.

2.3. Procedure

Participants viewed, and listened to, the visual and auditory stimuli, respectively; the monitor and speaker were about 0.46 m from the participant. Participants were not restricted from moving their head, but the experimenter monitored them to ensure that they remained approximately at the correct viewing distance from the monitor. Participants were instructed to press a mouse button when they thought that the object would hit or pass them had the object continued moving after it was no longer visible and audible. Feedback on response accuracy was not provided, in order to avoid response strategies based on the feedback. Practice trials were provided to familiarize participants with the task. TTC judgments were measured as the time interval between the last video frame of the visual stimulus, or the last audio sample of the auditory stimulus, and the time at which the participant pressed the mouse button. Trials in which the participant responded before the object disappeared were removed from the final analysis (less than 1.0% of the trials).

All participants completed the three modality conditions. Order was randomly assigned to and completely counter-balanced across participants. Trials were blocked by modality, with TTC and final distance intermixed within each block. In the audiovisual condition, TTC, final distance, and the difference between TTC_a and TTC_v were intermixed within a block.

3. Results

A boxplot of the data for each combination of TTC and modality condition showed that two participants had more outliers (using the Tukey criteria of 1.5 times the interquartile range above the third quartile) than the other participants and were thus excluded (the pattern of results was the same when these participants were retained). First, we compared the modality conditions by analyzing mean TTC estimates and the variability of the estimates. Then we used regression analyses to determine the relative weighting of auditory and visual information.

3.1. TTC Estimates

3.1.1. Mean TTC Estimates

Results are shown in Figs 2 and 3. Mean TTC estimates were analyzed with a 2 (Distance: near, far) × 4 (TTC: 1.0, 1.5, 2.0, 2.5 s) × 5 (Modality: auditory only, visual only, visual TTC = auditory TTC, visual TTC > auditory TTC, visual TTC < auditory TTC) repeated-measures analysis of variance (rmANOVA) using a univariate approach and Greenhouse–Geisser correction for the degrees of freedom. In the audiovisual conditions, the factor TTC was set to the value of the visual TTC. This analysis did not include the 0.5 s or 3.0 s TTC value because both were not included in all of the audiovisual conditions.

We replicated the well-established finding that mean estimated TTC increases as actual TTC increases (e.g., Brendel *et al.*, 2012; Oberfeld and Hecht, 2008; Schiff and Detwiler, 1979), indicated by a main effect of TTC,

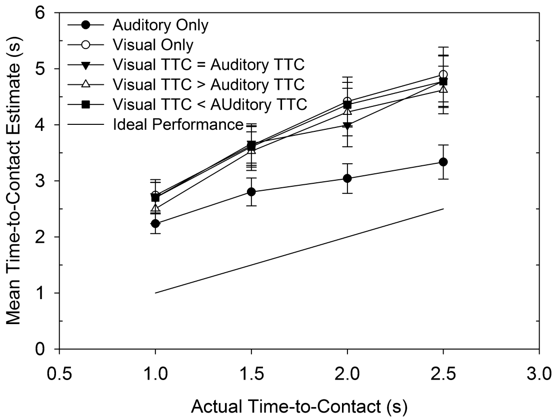

Figure 2. Mean TTC estimates as a function of actual time-to-contact and modality conditions. In the audiovisual conditions, the value on the x-axis specifies the visual TTC. The line without symbols represents ideal TTC estimation. Error bars represent ± 1 standard error of the mean (SEM). The mean actual TTC was 1.75 s.

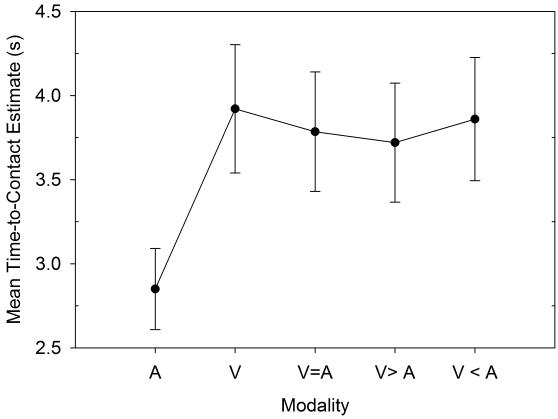

Figure 3. Mean TTC estimates as a function of modality conditions: auditory (A), visual (V), and audiovisual in which TTC_v was the same as TTC_a (V = A), TTC_v was greater than TTC_a (V > A), or TTC_v was less than TTC_a (V < A). Error bars represent ± 1 SEM.

$F(3, 63) = 74.06$, $p = 0.0001$, $\hat{\varepsilon} = 0.37$, $\eta_p^2 = 0.78$, and shown in Fig. 2. Participants overestimated TTC in all conditions, which is not surprising given our simple displays (Geri *et al.*, 2010; Gray and Regan, 1998) (see Note 3). Mean TTC judgments were greater for the farther distance than the closer distance, $F(1, 21) = 136.88$, $p = 0.0001$, $\eta_p^2 = 0.87$. The interaction between TTC and distance also was significant: $F(3, 63) = 10.67$, $p = 0.0001$, $\hat{\varepsilon} = 0.80$, $\eta_p^2 = 0.34$.

Of most interest are the effects involving modality. The main effect of modality was significant, $F(4, 84) = 7.21$, $p = 0.0024$, $\hat{\varepsilon} = 0.48$, $\eta_p^2 = 0.26$. Mean TTC estimates are displayed in Fig. 3. Tukey's HSD tests indicated that the mean TTC judgment was smaller when the auditory object was presented alone than when the visual object was presented alone or when visual and auditory stimuli were presented concurrently, $p < 0.05$. It also is apparent from Fig. 2 that, in terms of the deviation from the veridical value, the mean TTC judgments in the auditory condition were more accurate compared to the other conditions (compare means to ideal performance).

There was a significant interaction between modality and TTC, $F(12, 252) = 8.07$, $p = 0.0001$, $\hat{\varepsilon} = 0.44$, $\eta_p^2 = 0.28$. As can be seen in Fig. 2, the slope of the function relating estimated TTC and actual TTC was smaller in the auditory-only condition than in the other conditions. Separate one-way rmANOVAs and Tukey's HSD tests were conducted to determine the effects of modality at each level of TTC. As represented in Fig. 3, the mean TTC judgment was shorter in the auditory condition than in any of the other conditions (Fs > 8.26, ps < 0.0006). This difference was significant except when TTC was 1 s. This may have occurred because the perceived TTC is more compressed in the auditory condition than in the other conditions. It has been proposed that people rely more on cognitive processes when TTC is long and rely more on τ information when TTC is short (DeLucia, 2013, 2015; Tresilian, 1995).

The difference between the visual condition and the audiovisual conditions was not significant. This is compatible with Schiff and Oldak (1990) who found that mean TTC estimates in the audiovisual condition were closer to the visual-only than to the auditory-only condition. This is noteworthy because their stimuli (films of approaching cars) were much more realistic than ours. However, as will be shown with the regression analysis, it is not the case that TTC estimates were based solely on visual information. Both types of information were influential.

3.1.2. Variability of TTC Estimates

Apart from the systematic deviations of the TTC estimates from the veridical value (accuracy), the data also provide information about the variability of the TTC estimates across presentations of the same stimulus (precision). This ‘variable error’ (VE) in terms of Fechner (1860) is closely related to the just-noticeable difference estimated for example from psychometric functions (e.g., Treisman, 1963). For each combination of participant, modality condition, final distance, and TTC, we computed the VE as the standard deviation of the TTC estimates across the five replicates. The bias in the mean estimates (see Fig. 2) renders the interpretation of the VE difficult. The variability of TTC estimates in a prediction motion task is known to increase with the TTC (e.g., Oberfeld and Hecht, 2008). This effect might be due, for instance, to the necessity of timing a longer interval between the disappearance of the stimulus and the button press (e.g., Baurès *et al.*, 2011; Tresilian, 1995), because the variability of produced time intervals increases with the interval duration (e.g., Wearden and Lejeune, 2008). Thus, for example a higher VE in the visual compared to the auditory condition might in part be due to the longer TTC estimates produced in the former condition (see Fig. 2). We accounted for the effect of differences in the mean estimates by analyzing the coefficient of variation (CV = standard deviation divided by mean) rather than the standard deviation (VE) directly. The CV is similar to a Weber fraction in an experiment measuring JNDs (e.g., Treisman, 1963). As Fig. 4 shows, the CV was similar for the auditory and the visual condition for the far distance scenarios, and smaller for the visual than for the auditory condition for the near distance scenarios. The CVs were analyzed with a 2 (Distance: near, far) $\times$ 4 (TTC: 1.0, 1.5, 2.0, 2.5 s) $\times$ 5 (Modality: auditory only, visual only, visual TTC = auditory TTC, visual TTC > auditory TTC, visual TTC < auditory TTC) rmANOVA using a univariate approach and Greenhouse–Geisser correction for the degrees of freedom. The effect of modality condition on the

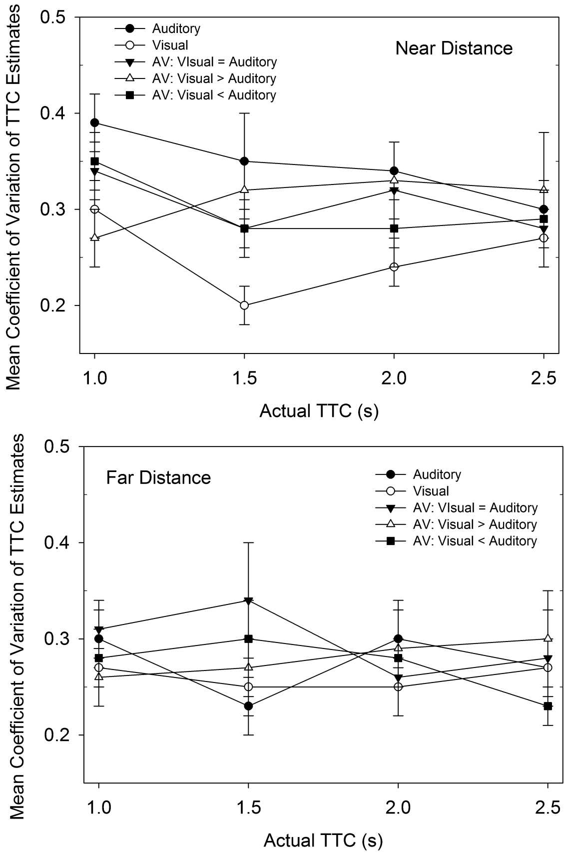

Figure 4. Mean coefficient of variation as a function of modality condition, final distance and TTC. In the audiovisual conditions, the value on the x-axis specifies the visual TTC. Error bars show $\pm$ 1 SEM. Top: Near distance. Bottom: Far distance.

CV was not significant, $F(4, 84) = 1.45$, $p = 0.25$, $\hat{\varepsilon} = 0.50$, $\eta_p^2 = 0.06$. The interaction between modality condition and distance was also not significant, $F(4, 84) = 2.60$, $p = 0.055$, $\hat{\varepsilon} = 0.80$, $\eta_p^2 = 0.11$. The effect of distance on the CV was significant, $F(1, 21) = 7.31$, $p = 0.013$, $\eta_p^2 = 0.26$. The remaining effects were not significant (ps > 0.19).

As discussed above, in the study by Zhou *et al.* (2007) the just-noticeable difference, which is a measure of precision just as the CV, was slightly smaller in the audiovisual condition than in the unimodal conditions. However, the audiovisual precision fell between the two unimodal conditions when one participant who showed a higher JND in the visual than in the auditory condition was excluded (see Note 1). In our results, the CV provided no evidence for a multimodal advantage in the sense of lower variability in the audiovisual condition compared to the unimodal conditions.

3.2. Regression Analyses

In the audiovisual condition, TTC_v and TTC_a were varied. Therefore, it was possible to quantify the impact of these two sources of information on TTC estimates. Using multiple linear regression analyses (e.g., Gray and Regan, 1998), we determined whether and to which extent participants relied on exact cues (auditory and visual τ, which specified auditory and visual TTC) and heuristic cues (final optical size and final sound pressure level).

The multiple linear regression model contained an intercept term and the predictors TTC_a and TTC_v, the inverse of the final optical size, θ_{final} (i.e., optical size on the final frame in degrees of visual angle) (e.g., DeLucia, 1991), and the inverse of the final sound pressure level (SPL_{final}, measured in dB SPL). The inverse values were used because they are positively and more linearly related to the actual TTCs. This model was fitted separately to each of the 22 individual data sets (audiovisual condition only), each of which contained 140 trials. The predictors were entered simultaneously. For a given cue (e.g., TTC_a), a regression coefficient equal to zero implies that the value of this cue had no influence on the TTC estimate. A regression coefficient greater than zero implies that the TTC estimate increased as the value of the cue increased. A regression coefficient smaller than zero indicates the opposite relation. Note that in our experiment TTC_a and TTC_v were correlated, but not linearly dependent. The inverse final size and final SPL were also correlated to the TTCs, as for all real-world approaching objects (e.g., Oberfeld *et al.*, 2011). However, according to the Gauß–Markov theorem (Gauß, 1821) the estimates provided by the multiple regression analysis will remain unbiased. This is another important advantage of our reverse correlation approach compared to other methods for studying multisensory integration, which are typically negatively affected by correlated predictors. As we noted earlier, example, the classical approach, in which the inverse variance of the estimates based on single cues is compared to the inverse variance based on the combined cues, assumes that the cues are uncorrelated.

The goodness of fit in terms of R^2 ranged between 0.11 and 0.83, and was, on average, $R^2 = 0.58$ (SD = 0.19), which is a moderately high proportion of variance accounted for. For all but four participants, R^2 was higher than 0.5. Individual Q–Q plots of the residuals of estimated TTC showed no systematic deviations from normality, and individual plots of estimated TTC as a function of the predictors showed no severe deviations from linearity. Across the 22 participants, 14 of the regression weights for TTC_v were significantly different from 0 ($p < 0.05$, two-tailed). In contrast, only two regression weights for TTC_a were significantly different from 0. The regression coefficients for the inverse final size and inverse final SPL were significant for six participants and five participants, respectively. Table 2 shows the average estimated regres-

Table 2.
Audiovisual condition: mean estimated regression parameters (and 95% confidence intervals). Model: $\text{TTC}_{\text{est}} = \beta_0 + \beta_{\text{TTC}_\text{a}}\text{TTC}_\text{a} + \beta_{1/\text{SPL}}1/\text{SPL}_{\text{final}} + \beta_{\text{TTC}_\text{v}}\text{TTC}_\text{v} + \beta_{1/\theta}1/\theta_{\text{final}}$. Units: TTCs measured in seconds, $\text{SPL}_{\text{final}}$ in dB SPL, θ_{final} in degrees of visual angle

Parameter	Estimate	95% CI
β_0	−2.99	[−4.08, −1.89]
$\beta_{\text{TTC}_\text{a}}$	−0.37	[−0.51, −0.23]
$\beta_{1/\text{SPL}}$	239.5	[158.21, 320.80]
$\beta_{\text{TTC}_\text{v}}$	1.19	[0.86, 1.52]
$\beta_{1/\theta}$	0.83	[0.41, 1.25]

sion parameters. Averaged across participants, the regression coefficients were significantly different from 0 for all of the four cues. Interestingly, the auditory TTC received, on average, a small negative weight.

What can be concluded concerning the relative importance of the four cues for the TTC estimates? It is obvious that because the four cues are measured on different scales (e.g., seconds for the TTCs, 1/[dB SPL] for the inverse final SPL), the relative size of the regression coefficients cannot be used for this purpose. We conducted two different analyses addressing the issue of relative importance.

In the first analysis, the same type of multiple regression analysis as above was conducted, but all predictors (cues) were z-standardized. For each experimental condition (auditory, visual, audiovisual), the mean, M, and standard deviation, SD, for each of the four cues (TTC_a, TTC_v, $1/\theta_{\text{final}}$, $1/\text{SPL}_{\text{final}}$) were computed across all trials presented to the participant. Next, the z-standardized values were computed as, for example, $z_{\text{TTC}_\text{a}} = (\text{TTC}_\text{a} - M_{\text{TTC}_\text{a}})/\text{SD}_{\text{TTC}_\text{a}}$. The four z-standardized cues were entered as predictors in the multiple regression analysis. The resulting regression coefficients show by how much the TTC estimates changed when the given cue changed by one standard deviation. Thus, this analysis provides information about the relative influence of the four cues on the TTC estimates, in relation to the variation of each cue within each experimental condition. The goodness-of-fit is identical to the model using the unstandardized predictors (Table 2).

Figure 5 displays the mean perceptual weights. Surprisingly, on average the auditory TTC received a weak negative weight, which means that the TTC estimates showed a slight decrease with increases in TTC_a. However, the inverse final SPL was clearly positively related to the TTC estimates. Thus, the participants did use auditory cues, but did not rely on auditory TTC. Instead, they used a heuristic cue which is an analog of the visual 'final size' cue reported to have a strong influence on TTC estimates in the visual domain, that is, we

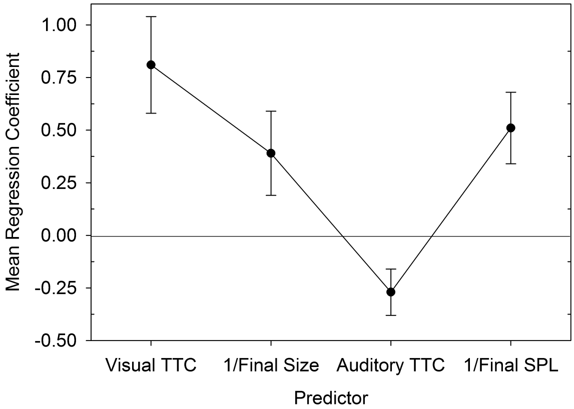

Figure 5. Mean regression coefficients as a function of z-standardized predictor (cue) in the audiovisual condition. TTC denotes time-to-collision. 1/Final Size denotes the inverse of the visual object's optical size on the final frame. 1/Final SPL denotes the inverse of the auditory object's sound pressure level on the final frame. Error bars show 95% confidence intervals.

demonstrated an auditory version of the size-arrival effect (DeLucia, 1991). Because we observed some negative regression weights, the absolute values of the regression weights for the z-standardized predictors were analyzed in order to decide which cues had a stronger influence on the TTC estimates, using an rmANOVA with the within-subjects factors modality (auditory, visual) and cue type (TTC_v or TTC_a versus $1/\theta_{final}$ and $1/SPL_{final}$). There was a significant effect of modality, $F(1, 21) = 6.74$, $p = 0.017$, $\eta_p^2 = 0.24$, reflecting the higher average value of the regression coefficients for the two visual compared to the two auditory cues. There was no significant main effect of cue type, $F(1, 21) = 1.20$. In the visual domain, the regression coefficient for TTC was higher than for $1/\theta_{final}$, while in the auditory domain we found a higher absolute value of the regression coefficient for $1/SPL_{final}$ than for TTC_a, confirmed by a significant modality × cue type interaction, $F(1, 21) = 14.45$, $p = 0.001$, $\eta_p^2 = 0.41$.

This analysis based on z-standardized predictors allows us to compare regression coefficients between cues measured on the same scale because all predictors are now expressed relative to the range of variation presented during the experiment. At the same time, this has the consequence that the estimated weights cannot be used to predict weights for a different experiment presenting different ranges of variation. For example, if one would conduct a follow-up experiment where TTC_v varies much less than TTC_a, then for the same participant using the same TTC estimation strategy (i.e., the same set of perceptual weights) the estimated weight for TTC_v would be lower than in the present experiment, because a change in TTC_v by one SD would correspond to a much smaller change on the physical scale (seconds). In contrast, the regression coefficients reported in Table 2 for the unstandardized cues can be used to predict the estimated TTC in a future experiment, regardless of the range of variation of the cues presented in the latter experiment.

In the second analysis, we used a definition of relative importance based on the proportion of variance accounted for by the four different cues. Given the presented range of variation in TTC_a, TTC_v, $1/SPL_{final}$, and $1/\theta_{final}$, which cue accounted for the highest amount of variance in the TTC estimates? For this analysis of relative importance, when one refers to the contribution a variable makes to the prediction of a criterion variable by itself, and in combination with other predictor variables, it is important to consider the correlations among the four predictors (for a detailed discussion of these issues see Tonidandel and LeBreton, 2011). We used the 'dominance analysis' approach proposed by Budescu (1993), which was shown to be a useful measure both on theoretical grounds and in simulation studies (LeBreton *et al.*, 2004; Tonidandel and LeBreton, 2011). Dominance analysis provides a quantitative measure of relative importance by examining the change in the variance-accounted-for (R^2) resulting from adding a predictor to all possible regression models containing subsets of the predictors. For example, for $K = 2$ predictors, there are three possible models containing subsets of the two predictors (null model containing only an intercept term, model containing only first predictor plus intercept, model containing only second predictor plus intercept). A predictor's *general dominance weight* (Azen and Budescu, 2003) is found by averaging the squared semipartial correlations across all of the possible models. This measure indexes a variable's contribution to the prediction of the dependent variable, by itself and in combination with the other predictors. The general dominance weights were computed for each participant using a SAS macro by Azen and Budescu (2003).

Figure 6 shows the mean dominance weights. Consistent with the regression analysis using z-standardized predictors reported above (Fig. 5), the dominance weight was smaller for TTC_a than for TTC_v, and higher for $1/SPL_{final}$ than for TTC_a. Thus, the analysis again showed that both auditory and visual cues were important for the TTC estimates. However, the ordering of the relative importance of TTC_v and $1/\theta_{final}$ in terms of the dominance weight was reversed compared to the regression analysis using z-standardized predictors. The dominance weight was higher for the inverse final size than for TTC_v, thus providing even stronger evidence for the use of 'heuristic' cues than the preceding analysis. As explained by Tonidandel and LeBreton (2011), standardized regression weights do not appropriately partition variance when predictors are correlated; consequently, the assessment of relative importance should be based on the dominance weights.

The dominance weights were analyzed with an rmANOVA with the within-subjects factors modality (auditory, visual) and cue type (TTC_v or TTC_a

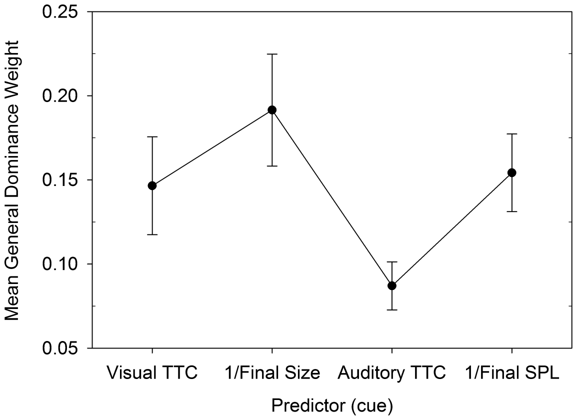

Figure 6. Mean general dominance weight as a function of predictor (cue) in the audiovisual condition. TTC denotes time-to-collision. 1/Final Size denotes the inverse of the visual object's optical size on the final frame. 1/Final SPL denotes the inverse of the auditory object's sound pressure level on the final frame. Error bars show 95% confidence intervals.

versus $1/\theta_{\text{final}}$ and $1/\text{SPL}_{\text{final}}$). There was a significant effect of modality, $F(1, 21) = 69.6$, $p < 0.001$, $\eta_p^2 = 0.76$, reflecting the higher average dominance weights assigned to the two visual compared to the two auditory cues. The average dominance weight for the TTCs was significantly smaller than for the heuristic cues ($1/\theta_{\text{final}}$ and $1/\text{SPL}_{\text{final}}$), $F(1, 21) = 21.3$, $p < 0.001$, $\eta_p^2 = 0.50$. In the auditory domain, the higher reliance on $1/\text{SPL}_{\text{final}}$ compared to TTC_{a} was more pronounced than the difference between the weights for these two types of cues in the visual domain, confirmed by a significant modality × cue type interaction, $F(1, 21) = 9.0$, $p < 0.001$, $\eta_p^2 = 0.30$.

Again, it should be noted that the dominance weights depend on the actual range of cue values presented in the experiment, as does the regression analysis using z-standardized predictors. For example, imagine that a participant applies the same regression weights for TTC_{a} and TTC_{v} (unstandardized) in two experimental conditions. If the range of presented auditory TTCs is similar to the range of visual TTCs in one condition, but much lower than for visual TTC in the second condition, then the variance accounted for by TTC_{a} will be much lower in the second condition. As a consequence, the dominance weight for TTC_{a} will be smaller in the condition presenting the smaller range of variation in TTC_{a}.

To investigate whether the observed information integration strategy is specific to the audiovisual condition, we fitted separate multiple regression models to the data from the auditory and visual conditions. For the auditory condition, the predictors were TTC_{a} and the inverse final SPL. For the visual condition, TTC_{v} and the inverse final size served as predictors. Table 3 shows

Table 3.
Auditory condition: mean estimated regression parameters (and 95% confidence intervals). Model: $TTC_{est} = \beta_0 + \beta_{TTC_a} TTC_a + \beta_{1/SPL} 1/SPL_{final}$. Units: TTCs measured in seconds, SPL_{final} in dB SPL

Parameter	Estimate	95% CI
β_0	−6.62	[−8.14, −5.10]
β_{TTC_a}	−0.004	[−0.15, 0.14]
$\beta_{1/SPL}$	487.73	[389.60, 585.86]

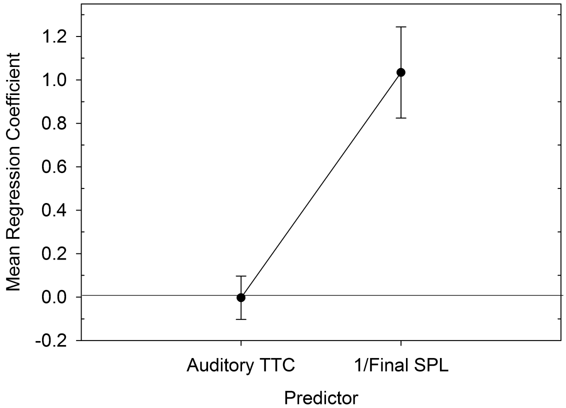

Figure 7. Mean regression coefficients as a function of z-standardized predictor (cue) in the auditory condition. TTC denotes time-to-collision. 1/Final SPL denotes the inverse of the auditory object's sound pressure level on the final frame. Error bars show 95% confidence intervals.

the average regression parameters with unstandardized predictors for the auditory condition, and Fig. 7 displays the mean perceptual weights computed with z-standardized predictors. The figure clearly demonstrates that in the auditory condition participants also based their TTC estimates on the final SPL rather than on the auditory TTC. The dominance weights are shown in Fig. 8. Not surprisingly, both analyses showed a higher importance of $1/SPL_{final}$ than of TTC_a. The average ratio of the dominance weights for final SPL and auditory TTC was even higher in the auditory-only condition ($M = 3.6$, $SD = 1.4$) than in the audiovisual condition ($M = 1.9$, $SD = 0.7$).

The mean regression parameters with unstandardized predictors for the visual condition are displayed in Table 4. The mean perceptual weights computed with z-standardized predictors and the mean dominance weights are shown in Figs 9 and 10, respectively. Both analyses showed that the participants, on average, assigned a slightly higher weight to the final optical size

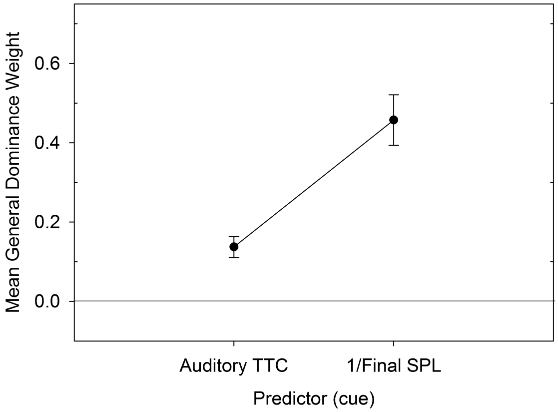

Figure 8. Mean general dominance weight as a function of predictor (cue) in the auditory condition. TTC denotes time-to-collision. 1/Final SPL denotes the inverse of the auditory object's sound pressure level on the final frame. Error bars show 95% confidence intervals.

Table 4.
Visual condition: Mean estimated regression parameters (and 95% confidence intervals). Model: $TTC_{est} = \beta_0 + \beta_{TTC_v} TTC_v + \beta_{1/\theta} 1/\theta_{final}$. Units: TTCs measured in seconds, θ_{final} in degrees of visual angle

Parameter	Estimate	95% CI
β_0	0.6	[0.33, 0.87]
β_{TTC_v}	1.07	[0.74, 1.40]
$\beta_{1/\theta}$	1.71	[1.36, 2.06]

than to the visual TTC, but the difference was neither significant for the regression weights, $t(21) = 0.36$, $p = 0.74$, nor for the dominance weights, $t(21) = 1.41$, $p = 0.121$. The average ratio of the dominance weights for final size and visual TTC was similar in the visual-only condition ($M = 1.2$, SD = 0.5) and in the audiovisual condition ($M = 1.4$, SD = 0.6).

4. General Discussion

The results of the current study have several important implications for understanding audiovisual integration of TTC information in judgments of approaching objects. First, TTC judgments of an approaching auditory object are shorter than for an approaching visual object with the same actual TTC. This pattern was also reported in earlier studies (Hassan, 2012; Schiff and Oldak, 1990; but see Zhou *et al.*, 2007) and suggests that the manner in which people estimate TTC depends on the modality in which TTC information is

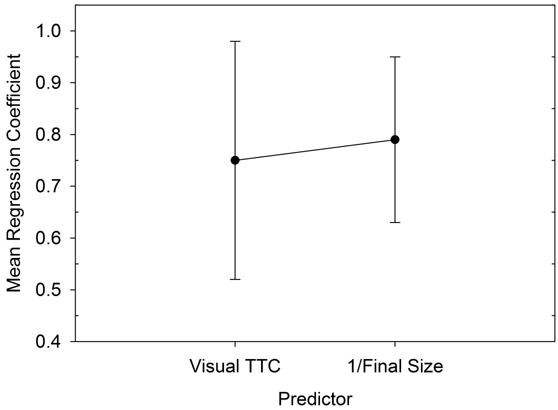

Figure 9. Mean regression coefficients as a function of z-standardized predictor (cue) in the visual condition. TTC denotes time-to-collision. 1/Final Size denotes the inverse of the visual object's optical size on the final frame. Error bars show 95% confidence intervals.

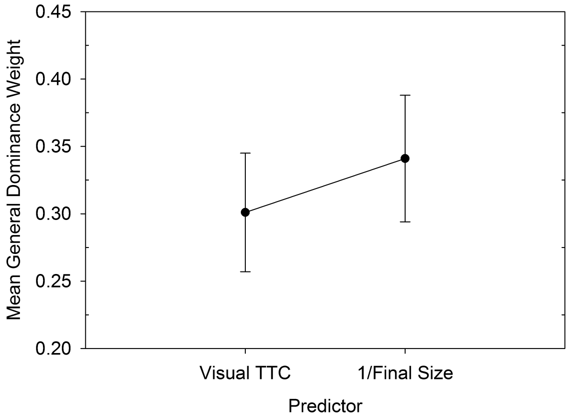

Figure 10. Mean general dominance weight as a function of predictor (cue) in the visual condition. TTC denotes time-to-collision. 1/Final Size denotes the inverse of the visual object's optical size on the final frame. Error bars show 95% confidence intervals.

presented. It has been suggested (and reported by study participants) that visual TTC judgments involve mental extrapolation or visual imagery (DeLucia and Liddell, 1998; Schiff and Oldak, 1990). In other words, people visualize the object approaching after it disappears. It is not obvious how people extrapolate an object's motion when only auditory information is presented, but it is reasonable to expect that visual and motion extrapolation rely on different modality-specific processes (Schmiedchen *et al.*, 2013). Another alternative is to first extract a TTC estimate from the visual or auditory cues, and then to time the motor response to coincide with the estimated TTC (Baurès, *et al.*, 2011; Tresilian, 1995). One admittedly speculative possibility is that observers use mental extrapolation in the visual modality and the timing mechanism in the auditory modality, potentially because it is easier to visualize (i.e., see in the mind's eye) changes in distance than it is to 'auditorialize' (i.e., hear in the mind's ear) such changes. Another possibility is that TTC estimates involve a timing mechanism in both modalities, and that this mechanism is similar to that used in time reproduction tasks (Grondin, 2010); but the time interval to be reproduced is estimated on the basis of different cues in the two modalities, resulting in different TTC estimates.

Second, when people estimate the TTC of an approaching object that is represented in both the auditory and visual modalities, judgments are closer to that obtained during judgments of the visual object than judgments of the auditory object. The implication is that participants relied primarily on visual information when both auditory and visual information were available, compatible with results by Zhou *et al.* (2007). This conclusion was clearly supported by the regression analyses, which quantified the influence of auditory and visual cues on the TTC estimates. It remains to be shown whether the finding of 'visual dominance' can be attributed to a ventriloquist effect, that is, the shift of the perceived distance of an auditory stimulus towards the position of a concurrently presented visual stimulus (Bertelson and Aschersleben, 1998; Hládek *et al.*, 2013).

Third, both auditory and visual TTC judgments were influenced by heuristic cues other than TTC. In the visual domain, the final size (and potentially distance; Rushton, 2004) of the object had a significant influence on the TTC estimates, compatible with previous reports of size-arrival effects (DeLucia, 1991). In the auditory domain, we observed an even more extreme pattern, with near-zero weight assigned to the auditory TTC whereas the TTC estimates strongly depended on the final SPL. This pattern is compatible with the report of Zhou *et al.* (2007) that participants were more sensitive to the final distance of the auditory objects than to its TTC because, owing to the inverse square law, $I(r) \propto 1/r^2$, the final SPL is negatively correlated with final distance. It would be interesting to further explore these relations by varying the SPL of the sound source in future experiments. In the audiovisual condition, the TTC estimates were virtually uninfluenced by auditory TTC, but depended on the final SPL, showing evidence for audiovisual integration. The strong use of the final size/SPL cues also explains the effects of distance on mean TTC estimates. On a more general level, our results indicate the importance of considering cues other than TTC or τ when studying audiovisual integration, consistent with conclusions of many studies of visual TTC estimation (e.g., DeLucia, 2015).

In terms of the deviation of the mean TTC estimates from the veridical values, the addition of auditory information caused only a very weak increase in accuracy compared with only visual information (see Fig. 2), even though judgments of auditory objects were more accurate than judgments of visual objects when each was presented alone. This pattern corroborates the conclusion from the regression analyses that TTC estimates are dominated by visual information. We analyzed the precision of the TTC estimates in terms of the coefficient of variation that accounts for potential effects of the observed biases in mean TTC estimates on the variable error. There was no evidence for a multimodal advantage in the sense of less variable estimates in the audiovisual condition compared to the unimodal conditions. Generally, higher variability in an audiovisual condition was reported by Prime and Harris (2010) who also presented auditory and visual motion stimuli that were slightly shifted relative to each other. These authors did not measure TTC estimation, however, but asked their participants to point to the spatial position of the object at a signaled point in time. That is, they predicted the position of a moving object after it disappeared, rather than when it would reach a certain location.

In our experiment, we restricted the difference between the auditory and visual TTC to 0.5 s, so that it could be assumed that the auditory and visual stimuli were perceived as one, unitary object. In fact, on a post-experiment questionnaire 18 of the 22 participants reported that the object's auditory and visual information seemed consistent with each other. In other words, they did not consciously perceive the slight TTC mismatch between modalities, although the data show that the mismatch had a systematic effect on the TTC estimates. It remains for future research to show how the auditory and visual cues are weighted if the auditory and the visual stimulus are perceived as two separate objects (for example due to a large TTC difference), or how TTC estimates change when, for example the auditory stimulus is 'neutral' because it contains no information concerning TTC.

Several limitations of our study should be recognized. One limitation was that the auditory objects changed only in intensity as they approached, and the visual objects changed only in optical size. This simplified the simulation of the approaching objects and allowed direct comparisons of effects of auditory and visual tau, but resulted in limited external validity. Nevertheless, on our post-experiment questionnaire, 21 of the 22 s reported that the objects looked like they moved in depth toward them, and 19 of the 22 s reported that the objects sounded like they moved in depth toward them. Moreover, the absence of a multimodal advantage obtained with our simple displays is compatible with results from studies using simulations with more realism (Schiff and Oldak, 1990; Zhou *et al.*, 2007) and even real traffic environments (Hassan, 2012), and with studies of the integration of visual and vestibular information in judgments of heading (Butler *et al.*, 2010; De Winkel *et al.*, 2010). A related limitation was that we used a limited range of discrepancies between auditory and visual TTC information (0.5 s) so that participants perceived a single approaching object rather than two separate approaching objects. Additional studies are needed to determine the generalizability of our results to discrepancies that are larger but which continue to result in the perception of a single approaching object.

In conclusion, results indicate that both auditory and visual cues are used in TTC judgments, but they are not weighted equally. People use auditory TTC information but rely primarily on visual TTC information when both are available. Moreover, a reliance on τ compared to heuristic cues differs between the two modalities. The implication is that different sources of information are weighted differently in auditory and visual TTC judgments. Finally, the absence of an audiovisual advantage suggests that providing auditory information to supplement visual information may not necessarily result in improved performance. This has practical relevance when designing multimodal warning systems, for example, when alerting a driver of a danger (e.g., Ho *et al.*, 2013), and when designing assistive technologies for the visually-impaired.

Acknowledgements

We are grateful to Dewayne Paschall and Jeremy Donai for assistance with calibrations of auditory stimuli and for helpful discussions during auditory scene development.

Notes

1. In Zhou *et al.* (2007), the average JNDs for TTC reported in their Table 3 differ from the arithmetic means of the individual values (excluding participant 4 as described by the authors) reported in their Table 1. We based our statements on Table 1 in Zhou *et al.* (2007).

2. We initially developed the stimuli using a broadband signal. However, when we piloted the study, it was difficult to perceive the sound as an approaching object. The perception of approach motion was more compelling with the tone, compatible with results by Neuhoff (1998).

3. The percentage error we obtained ranged from 96.0% to 174.0% for visual-only displays (33.0% to 124.0% for auditory-only displays). This is substantially greater than that reported for monocular displays in Gray and Regan (1998), which ranged from about 2.0% to 12.0% overall. We attribute this difference to the methods and displays. For example, we presented an untextured approaching square on a plain white background whereas Gray and Regan presented an untextured approaching spot against a background of 200 large black dots. Their participants judged whether the object would reach them before or after the onset of a click, whereas our participants pressed a button exactly when TTC occurred. The accu-

racy in our study more closely resembles a study by Geri *et al.* (2010), who presented an untextured approaching spot surrounded by a plain dark 'sky' and textured ground plane. They reported mean percentage errors between 50.0% and 170.0% for monocular displays when TTC was between 0.75 s and 3.0 s. Similarly, with simple approaching outline circles, Heuer (1993) reported a mean percentage error of 50.0% for monocular displays when actual TTC was 2.0 s. Geri *et al.* noted the importance of context, task and other methodological factors in whether participants overestimate or underestimate TTC. We replicated overestimation errors in two subsequent studies using the same scenes, which indicates a reliable pattern.

References

Ahumada, A. J. and Lovell, J. (1971). Stimulus features in signal detection, *J. Acoust. Soc. Am.* **49**, 1751–1756.

Ashmead, D. H., Guth, D., Wa, R. S., Long, R. G. and Ponchillia, P. E. (2005). Street crossing by sighted and blind pedestrians at a modern roundabout, *J. Transp. Eng.* **131**, 812–821.

Ashmead, D. H., Grantham, D. W., Maloff, E. S., Hornsby, B., Nakamura, T., Davis, T. J., Pampel, F. and Rushing, E. G. (2012). Auditory perception of motor vehicle travel paths, *Hum. Factors* **54**, 437–453.

Azen, R. and Budescu, D. V. (2003). The dominance analysis approach for comparing predictors in multiple regression, *Psychol. Meth.* **8**, 129–148.

Baurès, R., Oberfeld, D. and Hecht, H. (2011). Temporal-range estimation of multiple objects: evidence for an early bottleneck, *Acta Psychol.* **137**, 76–82.

Baurès, R., Oberfeld, D., Tournier, I., Hecht, H. and Cavallo, V. (2014). Arrival-time judgments on multiple-lane streets: the failure to ignore irrelevant traffic, *Accid. Anal. Prev.* **65**, 72–84.

Beard, B. L. and Ahumada, A. J. (1998). A technique to extract relevant image features for visual tasks, in: *Human Vision and Electronic Imaging III*, *SPIE Proc.*, Vol. 3299, B. E. Rogowitz and T. N. Pappas (Eds), pp. 79–85.

Begault, D. R. and Wenzel, E. M. (2001). Direct comparison of the impact of head tracking, reverberation, and individualized head-related transfer functions on the spatial perception of a virtual speech source, *J. Audio Eng. Soc.* **49**, 904–916.

Berg, B. G. (1989). Analysis of weights in multiple observation tasks, *J. Acoust. Soc. Am.* **86**, 1743–1746.

Bertelson, P. and Aschersleben, G. (1998). Automatic visual bias of perceived auditory location, *Psychonom. Bull. Rev.* **5**, 482–489.

Blauert, J. (1996). *Spatial Hearing. The Psychophysics of Human Sound Localization*, revised edn. MIT Press, Cambridge, MA, USA.

Brendel, E., DeLucia, P. R., Hecht, H., Stacy, R. L. and Larsen, J. T. (2012). Threatening pictures induce shortened time-to-contact estimates, *Atten. Percept. Psychophys.* **74**, 979–987.

Budescu, D. V. (1993). Dominance analysis: a new approach to the problem of relative importance of predictors in multiple regression, *Psychol. Bull.* **114**, 542–551.

Butler, J. S., Smith, S. T., Campos, J. L. and Bülthoff, H. H. (2010). Bayesian integration of visual and vestibular signals for heading, *J. Vis.* **10**, 23. DOI:10.1167/10.11.23.

Calabro, F. J., Beardsley, S. A. and Vaina, L. M. (2011). Different motion cues are used to estimate time-to-arrival for frontoparallel and looming trajectories, *Vis. Res.* **51**, 2378–2385.

Cappe, C., Thut, G., Romei, V. and Murraya, M. M. (2009). Selective integration of auditory–visual looming cues by humans, *Neuropsychologia* **47**, 1045–1052.

Conrad, V., Kleiner, M., Bartels, A., O'Brien, J. H., Bülthoff, H. H. and Noppeney, U. (2013). Naturalistic stimulus structure determines the integration of audiovisual looming signals in binocular rivalry, *Plos One* **8**, e70710. DOI:10.1371/journal.pone.0070710.

De Winkel, K. N., Weesie, J., Werkhoven, P. J. and Groen, E. L. (2010). Integration of visual and inertial cues in perceived heading of self-motion, *J. Vis.* **10**, 1. DOI:10.1167/10.12.1.

DeLucia, P. R. (1991). Pictorial and motion-based information for depth perception, *J. Exp. Psychol. Hum. Percept. Perform.* **17**, 738–748.

DeLucia, P. R. (2004). Multiple sources of information influence time-to-contact judgments: do heuristics accommodate limits in sensory and cognitive processes?, in: *Time-to-Contact*, H. Hecht and G. J. P. Savelsbergh (Eds), pp. 243–286. Elsevier Science Publishers, Amsterdam, The Netherlands.

DeLucia, P. R. (2013). Effects of size on collision perception and implications for perceptual theory and transportation safety, *Curr. Dir. Psychol. Sci.* **22**, 199–204.

DeLucia, P. R. (2015). Perception of collision, in: *The Cambridge Handbook of Applied Perception Research*, R. R. Hoffman, P. A. Hancock, M. Scerbo, R. Parasuraman and J. L. Szalma (Eds), pp. 568–591. Cambridge University Press, Cambridge, UK.

DeLucia, P. R. and Liddell, G. W. (1998). Cognitive motion extrapolation and cognitive clocking in prediction motion tasks, *J. Exp. Psychol. Hum. Percept. Perform.* **24**, 901–914.

DeLucia, P. R., Kaiser, M. K., Bush, J. M., Meyer, L. E. and Sweet, B. T. (2003). Information integration in judgements of time to contact, *Q. J. Exp. Psychol. A* **56**, 1165–1189.

Emerson, R. W., Naghshineh, K., Hapeman, J. and Wiener, W. (2011). A pilot study of pedestrians with visual impairments detecting traffic gaps and surges containing hybrid vehicles, *Transp. Res. Part F Traffic Psychol. Behav.* **14**, 117–127.

Fechner, G. T. (1860). *Elemente der Psychophysik*. Breitkopf und Härtel, Leipzig, Germany.

Florentine, M. (1986). Level discrimination of tones as a function of duration, *J. Acoust. Soc. Am.* **79**, 792–798.

Freiberg, K., Tually, K. and Crassini, B. (2001). Use of an auditory looming task to test infants' sensitivity to sound pressure level as an auditory distance cue, *Br. J. Dev. Psychol.* **19**, 1–10.

Gauß, C. F. (1821). Theoria combinationis observationum erroribus minimis obnoxiae, *Commentationes Societatis Regiae Scientiarum Gottingensis recentiores* **5**, 33–90.

Geri, G. A., Gray, R. and Grutzmacher, R. (2010). Simulating time-to-contact when both target and observer are in motion, *Displays* **31**, 59–66.

Gonzalez, E. G., Allison, R. S., Ono, H. and Vinnikov, M. (2010). Cue conflict between disparity change and looming in the perception of motion in depth, *Vis. Res.* **50**, 136–143.

Gordon, M. S. and Rosenblum, L. D. (2005). Effects of intrastimulus modality change on audiovisual time-to-arrival judgments, *Percept. Psychophys.* **67**, 580–594.

Gordon, M. S., Russo, F. A. and MacDonald, E. (2013). Spectral information for detection of acoustic time to arrival, *Atten. Percept. Psychophys.* **75**, 738–750.

Gray, R. and Regan, D. (1998). Accuracy of estimating time to collision using binocular and monocular information, *Vis. Res.* **38**, 499–512.

Gray, R. and Regan, D. (1999). Do monocular time-to-collision estimates necessarily involve perceived distance? *Perception* **28**, 1257–1264.

Grondin, S. (2010). Timing and time perception: a review of recent behavioral and neuroscience findings and theoretical directions, *Atten. Percept. Psychophys.* **72**, 561–582.

Guski, R. (1992). Acoustic tau: an easy analogue to visual tau? *Ecol. Psychol.* **4**, 189–197.

Hartmann, W. M. (2005). *Signals, Sound, and Sensation*, 5th edn. Springer, New York, NY, USA.

Hassan, S. E. (2012). Are normally sighted, visually impaired, and blind pedestrians accurate and reliable at making street crossing decisions? *Invest. Ophthalmol. Vis. Sci.* **53**, 2593–2600.

Hassan, S. E. and Massof, R. W. (2012). Measurements of street-crossing decision-making in pedestrians with low vision, *Accid. Anal. Prev.* **49**, 410–418.

Hecht, H. and Savelsbergh, G. J. P. (Eds) (2004). *Time-to-Contact*. Elsevier Science Publishers, Amsterdam, The Netherlands.

Hellmann, A. (1996). Auditory perception of approaching sound sources, and time-of-arrival judgements. *PhD Thesis*, Ruhr-Universität Bochum, Bochum, Germany.

Heuer, H. (1993). Estimates of time to contact based on changing size and changing target vergence, *Perception* **22**, 549–563.

Hills, B. L. (1975). Some studies of movement perception, age and accidents, Report No. SR137, Department of the Environment, Transport and Road Research Laboratory, Crowthorne, UK.

Hládek, L., Le Dantec, C. C., Kopčo, N. and Seitz, A. (2013). Ventriloquism effect and aftereffect in the distance dimension, *Proc. Meet. Acoust.* **19**, 050042. DOI:10.1121/1.4799881.

Ho, C., Gray, R. and Spence, C. (2013). Role of audiovisual synchrony in driving head orienting responses, *Exp. Brain Res.* **227**, 467–476.

Hoyle, F. (1957). *The Black Cloud*. Heinemann, London, UK.

Jenison, R. L. (1997). On acoustic information for motion, *Ecol. Psychol.* **9**, 131–151.

Kaczmarek, T. (2005). Auditory perception of sound source velocity, *J. Acoust. Soc. Am.* **117**, 3149–3156.

Kaczmarek, T. and Niewiarowicz, M. (2013). Auditory motion perception in normal hearing and in hearing impaired people, *Acta Acust. United Acust.* **99**, 283–291.

Khuu, S. K., Lee, T. C. P. and Hayes, A. (2010). Object speed derived from the integration of motion in the image plane and motion-in-depth signaled by stereomotion and looming, *Vis. Res.* **50**, 904–913.

LeBreton, J. M., Ployhart, R. E. and Ladd, R. T. (2004). A Monte Carlo comparison of relative importance methodologies, *Organ. Res. Methods* **7**, 258–282.

Lee, D. N. (1976). Theory of visual control of braking based on information about time-to-collision, *Perception* **5**, 437–459.

Lutfi, R. A. and Wang, W. (1999). Correlational analysis of acoustic cues for the discrimination of auditory motion, *J. Acoust. Soc. Am.* **106**, 919–928.

Maier, J. X., Neuhoff, J. G., Logothetis, N. K. and Ghazanfar, A. A. (2004). Multisensory integration of looming signals by Rhesus monkeys, *Neuron* **43**, 177–181.

Murray, R. F. (2011). Classification images: a review, *J. Vis.* **11**, 1–25.

Neuhoff, J. G. (1998). Perceptual bias for rising tones, *Nature* **395**(6698), 123–124.

Oberfeld, D. and Hecht, H. (2008). Effects of a moving distractor object on time-to-contact judgments, *J. Exp. Psychol. Hum. Percept. Perform.* **34**, 605–623.

Oberfeld, D., Hecht, H. and Landwehr, K. (2011). Effects of task-irrelevant texture motion on time-to-contact judgments, *Atten. Percept. Psychophys.* **73**, 581–596.

Oberfeld, D., Klöckner-Nowotny, F., Reinhard, R. and DeLucia, P. R. (2014). Auditory detection of gradual changes in intensity, in: *Abstracts of the 56th Conference of Experimental Psychologists*, A. C. Schütz, K. Drewing and K. R. Gegenfurtner (Eds), p. 192. Pabst Science Publishers, Gießen, Germany.

Oruc, I., Maloney, L. T. and Landy, M. S. (2003). Weighted linear cue combination with possibly correlated error, *Vis. Res.* **43**, 2451–2468.

Porschmann, C. and Storig, C. (2009). Investigations into the velocity and distance perception of moving sound sources, *Acta Acust. United Acust.* **95**, 696–706.

Prime, S. L. and Harris, L. R. (2010). Predicting the position of moving audiovisual stimuli, *Exp. Brain Res.* **203**, 249–260.

Rosenblum, L. D., Carello, C. and Pastore, R. E. (1987). Relative effectiveness of three stimulus variables for locating a moving sound source, *Perception* **16**, 175–186.

Rosenblum, L. D., Wuestefeld, A. P. and Saldana, H. M. (1993). Auditory looming perception: influences on anticipatory judgments, *Perception* **22**, 1467–1482.

Rushton, S. K. (2004). Interception of projectiles, from when and where to where once, in: *Time-to-Contact*, H. Hecht and G. J. P. Savelsbergh (Eds), pp. 327–353. Elsevier Science Publishers, Amsterdam, The Netherlands.

Rushton, S. K. and Wann, J. P. (1999). Weighted combination of size and disparity: a computational model for timing a ball catch, *Nat. Neurosci.* **2**, 186–190.

Schiff, W. and Detwiler, M. L. (1979). Information used in judging impending collision, *Perception* **8**, 647–658.

Schiff, W. and Oldak, R. (1990). Accuracy of judging time to arrival: effects of modality, trajectory, and gender, *J. Exp. Psychol. Hum. Percept. Perform.* **16**, 303–316.

Schmiedchen, K., Freigang, C., Rübsamen, R. and Richter, N. (2013). A comparison of visual and auditory representational momentum in spatial tasks, *Atten. Percept. Psychophys.* **75**, 1507–1519.

Shaw, B. K., McGowan, R. S. and Turvey, M. T. (1991). An acoustic variable specifying time-to-contact, *Ecol. Psychol.* **3**, 253–261.

Tonidandel, S. and LeBreton, J. M. (2011). Relative importance analysis: a useful supplement to regression analysis, *J. Bus. Psychol.* **26**, 1–9.

Treisman, M. (1963). Temporal discrimination and the indifference interval: implications for a model of the 'internal clock', *Psychol. Monogr.* **77**, 1–31.

Tresilian, J. R. (1994). Approximate information sources and perceptual variables in interceptive timing, *J. Exp. Psychol. Hum. Percept. Perform.* **20**, 154–173.

Tresilian, J. R. (1995). Perceptual and cognitive processes in time-to-contact estimation: analysis of prediction-motion and relative judgment tasks, *Percept. Psychophys.* **57**, 231–245.

von Hofsten, C. and Lee, D. N. (1985). Dialogue on perception and action, in: *Persistence and Change: Proceedings of the First International Conference on Event Perception*, W. H. Warren Jr and R. E. Shaw (Eds), pp. 231–242. Erlbaum, Hillsdale, NJ, USA.

Wearden, J. H. and Lejeune, H. (2008). Scalar properties in human timing: conformity and violations, *Q. J. Exp. Psychol.* **61**, 569–587.

Wuerger, S. M., Hofbauer, M. and Meyer, G. F. (2003). The integration of auditory and visual motion signals at threshold, *Percept. Psychophys.* **65**, 1188–1196.

Yan, J. J., Lorv, B., Li, H. and Sun, H. J. (2011). Visual processing of the impending collision of a looming object: time to collision revisited, *J. Vis.* **11**, 7. DOI:10.1167/11.12.7.

Zahorik, P. (2002). Assessing auditory distance perception using virtual acoustics, *J. Acoust. Soc. Am.* **111**, 1832–1846.

Zakarauskas, P. and Cynader, M. S. (1991). Aural intensity for a moving source, *Hear. Res.* **52**, 233–244.

Zhou, L., Yan, J., Liu, Q., Li, H., Xie, C., Wang, Y., Campos, J. L. and Sun, H.-J. (2007). Visual and auditory information specifying an impending collision of an approaching object, in: *Human–Computer Interaction, Pt 2, HCII 2007*, *LNCS*, Vol. 4551, J. Jacko (Ed.), pp. 720–729. Springer-Verlag, Berlin, Germany.