Cognition 227 (2022) 105204

Contents lists available at ScienceDirect

Cognition

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

No selective integration required: A race model explains responses to audiovisual motion-in-depth

S.F. Andrew Chua *, Yue Liu, Julie M. Harris, Thomas U. Otto *

School of Psychology & Neuroscience, University of St Andrews, St Mary's Quad, South Street, St Andrews KY16 9JP, United Kingdom

ARTICLE INFO

Keywords:
Multisensory integration
Audio-visual motion
Looming bias
Perceptual decision making
Race model
Logic OR gate
Probability summation
Model selection

ABSTRACT

Looming motion is an ecologically salient signal that often signifies danger. In both audition and vision, humans show behavioral biases in response to perceiving looming motion, which is suggested to indicate an adaptation for survival. However, it is an open question whether such biases occur also in the combined processing of multisensory signals. Towards this aim, Cappe, Thut, Romei, and Murraya (2009) found that responses to audiovisual signals were faster for congruent looming motion compared to receding motion or incongruent combinations. They considered this as evidence for selective integration of multisensory looming signals. To test this proposal, here, we successfully replicate the behavioral results by Cappe et al. (2009). We then show that the redundant signals effect (RSE - a speedup of multisensory compared to unisensory responses) is not distinct for congruent looming motion. Instead, as predicted by a simple probability summation rule, the RSE is primarily modulated by the looming bias in audition, which suggests that multisensory processing inherits a unisensory effect. Finally, we compare a large set of so-called race models that implement probability summation, but that allow for interference between auditory and visual processing. The best-fitting model, selected by the Akaike Information Criterion (AIC), virtually perfectly explained the RSE across conditions with interference parameters that were either constant or varied only with auditory motion. In the absence of effects jointly caused by auditory and visual motion, we conclude that selective integration is not required to explain the behavioral benefits that occur with audiovisual looming motion.

1. Introduction

Looming motion is motion towards oneself. It is potentially dangerous as it may indicate an imminent attack or collision. It has been proposed that there is a looming bias, thought to speed up an appropriate evasive response (e.g., Neuhoff, 1998, 2001). In contrast, there should be no bias towards perceiving receding motion, because motion away from oneself suggests disengagement or a de-escalation of danger and does not require an urgent response (Tyll et al., 2013).

Taking such real-life ideas to the lab, a wide range of studies have investigated responses to looming motion using unisensory signals. In auditory experiments, looming is typically represented by tones of rising intensity (Bach, Neuhoff, Perrig, & Seifritz, 2009; Rosenblum, Carello, & Pastore, 1987; for additional cues, see Baumgartner et al., 2017), which can increase neural activity linked to sensory processing, attention, and arousal (Bach et al., 2008; Maier & Ghazanfar, 2007; Seifritz et al., 2002). Tones are perceived to change more when their intensity increases, compared with a symmetrical intensity decrease (Neuhoff, 1998, 2001). Looming tones are also judged to emanate from sources closer to oneself than they are in reality (Neuhoff, 2001; Rosenblum, Wuestefeld, & Saldana, 1993), and are related to faster motor responses when judged as nearer (Camponogara, Komeilipoor, & Cesari, 2015a). Similarly, in experiments using visual stimuli, expanding shapes are simple cues to looming motion, which can capture attention and attract hand movements (Franconeri & Simons, 2003; Moher, Sit, & Song, 2015). In both modalities, human infants present avoidance behaviors to looming signals, but not to the receding counterparts (Ball & Tronick, 1971; Freiberg, Tually, & Crassini, 2001), suggesting that biases to looming motion are innate. In summary, there is ample evidence demonstrating biases towards looming motion with unisensory signals.

The study of looming biases reaches another level when signals from vision and audition are considered together. The central question in multisensory research is how signals from two (or more) modalities are combined into a coherent percept to guide action (Alais, Newell, & Mamassian, 2010; Burr & Alais, 2006; Driver & Noesselt, 2008; Ernst & Bülthoff, 2004). When signals across modalities originate from the same

* Corresponding authors.
E-mail addresses: sfac@st-andrews.ac.uk (S.F.A. Chua), to7@st-andrews.ac.uk (T.U. Otto).

https://doi.org/10.1016/j.cognition.2022.105204
Received 10 November 2021; Received in revised form 2 June 2022; Accepted 8 June 2022
Available online 23 June 2022

external event or object of interest, they can be redundant because either signal is sufficient to detect the event or to describe the object. Combination of redundant signals can aid perception and is typically beneficial for behavior. Given the presence of looming biases in both audition and vision, the question arises if, and in what way, such biases also occur in the combined processing of redundant audiovisual looming signals. Do multisensory responses inherit looming biases present in the unisensory components, or is there selective integration of audio-visual looming signals, which would manifest as a looming bias in multisensory processing?

This question was addressed by Cappe et al. (2009); Cappe, Thelen, Romei, Thut, and Murray (2012), who tested behavioral responses to audiovisual motion-in-depth using the classic redundant signals paradigm. The basic task asks participants to respond to auditory, visual, and combined audiovisual signals with a single button press (Fig. 1a). In either modality, the signal could be looming or receding (via changes in auditory intensity or visual size, Fig. 1b, c). By combining different auditory and visual motion directions, Cappe et al. (2009) tested various multisensory conditions, with congruent audiovisual looming motion being the key condition of interest (Fig. 1d). As a main result, Cappe et al. (2009) found that average response times (RTs)[1] to audiovisual signals were faster compared to unisensory signals, which is known as the redundant signals effect (RSE; e.g., Giard & Peronnet, 1999; Hershenson, 1962; Kinchla, 1974; Miller, 1982; Todd, 1912). Critically, among the different audiovisual conditions, the central finding was that average RTs were faster for congruent looming signals compared to the other combinations. Such signals were also subjectively rated as moving the most. In addition, a later published analysis found super-additive effects (where the multisensory response is larger than the summed unisensory responses) in event related potentials for congruent audiovisual looming signals (Cappe et al., 2012). Altogether, Cappe et al. (2009, 2012) used these findings to argue for a looming bias in multisensory processing mediated by selective integration of audiovisual looming (and not receding or incongruent) signals.

We would like to argue that the patterns of behavioral data observed by Cappe et al. (2009) do not necessarily evidence a selective integration mechanism. First, given unisensory looming biases in both audition and vision, unisensory RTs are likely to be faster for looming compared to receding signals. If multisensory responses inherit from their unisensory components, generally speaking, it is expected that the combination of unisensory signals with relatively fast RTs should produce faster multisensory RTs as compared to a combination of signals with relatively slow RTs (e.g., Otto, Dassy, & Mamassian, 2013). Hence, finding on average fastest RTs in audiovisual looming conditions among the other multisensory conditions does not necessarily show that the combination of congruent looming signals is preferentially processed. Second, to support the selective integration account, it would be important to translate the verbal term “selective integration” into a computational model that connects uni- and multisensory processing by an explicit combination rule. Strong evidence for selective integration would emerge if a different combination rule or model parameterization is needed to explain the data with congruent looming signals compared to the other motion combinations. As Cappe et al. (2009) focused on average RTs and without a testable “selective integration” model, we conclude that the question of how looming biases manifest in the multisensory processing of audiovisual looming signals remains open.

1.1. Race models as a theoretical framework for multisensory RTs

Historically, so-called race models provide an elaborated theoretical framework to understand multisensory processing, and in particular the advantage of having more than one sensory input (Raab, 1962). In experiments using the redundant signals paradigm, participants are asked to respond with the same motor act to uni- and multisensory signals. Hence, multisensory signals are redundant here in that detecting one of the two is sufficient for a correct response. However, how is this redundancy converted into a behavioral benefit? In race models, two perceptual decision units are arranged in parallel, one for audition and one for vision, which are then coupled by a logic OR gate. This architecture leads to better performance (a redundancy gain) when both signals are present because a multisensory response can be triggered by the faster of the two decision units (Raab, 1962). The mechanism is like playing dice and aiming for a small number (corresponding to a faster RT). When a player is allowed to roll two dice and then picks the one with the smaller number, the probability of obtaining a small number is increased compared to when rolling only one die. This can be precisely quantified using probability summation as a computational combination rule.

While the race model approach is compelling, substantial research on the RSE has shown that care must be taken when using probability summation to predict redundancy gains with multisensory signals. The reason is that the framework requires two key assumptions, which are not necessarily correct. First, the probability summation rule in its most basic form assumes that RTs measured across trials are *statistically independent* (e.g., Ashby & Townsend, 1986; Luce, 1986; Miller, 1982; Otto & Mamassian, 2017). This includes the assumption that a RT on a given trial is not affected by stimulation and/or performance on a preceding trial. However, this assumption can be violated in experiments due to modality switch costs (e.g., Gondan, Lange, Rosler, & Roder, 2004; Innes & Otto, 2019; Miller, 1982; Otto & Mamassian, 2012; Shaw et al., 2020). For example, the response to an auditory signal can be faster when the auditory modality is repeated from a previous trial compared to when it is switched from vision. Critically, such dependencies can hugely affect predictions based on probability summation. Considering this, Miller (1982) derived an upper bound for redundancy gains that are in line with probability summation (in technical terms, this so-called Miller's bound assumes a maximal negative correlation instead of statistical independence; Colonius, 1990). As a milestone finding, multisensory RTs typically violate Miller's bound, which has led to a broad rejection of race models as an explanation of the RSE (see the many replications of Miller, 1982).

At this point, it is however crucial to note a second and often neglected assumption called *context invariance* (e.g., Ashby & Townsend, 1986; Liu & Otto, 2020; Luce, 1986; Otto & Mamassian, 2017; Townsend, Liu, Zhang, & Wenger, 2020; Townsend & Wenger, 2004; Yang, Altieri, & Little, 2018). This assumption states that unisensory processing in one modality does not interfere with processing in the other modality (e.g., the auditory decision unit is not affected by the presence or absence of a visual target signal). This assumption is implicitly made when unisensory conditions are used to derive predictions for multisensory conditions. Violations of Miller's bound are often used to rule out the race model as a framework for understanding multisensory processing. However, such violations can be understood as showing that the context invariance assumption is incorrect (for a systematic analysis of this argument, see Otto & Mamassian, 2017).

Considering these issues, Otto and Mamassian (2012) proposed a modified race model, which builds on the basic architecture with two unisensory decision units that are coupled by a logic OR gate (Fig. 2a). Each decision unit is implemented as an instance of the Linear Approach to Threshold with Ergodic Rate (LATER) model, which belongs to the larger class of models assuming that sensory evidence is accumulated over time until a threshold is reached and a categorical decision to trigger a response is made (e.g., Carpenter & Williams, 1995b; Gold & Shadlen, 2007; Noorani & Carpenter, 2016). To describe unisensory RTs, each LATER unit uses two parameters, an accumulation rate μ and its variability σ. If an experimental manipulation increases for example

[1] Non-standard abbreviations: Akaike information criterion (AIC), auditory looming/receding (AL, AR), Bayesian information criterion (BIC), linear approach to threshold with ergodic rate (LATER) model, redundant signals effect (RSE), response time (RT), and visual looming/receding (VL, VR).

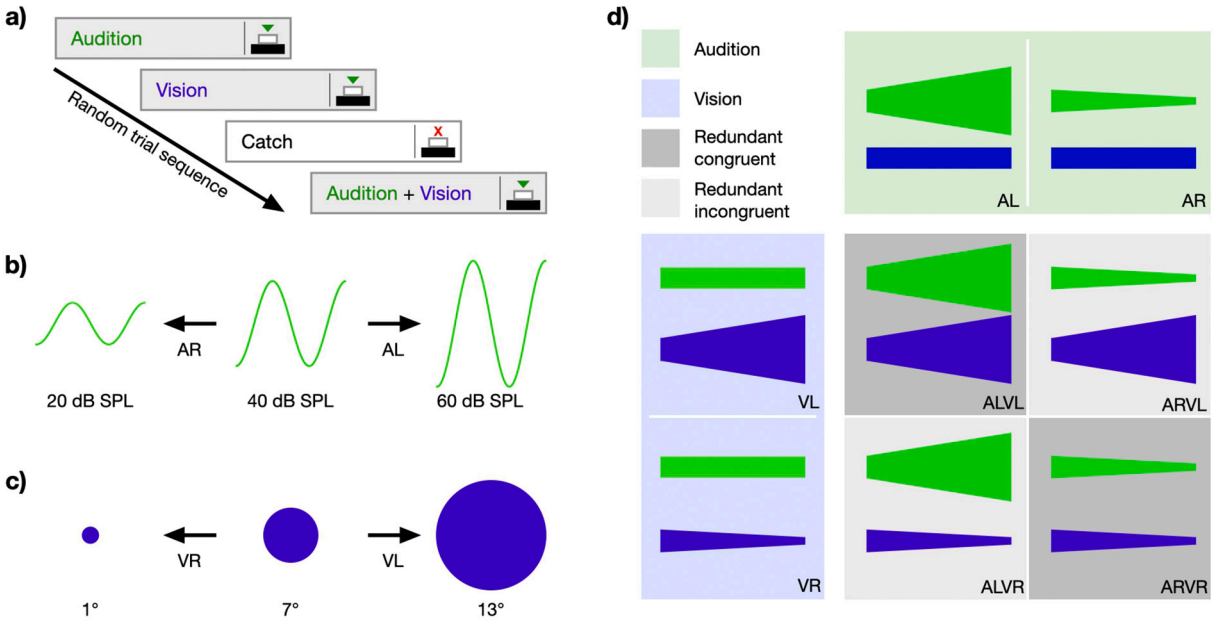

Fig. 1. Experimental design replicating Cappe et al. (2009). a) Redundant signals paradigm. Participants responded with the same button press on presentation of auditory, visual, and redundant audio-visual motion signals. No response was required on catch trials with static signals. b) Auditory motion stimulus. 1000 Hz pure tones were presented with a starting intensity of 40 dB SPL. For receding motion (AR), the intensity decreased to 20 dB SPL. For looming motion (AL), the intensity increased to 60 dB SPL. c) Visual motion stimulus. A filled circle was presented with a starting diameter of 7°. For receding motion (VR), the circle reduced in diameter to 1°. For looming motion (VL), the circle increased in diameter to 13°. d) Stimulus conditions. Audio-only conditions are shown in green boxes, visual-only in blue. For redundant conditions shown in gray boxes, auditory and visual motion signals were combined using a 2 × 2 design. Combinations were either congruent (ALVL, ARVR; mid-gray) or incongruent (ALVR, ARVL; light gray). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

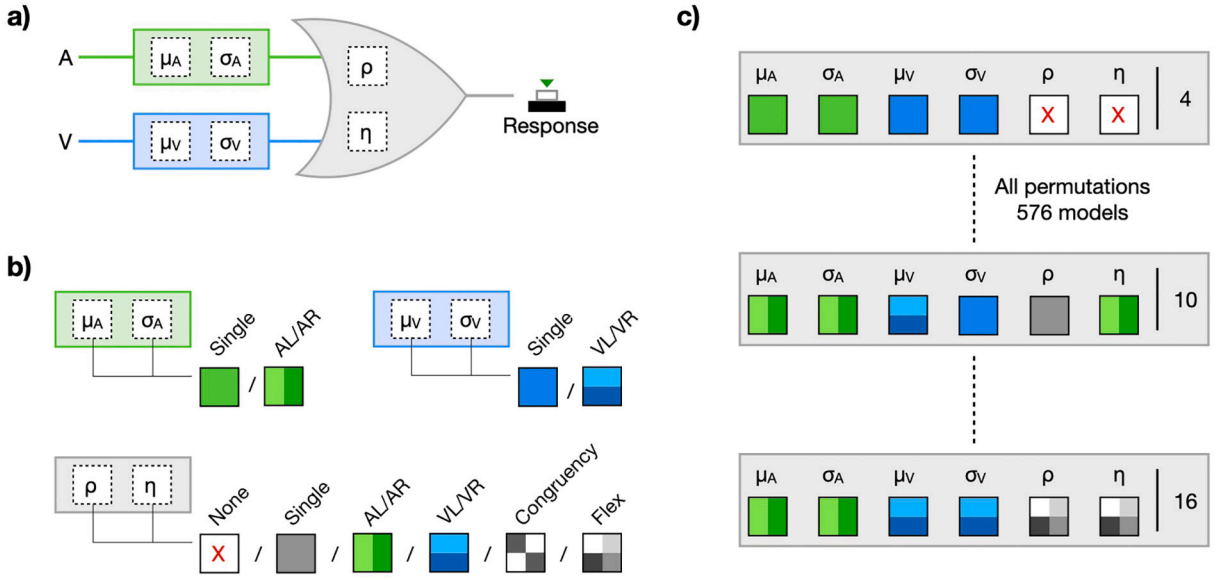

Fig. 2. Model design. a) To analyze the redundant signals effect (RSE), Otto and Mamassian (2012) used a race model consisting of two parallel unisensory decision units (green and blue boxes) coupled by a logical OR gate (gray arrow shape) to trigger a response. In its basic version, the model has 6 free parameters to fit the RT distributions in the 3 basic conditions (audition, vision, and redundant; Fig. 1a). b) To extend the basic model to the 8 stimulus conditions (Fig. 1d), parameters of the unisensory decision units (μ, σ) could take a single value irrespective of the motion direction or 2 different values to account for looming and receding motion, respectively. The interference parameters (ρ, η) could follow one of six options: not used (None), 1 value (Single), 2 values varying with auditory motion (AL/AR), 2 values varying with visual motion (VL/VR), 2 values varying with motion congruency (Congruency), or 4 values varying fully flexibly with the 4 audio-visual combinations (Flex). c) To generate a comprehensive set of candidate models, we tested all 576 permutations of these parameter options. Here we show three candidate models ranging from the simplest (4 free parameters) to the most complex (16 free parameters). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

the signal strength, the model assumes a faster accumulation rate which leads to shorter RTs. On presentation of redundant audio-visual signals, the OR gate selects the faster of the two unisensory units to trigger a response. To describe the multisensory RTs, the model includes two further parameters, each considering one of the assumptions outlined above. First, the model does not assume statistical independence, instead it uses a free correlation parameter ρ, which accounts for modality switch costs. Second, the model does not assume context invariance, instead it uses an additive noise parameter η, which can be understood as an unspecific interference that increases the variability in the unisensory decision units when both stimuli are processed simultaneously. This latter parameter allows the model to violate Miller's bound. With these two parameters, the model can account for the RSE virtually perfectly (Otto & Mamassian, 2012). This computational modeling approach thus provides an elaboration on early, more limited, race models. Importantly, it provides an explicitly formulated combination rule that will be used here to test for selective integration of looming signals in multisensory processing.

1.2. Using race models to test for a multisensory looming bias

To test for a multisensory looming bias, in this study, we replicated the behavioral experiment by Cappe et al. (2009). As with the original study, we used the redundant signals paradigm (Fig. 1a) and simulated looming or receding motion via auditory intensity (Fig. 1b) and visual size changes (Fig. 1c). Thus, with the factors auditory motion (looming, receding) and visual motion (looming, receding), the resulting 2 × 2 design allows us to investigate potential interactions between the two factors (Fig. 1d). To anticipate the results (presented in section 3.1), we successfully replicated the basic finding by Cappe et al. (2009): average RTs were fastest for congruent looming signals compared to receding and incongruent combinations. We then subject this data to a systematic analysis and modeling approach as outlined in the following three steps.

As a first step (section 3.2), we moved beyond the analysis of average RTs and used instead an approach that measures the actual RSE (how much are multisensory RTs faster compared to the unisensory component RTs?). As a key benefit, this approach allows comparison of the empirical RSE to parameter-free predictions obtained from the basic probability summation rule (Raab, 1962; i.e., assuming both statistical independence and context invariance). When analyzing the RSE, the inclusion of such quantitative predictions is very informative as they reveal changes in RSE that are simply expected from performance changes caused, for example, by unisensory looming biases. To use a more statistical language, if the RSE is modulated by unisensory looming biases, we expect to find main effects of auditory and/or visual motion. Critically, a specific multisensory looming bias would manifest itself as an interaction effect involving both motion directions.

As a second step (section 3.3), we test the assumptions of statistical independence and context invariance, which the basic probability summation rule relies on. For the former, we test for modality switch costs, which if present imply that RTs cannot be assumed to be statistically independent. For the latter, we check for violations of Miller's bound, which if present imply that context invariance cannot safely be assumed.

As a final step (section 3.4), we engage in a large-scale model comparison to investigate how multisensory motion processing is affected by looming biases. For this, we elaborated on the model by Otto and Mamassian (2012) that can incorporate violations of the two key assumptions by adapting it to the eight stimulus conditions of the 2 × 2 design. For the auditory decision unit, we allowed both parameters (μ_A, σ_A) to use either a single value independent of the auditory motion direction, or to use two different values to account for performance differences between looming and receding motion (e.g., by allowing two free parameters μ_{AL} and μ_{AR} instead of only μ_A; Fig. 2b). For the visual decision unit (μ_V, σ_V), we allowed for the same flexibility in response to the visual motion manipulation. For each of the two interference parameters (ρ and η), we allowed for six different options to account for performance differences in the multisensory conditions. These ranged from not using the interference parameters (i.e., the model assumptions of statistical independence and context invariance would be upheld) to allowing a different value for each motion combination (Fig. 2b). By considering all permutations of parameter options (2^2 auditory * 2^2 visual * 6^2 multisensory), we obtained a comprehensive set of 576 candidate models that range between having four and 16 free parameters (Fig. 2c). We then selected the best-fitting model using the Akaike Information Criterion (AIC; Heathcote, Brown, & Mewhort, 2002; Lewandowsky & Farrell, 2011; Wagenmakers & Farrell, 2004; see section 2.7. for a full technical description of the modeling approach). The key question is whether this approach selects a candidate model with interference parameters (ρ and η) that are either constant across motion combinations or that vary only with unisensory factors. The latter would be evidence that biases must be unisensory. Alternatively, the approach may identify a candidate model with a more complex set of interference parameters, which would point to a specific looming bias in multisensory processing. We start by describing our experiment, and then move to describe how the modeling allows us to understand the nature of looming biases in multisensory RTs.

2. Methods

2.1. Participants

20 participants were recruited (mean age: 19.9 years; 19 right-handed; all with self-declared normal or corrected-to-normal hearing and eyesight). Informed consent was obtained before the experiment. Participant time was compensated with financial reward (£5/h). All procedures were in accordance with the *Code of Human Research Ethics* (The-British-Psychological-Society, 2014). The study was approved by the *University Teaching and Research Ethics Committee* (UTREC, University of St Andrews; approval code PS12994).

2.2. Apparatus

To control and run the experiment, we used a HP Z240 desktop workstation running Windows 10 (Microsoft Corporation), Matlab (The MathWorks, Inc.), and the Psychophysics Toolbox (Brainard, 1997; Kleiner, Brainard, & Pelli, 2007; Pelli, 1997). Visual stimuli were presented on a 27-in. LCD display (Dell U2713HM; 1920 × 1080 pixel resolution; 60 Hz refresh rate). A custom-made chinrest maintained a viewing distance of 570 mm. Auditory stimuli were presented at a sampling rate of 48 kHz via over-ear headphones (Sennheiser HD280 Pro). Sounds were calibrated using a sound level meter (Brüel & Kjær Type 2250) with an artificial ear (Type 4153). Participants responded with their thumb using a custom-made handheld button connected to a RTbox v5/6 (Li, Liang, Kleiner, & Lu, 2010). Prior to data collection, the RTbox was also used to calibrate audiovisual timing, such that auditory and visual signals were physically synchronous with an onset jitter below 1 ms.

2.3. Experimental design

Following procedures by Cappe et al. (2009), we adopted the classic redundant signals paradigm (Fig. 1a; e.g. Giard & Peronnet, 1999; Hershenson, 1962; Kinchla, 1974; Miller, 1982; Todd, 1912). In two unisensory conditions, we presented either auditory or visual motion-in-depth. In a third condition with redundant signals, both motion signals were presented simultaneously. On catch trials, no motion signal was presented (i.e., auditory and visual stimuli displayed no motion). The task was to respond with a button press as soon as a motion signal was detected in either modality.

We extended the classic redundant signals paradigm by using a 2 × 2 design, combining auditory motion (looming or receding) with visual

motion (looming or receding). This yielded four redundant signals conditions (redundant congruent: ALVL, ARVR; redundant incongruent: ALVR, ARVL). Thus, together with the unisensory conditions (auditory: AL, AR; visual: VL, VR), there were a total of eight motion signal conditions (Fig. 1d).

2.4. Stimuli

Auditory stimulation was based on 1000 Hz pure tones, which had an initial level of 40 dB SPL. To simulate looming motion, the intensity increased linearly over a period of 0.5 s to a final level of 60 dB SPL (Fig. 1b). To simulate receding motion, the intensity linearly decreased over a period of 0.5 s to a final level of 20 dB SPL. Cappe et al. (2009) used an initial intensity of 77 dB SPL rising up to 87 dB SPL. We used lower intensities here as such levels played via the over-ear headphones felt uncomfortable.

Visual stimulation was based on black disks presented centrally on a gray background. The initial diameter was 7°. To simulate looming motion, the disk linearly increased in size over a period of 0.5 s to a final diameter of 13° (Fig. 1c). To simulate receding motion, the disk linearly decreased in size over a period of 0.5 s to a final diameter of 1°.

2.5. Procedures

Each trial began with a foreperiod, which had a minimum duration of 0.75 s plus a random component sampled from an exponential distribution with a mean of 0.25 s. During the foreperiod, static auditory and visual stimuli were presented at the initial level (intensity/size respectively). This deviates from the design used by Cappe et al. (2009), who did not present a static foreperiod (for a replication experiment using the original procedures, see Supplementary Material). Then, one of the eight motion signals started. After moving for 0.5 s, signals remained static at the final level (intensity/size) for the rest of the trial. In single signal conditions, the signal in the other modality was static at the initial level. In catch trials, both signals were static.

When a response was recorded, or at the end of the response period of 1.5 s, all stimulation stopped, and feedback was given for 0.5 s. On motion trials, a response within the response period was considered a hit, which was indicated by showing 'Correct' in green. If no response was recorded, the motion trial was counted as a miss, which was indicated by showing 'Miss' in red. On catch trials, the green 'Correct' appeared if no response was made, whereas responding resulted in a red 'False alarm' being shown. After feedback, a blank gray screen was shown, and the next trial started after 0.5 s.

The experiment was organized into 20 blocks. A block lasted about three minutes. The entire experiment took approximately 90 min to complete. Each block consisted of 60 trials, made from a random sequence of 20 catch trials and 40 motion signal trials (five trials of each of the eight motion conditions). Hence, across all blocks, we presented 100 trials per motion condition per participant. In total, with 20 participants, we recorded a data set of 24,000 trials, of which 16,000 trials contained motion signals.

2.6. Data analysis

As a first processing step, we filtered our data for inappropriate responses. For both signal and catch trials, we removed trials where a false alarm occurred during the foreperiod from further analysis (Table 1; Foreperiod FA). Further, we submitted RTs to an outlier correction. For this, we transformed RTs into rates (1/RT) to account for the skewed nature of RT distributions (Carpenter & Williams, 1995a; Noorani & Carpenter, 2016). Then, separately for each condition and participant, we determined fast and slow bounds at ±1.4826 * 3 median absolute deviations around the median rate, which corresponds to ±3 standard deviations around the mean if rates are normally distributed (Leys, Ley, Klein, Bernard, & Licata, 2013; Otto, 2019). We removed trials with rates outside these bounds from further analysis (Table 1; Outlier). With these procedures, 15,747 valid trials remained for the planned main analysis of RTs. The research data underpinning this publication can be accessed at https://doi.org/10.17630/362541c2-ac2a-40d5-8d27-26d47a464e12 (Chua, Liu, Harris, & Otto, 2021).

The main analysis examined RTs, and specifically the RSE as a function of the different combinations of audiovisual motion. On the level of RT distributions, we used a geometric measure to quantify the RSE (for a detailed description, see Otto, 2019). The use of RT distributions is beneficial as these directly allow for parameter-free predictions of the RSE using probability summation (i.e., predictions of Raab (1962) independent race model), which can be compared to the empirical RSE. In addition, RT distributions allow us to test for violations of Miller's (1982) bound, which is a standard test in multisensory research to test for potential processing interference (Otto & Mamassian, 2017). As the geometric analysis benefits from equal numbers of RTs in all conditions, we extracted 50 RT quantiles out of ~98 RTs per participant and condition. Cumulative RT distributions were then obtained by plotting these quantiles against the corresponding cumulative probabilities. Corresponding analysis steps are described in full detail as part of the *RSE-box*, which is implemented in Matlab (Otto, 2019; relevant functions: *sampleDown*, *getCP*, *getGain*, *getRaab*, and *getViolation*).

For statistical tests, we used within-subjects ANOVAs with factors auditory (AR, AL) and visual motion (VR, VL). The interaction of these factors would point towards an audiovisual congruency effect. We used the Greenhouse-Geisser correction when the sphericity assumption was violated (Mauchly's test). The alpha-level for all statistical tests was 0.05.

2.7. Model fitting and comparison

To identify audiovisual interference effects that may point to a multisensory looming bias in the processing of audio-visual motion-in-depth, we applied an approach using a race model architecture (Otto & Mamassian, 2012; Raab, 1962; for an analogous modeling approach on

Table 1
Performance summary. Mean % performance and RTs (±SEM) across 20 participants.

Sensory condition	Trials (#)	Foreperiod FA (%)	FA (%)	Hit (%)	Outlier (%)	Valid RTs (#)	Median RT (s)
AR	100	0.00 ± 0.00	–	98.79 ± 0.31	1.35 ± 0.34	97.45 ± 0.39	0.596 ± 0.020
AL	100	0.25 ± 0.10	–	99.69 ± 0.13	1.55 ± 0.41	97.90 ± 0.42	0.395 ± 0.012
VR	100	0.25 ± 0.12	–	99.90 ± 0.07	0.85 ± 0.22	98.80 ± 0.26	0.305 ± 0.009
VL	100	0.60 ± 0.20	–	100.00 ± 0.00	1.11 ± 0.31	98.30 ± 0.40	0.293 ± 0.010
ARVR	100	0.10 ± 0.07	–	99.95 ± 0.05	0.95 ± 0.26	98.90 ± 0.26	0.301 ± 0.009
ARVL	100	0.35 ± 0.11	–	99.90 ± 0.07	1.41 ± 0.42	98.15 ± 0.48	0.294 ± 0.011
ALVR	100	0.20 ± 0.09	–	100.00 ± 0.00	0.95 ± 0.19	98.85 ± 0.22	0.284 ± 0.009
ALVL	100	0.20 ± 0.09	–	100.00 ± 0.00	0.80 ± 0.31	99.00 ± 0.33	0.274 ± 0.009
asvs	400	0.28 ± 0.05	1.35 ± 0.20	–	–	–	–

Number of trials. FA = false alarm rate.

audiovisual location, see Liu, Chua, & Otto, 2022). To implement a pair of parallel unisensory decision units, we used the LATER model (Carpenter & Williams, 1995a; Noorani & Carpenter, 2016). According to this model, the empirical RT distribution with unisensory signals can be described by a reci-normal distribution, which is the distribution of a random variable X whose reciprocal $Y = 1/X$ is normally distributed with mean rate μ and standard deviation σ. On presentation of redundant audiovisual signals, the model selects the unit to detect a signal first (i.e., here the unit with the higher rate). The exact distribution can be computed using the maximum distribution of two Gaussian random numbers (Nadarajah & Kotz, 2008).

To fit empirical distributions with redundant signals, the model allows for two interference parameters (Otto & Mamassian, 2012). First the correlation ρ is linked to trial sequence effects, which occur when RTs on a given trial depend on the signal/response of a previous trial, for example, due to modality switch costs. In the presence of such effects, RTs are not statistically independent (as assumed by the predictions using probability summation, see above). Second, the additional noise η increases the variability of rates in redundant, compared to single signal, conditions. Otto and Mamassian (2012) introduced this parameter to account for an increased variability of multisensory RTs compared to using a model with only ρ as a free parameter. Hence, to fit RT distributions in the three conditions of the RSE paradigm (auditory, visual, redundant; Fig. 1a), this basic race model has six free parameters (four unisensory, two interference parameters; Fig. 2a). The model is implemented in Matlab as part of the *RSE-box* (Otto, 2019).

To extend the basic model to the eight tested conditions (two auditory, two visual, and four redundant), we created a set of nested models that allowed the six basic parameters to vary with the experimental factors (Fig. 2b). First, LATER parameters could vary with the unisensory motion direction. For example, a nested model version could have either a single parameter μ_A or two parameters μ_{AL} and μ_{AR} to account for looming / receding motion. Second, interference parameters could vary with any experimental factor. For example, the correlation ρ could follow one of six settings: (1) ρ not used, i.e. assuming statistical independence, (2) ρ identical in all four redundant conditions, (3) ρ_{AL} and ρ_{AR} to vary with auditory motion direction, (4) ρ_{VL} and ρ_{VR} to vary with visual motion direction, (5) ρ_C and ρ_I to vary with congruent / incongruent motion directions, and (6) ρ_{ALVL}, ρ_{ARVR}, ρ_{ALVR}, ρ_{ARVL} to vary with each of the four redundant signals conditions. Then, to generate a comprehensive set of candidate models, we used all permutations of parameter settings. The simplest model was an independent race model with only four free parameters (μ_A, μ_V, σ_A, σ_V), which assumes no looming bias or interference. The most complex model had 16 free parameters (μ_{AL}, μ_{AR}, μ_{VL}, μ_{VR}, σ_{AL}, σ_{AR}, σ_{VL}, σ_{VR}, ρ_{ALVL}, ρ_{ARVR}, ρ_{ALVR}, ρ_{ARVL}, η_{ALVL}, η_{ARVR}, η_{ALVR}, η_{ARVL}). In total, this procedure created a comprehensive set of 576 nested models (Fig. 2c).

For model fitting, we used quantile maximum probability estimation (Heathcote et al., 2002; Heathcote, Brown, & Cousineau, 2004). First, we transformed the valid RT data of each condition into quintiles and a count of RTs falling into the corresponding bins. As with continuous maximum likelihood estimation, we then searched for parameter values that maximize the quantile probability. Using Matlab's *fmincon* function, we achieved this by minimizing the model deviance given by twice the negative log quantile likelihood summed across all eight conditions. To avoid local minima, model fitting was performed using multiple sets of start values. For μ parameters, we used the best-fitting estimates obtained separately in corresponding unisensory conditions plus values falling $\pm 2\%$ on either side. For σ parameters, we also used the best-fitting unisensory estimates plus values falling $\pm 2.5\%$ and $\pm$ 5% on either side. For ρ parameters, we used 10 start values that were evenly spaced between -0.9 and 0.9. For η parameters, we used four start values that were evenly spaced between 0% and 30% of the best-fitting σ estimates in unisensory conditions. Using all combinations, fitting of each model was initiated with up to 600 start value sets (fewer sets were used for model versions that did not use ρ or η).

Selection of the 'best' model was achieved using the Akaike Information Criterion (AIC; Akaike, 1973), which balances the model fit with a penalty for model complexity (given by the number of free parameters). We fitted each model individually for each of the 20 participants. To select the best model for the group as a whole, we used the group AIC, which is given by summing individual AIC values across participants for each candidate model (Rae, Heathcote, Donkin, Averell, & Brown, 2014). We then computed group AIC weights across all 576 candidate models and selected the model with the highest group AIC weight (Lewandowsky & Farrell, 2011; Wagenmakers & Farrell, 2004). As a sanity check, we tested the model selection procedures in model recovery simulations, which are presented as Supplementary Material. To complete the picture, we also used the Bayesian Information Criterion (BIC; Schwarz, 1978). The BIC tends to select models with fewer free parameters compared to the AIC (Lewandowsky & Farrell, 2011; Wagenmakers & Farrell, 2004).

3. Results

3.1. Experimental replication results

To understand how motion-in-depth signals are processed between auditory and visual modalities, we replicated the study by Cappe et al. (2009). In an initial experiment, we closely replicated the original procedures, which did not present static signals during the foreperiod. Hence, all signal and catch trials contained a transient stimulus onset that coincided in time with the task-relevant motion onset in signal trials. As the planned modeling work assumes ceiling performance, we checked whether performance was close to perfect before engaging in further analysis and modeling. Performance in the initial experiment was not at ceiling (Supplementary Results). We suspected the transient stimulus might be a factor that was detrimental to performance (Supplementary Fig. S1). Hence, we decoupled stimulus and motion onsets by introducing static signals during the foreperiod in our main experiment, the result of which are described below.

To check for ceiling performance in the main experiment, we assessed general performance measures (Table 1). First, overall hit rates were close to perfect at 99.8% ($\pm$0.146%, SEM). We analyzed hit rates as a function of modality (A, V, and AV; averaged across motion directions). A one-way ANOVA showed an effect of modality, $F(1.02,19.4) = 14.2$, $p = 0.001$, $\eta_p^2 = 0.428$, Greenhouse-Geisser corrected. Post-hoc tests revealed that auditory hit rates were significantly lower than those for both visual (mean difference: 0.708% $\pm$0.188%, $p = 0.004$) and redundant conditions (mean difference: 0.721% $\pm$0.190%, $p = 0.004$). However, the auditory hit rate was still close to ceiling (99.2% $\pm$0.205%). Hit rates in visual and redundant conditions were not significantly different. Second, the false alarm rate was low at 1.35% ($\pm$0.196%). Thus, the overall rate of correct responding (hits, correct rejections) was 99.4% ($\pm$0.092%). Performance in this experiment was therefore at ceiling levels, which allows us to proceed with the main analysis on RTs as planned.

Given the modifications to the experimental stimuli, the next analysis tested if we still replicated the core findings reported by Cappe et al. (2009). First, we confirmed that median RTs in redundant conditions were faster compared to unisensory conditions (Table 1). Averaged across different motion directions, a one-way ANOVA on median RTs showed a significant effect of modality $F(1.04,19.7) = 493$, $p < 0.001$, $\eta_p^2 = 0.963$, Greenhouse-Geisser corrected. Post-hoc tests revealed that responses to redundant signals were significantly faster than to both visual (mean difference: 0.011 s $\pm$ 0.002 s, $p < 0.001$) and auditory signals (mean difference: 0.207 s $\pm$ 0.009 s, $p < 0.001$). In addition, responses to visual signals were faster compared to auditory signals (mean difference: 0.197 s $\pm$ 0.009 s, $p < 0.001$). Second, and most critically, we checked for RT differences between congruent looming signals (ALVL) and the other redundant motion combinations (ALVR, ARVL, and ARVR). As RTs in the ALVL condition were significantly

Table 2
RT differences between redundant looming (ALVL) and the other motion combinations.

Pair	Mean RT difference (±SEM)	Paired *t*-test
ARVR - ALVL	0.027 s ± 0.003 s	$t_{(19)} = 10.0$, $p < 0.001$
ARVL - ALVL	0.019 s ± 0.003 s	$t_{(19)} = 7.21$, $p < 0.001$
ALVR - ALVL	0.010 s ± 0.002 s	$t_{(19)} = 4.02$, $p = 0.001$

faster compared to all other conditions (paired-samples *t*-tests, see Table 2), we successfully replicated the central finding of the original study. Cappe et al. (2009) used these findings to propose a selective integration mechanism for congruent audiovisual looming signals. Next, as outlined in the introduction, we follow a systematic analysis and modeling approach to study whether we have evidence to support this proposal.

3.2. Predicted and empirical RSE

Congruent audiovisual looming motion (ALVL) yielded fastest median RTs. However, this does not necessarily demonstrate a multisensory looming bias in the combined processing of audiovisual motion-in-depth. An alternative account could be that the condition with congruent looming signals inherit effects due to looming biases in the unisensory components.

To address this issue, we measured the RSE (i.e., the speedup of RTs in redundant compared to unisensory conditions) on the level of RT distributions (for example data in one audiovisual condition from one participant, see Fig. 3a). The RSE is here given by the shaded area enclosed by the redundant distribution (black) and the faster of the two single signal distributions, which is vision in this example (blue). Compared to median RTs, a key advantage of RT distributions is that they directly allow for parameter-free predictions of the RSE using probability summation as a simple combination rule (Otto, 2019; Raab, 1962). This analysis thus informs about potential changes in RSE that are expected from the unisensory components. The predicted distribution (red) derives from the independent race model (assuming statistical independence and context invariance), which uses the unisensory RT distributions and probability summation to compute the RT distribution with redundant signals. Using the predicted distribution for each participant in each of the four redundant conditions, we measured the predicted RSE (Fig. 3b) analogously to the empirical RSE (Fig. 3c).

To examine how the RSE varies in the four redundant conditions (Figs. 3b, c), we performed a 2x2x2 ANOVA with factors auditory motion (AR, AL), visual motion (VR, VL), and data type (predicted, empirical). There was a main effect of auditory motion, $F(1,19) = 87.1$, $p < 0.001$, $\eta_p^2 = 0.821$. Post-hoc tests revealed that the RSE was significantly larger for conditions including auditory looming compared to auditory receding motion (mean difference: 0.016 s ± 0.002 s). There was no main effect of vision, nor any interaction effect involving vision.

There was a main effect of data type, $F(1,19) = 29.5$, $p < 0.001$, $\eta_p^2 = 0.608$. The empirical RSE was larger than predicted by probability summation (mean difference: 0.008 s ± 0.001 s), suggesting that there is more going on than the simple probability summation prediction. There was also a significant interaction effect between auditory motion and data type, $F(1,19) = 18.0$, $p < 0.001$, $\eta_p^2 = 0.486$. To understand this effect, we analyzed the corresponding simple effects.

Regarding data type, the empirical RSE was significantly larger than predicted for both AR ($t_{(19)} = 2.29$, $p = 0.034$, mean difference: 0.003 s ± 0.001 s) and AL conditions ($t_{(19)} = 5.76$, $p < 0.001$, mean difference: 0.013 s ± 0.002 s). Regarding auditory motion, the RSE in AL conditions was significantly larger than in AR conditions, both in prediction ($t_{(19)} = 5.45$, $p < 0.001$, mean difference: 0.011 s ± 0.002 s) and empirically ($t_{(19)} = 9.98$, $p < 0.001$, mean difference: 0.021 s ± 0.002 s). Hence, both main effects (auditory motion and data type) were confirmed on the level of simple effects. The interaction effect can be understood in that the difference between empirical and predicted RSE is larger for conditions involving auditory looming compared to receding signals.

The main message from this analysis is that the modulation of the RSE in this experiment was driven by the characteristics of the auditory motion signal, with a larger RSE when redundant conditions involved auditory looming compared to auditory receding motion. Hence, the RSE here inherited an effect caused by one of the unisensory components.

In line with previous research and the proposal of a multisensory looming bias, we had expected to find a larger RSE with congruent audiovisual looming motion. However, as there was no interaction effect involving auditory and visual motion, we cannot conclude that the RSE was special with congruent audiovisual looming motion.

As a final point, the empirical RSE was consistently larger than predicted by the independent race model. Yet, the parameter-free model correctly predicted the main effect caused by auditory motion (compare Fig. 3b and c). In addition, the model accounts to a large extent for the variability in RSE across participants as demonstrated by a significant correlation between predicted and empirical RSE ($r = 0.662$, $p = 0.001$; considering individual data points of all four multisensory combinations). This correspondence between empirical RSE and prediction, although not perfect, shows a strong explanatory power of probability

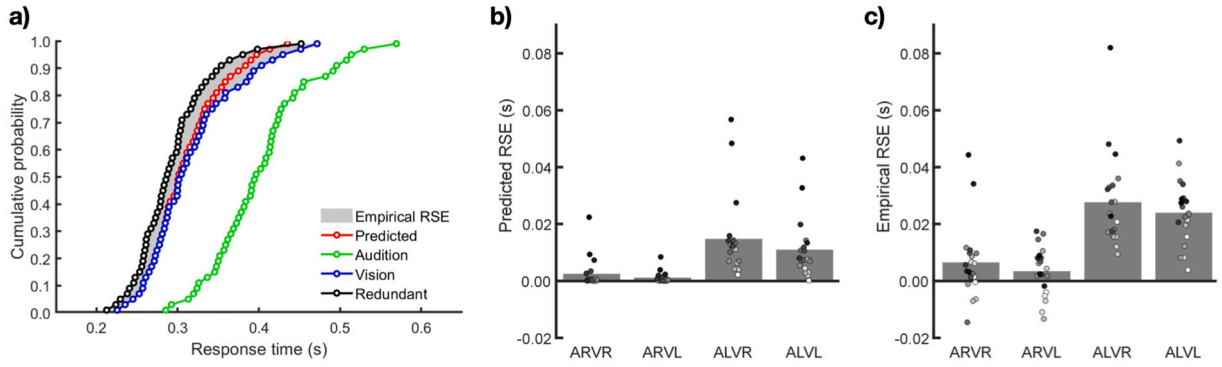

Fig. 3. Predicted and empirical RSE. a) Cumulative RT distributions for example data from one participant in the ALVL condition. Green, blue, and black circles show empirical distributions in the three signal combinations (50 quantiles per distribution). The RSE is given by the shaded area enclosed by the redundant distribution (black) and the faster of the two single signal distributions, which is vision in this example (blue). In addition, the two single signal distributions provide a parameter-free prediction of the redundant distribution using probability summation (red circles; note that the prediction in this example is partially occluded by the visual distribution). This prediction can then be used to measure the predicted RSE analogously to the empirical RSE. b) Predicted RSE. c) Empirical RSE. Bars show the mean across 20 participants, dots indicate individual data (gray levels follow the rank order averaged across conditions). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

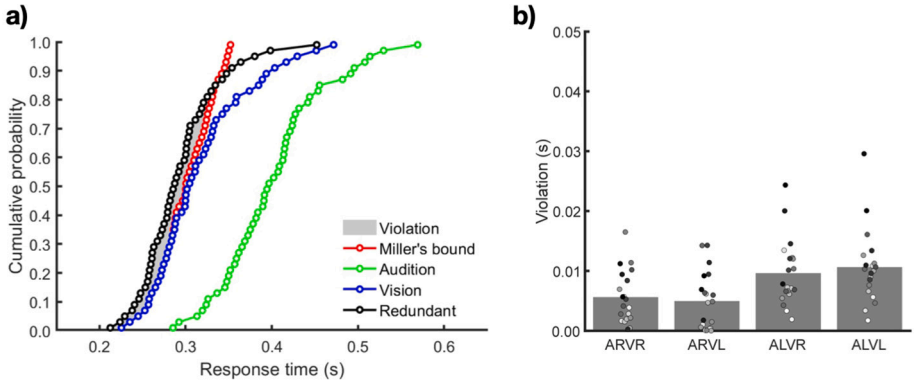

Fig. 4. Violation of Miller's bound. a) Quantification of Miller's bound violation using the area, for which the RT distribution with redundant signals exceeds Miller's bound (shaded gray). Miller's bound is computed by the sum of the two single signal RT distributions. Example data as in Fig. 3a. b) Violations of Miller's bound (area between the curves) as a function of signal combination. Bars show the mean violation across 20 participants, dots indicate individual data (gray levels follow the rank order averaged across conditions).

summation and race models. The parameter-free predictions rely on the assumptions of statistical independence and context invariance, which could be incorrect. Consequently, we next investigate processing interference beyond the independent race model, to allow us to search more thoroughly for a specific multisensory looming bias.

3.3. Processing interference effects

As a reminder, two assumptions are made by the independent race model. First, statistical independence states that the RT on a given trial is not affected by the RT on another trial (Ashby & Townsend, 1986; Colonius, 1990; Luce, 1986; Miller, 1982; Otto & Mamassian, 2017). Second, context invariance states that unisensory processing components act in exactly the same way whether there is a concurrent signal in the other modality or not (Ashby & Townsend, 1986; Liu & Otto, 2020; Luce, 1986; Otto & Mamassian, 2017; Townsend et al., 2020; Townsend & Wenger, 2004). The difference between empirical and predicted RSE could be explained by multisensory processing interference that violate one or the other assumption.

To test the statistical independence assumption, we examined potential sequential effects. Specifically, in single signal conditions we tested if the RT on a given trial depends on what was presented on the previous trial. For both audition and vision, we computed mean RTs on repetition (e.g., audition preceded by audition) and switch trials (e.g., audition preceded by vision). The *modality switch cost* is defined as the RT difference between repetition and switch trials. We calculated each participant's modality switch cost, separately for audition and vision, and then averaged these values to obtain an overall estimate of modality switch cost (the analysis here focuses only on signal modality, additional analysis including modality and motion direction did not reveal further effects, data not shown). A one-sample *t*-test showed that the overall modality switch cost (0.010 s ± 0.004 s) was significantly different from zero, $t_{(19)} = 2.19$, $p = 0.041$. While the effect here is smaller compared to previous RSE studies (e.g., Gondan et al., 2004; Innes & Otto, 2019; Miller, 1982; Otto & Mamassian, 2012; Shaw et al., 2020), the finding critically shows that the previously presented modality has opposing effects on auditory and visual RTs (e.g., a previous auditory trial brings a modality repetition for an auditory trial but a modality switch for a visual trial). It follows that RTs in single signal conditions, which are used for model predictions, are in fact not statistically independent.

Following from above, we must drop the assumption of statistical independence. For this case, Miller (1982) developed an upper bound that provides the largest possible RSE as predicted by a race model and potentially correlated RTs. If redundant signals violate any part of this bound, the implication is that multisensory processing interference beyond a race process with correlated RTs must have taken place. Consequently, as a second examination, we tested our data for potential violations of Miller's (1982) bound, which we quantified similar to the RSE (Fig. 4a; Otto, 2019).

To analyze violations in the four redundant signals conditions (Fig. 4b), we performed a 2 × 2 ANOVA with factors auditory (AR, AL) and visual motion (VR, VL). We found a significant intercept, $F(1,19) = 119$, $p < 0.001$, $\eta_p^2 = 0.862$, meaning that violations of Miller's bound were significantly different from zero. We performed additional bootstrap simulations showing that violations were significant on the group level in each of the four conditions (on the level of individual participants, violations were significant for all participants in conditions involving auditory looming motion and for 14/12 of the 20 participants in ARVL and ARVR, respectively). We also found a main effect of auditory motion, $F(1,19) = 22.3$, $p < 0.001$, $\eta_p^2 = 0.539$. Post-hoc tests revealed larger violations in conditions with auditory looming compared to conditions with auditory receding conditions (mean difference: 0.005 s ± 0.001 s). There was neither a main effect of visual motion nor an interaction effect between auditory and visual motion. Hence, like the RSE (Fig. 3), violations of Miller's bound were modulated by the auditory motion signal, specifically, violations were larger for conditions with auditory looming compared to auditory receding signals. To support the proposed multisensory looming bias (Cappe et al., 2009), we had expected to find larger violations in conditions with congruent audiovisual looming motion. However, as there was no interaction effect between auditory and visual motion, we cannot conclude that violations of Miller's bound were in any form special with congruent audiovisual looming motion.

As a final point, this section on multisensory processing interference has shown that the assumption of statistical independence, as made for predictions using the independent race model (Fig. 3), is incorrect due to there being modality switch costs. Moreover, when potential correlations are considered, we found significant violations of Miller's bound (Fig. 4). Given that Miller's bound, like the independent race model, assumes context invariance (Ashby & Townsend, 1986; Liu & Otto, 2020; Luce, 1986; Otto & Mamassian, 2017; Townsend et al., 2020; Townsend & Wenger, 2004), a direct explanation for the violations is that also the context invariance assumption is likely incorrect. At least one processing component must act differently on redundant compared to unisensory trials, which points to a potential second processing interference with audiovisual motion-in-depth. Putting all the pieces together, this implies that the RSE could be fully explained by a model with three components: a race mechanism, modality switch effects, and one further form of processing interference. We explored this possibility in the next section.

3.4. Model comparison

Here we aimed to provide a mathematical explanation for the RSE with audiovisual motion-in-depth signals that should capture all the effects found in the data. As the central question, we ask if multisensory responses inherit looming biases present in the unisensory components, or if there is a looming bias in multisensory processing.

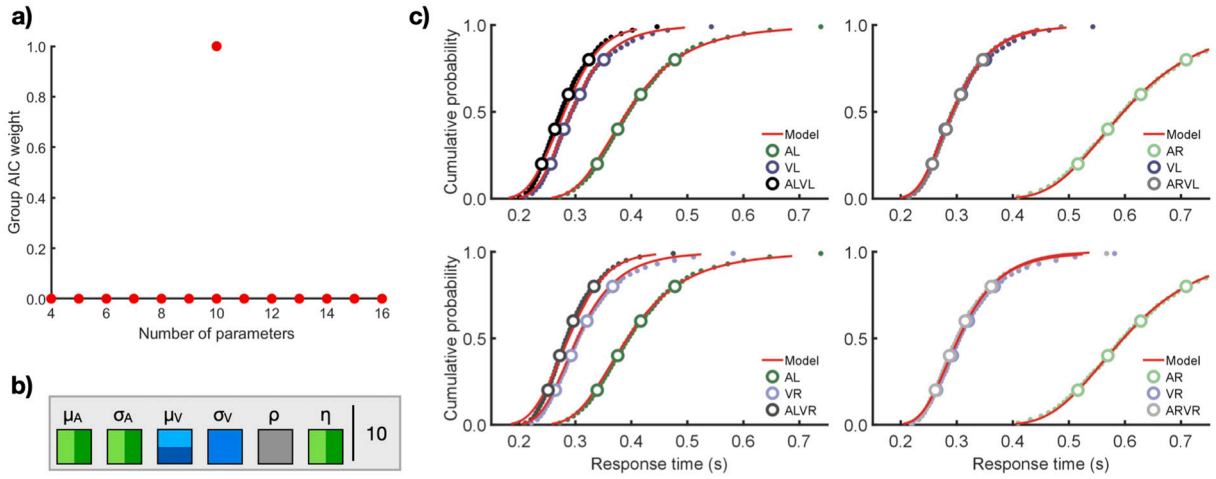

Fig. 5. Model comparison using the Akaike information criterion (AIC). a) Group AIC weights as a function of the number of free parameters for all 576 tested models. A 10-parameter model provided the highest weight. b) Best-fitting model. 3 of 4 unisensory parameters (not σ_V) use 2 values to account for the different motion directions. The model has a single parameter value for the correlation ρ, but 2 values for the noise η, which varies with the auditory motion direction. c) Model fit. Group-averaged RT distributions generated by the best-fitting model (red), compared to the empirical RT distributions. Each plot shows a redundant condition and its unisensory constituents (hence, each unisensory distribution is shown twice). Small dots show Vincentized group distributions, large circles show quantile estimates used for model fitting. For the underlying individual fits, per participant, see Supplemental Fig. S2. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

To explain the RSE on the level of RT distributions, we used a modeling approach that incorporates the race model architecture, but that factors-in modality switching and that allows for additional processing interference (Fig. 2a; Otto & Mamassian, 2012). To extend the model to the eight conditions tested here (two auditory, two visual, and four redundant; Fig. 1d), we allowed parameters to vary with experimental stimulus factors (i.e., the auditory motion direction, the visual motion direction, or combinations thereof; Fig. 2b). By considering all permutations of parameter options, we obtained a comprehensive set of 576 nested candidate models that ranged between four and 16 free parameters (Fig. 2c). We fitted each candidate model to the data of each participant using quantile maximum probability estimation (Heathcote et al., 2004; Heathcote et al., 2002; see Methods for more details). We then selected the best-fitting model using group AIC weights. To answer the central question above, it is of particular interest whether the interference parameters (ρ, η) of the best-fitting model are constant across the four redundant conditions or are different for a particular subset of motion combinations.

The model selection procedure with group AIC weights found a model with 10 free parameters as best-fitting (Fig. 5a, b). The selected model fitted the RT distributions in all eight conditions, including violations of Miller's bound, virtually perfectly (Fig. 5c; for model fits to data from individual participants, see Supplementary Fig. S2; for best-fitting parameter estimates, see Table 3). Seven of the 10 free model parameters were used to describe performance differences between looming and receding motion in the four unisensory conditions.

For processing within the auditory decision unit, both parameters μ_A and σ_A varied with the auditory motion direction, which captures the major RT differences between looming and receding conditions (e.g., the average RT difference between AL and AR is about 0.2 s, see Table 1). Hence, the model selection procedures capture a strong unisensory looming bias with auditory signals.

For processing within the visual decision unit, the rate parameter μ_V varied with the visual motion direction. The spread parameter σ_V used a single value across conditions, which may reflect the smaller RT differences between VR and VL conditions (approximately 1/17 compared to audition, see Table 1). Still, the model selection procedures capture also a unisensory looming bias with visual signals.

To understand the effect of multisensory motion-in-depth on the RSE, the remaining three free model parameters are most relevant. First, the correlation parameter ρ was used but did not vary across the four combinations of audiovisual motion directions. A one-sample *t*-test found that the ρ estimate (-0.394 ± 0.109) was significantly smaller than zero ($t_{(19)} = -3.62$, $p = 0.002$), thus supporting the notion that the assumption of statistical independence is incorrect. The negative correlation can be understood in that a previously presented auditory signal, for example, has opposite effects on RTs to auditory and visual signals on a present trial (Otto & Mamassian, 2012). Critically, as only one parameter is used, the model selection procedures do not reveal any multisensory looming bias here.

More interestingly, the additional noise parameter η varied with the auditory motion direction, as expressed by the parameters η_{AR} and η_{AL} both featuring in the best-fitting model. These parameters demonstrate that our dataset violates the context invariance assumption. The inclusion of the noise parameters enables the model to violate Miller's bound as reported above. A paired-samples t-test showed that η_{AL} was significantly larger than η_{AR} ($t_{(19)} = 7.72$, $p < 0.001$, mean difference: $0.198 \pm 0.026\ s^{-1}$), demonstrating that the model assumes more additional noise in conditions with auditory looming compared to auditory receding signals. As the two parameter values vary with the auditory motion direction, the model selection procedures reveal that multisensory processing inherits a unisensory looming bias here.

To sum up, along with the seven unisensory parameters, only three interference parameters are required to fit the RT distributions in the four redundant conditions. Critically, with respect to the main question regarding a selective integration of audiovisual looming signals, neither of the interference parameters varied with audiovisual motion congruency. Moreover, model recovery simulations show that our fitting procedures and sample size are sufficient to recover the 10-parameter model as best-fitting and that more complex models risk overfitting the data (see Supplementary Material). Thus, following this model-based analysis, we conclude that the RSE with audiovisual motion-in-depth signals does not require a selective integrative process for congruent looming motion.

As there are multiple ways to choose the best model for a dataset, for completeness, we also employed a second method using the BIC (Schwarz, 1978). This criterion selected a model with only six free parameters as best-fitting (Table 3, Supplementary Figs. S3 and S4). Regarding the unisensory decision units, only auditory parameters μ_A and σ_A varied with motion direction, just as was found for the best model

Table 3
Best-fitting model parameters. Mean and SEM of 20 participants.

Criterion	Parameters	μ_{AL} (s^{-1})	μ_{AR} (s^{-1})	μ_{VL} (s^{-1})	μ_{VR} (s^{-1})	σ_{AL} (s^{-1})	σ_{AR} (s^{-1})	σ_{V} (s^{-1})	ρ	η_{AL} (s^{-1})	η_{AR} (s^{-1})
AIC	10	2.59 ± 0.08	1.71 ± 0.058	3.49 ± 0.11	3.35 ± 0.10	0.53 ± 0.03	0.32 ± 0.02	0.62 ± 0.02	−0.39 ± 0.11	0.19 ± 0.03	−0.01 ± 0.01
BIC	6	2.60 ± 0.09	1.71 ± 0.058	3.46 ± 0.11	=μ_{VL}	0.57 ± 0.04	0.32 ± 0.02	0.66 ± 0.02	x	x	x

selected by the AIC. However, unlike the AIC approach, the BIC favored greater simplicity by using a single value in both visual parameters μ_V and σ_V. Interestingly, the best-fitting model according to the BIC did not use any interference parameter (i.e., neither ρ nor η). In other words, the BIC selected the independent race model as best-fitting (though note that the fit of the BIC-selected model showed some systematic deviations from the data, which are not noticeable with the AIC-selected model, compare Fig. 5 to Supplementary Fig. S3). Importantly, as with the analysis based on the AIC, the BIC did not select a model that had parameters varying with motion congruency, thus agreeing that selective integrative processing of audiovisual looming signals is not required to explain the data.

4. Discussion

4.1. Redundant signals effect (RSE) and potential sensory biases

Behavioral biases towards perceiving looming motion occur both in audition (e.g., Bach et al., 2008; Bach et al., 2009; Baumgartner et al., 2017; Freiberg et al., 2001; Neuhoff, 1998, 2001; Rosenblum et al., 1987; Rosenblum et al., 1993) and vision (e.g., Ball & Tronick, 1971; Franconeri & Simons, 2003; Lin, Franconeri, & Enns, 2008; Moher et al., 2015; Skarratt, Gellatly, Cole, Pilling, & Hulleman, 2014). Such biases have been suggested to be adaptive because they facilitate faster evasive actions, which likely improve survival odds when faced with an imminent collision or attack (e.g., Neuhoff, 2001; Rosenblum et al., 1993; Skarratt et al., 2014). We were interested in whether such biases also occur in the multisensory processing of audiovisual motion-in-depth. We have here successfully replicated findings by Cappe et al. (2009) showing that RTs to congruent looming motion are faster compared to congruent receding motion and any incongruent combination (Table 2). But our additional analyses lead us to a rather different conclusion. Based on a systematic analysis comparing data from both congruent and incongruent motion combinations and a model-based approach, we conclude that a selective integration mechanism is not required to explain the effect.

As a key metric to quantify multisensory benefits, we measured the RSE, which is the speedup of responses to redundant signals compared to the fastest of the constituent unisensory responses (Fig. 3). For there to be a looming bias in multisensory processing, with the 2 × 2 design, we would expect an interaction effect between auditory and visual motion directions, to the effect of the RSE being the largest of all for congruent audiovisual looming motion. However, we only found a main effect of the auditory motion direction. The RSE was generally larger when auditory motion was looming, compared to when it was receding. Hence, the RSE was largely modulated by an effect inherited from audition as one of the two unisensory processing components.

The modulation of the RSE by the auditory component was correctly predicted by probability summation (although the overall RSE magnitude was smaller than predicted; Fig. 3). The prediction can be easily understood from the perspective of models, which explain the RSE by a race between unisensory processes. To produce a redundancy gain, these models require the two “racers” to be competitive (i.e., both have a chance of winning). In contrast, if the same “racer” always wins, let's say vision, the race mechanism predicts no speed-up for redundant conditions. An implication of probability summation is consequently the *principle of congruent effectiveness*, which states that multisensory facilitation is larger when performance in unisensory conditions is more similar (Otto et al., 2013).

With respect to the present study, we found exactly this pattern of facilitation. RTs to visual signals are about 0.2 s faster than RTs to auditory signals (Table 1). Critically, this difference is much smaller for auditory looming (the difference between AL and V is about 0.1 s) compared to auditory receding motion (the difference between AR and V is about 0.3 s). As visual RTs are more similar to AL than to AR, the RSE is expected to be larger in conditions including auditory looming

motion, which is what we found (Fig. 3). In other words, the race is more competitive when involving auditory looming compared to receding signals, which directly explains the observed modulation of the RSE.

The model-based analysis can also help us understand why average RTs are fastest in conditions with congruent looming signals (Cappe et al., 2009; Table 2). This finding is explained in that race models sample the RT on a given trial from the faster of two parallel unisensory decision processes. Consequently, when the two unisensory processes are rather slow, the sampled multisensory RT is rather slow. When the two unisensory RTs are rather fast, the sampled multisensory RT is rather fast, too. Now, with respect to the present study, the unisensory conditions with the fastest average RTs were auditory looming and visual looming, respectively. Consequently, it is no surprise to find the redundant condition that combines these two conditions to be overall the fastest.

The last paragraphs demonstrated the importance of considering the RSE in relation to the unisensory component signals, and together with race model predictions. In fact, this issue concerns not only motion-in-depth but multisensory processing more generally. For example, the physical synchrony of signals is considered a principle of multisensory integration, benefits from combining synchronous signals will be larger (c.f., the temporal rule; Meredith & Stein, 1983; Stein & Meredith, 1993; Stein & Stanford, 2008). In behavioral RSE experiments, this rule is tested by introducing a physical delay to one or the other unisensory signal (e.g., Hershenson, 1962; Leone & McCourt, 2013; Miller, 1986; Zumer, White, & Noppeney, 2020). It is then found that multisensory RTs follow a U-shape function and are fastest in conditions without delay added (i.e., when signals are synchronous). Using the same explanation as developed above, this outcome is not surprising from the perspective of race models. By delaying a unisensory component, the resultant multisensory RT will be delayed, too. If one considers RT differences between sensory modalities, probability summation and the principle of congruent effectiveness predict that the RSE will be largest when the faster unisensory component is delayed by an amount equal to the RT difference between signals (i.e. when the added delay makes unisensory responses synchronous). This prediction has been confirmed by empirical studies (Colonius & Diederich, 2004; Hershenson, 1962; Otto et al., 2013). In the aggregate, these considerations add to the strong predictive power of race models (Otto & Mamassian, 2017).

However, pure race models and probability summation alone are often not sufficient to account for the RSE, and they were not sufficient to account for the data of this study. Most commonly, the RSE is found to violate Miller's (1982) bound. In other words, the fastest multisensory RTs are slightly faster than expected by a pure race mechanism. Such race model violations have led to a widespread rejection of race models as an explanation of the RSE (see the many follow-up studies of Miller, 1982) and are sometimes interpreted as evidence for integrative processing (but see Otto & Mamassian, 2017 for a systematic critique of this argument). Following this interpretation, to demonstrate selective integrative processing in the condition with congruent looming signals, we would expect to observe particularly large violations with these signals and potentially no violations in the other motion combinations. However, we found violations in all conditions (Fig. 4). Moreover, the size of the violations was determined by the auditory motion direction (larger violations in conditions involving auditory looming compared to auditory receding motion). Hence, following Miller's (1982) argument, our data do not provide support for the notion of a selective integration mechanism for multisensory looming signals.

Our analysis is not the only argument questioning a selective integration mechanism for multisensory looming signals. In a related study, Huygelier, van Ee, Lanssens, Wagemans, and Gillebert (2021) assessed attention capture (exogenous), attention orientation (endogenous), and sustained attention towards synchronous and asynchronous audiovisual looming signals using a variety of experimental paradigms. Critically, none of these experiments found evidence for an attentional benefit with synchronous compared to either asynchronous audiovisual or unisensory looming signals. As an explanation, Huygelier et al. (2021) suggested that their results could point to a lack of selective multisensory integration for synchronous looming signals. In another study, Grassi and Pavan (2012) investigated the subjective duration of audiovisual looming, receding, and stationary stimuli and found that the combination of any congruent signals is consistent with predictions using maximum likelihood estimation. While the study did not include incongruent combinations, the analysis still suggest that the same combination rule is used both for looming and receding audio-visual motion. Hence, both studies match the conclusion of our model-based analysis of the RSE.

4.2. Large scale model comparison

As a final step of our study, we engaged in a large-scale model comparison, with the aim of delivering a model that could best-fit our data, from a choice of many models, and so that we could understand why particular patterns occurred in the data. Otto and Mamassian (2012) proposed an interactive race model, which typically accounts for the RSE on the level of RT distributions including violations of Miller's bound (e.g., Harrar et al., 2014; Innes & Otto, 2019; Mercier & Cappe, 2020). To comprehensively explore the multisensory processing of audiovisual motion signals, we fitted our data using 576 model versions, the set of models representing all permutations of free parameters for unisensory decision units and potential interferences (Fig. 2). Critically, we found that the best-fitting model according to the AIC was a 10-parameter model which featured three interference parameters: a single negative correlation ρ and two noise parameters η_{AL} and η_{AR}, which varied with auditory motion. This 10-parameter model offers a virtually perfect fit to our empirical data (Fig. 5). Importantly, the model that best fit the data did not have the pattern of parameters expected if there were a distinct process that is selective towards congruent audiovisual looming motion. The important result here was that a relatively simple model, based on probability summation, plus a small set of interference parameters, is sufficient to explain behavior across all four combinations of audiovisual motions-in-depth, including congruent audiovisual looming.

Let us consider in detail, what key parameter emerges from the best-fit model, and why. The best-fitting parameter estimate of the additional noise (η) is larger for auditory looming compared to auditory receding motion (Table 3). One explanation could be that the noise interference is largest when the two unisensory decision processes are temporally overlapping, which is more likely the case when unisensory RTs are similar (this issue mirrors the above discussion regarding the principle of congruent effectiveness). In our data, visual RTs were closer to RTs from auditory looming conditions than from auditory receding conditions. Hence, auditory and visual processes are more likely to interfere when the auditory signal is looming. Similarly, in experiments that manipulated signal intensity and temporal delays, the noise interference follows an inverted U-shaped function and is largest in conditions with the smallest difference in unisensory RTs (Otto et al., 2013, their Fig. 7). In this respect, the main effect of auditory motion direction on the noise parameter found here may be more broadly understood as a general effect that emerges when unisensory decisions are manipulated, for example by signal intensity.

One point that should be noted is that a different rule for model selection did result in a slightly different best-fitting model. In addition to the AIC model selection, we also applied the BIC. One characteristic of the BIC is that it is more punitive than the AIC towards more free parameters (e.g., Lewandowsky & Farrell, 2011; Wagenmakers & Farrell, 2004). Hence, as expected, the BIC selected a model with fewer free parameters than the 10-parameter one chosen by the AIC. Intriguingly, the BIC selected a 6-parameter model that did not feature any interference parameter, which is therefore the independent race model. A possible explanation for why the BIC selected the independent race model might be that neither of the two interference parameters (ρ and η)

alone is sufficient to capture the processing that must take place in addition to the parallel race process. It remains therefore a task for future research to identify a single interference parameter, eventually replacing both ρ and η, that would provide a more accurate characterization of multisensory processing interference including modality switching. With such a model absent, the BIC, which favors simplicity, selected a too simple model.

Beyond the analysis of behavioral data, we hope that our modeling approach can inform the interpretation of neurophysiological studies as promoted in the area of model-based cognitive neuroscience (Forstmann & Wagenmakers, 2015). For example, there are highly relevant neuroimaging and electrophysiological studies investigating responses to multisensory looming signals (e.g., Maier, Chandrasekaran, & Ghazanfar, 2008; Tyll et al., 2013). Unfortunately, these studies used behavioral tasks that were not directly related to the audio-visual motion signals, which makes it at present difficult to discuss neuronal correlates of our behavioral data using the redundant signals paradigm with audio-visual motion-in-depth signals.

4.3. Considering the stimuli for motion in depth studies

A final point to consider is whether our stimuli were valid representations of looming motion. We used auditory intensity and visual size changes to represent motion-in-depth. However, there are many other cues to motion-in-depth in the real world. For audition, motion-in-depth can be signaled by changes in intensity, direct-to-reverberant energy ratio, spectra, interaural time difference, interaural level difference, as well as dynamic cues like the Doppler effect (Zahorik, Brungart, & Bronkhorst, 2005). For vision, motion-in-depth is signaled by changes in size, clarity, occlusion, stereopsis, eye convergence and accommodation (Paquier, Cote, Devillers, & Koehl, 2016). Additionally, the concern is not just stimulus simplicity per se. A multitude of studies suggest that the looming response is strong only if the object looms close (to the peripersonal space; e.g., Camponogara, Komeilipoor, & Cesari, 2015b; Canzoneri, Magosso, & Serino, 2012; Graziano, Reiss, & Gross, 1999; Graziano, Yap, & Gross, 1994; Serino, Annella, & Avenanti, 2009). With purely auditory intensity change on a synthetic tone via headphones, and visual size change on an abstract shape, the distance to the simulated looming object is unknown (Bronkhorst & Houtgast, 1999; Zahorik et al., 2005). Presentation equipment such as headphones and monitors complicate things further because such presentation equipment anchors the distance even if the stimulus itself simulates motion-in-depth (Paquier et al., 2016). There is some comfort in that past looming studies have typically used only auditory intensity change (e.g., Freiberg et al., 2001; Neuhoff, 1998, 2001; Seifritz et al., 2002) or visual size change (e.g., Franconeri & Simons, 2003; Lin et al., 2008; Moher et al., 2015; Skarratt et al., 2014), and such studies have found behaviors that are appropriate to looming. Auditory intensity change was also found to be the dominant cue for predicting time-to-arrival, compared to interaural temporal difference and the Doppler effect (Rosenblum et al., 1987). Furthermore, several studies have found similar fear responses between single-cue looming and multi-cue looming, in both vision (e.g., Ball & Tronick, 1971) and audition (e.g., Bach et al., 2008). Hence, our simplistic stimuli were sufficient, at least in terms of looming-induced fear. It remains an open question whether inducing presumably more fear from unambiguously close looming would further our finding: probability summation with processing interference explains the response towards audiovisual motion-in-depth.

5. Conclusion

Biases in the perception of looming, using auditory or visual motion signals, are established phenomena. It was previously proposed that there is also a looming bias for multisensory processing using congruent audiovisual looming signals. We conducted an experiment using the classic redundant signals paradigm and found the well-known redundant signals effect (RSE): there is a speedup of redundant signals RT compared to RT's for constituent unisensory stimuli. Critically, using a large-scale model comparison, we discovered that the RSE is not distinct for congruent audiovisual looming. In fact, the differences in RSE were driven by auditory motion alone, and probability summation with processing interference explains behavioral performance equally in all of the audiovisual conditions tested. To sum up, we have found no evidence for a looming bias in multisensory processing.

CRediT authorship contribution statement

S.F. Andrew Chua: Conceptualization, Methodology, Data curation, Software, Formal analysis, Visualization, Writing – original draft. **Yue Liu:** Software, Writing – review & editing. **Julie M. Harris:** Conceptualization, Writing – review & editing, Supervision. **Thomas U. Otto:** Conceptualization, Methodology, Data curation, Software, Writing – review & editing, Supervision.

Acknowledgements

T.U.O. was supported by the Biotechnology and Biological Sciences Research Council (BBSRC, grant number: BB/N010108/1). S.F.A.C was partially funded by an Experimental Psychology Society Undergraduate Research Bursary.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cognition.2022.105204.

References

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In B. N. Petrox, & F. Caski (Eds.), *Second international symposium on information theory* (pp. 267–281). Budapest: Akademiai Kiado.

Alais, D., Newell, F. N., & Mamassian, P. (2010). Multisensory processing in review: From physiology to behaviour. *Seeing and Perceiving, 23*(1), 3–38. https://doi.org/10.1163/187847510X488603

Ashby, F. G., & Townsend, J. T. (1986). Varieties of perceptual Independence. *Psychological Review, 93*(2), 154–179. https://doi.org/10.1037//0033-295x.93.2.154

Bach, D. R., Neuhoff, J. G., Perrig, W., & Seifritz, E. (2009). Looming sounds as warning signals: The function of motion cues. *International Journal of Psychophysiology, 74*(1), 28–33. https://doi.org/10.1016/j.ijpsycho.2009.06.004

Bach, D. R., Schachinger, H., Neuhoff, J. G., Esposito, F., Di Salle, F., Lehmann, C., … Seifritz, E. (2008). Rising sound intensity: An intrinsic warning cue activating the amygdala. *Cerebral Cortex, 18*(1), 145–150. https://doi.org/10.1093/cercor/bhm040

Ball, W., & Tronick, E. (1971). Infant responses to impending collision: Optical and real. *Science, 171*(3973), 818–820. https://doi.org/10.1126/science.171.3973.818

Baumgartner, R., Reed, D. K., Toth, B., Best, V., Majdak, P., Colburn, H. S., & Shinn-Cunningham, B. (2017). Asymmetries in behavioral and neural responses to spectral cues demonstrate the generality of auditory looming bias. *Proceedings of the National Academy of Sciences of the United States of America, 114*(36), 9743–9748. https://doi.org/10.1073/pnas.1703247114

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision, 10*(4), 433–436. https://doi.org/10.1163/156856897X00357

Bronkhorst, A. W., & Houtgast, T. (1999). Auditory distance perception in rooms. *Nature, 397*(6719), 517–520. https://doi.org/10.1038/17374

Burr, D., & Alais, D. (2006). Combining visual and auditory information. *Progress in Brain Research, 155*, 243–258. https://doi.org/10.1016/S0079-6123(06)55014-9

Camponogara, I., Komeilipoor, N., & Cesari, P. (2015a). When distance matters: Perceptual bias and behavioral response for approaching sounds in peripersonal and extrapersonal space. *Neuroscience, 304*, 101–108. https://doi.org/10.1016/j.neuroscience.2015.07.054

Camponogara, I., Komeilipoor, N., & Cesari, P. (2015b). When distance matters: Perceptual bias and behavioral response for approaching sounds in peripersonal and extrapersonal space. *Neuroscience, 304*, 101–108. https://doi.org/10.1016/j.neuroscience.2015.07.054

Canzoneri, E., Magosso, E., & Serino, A. (2012). Dynamic sounds capture the boundaries of peripersonal space representation in humans. *PLoS One, 7*(9), Article e44306. https://doi.org/10.1371/journal.pone.0044306

Cappe, C., Thelen, A., Romei, V., Thut, G., & Murray, M. M. (2012). Looming signals reveal synergistic principles of multisensory integration. *Journal of Neuroscience, 32*(4), 1171–1182. https://doi.org/10.1523/Jneurosci.5517-11.2012

Cappe, C., Thut, G., Romei, V., & Murraya, M. M. (2009). Selective integration of auditory-visual looming cues by humans. *Neuropsychologia, 47*(4), 1045–1052. https://doi.org/10.1016/j.neuropsychologia.2008.11.003

Carpenter, R. H. S., & Williams, M. L. (1995a). Neural computation of log likelihood in control of saccadic eye movements. *Nature, 377*(6544), 59–62. https://doi.org/10.1038/377059a0

Carpenter, R. H. S., & Williams, M. L. L. (1995b). Neural computation of log likelihood in control of saccadic eye movements. *Nature, 377*(6544), 59–62. https://doi.org/10.1038/377059a0

Chua, S. F. A., Liu, Y., Harris, J. M., & Otto, T. U. (2021). No selective integration required: A race model explains responses to audiovisual motion-in-depth (dataset). *University of St Andrews Research Portal.*. https://doi.org/10.17630/362541c2-ac2a-40d5-8d27-26d47a464e12

Colonius, H. (1990). Possibly dependent probability summation of reaction-time. *Journal of Mathematical Psychology, 34*(3), 253–275. https://doi.org/10.1016/0022-2496(90)90032-5

Colonius, H., & Diederich, A. (2004). Multisensory interaction in saccadic reaction time: A time-window-of-integration model. *Journal of Cognitive Neuroscience, 16*(6), 1000–1009. https://doi.org/10.1162/0898929041502733

Driver, J., & Noesselt, T. (2008). Multisensory interplay reveals crossmodal influences on 'sensory-specific' brain regions, neural responses, and judgments. *Neuron, 57*(1), 11–23. https://doi.org/10.1016/j.neuron.2007.12.013

Ernst, M. O., & Bülthoff, H. H. (2004). Merging the senses into a robust percept. *Trends in Cognitive Sciences, 8*(4), 162–169. https://doi.org/10.1016/j.tics.2004.02.002

Forstmann, B. U., & Wagenmakers, E.-J. (2015). *An introduction to model-based cognitive neuroscience*. New York: Springer.

Franconeri, S. L., & Simons, D. J. (2003). Moving and looming stimuli capture attention. *Perception & Psychophysics, 65*(7), 999–1010. https://doi.org/10.3758/bf03194829

Freiberg, K., Tually, K., & Crassini, B. (2001). Use of an auditory looming task to test infants' sensitivity to sound pressure level as an auditory distance cue. *British Journal of Developmental Psychology, 19*(1), 1–10. https://doi.org/10.1348/026151001165903

Giard, M. H., & Peronnet, F. (1999). Auditory-visual integration during multimodal object recognition in humans: A behavioral and electrophysiological study. *Journal of Cognitive Neuroscience, 11*(5), 473–490. https://doi.org/10.1162/089892999563544

Gold, J. I., & Shadlen, M. N. (2007). The neural basis of decision making. *Annual Review of Neuroscience, 30*, 535–574. https://doi.org/10.1146/annurev.neuro.29.051605.113038

Gondan, M., Lange, K., Rosler, F., & Roder, B. (2004). The redundant target effect is affected by modality switch costs. *Psychonomic Bulletin & Review, 11*(2), 307–313. https://doi.org/10.3758/Bf03196575

Grassi, M., & Pavan, A. (2012). The subjective duration of audiovisual looming and receding stimuli. *Attention, Perception, & Psychophysics, 74*(6), 1321–1333. https://doi.org/10.3758/s13414-012-0324-x

Graziano, M. S., Reiss, L. A., & Gross, C. G. (1999). A neuronal representation of the location of nearby sounds. *Nature, 397*(6718), 428–430. https://doi.org/10.1038/17115

Graziano, M. S., Yap, G. S., & Gross, C. G. (1994). Coding of visual space by premotor neurons. *Science, 266*(5187), 1054–1057. https://doi.org/10.1126/science.7973661

Harrar, V., Tammam, J., Perez-Bellido, A., Pitt, A., Stein, J., & Spence, C. (2014). Multisensory integration and attention in developmental dyslexia. *Current Biology, 24*(5), 531–535. https://doi.org/10.1016/j.cub.2014.01.029

Heathcote, A., Brown, S., & Cousineau, D. (2004). QMPE: Estimating lognormal, Wald, and Weibull RT distributions with a parameter-dependent lower bound. *Behavior Research Methods, Instruments, & Computers, 36*(2), 277–290. https://doi.org/10.3758/bf03195574

Heathcote, A., Brown, S., & Mewhort, D. J. K. (2002). Quantile maximum likelihood estimation of response time distributions. *Psychonomic Bulletin & Review, 9*(2), 394–401. https://doi.org/10.3758/Bf03196299

Hershenson, M. (1962). Reaction time as a measure of intersensory facilitation. *Journal of Experimental Psychology, 63*, 289–293. https://doi.org/10.1037/h0039516

Huygelier, H., van Ee, R., Lanssens, A., Wagemans, J., & Gillebert, C. R. (2021). Audiovisual looming signals are not always prioritised: Evidence from exogenous, endogenous and sustained attention. *Journal of Cognitive Psychology, 33*(3), 282–303. https://doi.org/10.1080/20445911.2021.1896528

Innes, B. R., & Otto, T. U. (2019). A comparative analysis of response times shows that multisensory benefits and interactions are not equivalent. *Scientific Reports, 9*, 2921. https://doi.org/10.1038/s41598-019-39924-6

Kinchla, R. A. (1974). Detecting target elements in multielement arrays: A confusability model. *Perception & Psychophysics, 15*(1), 149–158. https://doi.org/10.3758/BF03205843

Kleiner, M., Brainard, D. H., & Pelli, D. G. (2007). What's new in Psychtoolbox-3? *Perception, 36* (ECVP Abstract Supplement).

Leone, L. M., & McCourt, M. E. (2013). The roles of physical and physiological simultaneity in audiovisual multisensory facilitation. *Iperception, 4*(4), 213–228. https://doi.org/10.1068/i0532

Lewandowsky, S., & Farrell, S. (2011). *Computational modeling in cognition: Principles and practice*. SAGE publications. https://doi.org/10.4135/9781483349428

Leys, C., Ley, C., Klein, O., Bernard, P., & Licata, L. (2013). Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. *Journal of Experimental Social Psychology, 49*(4), 764–766. https://doi.org/10.1016/j.jesp.2013.03.013

Li, X. R., Liang, Z., Kleiner, M., & Lu, Z. L. (2010). RTbox: A device for highly accurate response time measurements. *Behavior Research Methods, 42*(1), 212–225. https://doi.org/10.3758/Brm.42.1.212

Lin, J. Y., Franconeri, S., & Enns, J. T. (2008). Objects on a collision path with the observer demand attention. *Psychological Science, 19*(7), 686–692. https://doi.org/10.1111/j.1467-9280.2008.02143.x

Liu, Y., Chua, S. F. A., & Otto, T. U. (2022) (under review). *The spatial rule applies to multisensory reaction times, but it is due to sequential effects*.

Liu, Y., & Otto, T. U. (2020). The role of context in experiments and models of multisensory decision making. *Journal of Mathematical Psychology, 96*, Article 102352. https://doi.org/10.1016/j.jmp.2020.102352

Luce, R. D. (1986). *Response times: Their role in inferring elementary mental organization*. New York: Oxford University Press.

Maier, J. X., Chandrasekaran, C., & Ghazanfar, A. A. (2008). Integration of bimodal looming signals through neuronal coherence in the temporal lobe. *Current Biology, 18*(13), 963–968. https://doi.org/10.1016/j.cub.2008.05.043

Maier, J. X., & Ghazanfar, A. A. (2007). Looming biases in monkey auditory cortex. *Journal of Neuroscience, 27*(15), 4093–4100. https://doi.org/10.1523/JNEUROSCI.0330-07.2007

Mercier, M. R., & Cappe, C. (2020). The interplay between multisensory integration and perceptual decision making. *Neuroimage, 222*. https://doi.org/10.1016/j.neuroimage.2020.116970

Meredith, M. A., & Stein, B. E. (1983). Interactions among converging sensory inputs in the superior colliculus. *Science, 221*(4608), 389–391. https://doi.org/10.1126/science.6867718

Miller, J. (1982). Divided attention: Evidence for coactivation with redundant signals. *Cognitive Psychology, 14*(2), 247–279. https://doi.org/10.1016/0010-0285(82)90010-X

Miller, J. (1986). Timecourse of coactivation in bimodal divided attention. *Perception & Psychophysics, 40*(5), 331–343. https://doi.org/10.3758/Bf03203025

Moher, J., Sit, J., & Song, J. H. (2015). Goal-directed action is automatically biased towards looming motion. *Vision Research, 113*, 188–197. https://doi.org/10.1016/j.visres.2014.08.005

Nadarajah, S., & Kotz, S. (2008). Exact distribution of the max/min of two Gaussian random variables. *IEEE Transactions on Very Large Scale Integration (VLSI) Systems, 16*(2), 210–212. https://doi.org/10.1109/tvlsi.2007.912191

Neuhoff, J. G. (1998). Perceptual bias for rising tones. *Nature, 395*(6698), 123–124. https://doi.org/10.1038/25862

Neuhoff, J. G. (2001). An adaptive bias in the perception of looming auditory motion. *Ecological Psychology, 13*(2), 87–110. https://doi.org/10.1207/S15326969ECO1302_2

Noorani, I., & Carpenter, R. H. (2016). The LATER model of reaction time and decision. *Neuroscience and Biobehavioral Reviews, 64*, 229–251. https://doi.org/10.1016/j.neubiorev.2016.02.018

Otto, T. U. (2019). RSE-box: An analysis and modelling package to study response times to multiple signals. *The Quantitative Methods for Psychology, 15*(2), 112–133. https://doi.org/10.20982/tqmp.15.2.p112

Otto, T. U., Dassy, B., & Mamassian, P. (2013). Principles of multisensory behavior. *Journal of Neuroscience, 33*(17), 7463–7474. https://doi.org/10.1523/JNEUROSCI.4678-12.2013

Otto, T. U., & Mamassian, P. (2012). Noise and correlations in parallel perceptual decision making. *Current Biology, 22*(15), 1391–1396. https://doi.org/10.1016/j.cub.2012.05.031

Otto, T. U., & Mamassian, P. (2017). Multisensory decisions: The test of a race model, its logic, and power. *Multisensory Research, 30*(1), 1–24. https://doi.org/10.1163/22134808-00002541

Paquier, M., Cote, N., Devillers, F., & Koehl, V. (2016). Interaction between auditory and visual perceptions on distance estimations in a virtual environment. *Applied Acoustics, 105*, 186–199. https://doi.org/10.1016/j.apacoust.2015.12.014

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision, 10*(4), 437–442. https://doi.org/10.1163/156856897x00366

Raab, D. H. (1962). Statistical facilitation of simple reaction times. *Transactions of the New York Academy of Sciences, 24*(5), 574–590.

Rae, B., Heathcote, A., Donkin, C., Averell, L., & Brown, S. (2014). The hare and the tortoise: Emphasizing speed can change the evidence used to make decisions. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 40*(5), 1226–1243. https://doi.org/10.1037/a0036801

Rosenblum, L. D., Carello, C., & Pastore, R. E. (1987). Relative effectiveness of three stimulus variables for locating a moving sound source. *Perception, 16*(2), 175–186. https://doi.org/10.1068/p160175

Rosenblum, L. D., Wuestefeld, A. P., & Saldana, H. M. (1993). Auditory looming perception: Influences on anticipatory judgments. *Perception, 22*(12), 1467–1482. https://doi.org/10.1068/p221467

Schwarz, G. E. (1978). Estimating the dimension of a model. *Annals of Statistics, 6*(2), 461–464. https://doi.org/10.1214/aos/1176344136

Seifritz, E., Neuhoff, J. G., Bilecen, D., Scheffler, K., Mustovic, H., Schachinger, H., … Di Salle, F. (2002). Neural processing of auditory looming in the human brain. *Current Biology, 12*(24), 2147–2151. https://doi.org/10.1016/s0960-9822(02)01356-8

Serino, A., Annella, L., & Avenanti, A. (2009). Motor properties of peripersonal space in humans. *PLoS One, 4*(8), Article e6582. https://doi.org/10.1371/journal.pone.0006582

Shaw, L. H., Freedman, E. G., Crosse, M. J., Nicholas, E., Chen, A. M., Braiman, M. S., … Foxe, J. J. (2020). Operating in a multisensory context: Assessing the interplay between multisensory reaction time facilitation and inter-sensory task-switching effects. *Neuroscience, 436*, 122–135. https://doi.org/10.1016/j.neuroscience.2020.04.013

Skarratt, P. A., Gellatly, A. R., Cole, G. G., Pilling, M., & Hulleman, J. (2014). Looming motion primes the visuomotor system. *Journal of Experimental Psychology: Human Perception and Performance, 40*(2), 566–579. https://doi.org/10.1037/a0034456

Stein, B. E., & Meredith, M. A. (1993). *The merging of the senses*. The MIT Press.

Stein, B. E., & Stanford, T. R. (2008). Multisensory integration: Current issues from the perspective of the single neuron. *Nature Reviews Neuroscience, 9*(4), 255–266. https://doi.org/10.1038/nrn2331

The-British-Psychological-Society. (2014). Code of human research ethics. St Andrews House, 48 Princess Road East, Leicester, ISBN 978–1–85433-762-7. Retrieved from https://www.bps.org.uk/sites/bps.org.uk/files/Policy%20-%20Files/BPS%20Code%20of%20Human%20Research%20Ethics.pdf.

Todd, J. (1912). *Reaction to multiple stimuli* (Vol. 3). New York City: The Science Press. https://doi.org/10.1037/13053-000

Townsend, J. T., Liu, Y., Zhang, R., & Wenger, M. J. (2020). Interactive parallel models: No Virginia, violation of Miller's race inequality does not imply coactivation and yes Virginia, context invariance is testable. *The Quantitative Methods for Psychology, 16*(2), 192–212. https://doi.org/10.20982/tqmp.16.2.p192

Townsend, J. T., & Wenger, M. J. (2004). A theory of interactive parallel processing: New capacity measures and predictions for a response time inequality series. *Psychological Review, 111*(4), 1003–1035. https://doi.org/10.1037/0033-295x.111.4.1003

Tyll, S., Bonath, B., Schoenfeld, M. A., Heinze, H. J., Ohl, F. W., & Noesselt, T. (2013). Neural basis of multisensory looming signals. *Neuroimage, 65*, 13–22. https://doi.org/10.1016/j.neuroimage.2012.09.056

Wagenmakers, E. J., & Farrell, S. (2004). AIC model selection using Akaike weights. *Psychonomic Bulletin & Review, 11*(1), 192–196. https://doi.org/10.3758/Bf03206482

Yang, C. T., Altieri, N., & Little, D. R. (2018). An examination of parallel versus coactive processing accounts of redundant-target audiovisual signal processing. *Journal of Mathematical Psychology, 82*, 138–158. https://doi.org/10.1016/j.jmp.2017.09.003

Zahorik, P., Brungart, D. S., & Bronkhorst, A. W. (2005). Auditory distance perception in humans: A summary of past and present research. *Acta Acustica united with Acustica, 91*(3), 409–420.

Zumer, J. M., White, T. P., & Noppeney, U. (2020). The neural mechanisms of audiotactile binding depend on asynchrony. *The European Journal of Neuroscience*. https://doi.org/10.1111/ejn.14928