Dynamic Weighting of Time-Varying Visual and Auditory Evidence During Multisensory Decision Making

Tuip, R.R.M.; van der Ham, Wessel; Lorteije, J.A.M.; van Opstal, F.

DOI
10.1163/22134808-bja10088

Publication date
2023

Document Version
Final published version

Published in
Multisensory Research

BRILL

Multisensory Research 36 (2023) 31–56

MULTISENSORY RESEARCH

brill.com/msr

Dynamic Weighting of Time-Varying Visual and Auditory Evidence During Multisensory Decision Making

Rosanne R. M. Tuip [1,2,*], **Wessel van der Ham** [1], **Jeannette A. M. Lorteije** [1,3,**] **and Filip Van Opstal** [2,*,**]

[1] Swammerdam Institute for Life Sciences, Center for Neuroscience, Faculty of Science, University of Amsterdam, 1098 XH Amsterdam, The Netherlands

[2] Department of Psychology, Brain and Cognition, University of Amsterdam, 1098 XH Amsterdam, The Netherlands

[3] Animal Welfare Body, Radboud University/UMC, 6525 EZ Nijmegen, The Netherlands

[*] Corresponding authors; e-mail: R.R.M.Tuip@uva.nl; F.vanOpstal@uva.nl

[**] These authors contributed equally.

ORCID iD: Tuip: 0000-0002-3896-824X

Received 21 February 2022; accepted 22 November 2022; published online 1 December 2022; published in print 11 January 2023

Abstract

Perceptual decision-making in a dynamic environment requires two integration processes: integration of sensory evidence from multiple modalities to form a coherent representation of the environment, and integration of evidence across time to accurately make a decision. Only recently studies started to unravel how evidence from two modalities is accumulated across time to form a perceptual decision. One important question is whether information from individual senses contributes equally to multisensory decisions. We designed a new psychophysical task that measures how visual and auditory evidence is weighted across time. Participants were asked to discriminate between two visual gratings, and/or two sounds presented to the right and left ear based on respectively contrast and loudness. We varied the evidence, i.e., the contrast of the gratings and amplitude of the sound, over time. Results showed a significant increase in performance accuracy on multisensory trials compared to unisensory trials, indicating that discriminating between two sources is improved when multisensory information is available. Furthermore, we found that early evidence contributed most to sensory decisions. Weighting of unisensory information during audiovisual decision-making dynamically changed over time. A first epoch was characterized by both visual and auditory weighting, during the second epoch vision dominated and the third epoch finalized the weighting profile with auditory dominance. Our results suggest that during our task multisensory improvement is generated by a mechanism that requires cross-modal interactions but also dynamically evokes dominance switching.

DOI:10.1163/22134808-bja10088

Keywords
multisensory integration, evidence accumulation, sensory weighting, temporal integration, perceptual decision-making

1. Introduction

Our brain combines information from the environment to form a coherent representation of the world. This involves combining sensory signals originating from different sources. Sensory evidence is not always instantaneously clear, but instead it can be noisy as it can consist of very subtle and sometimes even contradictory brief events that vary dynamically. For example, imagine a picnic with a group of friends in the park on a cloudy spring day. When you are having a conversation with the person across from you, you must integrate the fragmented movements of the lips as belonging to a single origin while the scene constantly changes in light intensity. It is therefore not surprising that studies in the field of perceptual decision-making have impinged on the notion that we need to continuously accumulate sensory evidence across time (Drugowitsch *et al*., 2014; Gold and Shadlen, 2007; Ratcliff *et al*., 2016).

Different strategies have been revealed for studying sensory evidence accumulation. A number of studies using fluctuating visual information (i.e., where the evidence changes over time) have demonstrated that observers tend to weight early sensory information more heavily than late information (Booras *et al*., 2021; Huk and Shadlen, 2005; Kiani *et al*., 2008; Nienborg and Cumming, 2009; Odoemene *et al*., 2018; Zylberberg *et al*., 2012). However, late sensory information integration strategies (Bronfman *et al*., 2016; Cheadle *et al*., 2014; Levi *et al*., 2018) and flat weighting profiles (Bronfman *et al*., 2016; Odoemene *et al*., 2018) have also been observed. These differences in weighting profiles might relate to task specifics and stimulus features. Early profiles are observed when information throughout the trial is equally informative, while late profiles are related to instances where integrating early information is not sufficient to solve the trial and thus late integration is necessary (Bronfman *et al*., 2016; Levi *et al*., 2018; Talluri *et al*., 2021).

Besides dealing with noisy information, our brain receives and integrates sensory information originating from different modalities. Early work by Meredith and Stein (1986) has demonstrated that in the cat superior colliculus, multisensory integration is associated with dynamic neuronal responses such as enhanced responses to multisensory stimuli compared to unisensory stimuli. Later, it was revealed that multisensory stimuli are integrated in a statistically optimal fashion (Ernst and Banks, 2002) and additional brain regions as well as different processing stages have been proposed to be involved in

multisensory integration (for reviews see Bizley *et al*., 2016 and Mercier and Cappe, 2020). In a study by Raposo *et al*. (2012), rats and humans had to integrate time-varying audiovisual information to discriminate between high and low rate events. They found that rate categorization was better on audiovisual compared to unisensory trials in both humans and rats. It is important to point out that the variable that subjects needed to estimate in this study (i.e., rate) was dependent on time. Perceptual decisions, however, are not always of this nature and require the estimation of the quality of sensory information. Estimations of visual features such as contrast and colour and auditory features such as the loudness and tone are crucial to discriminate between real life events in space and time. Raposo and colleagues (2012) additionally showed that evidence integration in humans was characterized by an early-weighting profile. However, the differential contribution of visual and auditory information on audiovisual trials was not investigated.

Visual and auditory information streams are often processed with unequal weights. Visual dominance has been observed in numerous studies where participants relied more on a visual stimulus compared to an auditory stimulus during audiovisual decision-making tasks (Bertelson and Radeau, 1981; Colavita, 1974; Pick *et al*., 1969; Welch and Warren, 1980). How visual and auditory information individually contribute to audiovisual decisions over time, however, remains to be investigated. In this study we aimed to address this issue. We designed an experimental paradigm during which participants had to discriminate between two visual gratings, two sound sources, or a combination of both, based on contrast and loudness.

2. Experiment

2.1. Methods

2.1.1. Participants

Fifty-four female and 24 male participants (mean age = 20.26, standard deviation = 2.52, age range = 18–38) took part in this experiment. All participants were students at the University of Amsterdam and participated for course credits. They were recruited *via* the website of the Behavioural Science Lab. Participants were screened to exclude subjects with visual or auditory impairments, with the exception of corrected-to-normal vision and audition. They provided written consent and were naive regarding the experimental design and goal of the study. The study was approved by The Faculty Ethics Review Board of the Faculty of Social and Behavioural Sciences of the University of Amsterdam.

2.1.2. Stimuli and Procedure

Participants performed a two-alternative forced-choice decision task designed using the Psychtoolbox library (Brainard, 1997; Pelli, 1997) in MATLAB R2017a (The MathWorks, Natick, MA, USA). The task consisted of 600 trials in total, comprising 51 blocks of four visual trials, 51 blocks of four auditory trials and 48 blocks of four audiovisual trials. The experiment started with two blocks of visual practice trials and two blocks of auditory practice trials (always in the order of visual, auditory, visual and auditory) (Fig. 1A). These practice trials were introduced for participants to become familiar with the task, for example to respond within the maximum response time window of

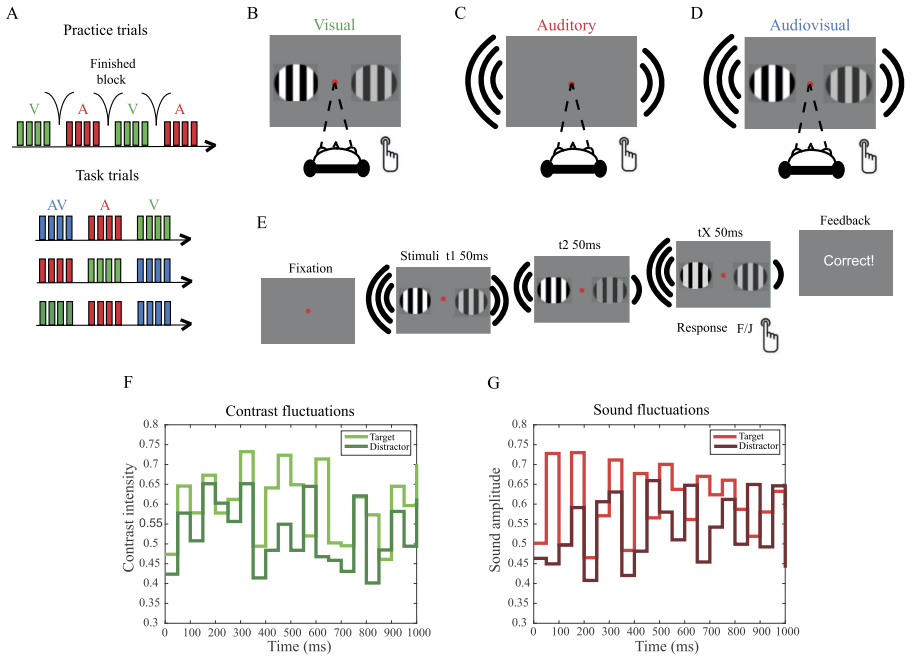

Figure 1. Design of the experimental task. (A) After practice blocks of visual and auditory trials, the task trials commenced, which also included audiovisual trials. (B) On visual trials, participants were requested to indicate whether the left or the right visual grating had the highest contrast. (C) During auditory trials participants had to indicate whether the left or right sound was louder. (D) On audiovisual trials, participants had to indicate on which side the sound was louder and the contrast higher. (E) Visual contrast and loudness randomly fluctuated every 48.6 ms around an average value which was fixed for the target stimuli, but differed for distractor values between the easy, intermediate and hard trials. Participants were required to answer as soon as they perceived the target side and received feedback regarding their answer after each trial. (F) Example of the contrast fluctuations of the visual target (bright green) and distractor (dark green) on an intermediate trial. (G) Example of the volume fluctuations of the auditory target (bright red) and distractor (dark red) on an intermediate trial.

2 s. After the practice blocks, visual, auditory and audiovisual blocks were presented in a random order.

Stimuli were presented on a 17.3 inch MSI Bravo gaming laptop with a screen refresh rate of 144 Hz which was gamma-corrected. Participants were at a viewing distance of roughly 60 cm from the centre of the screen. During visual trials, two Gabor gratings with different contrast intensities were shown on the screen on a grey background as depicted in Fig. 1B. The centres of the patches were 10 degrees to the left and right from the centre of the screen and had a random orientation each trial albeit the same for both. Participants were instructed to press the 'f' key if the stimulus with the highest contrast (i.e., the visual target) was presented on the left side of the screen and the stimulus with the lowest contrast (i.e., the visual distractor) on the right side. They had to press the 'j' key if the visual target was presented on the right side of the screen and the visual distractor on the left side. After each key press participants were provided with response feedback, a grey screen with the text 'correct!' or 'incorrect!' depending on the outcome (Fig. 1E). During auditory trials, two pink noise bursts starting at 7 kHz and linearly decaying to 32 kHz were presented to the left and right ear using headphones with a sampling rate of 44.1 kHz (MD-5000DR, IMG stageline) (Fig. 1C). Participants were asked to indicate at which side the auditory target was, i.e., where the sound was highest in amplitude (loudness), by pressing the same corresponding keyboard keys as during visual trials. During audiovisual trials the auditory and visual stimuli were presented simultaneously and the visual target and auditory target were always on the same side (Fig. 1D, E).

Trials could be of three difficulty levels: easy, intermediate and hard. The difficulty level was determined by changing the baseline intensity of the distractor stimulus while the baseline intensity of the target stimulus was always 60%. We used a one-down-two-up staircase procedure on intermediate unimodal visual and auditory trials to determine the visual and auditory distractor baseline intensities for these levels (García-Pérez, 1998). We designed the procedure in such a way as to obtain an accuracy of $\pm$71% on intermediate audiovisual trials. The staircase procedure started after the practice trials and continued until the end of the experiment. For the first two trials of each modality, the intensity level of the distractor stimulus was set at 80% of that of the target. From the third trial on, the intensity decreased with 1% if one unimodal intermediate trial was incorrect. This way, the difference between the target and distractor intensity (i.e., the evidence) was increased. If two consecutive unimodal intermediate trials were correct, the intensity increased with 1% to decrease the evidence. On easy trials, the distractor baseline intensity was decreased by multiplying the difference between the target and distractor on the previous intermediate trials with a factor of 1.5 and on hard trials the distractor baseline intensity was increased by multiplying the difference

with a factor of 0.5. This resulted in an average of 63.0 ± 4.0% accuracy on intermediate visual trials, an average of 62.3 ± 3.4% accuracy on intermediate auditory trials and an average of 71.2 ± 8.6% accuracy on intermediate audiovisual trials.

The contrast intensities of the visual target and distractor and the amplitude of the auditory target and distractor fluctuated every 48.6 ms (seven frames on a 144-Hz monitor) over a baseline value (Fig. 1F, G). These time-varying stimulus intensities were included to retrospectively test which moments in time significantly contributed to the decisions. To avoid clicks when the sound amplitude increased or decreased during these fluctuations, the sound level gradually approached the intensity of the next fluctuation over the last 10 ms period. The change in visual contrast was abrupt. The fluctuation range was 14% from baseline intensity resulting in fluctuation intensities between 46% and 74% for the target stimulus. Depending on the performance accuracy of the participant and the difficulty level, the values of target and distractor across the fluctuations could be very close to each other. For some participants, the fluctuations could cause the evidence for the distractor location to be stronger than evidence for the target location at random points in time (Fig. 1F, G). This ensured that participants had to evaluate evidence and integrate it over time in order to make a correct decision. The baseline intensity of the target, however, was always higher than that of the distractor and therefore the majority of the fluctuations were higher on the side of the target. Moreover, the fluctuation onsets of the target and the distractor stimuli were simultaneous but the fluctuation intensities were calculated randomly (i.e., they were asynchronous).

We aimed to identify sensory weighting profiles during perceptual decision-making in which decision times were under the control of the participants. The task script, however, did not allow for the stimulus to discontinue when the participants had responded. To circumvent this issue, the trial length was determined using the reaction times (RTs) of the participants. Early in the task, from the fifth trial up to the 20th, the average RTs of all of the previous trials was calculated. These trials could consist of trials from all modalities. Later in the task, after trial 20, the RTs of the 20 most recent trials were averaged. Subsequently, 65% was added to the average RT value with a lower limit of 500 ms and an upper limit of 2 s to obtain the stimulus length for each trial. Using this method, the trial length felt natural.

2.1.3. Data Analysis

Data of participants who experienced technical issues during the task or performed with an overall accuracy below 60% were removed from further analyses. This entailed excluding the data of three participants due to technical issues and of 11 participants who performed around chance level. The analyses were thus performed on the data of 64 participants.

A 3 (Modality: Visual, Auditory and Audiovisual) × 3 (Difficulty: Easy, Intermediate, Hard) repeated-measures analysis of variance (rmANOVA) on the median RTs on correct trials was done to test how Modality and Difficulty influenced the RTs. We performed *post-hoc* tests using the Holm method for multiple-comparison corrections. We carried out a generalized linear mixed-effects model (GLM) to examine how Modality and Difficulty affected performance accuracy. Decision outcomes (correct/incorrect) of all participants on all trials were used to model the the probability to make a correct decision with predictor variables Modality and Difficulty. The following equation describes the model:

$$Y = \left[1 + \exp\left\{-(\beta_0 + \beta_1 * \mathrm{Mod_{ij}} + \beta_2 * \mathit{Diff}_{ij} + \beta_3 * \mathrm{Mod}_{ij} * \mathit{Diff}_{ij} + b_j)\right\}\right]^{-1} \quad (1)$$

where Y is the response accuracy of the trial (i.e., correct or incorrect), β_0 is the intercept term and β_1 reflects the modality and β_2 is the difficulty on trial i for participant j. b_j, is a random-effects term comprising an intercept and slope for each participant j that accounts for potential participant-specific variation in task performance. We performed an ANOVA on the GLM model fit to determine if the coefficients of each fixed effect and interaction effect in these GLM models are equal to 0. This represents the main effect per fixed effect and interaction effect. For *post-hoc* comparisons of the significant main effect and interaction effect we performed an F-test to test the coefficients for similarity per level comparison (i.e., contrast) using the Wald test.

To investigate the dynamics of sensory integration and at which moments in time auditory and visual information contributed to the decision, we performed different GLMs for visual, auditory and audiovisual trials (MATLAB function *fitglme* with binomial distribution and logit link). As our task entailed discriminating the target from the distractor stimulus where the baseline as well as the fluctuation intensities of the two stimuli could be in close proximity (Fig. 1F, G), the task could only be solved by evaluating the stimulus intensities relative to one another. We thus calculated the difference between the target and distractor intensity for every fluctuation sample and standardized these values by z-scoring per time point and per trial. We used these standardized visual and auditory evidence values to subsequently model the probability to make a correct decision with predictor variables separately for the evidence of the visual stimulus (V_{ev}) and the auditory stimulus (A_{ev}). We only included the evidence samples before the participant responded by taking the RT on each trial as a cut-off point. For each evidence sample with a time period t of 48.6 ms we used the following equation for visual trials:

$$Y = \left[1 + \exp\left\{-(\beta_0 + \beta_{1,t} * V_{\mathrm{ev}_{ij,t}} + b_{j,t})\right\}\right]^{-1} \quad (2)$$

for auditory trials:

$$Y = \left[1 + \exp\left\{-(\beta_0 + \beta_{1,t} * A_{\mathrm{ev}_{ij,t}} + b_{j,t})\right\}\right]^{-1} \quad (3)$$

and for audiovisual trials:

$$Y = \left[1 + \exp\left\{-(\beta_0 + \beta_{1,t} * V_{\mathrm{ev}_{ij,t}} + \beta_{2,t} * A_{\mathrm{ev}_{ij,t}} + b_{j,t})\right\}\right]^{-1} \quad (4)$$

We performed the Wald test (MATLAB function *waldtest*) to test whether the fixed effects in the models were significant. We implemented the false discovery rate (FDR) method (Benjamini and Hochberg, 1995) for multiple-comparison correction of the p values (MATLAB function *fdr_hb*) obtained with the Wald tests. To test whether auditory and visual evidence contributed to audiovisual decisions with similar weights over time, we tested the coefficients for similarity per time point (*coefTest*). We used the outcome of this test as a proxy for sensory dominance during evidence weighting. We repeated these GLMs and *post-hoc* analyses using the 25% fastest and 25% slowest trials separately to detect reaction time-dependent weighting profiles.

To investigate whether modality dominance in the audiovisual weighting profiles was dependent on difficulty, we performed an extra set of GLMs. We calculated the difference between the visual and auditory evidence to obtain predictor δAV, a term for modality dominance. This difference value was used as we found a significantly different contribution of visual evidence compared to auditory evidence and vice versa on some time points on audiovisual trials which we proposed to be a proxy for modality dominance. We modelled the probability to make a correct decision with predictor variables δAV, the difficulty level (*Diff*) and an interaction effect between the two (δAV*$Diff$). We performed these GLMs using the time points where we observed visual and/or auditory dominance on the stimulus and response-aligned profile with trials of all difficulties pooled. The equation for these GLMs:

$$\begin{aligned} Y = \Big[1 + \exp\Big\{-(&\beta_0 + \beta_{1,t} * \delta AV_{ij,t} + \beta_{2,t} * Diff_{ij,t} \\ &+ \beta_{3,t} * \delta AV_{ij,t} * Diff_{ij,t} + b_{j,t})\Big\}\Big]^{-1} \end{aligned} \quad (5)$$

As a last step to explicitly test the dependence of auditory dominance on difficulty, we performed an ANOVA on the GLM model fit to determine if the coefficients of the interaction effect in these GLM models were equal to 0.

2.1.4. Code Accessibility

All data and all codes used for testing and the analyses in the current study can be accessed on OSF (https://osf.io/qznyu/files/).

3. Results

3.1. Performance Is Enhanced on Multisensory Trials

Our first aim was to assess whether audiovisual stimuli elicited multisensory benefits in our task. A GLM was carried out to model the decision outcome (correct/incorrect) as the linear combination of fixed effects Modality and Difficulty, an interaction term between the two and a random-effects term to correct for any potential participant-specific effects on the performance. The GLM revealed a main effect of Modality, $F_{2,38\,391} = 43.994$, $p < 0.001$, of Difficulty, $F_{2,38\,391} = 193.49$, $p < 0.001$, and a significant interaction effect, $F_{4,38\,391} = 8.5982$, $p < 0.001$. The effects of Modality and Difficulty on performance accuracy are illustrated in Fig. 2A. *Post-hoc* tests to investigate the main effects of Modality and Difficulty showed that performance was increased on audiovisual trials compared to auditory trials, $F_{1,38\,391} = 75.630$, $p < 0.001$, and compared to visual trials, $F_{1,38\,391} = 39.397$, $p < 0.001$. Furthermore, the performance on visual trials was better than on auditory trials, $F_{1,38\,391} = 4.486$, $p < 0.05$. As expected by our experimental design, performance accuracy was better on easy trials compared to intermediate trials, $F_{1,38\,391} = 18.045$, $p < 0.001$, and hard trials, $F_{1,38\,391} = 39.1$, $p < 0.001$, as well as on intermediate trials *vs* hard trials, $F_{1,38\,391} = 11.567$, $p < 0.001$.

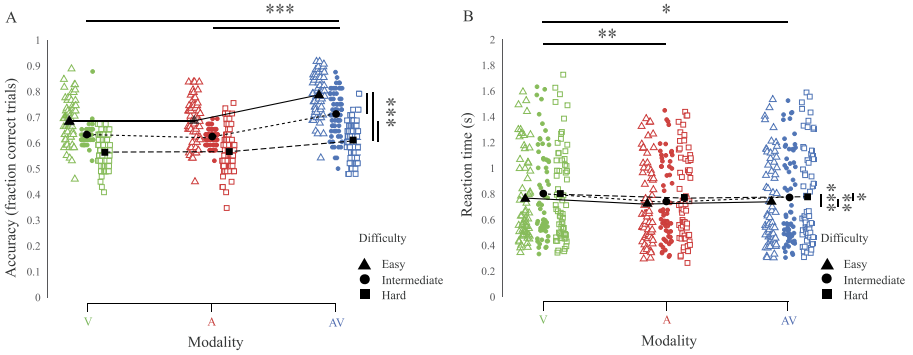

Figure 2. Performance is enhanced on audiovisual trials. (A) The accuracies of each participant and the averages for visual trials (green), auditory trials (red) and audiovisual trials (blue) for the three difficulty levels (easy: triangle, intermediate: circle, hard: square). Performance is significantly higher on audiovisual trials compared to visual trials and auditory trials and significantly different between the three difficulty levels. (B) The median correct reaction times of each participant and the average median reaction times for visual trials (green), auditory trials (red) and audiovisual trials (blue) for the three difficulty levels (easy: triangle, intermediate: circle, hard: square). Reaction times are significantly larger for visual trials compared to auditory and audiovisual trials. On easy trials reaction times are largest and on hard trials smallest. Significance is indicated by solid lines above and next to the plots: *, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$. Error bars denote the 95% confidence interval.

Post-hoc tests on the interaction effect revealed that at all difficulty levels, performance was better on audiovisual compared to visual and auditory trials (Table 1).

A 3 (Modality: Auditory, Visual and Audiovisual) × 3 (Difficulty: Easy, Intermediate and Hard) rmANOVA was carried out on the median RTs on correct trials of each participant. Sphericity was violated ($\varepsilon = 0.784$ for Modality, $\varepsilon = 0.777$ for Difficulty) and therefore Huyn–Feldt-corrected results are reported for the effects of Modality and Difficulty. This analysis revealed a main effect of Modality, $F_{1.6,100.816} = 6.257$, $p = 0.005$, a main effect of Difficulty, $_{1.586,99.891} = 15.054$, $p < 0.001$, but no interaction between the two factors. The effects of Modality and Difficulty on RTs are illustrated in Fig. 2B. *Post-hoc* tests showed that RTs were larger on visual trials compared to auditory trials, $t_{63} = 3.489$, $p < 0.01$, and compared to audiovisual trials $t_{63} = 2.249$, $p = 0.05$. RTs were smaller on easy trials compared to intermediate trials, $t_{63} = -3.095$, $p < 0.01$, and compared to hard trials, $t_{63} = -5.471$, $p < 0.001$. Furthermore, RTs on intermediate trials were smaller compared to hard trials, $t_{63} = -2.377$, $p < 0.05$.

3.2. Early-Weighting Profiles and Sequential Modality Dominance on Audiovisual Trials

The temporal dynamics of evidence integration during visual, auditory and audiovisual decision-making were determined by performing different GLMs. This provided us with beta coefficients of the target and distractor differences, i.e., the evidence, per fluctuation time point. The coefficient value reflects the weight of the evidence, which we plotted against time in Figs 3–6. We plotted the coefficient values up until the time point at which 75% of trials were finished. This was done as the number of trials that were solved at the end of the trial period of 2 s was low and a low number of data points might lead to noisy uninterpretable results. The significant coefficients ($p < 0.05$, FDR-corrected) that contribute significantly to the decision are marked with a coloured asterisk in the plots. We will use the term 'contribute' when we refer to coefficients that increase the probability of making a correct choice. We were also interested in the contribution of visual and auditory evidence on audiovisual trials relative to each other to reveal potential differential integration strategies based on modality dominance. Significant differences between visual and auditory coefficients weights ($p < 0.05$, FDR-corrected) are indicated with black asterisks. We term the significant differences as 'modality dominance'.

Relative to the stimulus onset, weighting profiles are apparent where early evidence contributes most to the decisions on auditory (W-stat range auditory evidence: 7.938–185.472) (Fig. 3A), visual (W- stat range visual evidence: 25.699–150.683) (Fig. 3B) and audiovisual (W-stat auditory evidence: 45.426, W-stat range visual evidence: 11.610–78.084) (Fig. 3C) trials. Focusing on

Table 1.
The results of the *post-hoc* tests for the effects of modality and difficulty on accuracy. Depicted are the modality comparisons per difficulty level, the degrees of freedom (df), the t statistic (t) and the p values (p). The test results show that multisensory improvement is present on each difficulty level

	Difficulty		
Contrast	Easy	Intermediate	Hard
V *vs* AV	$F_{1,38\,391} = 8.519$, $p < 0.005$	$F_{1,38\,391} = 9.737$, $p < 0.005$	$F_{1,38\,391} = 18{,}871$, $p < 0.001$
A *vs* AV	$F_{1,38\,391} = 52.687$, $p < 0.001$	$F_{1,38\,391} = 19{,}975$, $p < 0.001$	$F_{1,38\,391} = 26.662$, $p < 0.001$

A, auditory; AV, audio-visual; V, visual.

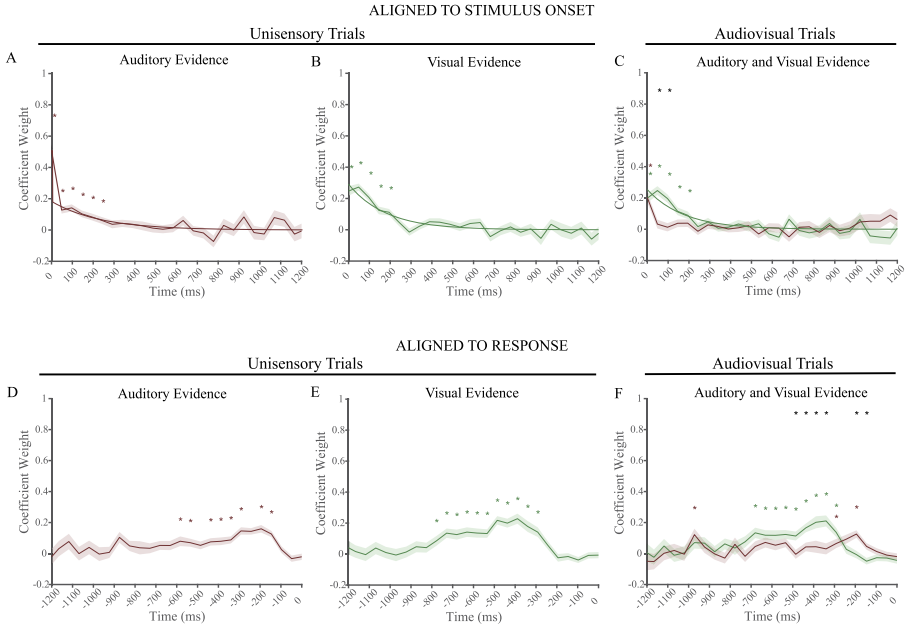

Figure 3. Visual and auditory evidence accumulation over all difficulty levels. Weighting profiles A–C are aligned to the stimulus onset and include an exponential fit, D–F are aligned to the response. (A) Early auditory evidence contributes to auditory decisions followed by some samples later in time ($n = 11\,840$ trials) (red asterisks and red exponential fit line). (B) Early visual evidence contributes to visual decisions ($n = 12\,278$ trials) (green asterisks and green exponential fit line). (C) The first auditory and visual sample contribute equally to audiovisual decisions ($n = 12\,117$ trials) (red and green asterisks respectively). During a subsequent short period, until 250 ms, visual evidence follows an early weighting profile (green exponential fit line) and is weighted significantly higher than auditory evidence (black asterisks). (D) Auditory evidence contributes to auditory decisions until 100 ms before the response (red asterisks). (E) Visual evidence contributes to visual decisions until 250 ms before the response (green asterisks). (F) On audiovisual trials, visual integration is largely dominant over auditory integration until 250 ms before the response after which auditory integration is dominant until the coefficient weights reach 0 around 100 ms before the response (black asterisks).

audiovisual trials, we observe that only the first sound sample (i.e., the sound onset) contributes significantly to the decision after which visual evidence is weighted exclusively. The coefficients of visual evidence on timepoints 50–150 ms were significantly higher than the auditory-evidence coefficients (F-stat range: 29.708–31.765) revealing visual dominance during this period (Fig. 3C, black asterisks). We quantified the shape of the weighting profile by fitting an exponential distribution to the coefficients of auditory evidence on auditory trials, visual evidence on visual trials and visual evidence on audiovisual trials. To capture the two different decay periods on auditory trials — the fast decay from the first to the second sample and the slower decay from

the contribution of the second time point onwards — we included a two-term exponential equation to model the unisensory auditory coefficients. Only the first time point of auditory evidence contributes to audiovisual decisions and therefore an early-weighting profile of auditory evidence on audiovisual trials is indisputable. Visual inspection of the fits and the model statistics suggest that exponential models mimic the decrease of the contribution of sensory evidence in our task (auditory weights on auditory trials: $R^2 = 0.89$, summed square of residuals (SSE): 0.03; visual weights on visual trials: $R^2 = 0.85$, SSE: 0.03; visual weights on audiovisual trials: $R^2 = 0.77$, SSE: 0.04).

Calculating the weights relative to the stimulus onset provides an accurate estimation of the contribution of the first samples. However, the contribution of the later samples close to the RTs are potentially underestimated as the RTs vary between trials and participants. Therefore, we sought to investigate whether evidence close to the response is weighted less than evidence right after stimulus onset as suggested by the results of our stimulus-locked analysis. Figure 3D–F show the weighting profiles relative to the response on auditory (W-stat range auditory evidence: 0.534–49.410) (Fig. 3D), visual (W-stat range visual evidence: 7.440–64.891) (Fig. 3E) and audiovisual (W-stat range auditory evidence: 7.998–34.194, W-stat range visual evidence: 12.972–35.154) (Fig. 3F) trials. The up-ramping weights reflect the spread and underestimation of the weight of the first sample as an expected consequence of this alignment. On auditory trials participants weight auditory evidence up until 100 ms before the response (Fig. 3D) while on visual trials this occurs until 250 ms (Fig. 3E). Interestingly, on audiovisual trials we observe a significantly dominant visual-weighting period from 450 up until 250 ms before the response (F-stat range: 8.851–17.208) followed by a significantly dominant auditory weighting period from 200 until 100 ms before the response (Fig. 3F) (F-stat: 9.61–17.076). The samples right before the responses do not contribute to the decisions (Fig. 3D–F).

Overall, we see that throughout the trial evidence integration is characterized by an exponential decrease in contribution of evidence and that evidence close to the response does not contribute to the decision. These findings argue for an overall early-weighting profile. Additionally the results we show here indicate that integration on audiovisual trials can be split into three epochs. The first epoch is characterized by the weighting of the first visual evidence sample as well as the first auditory-evidence sample which is most obvious when the coefficient weights are aligned to the stimulus onset. Aligned to the response, we observe that during the second epoch participants rely on visual information only until 300–250 ms before the response. Finally, auditory information is weighted during the third epoch until the coefficient weights reach zero around 100 ms before the response.

To control for the potential confounding influence of fast and slow RTs we performed the same analyses including the 25% fastest or the 25% slowest RTs. In general, we observe less significant time points on fast weighting profiles (Supplementary Fig. S1) compared to all trials pooled (Fig. 3), which could be a consequence of shorter evidence accumulation periods that require the integration of less evidence samples. Nevertheless, fast weighing profiles both aligned to the stimulus and response were quite similar during auditory (W-stat range auditory evidence aligned to stimulus: 27.329–115.434; aligned to response: 0.212–61.664), visual (W-stat range visual evidence aligned to stimulus: 13.099–151.332; aligned to response: 5.264–11.764) and audiovisual decision-making (W-stat auditory evidence aligned to stimulus: 35.318; aligned to response: 6.093–14.094; W-stat range visual evidence aligned to stimulus: 34.918–36.816; aligned to response: 11.365–62.865) (Supplementary Fig. S1A–F) compared to when we pooled all trials. Fast trials included the three epoch weighing profiles on audiovisual trials (F-stat visual dominance aligned to stimulus onset: 16.153; visual dominance aligned to response: 12.665–29.821; auditory dominance aligned to response: 5.985) (Supplementary Fig. S1C, F).

On slow auditory (W-stat auditory evidence aligned to stimulus: 18.897; aligned to response: none), visual (W-stat range visual evidence aligned to stimulus: none; aligned to response: 9.473–16.411) and audiovisual trials (W-stat auditory evidence aligned to stimulus: none; W-stat auditory evidence aligned to response: none; W-stat range visual evidence aligned to stimulus: none; W-stat visual evidence aligned to response: 19.034; W-stat visual dominance: 10.680) (Supplementary Fig. S2) we observe less time points or no time points that contribute to the decision compared to all trials pooled. When we focus on the shape of the weighting profiles, however, we find similarities compared to when we pool all trials. For example, the profiles for auditory integration were similar on auditory trials for both alignment methods (Supplementary Fig. S2A, D) compared to all trials pooled (Fig. 3A, D) even though no moments in time significantly contribute to the decision on the response-aligned profile (Supplementary Fig. S2D). On visual trials (Supplementary Fig. S2B, E), no time points significantly contribute to the decision relative to the stimulus while on the response-aligned profile a significant integration period is observed. There were likewise no significant early time points on stimulus-aligned weighting profiles on audiovisual trials (Supplementary Fig. S2C). However, the first 300 ms of the weighting profile are similar to the first part of the weighting profiles of the fast trials (Supplementary Fig. S1C) and all trials pooled (Fig. 3C). To elaborate, the first auditory and visual samples appear to contribute with an equal weight, after which the auditory weight decreases on time point two. The response-aligned profile on audiovisual trials (Supplementary Fig. S2F) has a similar shape as the all-trials-pooled profile

(Fig. 3F) but the auditory evidence peak right before the response does not significantly contribute to the decision. A potential reason for the decrease and/or lack of significance on the weighting profiles of slow trials could be that the multiple-comparison test using more time points for slow trials is more strict compared to multiple-comparison testing of less time points for fast trials and for all audiovisual trials pooled.

3.3. Audiovisual Integration Strategies per by Difficulty

Next, we tested how sensory integration was affected by difficulty. For all difficulties we observed overall early-weighting profiles during auditory (W-stat range auditory evidence easy trials: 12.175–40.515; intermediate trials: 0.0003–113.498; hard trials: 8.008–119.521), visual (W-stat range visual evidence easy trials: 5.927–59.155; intermediate trials: 8.747–41.722; hard trials: 8.435–56.107) and audiovisual decision-making (W-stat auditory evidence easy trials: 24.630; intermediate trials: 21.229; hard trials: 19.095; W-stat range visual evidence easy trials: 12.157–40.515; intermediate trials: 17.619–34.469; hard trials: 14.765–27.908) (Figs. 4–6). The profiles aligned to the stimulus onset of each modality condition was similar for easy (Fig. 4A–C) intermediate (Fig. 5A–C) and hard (Fig. 6A–C) trials, and similar compared to the profiles we observed for trials of all difficulties pooled (Fig. 3A–C). On audiovisual trials, for example, only the first auditory-evidence sample contributed to the audiovisual decision after which visual evidence continued to contribute for a short period. This period of visual dominance was significant for easy and intermediate trials (F-stat range easy trials: 17.746–18.069; intermediate trials: 12.525–16.703; hard trials: none).

We fitted exponential models to coefficients of each difficulty separately to inspect the shape of the weighting profiles on the different difficulty levels. All exponential models capture the coefficient decrease (auditory weights on easy auditory trials: $R^2 = 0.71$, SSE: 0.07; auditory weights on intermediate auditory trials: $R^2 = 0.60$, SSE: 0.15; auditory weights on hard auditory trials: $R^2 = 0.79$, SSE: 0.09; visual weights on easy visual trials: $R^2 = 0.70$, SSE = 0.05; visual weights on intermediate visual trials: $R^2 = 0.74$, SSE = 0.05; visual weights on hard visual trials: $R^2 = 0.81$, SSE = 0.06; visual weights on easy audiovisual trials: $R^2 = 0.56$, SSE = 0.1; visual weights on intermediate audiovisual trials: $R^2 = 0.55$, SSE = 0.99; visual weights on hard audiovisual trials: $R^2 = 0.57$, SSE = 1.92) indicating early weighting profiles.

Relative to the response, auditory weighting profiles and visual weighting profiles were also similar to the profiles we observed when we pooled trials of all difficulties and had the same characteristics on easy (Fig. 4D–F), intermediate (Fig. 5D–F) and hard (Fig. 6D–F) trials for auditory (W-stat range auditory

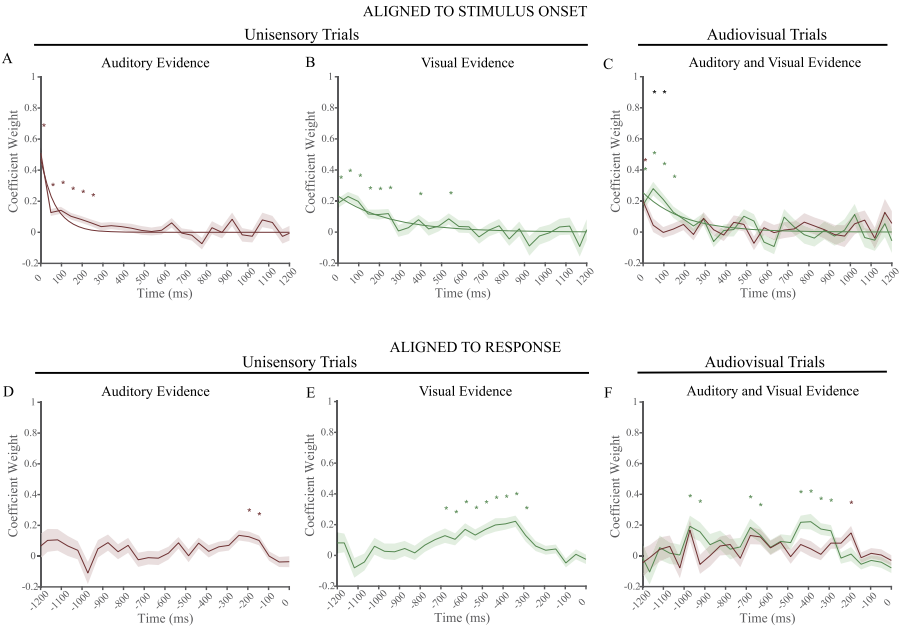

Figure 4. Visual and auditory evidence accumulation on easy trials. Weighting profiles A–C are aligned to the stimulus onset and include an exponential fit, D–F are aligned to the response (A) Early auditory evidence contributes to auditory decisions followed by some samples later in time ($n = 5925$ trials) (red asterisks and red exponential fit line). (B) Early visual evidence contributes to visual decisions ($n = 6149$ trials) (green asterisks and green exponential fit line). (C) The first auditory and visual sample contribute equally to audiovisual decisions ($n = 6061$ trials) (red and green asterisks respectively). During a subsequent short period of around 100 ms visual evidence follows an early weighting profile (green exponential fit line) and contributes significantly more to the decision than auditory evidence (black asterisks). (D) For 100 ms, auditory evidence contributes to auditory decisions until 100 ms before the response (red asterisks). (E) Visual evidence contributes to visual decisions until 250 ms before the response (green asterisks). (F) On audiovisual trials, visual evidence is integrated until 250 ms before the response after which auditory evidence solely contributes up until the coefficient weights reach 0 around 150 ms before the response.

evidence easy trials: 2.122–14.282; intermediate trials: 1.637–12.891; hard trials: 0.753, visual (W-stat range visual evidence easy trials: 7.218–36.491; intermediate trials: 5.780–27.094; hard trials: 7.492–34.835) and audiovisual decision-making (W-stat range auditory evidence easy trials: 18.551; intermediate trials: none; hard trials: 15.671; W-stat range visual evidence easy trials: 6.472–25.705; intermediate trials: 8.161–16.752; hard trials: 5.892–49.444). Visual dominance was observed on intermediate and hard trials (F-stats range intermediate trials: 7.421–8.388; hard trials: 6.959–9.151). The late auditory coefficients were significantly different from visual coefficients on the

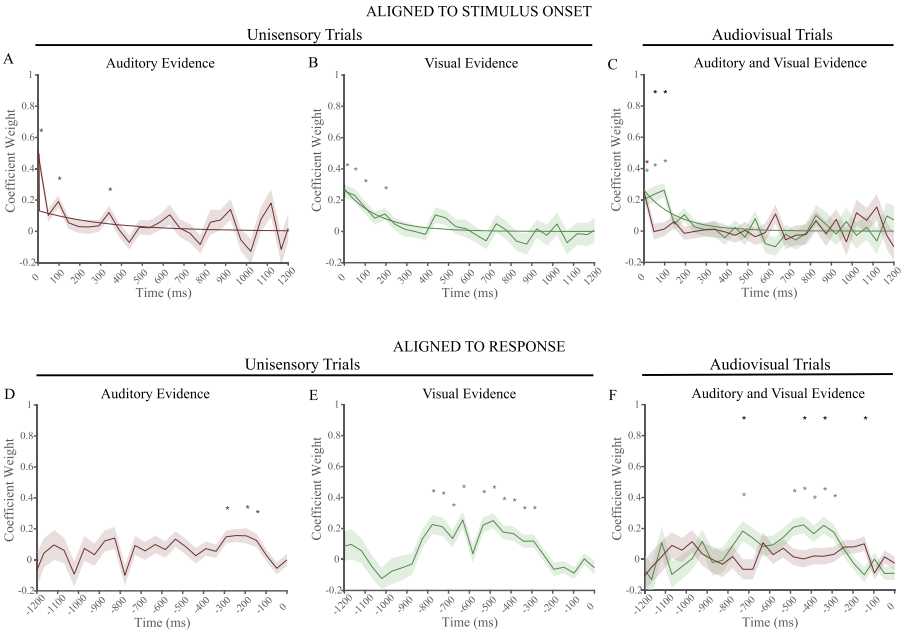

Figure 5. Visual and auditory evidence accumulation on intermediate trials weighting profiles. Weighting profiles A–C are aligned to the stimulus onset and include an exponential fit, D–F are aligned to the response. (A) Early auditory evidence contributes to auditory decisions followed by some samples later in time ($n = 2969$ trials) (red asterisks and red exponential fit line). (B) Early visual evidence contributes to visual decisions ($n = 3055$ trials) (green asterisks and green exponential fit line). (C) The first auditory and visual sample contribute equally to audiovisual decisions indicated ($n = 3026$ trials) (red and green asterisks respectively). During a subsequent short period of around 100 ms visual evidence follows an early weighting profile (green exponential fit line) and contributes significantly more to the decision than auditory evidence (black asterisks). (D) Auditory evidence contributes to auditory decisions until 100 ms before the response (red asterisks). (E) Visual evidence contributes to visual decisions until 350 ms before the response (green asterisks). (F) On audiovisual trials, visual integration is dominant over auditory integration for a discontinuous period after which auditory integration dominates (black asterisks).

intermediate and hard trials (F-stats intermediate trials: 9.809; hard trials: 7.709–8.766).

The modality dominance periods were absent on some stimulus- and response-aligned profiles of individual difficulty levels. To test whether the difficulty level significantly influenced modality dominance, we performed additional GLMs. The first 250 ms of evidence samples relative to the stimulus were included for testing the first visual dominance period. The time points encompassing the last 800 ms were further inspected for the visual dominance and auditory dominance observed on the response-aligned profiles. We took the difference between the visual and auditory evidence and difficulty level

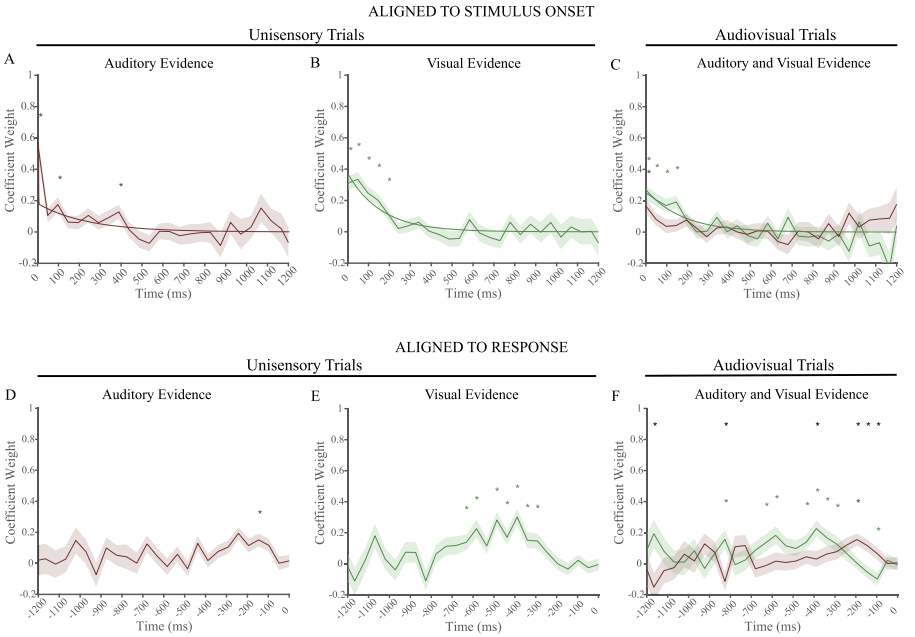

Figure 6. Visual and auditory evidence accumulation on hard trials. Weighting profiles A–C are aligned to the stimulus onset and include an exponential fit, D–F are aligned to the response. (A) Early auditory evidence contributes to auditory decisions followed by one sample later in time ($n = 2946$ trials) (red asterisks and red exponential fit line). (B) Early visual evidence contributes to visual decisions ($n = 3074$ trials) (green asterisks and green exponential fit line). (C) The first auditory and visual sample contribute equally to audiovisual decisions indicated ($n = 3030$ trials) (red and green asterisks respectively). During a subsequent short discontinuous period of around 100 ms visual evidence follows an early weighting profile (green exponential fit line) and contributes significantly more to the decision than auditory evidence (black asterisks). (D) Auditory evidence contributes to auditory decisions around 100 ms before the response (red asterisks). (E) Visual evidence contributes to visual decisions until 250 ms before the response (green asterisks). (F) On audiovisual trials, visual integration is dominant for some time points followed by auditory dominance and the negative contribution of one single visual evidence sample (black asterisks).

as predictors as well as an interaction term. We found no significant interaction effect between the evidence difference and difficulty level for the time points for the stimulus-aligned and response-aligned GLMs. The differences we observe in the weighting profiles of individual difficulty level compared to all difficulties pooled could potentially be explained by the lower number of data points when splitting the data into the different difficulty levels. These findings strongly suggest that audiovisual decision-making strategies are not transformed by difficulty level.

4. Discussion

This study investigated how time-varying visual and auditory evidence are weighted over time and used to discriminate between sensory events in the environment. Our analyses revealed three main characteristics in the temporal weighting profiles. First, early sensory information contributes strongest to decisions across all modalities and difficulties. Second, the processing of auditory information is characterized by a highly contributing stimulus onset. Third, improved audiovisual decision-making is associated with sequential modality dominance during which early visual and auditory evidence equally contribute to the decision followed by visual and auditory dominance switching.

The generalized linear mixed models revealed overall early-weighting profiles where early sensory information contributes most heavily to decision-making both when unisensory and when multisensory information is available. The overall early-weighting profiles are especially apparent when the coefficient weights were aligned to the stimulus onset which is reflected by the large coefficient weights of the first evidence samples and the exponentially decreasing weights of the later samples (Fig. 3A–C, Fig. 4A–C, Fig. 5A–C, Fig 6A–C). This conclusion is supported by the lack of a contribution of late evidence that is observed when aligning the weights to the response (Fig. 3D–F, Fig. 4D–F, Fig. 5D–F, Fig 6D–F). It should be stressed that even though the weighting profiles point to the weighting of early evidence, later evidence samples up until 100 ms still contribute to the decisions in some response-aligned weighting profiles. It has been demonstrated that different evidence integration strategies can be explained by specific differences in behavioural paradigms. Among important features that influence weighting profiles are the division of evidence during a trial (Bronfman *et al*., 2016; Levi *et al*., 2018; Raposo *et al*., 2012) and choice expectation (Booras *et al*., 2021; Talluri *et al*., 2021). When stimulus information is equally informative throughout the trial (Kiani *et al*., 2008; Levi *et al*., 2018) and the observer is able to report the decision at any time, like here, there is no need to integrate late information after the decision has been made. When stimulus durations are extended (Bronfman *et al*., 2016) or only late evidence is informative (Talluri *et al*., 2021; Levi *et al*., 2018) late-weighting profiles are observed. In this situation, early information would not reflect the state of the world relevant for the choice. Moreover, we show that RT can influence the weighting profiles. On slow trials, weighting profiles were less pronounced with less or no significant time points contributing to the decisions. Slow RTs have been associated with low task engagement (Qiao *et al*., 2018). This implies that internal features such as motivation and engagement could additionally shape sensory weighting.

Previous studies that have regressed decision outcomes to stimulus features in the framework of signal detection theory (SDT) have speculated how weighting profiles relate to sensory integration mechanisms (Gold and Shadlen, 2007; Kiani *et al*., 2008; Levi and Huk, 2020; Okazawa *et al*., 2018). According to this theory, weighting profiles of evidence accumulation most likely embody a combination of sensory integration and decision-making processes. The early-weighting profile could reflect a bounded accumulation process during which information is integrated until a decision bound has been reached and a decision has been made. Early information contributes heavily to the sensory integration process and the information that appears after the bound has been reached does not. In contrast, late-weighting profiles resemble leaky accumulation processes that are supported by a neural circuit that 'leaks' or 'forgets' information during the integration process. It should be noted that a one-on-one comparison of weighting profiles and decision-related processes might not be without error, as it can underestimate or ignore factors such as sensory weights, termination criterion of the decision, and the non-decision time (Okazawa *et al*., 2018).

A closer examination of the timescale of auditory evidence weighting revealed a particularly large weight of the first auditory evidence sample and a sharp decrease to the weight of the second sample. This salient onset effect was not as evident for visual-weighting profiles. One explanation could be that the auditory stimulus appeared in a silent background (i.e., a headphone covering the ears) compared to the visual stimulus which appeared in a background of light (i.e., a grey computer screen in a lit room). Salient auditory changes in the background can capture the exogenous attention of the observer (Huang and Elhilali, 2020). This attentional capture might cause an increase in the synchronization between neural populations which modulates the neural representation of a target stimulus (Elhilali *et al*., 2009). Moreover, louder sounds increase neural gain by evoking higher gamma band responses compared to sounds of lower intensities (Schadow *et al*., 2007). On intermediate and hard trials the distractor stimulus was louder than on easier trials, resulting in an overall higher volume level on intermediate and hard trials. The stimulus-driven attentional capture caused by loud(er) onsets of physically closely related stimuli could explain why the weight of the first sample is higher on auditory trials and increases with difficulty. Remarkably, the auditory onset contributed to the 25% slowest auditory decisions while the weighting profiles on slow trials were overall less pronounced, stressing the attentional capture effect of salient sound onsets.

Audiovisual decision-making was characterized by a strategy comprising sequential modality dominance. The onset of the auditory and visual evidence contributed equally to the decision after which participants relied only on visual evidence. These two epochs were followed by a period of auditory

evidence weighting. This points to visual dominance for an extended period during audiovisual decision-making followed by a switch to a short period of auditory dominance right before the response. Visual dominance has been shown in numerous other studies (Bertelson and Radeau, 1981; Pick *et al*., 1969; Welch and Warren, 1980). However, this is the first study that shows that the relative weights of unisensory information streams can change over time. The switching between modalities could be mediated by dynamically changing connection strength within and between early sensory areas and frontal and motor regions as demonstrated by Huang *et al*. (2015). Future studies should test how dominance switching aids sensory integration. For example, introducing a conflicting auditory target right before the response when auditory integration dominates, could reveal how crucial auditory integration is during this stage.

It is striking that although the contribution of auditory information is generally reduced on audiovisual trials compared to auditory trials, the addition of auditory information significantly increased performance accuracy and shortened RTs. The improved accuracy and shorter RTs on the one hand, and overall visual dominance on the other hand, argue against a neural circuit that processes both modalities completely separately where the decision is carried by the fastest modality (the auditory modality) as would be predicted by a race model (Raab, 1962; Townsend and Wenger, 2004, but also see Otto and Mamassian, 2017 for a review on the confusion and interpretation of race models and a new approach to study the redundant signals effect). It is more likely that visual and auditory information are combined during one of the stages of sensory processing and decision-making (Bizley *et al*., 2016; Mercier and Cappe, 2020). The unisensory information streams could be processed separately first in early sensory areas and later be integrated in regions as the superior colliculus (Meredith and Stein, 1983) or in higher-order cortical areas (Holdstock *et al*., 2009; Laurienti *et al*., 2003; Rolls and Baylis, 1994). A different processing scenario would hold that the prominent onset of the auditory stimulus boosts visual processing and that a dominantly visual representation of the target instead of an integrated multisensory estimate emerges during early processing. This is in line with the finding that a loud onset of an auditory stimulus is one of the most dominant features that is relayed from the primary auditory cortex (A1) to the primary visual cortex (V1) (Deneux *et al*., 2019). Additionally, it corroborates with the result that temporally coincident auditory input enhances the excitability of the visual cortex particularly after sound onset by increasing the connectivity between low-level visual and auditory areas (Lewis and Noppeney, 2010; Romei *et al*., 2007, 2009). The visual representation could be refined by late auditory information right before the response in a higher-order region.

Our experimental design relied on audiovisual trials that were asynchronous. This required that the fluctuation direction of the contrast of the visual target and distractor did not mimic that of the volume changes of the auditory target and distractor (Fig. 2F, G). These independent fluctuations allowed us to estimate the individual contribution of uncorrelated visual and auditory evidence. We hypothesized that participants integrate asynchronous fluctuations in a similar manner as synchronous fluctuations; that synchrony is not a confounding factor. To test this hypothesis we recruited additional participants to perform a control experiment (see Supplementary Material) where synchronous and asynchronous trials were presented (Supplementary Fig. S3A, B). We showed that audiovisual decision-making is not influenced by whether the auditory and visual stimuli are modulated synchronously over time (Supplementary Fig. S3C, D). This suggests that unlike temporal and spatial synchrony of unisensory information (Stein *et al*., 1988), modulation synchrony does not affect multisensory processing. The multisensory improvement that we showed in our main experiment and the findings of our control experiment imply that discriminating between events in space and time is more accurate when they are constructed by more than one sensory system even when the fluctuations of sensory information are out of sync. These results both corroborate with and expand on the knowledge of the field of multisensory research that has extensively and elegantly shown how multisensory information improves perceptual decision-making (Aller *et al*., 2015; Battaglia *et al*., 2003; Bizley *et al*., 2016; Li *et al*., 2015; Mercier and Cappe, 2020; Raposo *et al*., 2012; Teder-Sälejärvi *et al*., 2005; Todd, 1912).

To summarize, the results that we have demonstrated here suggest that early visual and auditory evidence is weighted most heavily during perceptual decision-making. Audiovisual decisions are generated by a mechanism that promotes cross-modal interactions while changing the gain of one modality over the other in a fast and dynamic way.

Acknowledgements

R.R.M.T. is supported by an Interdisciplinary Doctoral Agreement grant from the Institute for Interdisciplinary Studies of the University of Amsterdam. This work was further supported by Amsterdam Neuroscience grant CIA-2019-01. We thank Luis De La Cuesta Ferrer for his work on setting up the pilot experiments of this study. We thank Anna van Harmelen for her contribution to the task code and Myrthe Griffioen for gathering the data of the main experiment. We thank David Groppe for providing MATLAB code for the FDR multiple-comparison correction and John Hartman for providing MATLAB code for the Wald test using the estimated marginal (predicted) means from generalized linear mixed-effect models.

Supplementary Material

Supplementary material is available online at:
https://doi.org/10.6084/m9.figshare.21608418

References

Aller, M., Giani, A., Conrad, V., Watanabe, M. and Noppeney, U. (2015). A spatially collocated sound thrusts a flash into awareness, *Front. Integr. Neurosci.* **9**, 16. DOI:10.3389/fnint.2015.00016.

Battaglia, P. W., Jacobs, R. A. and Aslin, R. N. (2003). Bayesian integration of visual and auditory signals for spatial localization, *J. Opt. Soc. Am. A Opt. Image Sci. Vis.* **20**, 1391–1397. DOI:10.1364/josaa.20.001391.

Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing, *J. R. Stat. Soc. B Stat. Methodol.* **57**, 289–300. DOI:10.1111/j.2517-6161.1995.tb02031.x.

Bertelson, P. and Radeau, M. (1981). Cross-modal bias and perceptual fusion with auditory-visual spatial discordance, *Percept. Psychophys.* **29**, 578–584. DOI:10.3758/bf03207374.

Bizley, J. K., Jones, G. P. and Town, S. M. (2016). Where are multisensory signals combined for perceptual decision-making?, *Curr. Opin. Neurobiol.* **40**, 31–37. DOI:10.1016/j.conb.2016.06.003.

Booras, A., Stevenson, T., McCormack, C. N., Rhoads, M. E. and Hanks, T. D. (2021). Change point detection with multiple alternatives reveals parallel evaluation of the same stream of evidence along distinct timescales, *Sci. Rep.* **11**, 13098. DOI:10.1038/s41598-021-92470-y.

Brainard, D. H. (1997). The psychophysics toolbox, *Spat. Vis.* **10**, 433–436.

Bronfman, Z. Z., Brezis, N. and Usher, M. (2016). Non-monotonic temporal-weighting indicates a dynamically modulated evidence-integration mechanism, *PLoS Comput. Biol.* **12**, e1004667. DOI:10.1371/journal.pcbi.1004667.

Cheadle, S., Wyart, V., Tsetsos, K., Myers, N., de Gardelle, V., Herce Castañón, S. and Summerfield, C. (2014). Adaptive gain control during human perceptual choice, *Neuron* **81**, 1429–1441. DOI:10.1016/j.neuron.2014.01.020.

Colavita, F. B. (1974). Human sensory dominance, *Percept. Psychophys.* **16**, 409–412. DOI:10.3758/BF03203962.

Deneux, T., Harrell, E. R., Kempf, A., Ceballo, S., Filipchuk, A. and Bathellier, B. (2019). Context-dependent signaling of coincident auditory and visual events in primary visual cortex, *eLife* **8**, e44006. DOI:10.7554/eLife.44006.

Drugowitsch, J., DeAngelis, G. C., Klier, E. M., Angelaki, D. E. and Pouget, A. (2014). Optimal multisensory decision-making in a reaction-time task, *eLife* **3**, e03005. DOI:10.7554/eLife.03005.

Elhilali, M., Xiang, J., Shamma, S. A. and Simon, J. Z. (2009). Interaction between attention and bottom-up saliency mediates the representation of foreground and background in an auditory scene, *PLoS Biol.* **7**, e1000129. DOI:10.1371/journal.pbio.1000129.

Ernst, M. O. and Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion, *Nature* **415**, 429–433. DOI:10.1038/415429a.

García-Pérez, M. A. (1998). Forced-choice staircases with fixed step sizes: asymptotic and small-sample properties, *Vis. Res.* **38**, 1861–1881. DOI:10.1016/s0042-6989(97)00340-4.

Gold, J. I. and Shadlen, M. N. (2007). The neural basis of decision making, *Annu. Rev. Neurosci.* **30**, 535–574. DOI:10.1146/annurev.neuro.29.051605.113038.

Holdstock, J. S., Hocking, J., Notley, P., Devlin, J. T. and Price, C. J. (2009). Integrating visual and tactile information in the perirhinal cortex, *Cereb. Cortex* **19**, 2993–3000. DOI:10.1093/cercor/bhp073.

Huang, N. and Elhilali, M. (2020). Push-pull competition between bottom-up and top-down auditory attention to natural soundscapes, *eLife* **9**, e52984. DOI:10.7554/eLife.52984.

Huang, S., Li, Y., Zhang, W., Zhang, B., Liu, X., Mo, L. and Chen, Q. (2015). Multisensory competition is modulated by sensory pathway interactions with fronto-sensorimotor and default-mode network regions, *J. Neurosci.* **35**, 9064–9077. DOI:10.1523/JNEUROSCI.3760-14.2015.

Huk, A. C. and Shadlen, M. N. (2005). Neural activity in macaque parietal cortex reflects temporal integration of visual motion signals during perceptual decision making, *J. Neurosci.* **25**, 10420–10436. DOI:10.1523/JNEUROSCI.4684-04.2005.

Kiani, R., Hanks, T. D. and Shadlen, M. N. (2008). Bounded integration in parietal cortex underlies decisions even when viewing duration is dictated by the environment, *J. Neurosci.* **28**, 3017–3029. DOI:10.1523/JNEUROSCI.4761-07.2008.

Laurienti, P. J., Wallace, M. T., Maldjian, J. A., Susi, C. M., Stein, B. E. and Burdette, J. H. (2003). Cross-modal sensory processing in the anterior cingulate and medial prefrontal cortices, *Hum. Brain Mapp.* **19**, 213–223. DOI:10.1002/hbm.10112.

Levi, A. J. and Huk, A. C. (2020). Interpreting temporal dynamics during sensory decision-making, *Curr. Opin. Physiol.* **16**, 27–32. DOI:10.1016/j.cophys.2020.04.006.

Levi, A. J., Yates, J. L., Huk, A. C. and Katz, L. N. (2018). Strategic and dynamic temporal weighting for perceptual decisions in humans and macaques, *eNeuro* **5**, ENEURO.0169-18.2018. DOI:10.1523/ENEURO.0169-18.2018.

Lewis, R. and Noppeney, U. (2010). Audiovisual synchrony improves motion discrimination *via* enhanced connectivity between early visual and auditory areas, *J. Neurosci.* **30**, 12329–12339. DOI:10.1523/JNEUROSCI.5745-09.2010.

Li, Q., Yang, H., Sun, F. and Wu, J. (2015). Spatiotemporal relationships among audiovisual stimuli modulate auditory facilitation of visual target discrimination, *Perception* **44**, 232–242. DOI:10.1068/p7846.

Mercier, M. R. and Cappe, C. (2020). The interplay between multisensory integration and perceptual decision making, *NeuroImage* **222**, 116970. DOI:10.1016/j.neuroimage.2020.116970.

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

Meredith, M. A. and Stein, B. E. (1986). Visual, auditory, and somatosensory convergence on cells in superior colliculus results in multisensory integration, *J. Neurophysiol.* **56**, 640–662. DOI:10.1152/jn.1986.56.3.640.

Nienborg, H. and Cumming, B. G. (2009). Decision-related activity in sensory neurons reflects more than a neuron's causal effect, *Nature* **459**, 89–92. DOI:10.1038/nature07821.

Odoemene, O., Pisupati, S., Nguyen, H. and Churchland, A. K. (2018). Visual evidence accumulation guides decision-making in unrestrained mice, *J. Neurosci.* **38**, 10143–10155. DOI:10.1523/JNEUROSCI.3478-17.2018.

Okazawa, G., Sha, L., Purcell, B. A. and Kiani, R. (2018). Psychophysical reverse correlation reflects both sensory and decision-making processes, *Nat. Commun.* **9**, 3479. DOI:10.1038/s41467-018-05797-y.

Otto, T. U. and Mamassian, P. (2017). Multisensory decisions: the test of a race model, its logic, and power, *Multisens. Res.* **30**, 1–24. DOI:10.1163/22134808-00002541.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: transforming numbers into movies, *Spat. Vis.* **10**, 437–442. DOI:10.1163/156856897X00366.

Pick, H. L., Warren, D. H. and Hay, J. C. (1969). Sensory conflict in judgments of spatial direction, *Percept. Psychophys.* **6**, 203–205. DOI:10.3758/BF03207017.

Qiao, L., Xu, L., Che, X., Zhang, L., Li, Y., Xue, G., Li, H. and Chen, A. (2018). The motivation-based promotion of proactive control: the role of salience network, *Front. Hum. Neurosci.* **12**, 328. DOI:10.3389/fnhum.2018.00328.

Raab, D. H. (1962). Statistical facilitation of simple reaction times, *Trans. N. Y. Acad. Sci.* **24**, 574–590. DOI:10.1111/j.2164-0947.1962.tb01433.x.

Raposo, D., Sheppard, J. P., Schrater, P. R. and Churchland, A. K. (2012). Multisensory decision-making in rats and humans, *J. Neurosci.* **32**, 3726–3735. DOI:10.1523/JNEUROSCI.4998-11.2012.

Ratcliff, R., Smith, P. L., Brown, S. D. and McKoon, G. (2016). Diffusion decision model: current issues and history, *Trends Cogn. Sci.* **20**, 260–281. DOI:10.1016/j.tics.2016.01.007.

Rolls, E. T. and Baylis, L. L. (1994). Gustatory, olfactory, and visual convergence within the primate orbitofrontal cortex, *J. Neurosci.* **14**, 5437–5452. DOI:10.1523/JNEUROSCI.14-09-05437.1994.

Romei, V., Murray, M. M., Merabet, L. B. and Thut, G. (2007). Occipital transcranial magnetic stimulation has opposing effects on visual and auditory stimulus detection: implications for multisensory interactions, *J. Neurosci.* **27**, 11465–11472. DOI:10.1523/JNEUROSCI.2827-07.2007.

Romei, V., Murray, M. M., Cappe, C. and Thut, G. (2009). Preperceptual and stimulus-selective enhancement of low-level human visual cortex excitability by sounds, *Curr. Biol.* **19**, 1799–1805. DOI:10.1016/j.cub.2009.09.027.

Schadow, J., Lenz, D., Thaerig, S., Busch, N. A., Fründ, I. and Herrmann, C. S. (2007). Stimulus intensity affects early sensory processing: sound intensity modulates auditory evoked gamma-band activity in human EEG, *Int. J. Psychophysiol.* **65**, 152–161. DOI:10.1016/j.ijpsycho.2007.04.006.

Stein, B. E., Huneycutt, W. S. and Meredith, M. A. (1988). Neurons and behavior: the same rules of multisensory integration apply, *Brain Res.* **448**, 355–358. DOI:10.1016/0006-8993(88)91276-0.

Talluri, B. C., Urai, A. E., Bronfman, Z. Z., Brezis, N., Tsetsos, K., Usher, M. and Donner, T. H. (2021). Choices change the temporal weighting of decision evidence, *J. Neurophysiol.* **125**, 1468–1481. DOI:10.1152/jn.00462.2020.

Teder-Sälejärvi, W. A., Di Russo, F., McDonald, J. J. and Hillyard, S. A. (2005). Effects of spatial congruity on audio-visual multimodal integration, *J. Cogn. Neurosci.* **17**, 1396–1409. DOI:10.1162/0898929054985383.

Todd, J. W. (1912). *Reaction to Multiple Stimuli*. Science Press, New York, NY, USA.

Townsend, J. T. and Wenger, M. J. (2004). A theory of interactive parallel processing: new capacity measures and predictions for a response time inequality series, *Psychol. Rev.* **111**, 1003–1035. DOI:10.1037/0033-295X.111.4.1003.

Welch, R. B. and Warren, D. H. (1980). Immediate perceptual response to intersensory discrepancy, *Psychol. Bull* **88**, 638–667. DOI:10.1037/0033-2909.88.3.638.

Zylberberg, A., Ouellette, B., Sigman, M. and Roelfsema, P. R. (2012). Decision making during the psychological refractory period, *Curr. Biol.* **22**, 1795–1799. DOI:10.1016/j.cub.2012.07.043.