The interplay between multisensory integration and perceptual decision making

Manuel R Mercier, Céline Cappe

NeuroImage 222 (2020) 116970

Contents lists available at ScienceDirect

NeuroImage

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

The interplay between multisensory integration and perceptual decision making

Manuel R. Mercier [a,b,*], Celine Cappe [b]

[a] Institut de Neurosciences des Systèmes (INS), Inserm (UMR1106), Aix-Marseille University, 13005 Marseille, France
[b] Brain and Cognition Research Center (CerCo), CNRS, UMR 5549, University of Toulouse Paul Sabatier, Purpan Medical School Hospital, 31052, Toulouse, France

ARTICLE INFO

Keywords:
Multisensory integration
Perceptual decision making
Supervised machine learning
EEG decoding
Race model
Drift diffusion model

ABSTRACT

Facing perceptual uncertainty, the brain combines information from different senses to make optimal perceptual decisions and to guide behavior. However, decision making has been investigated mostly in unimodal contexts. Thus, how the brain integrates multisensory information during decision making is still unclear. Two opposing, but not mutually exclusive, scenarios are plausible: either the brain thoroughly combines the signals from different modalities before starting to build a supramodal decision, or unimodal signals are integrated during decision formation. To answer this question, we devised a paradigm mimicking naturalistic situations where human participants were exposed to continuous cacophonous audiovisual inputs containing an unpredictable signal cue in one or two modalities and had to perform a signal detection task or a cue categorization task. First, model-based analyses of behavioral data indicated that multisensory integration takes place alongside perceptual decision making. Next, using supervised machine learning on concurrently recorded EEG, we identified neural signatures of two processing stages: sensory encoding and decision formation. Generalization analyses across experimental conditions and time revealed that multisensory cues were processed faster during both stages. We further established that acceleration of neural dynamics during sensory encoding and decision formation was directly linked to multisensory integration. Our results were consistent across both signal detection and categorization tasks. Taken together, the results revealed a continuous dynamic interplay between multisensory integration and decision making processes (mixed scenario), with integration of multimodal information taking place both during sensory encoding as well as decision formation.

1. Introduction

Perceptual decision making has been described as a chain of computational steps. First, sensory signals are encoded by the neurons in sensory cortices. Thereafter, these inherently noisy (due to variability in ongoing neuronal activity) sensory signals are accumulated over time in parietal and frontal regions to form a decision. Lastly, if a decisional criterion is reached, a motor response is triggered (see for reviews (Gold and Shadlen, 2007; Heekeren et al., 2008)). Thus, formally perceptual decision making can be divided into sensory encoding and decision formation stages. So far, perceptual decision making has been studied mostly in unimodal contexts, whereas some evidence suggests that decision formation is *supramodal*[1] (O'Connell et al., 2012; Romo and de Lafuente, 2013). Moreover, the different sensory channels provide complementary information, the integration of which leads to more accurate and faster behavioral decisions (Stein and Meredith, 1993; Welch and Warren, 1980). Consequently, multisensory integration could be an integral part of perceptual decision making process, yet this point has been overlooked and investigations of multisensory integration and perceptual decision making remain two independent lines of research.

The neural basis of multisensory integration and its loci in the hierarchy of brain computations have been the focus of myriad of studies (see for

[1] The term "supramodal" is commonly used to qualify a mechanism shared between modalities. The prefix 'supra' (*i.e.* above) implies that it comes late in the stimulus processing hierarchy, presumably when modalities are merged and when information is somewhat "amodal". Also in the literature the term "supramodal" suggests neutrality as it does not consider integration between sensory channels; in such case we consider the term "multimodal" to be more accurate and thus preferable to indicate multisensory integration.

* Corresponding author. Brain and Cognition Research Center (CerCo), CNRS, UMR 5549, University of Toulouse Paul Sabatier, Purpan Medical School Hospital, 31052, Toulouse, France.
E-mail address: manuel.mercier@a3.epfl.ch (M.R. Mercier).

https://doi.org/10.1016/j.neuroimage.2020.116970
Received 31 August 2019; Received in revised form 23 March 2020; Accepted 15 May 2020
Available online 23 May 2020

reviews (Keil and Senkowski, 2018; Talsma et al., 2010; ten Oever et al., 2016)). It is now well established that multisensory integration starts early in the process chain (Foxe and Schroeder, 2005; Schroeder and Foxe, 2005). Both animal and human studies have demonstrated that the genesis of multisensory integration relies on cross-modal inputs to sensory cortices, which informs about the spatiotemporal co-occurrence of sensory cues (Atilgan et al., 2018; Bizley et al., 2007; Cappe et al., 2010; Kayser et al., 2008; Lakatos et al., 2007; Mercier et al., 2015, 2013). At a later stage, the integration of information from different modalities is related to congruency and reliability of multisensory inputs, as well as task relevance (Kayser et al., 2017; Rohe and Noppeney, 2016, 2015). However, it is unclear how multisensory information is processed during perceptual decision making: it is still an open question whether the observed behavioral benefits of multisensory inputs reflects a cumulative effect of multisensory integration at both sensory encoding and decision formation stages, and whether multisensory integration is at play during decision formation at all (Bizley et al., 2016). This question is best instantiated by the long-standing debate on how decision is formed in the context of redundant signals or two alternative choices. Signal processing in these situations can follow one of the two hypothetical paths: (1) signals from two modalities can be treated independently and be associated with two parallel decision formation processes (*e.g.* parallel race models); or (2) signals from two modalities are combined before being fed into a single *supramodal* decision formation process (Bizley et al., 2016; Bogacz, 2007; Otto and Mamassian, 2017; Smith and Ratcliff, 2004).

Given the two stages in perceptual decision making (sensory encoding and decision formation), the effect of multisensory integration could take place (Fig. 1 A): (1) during sensory encoding only, *i.e.* before a *supramodal* decision formation step, (2) during a multimodal decision formation (see note 1 for definitions), or (3) both during sensory encoding and decision formation. To test these alternative hypotheses, we employed behavioral modeling and a time-resolved decoding approach on human EEG while subjects detected or categorized an unpredictable unisensory-cue (auditory/visual) or a multisensory-cue embedded within a stream of audiovisual noise. To preview the results, behavior-based models alone indicated multisensory integration at multiple stages of perceptual decision making. This was further supported by the results from supervised machine learning analyses of EEG data under unisensory-cue conditions, which revealed the neural markers of sensory encoding and decision formation processes (unisensory classifier). Application of unisensory classifier onto multisensory-cue condition (cross-condition decoding generalized over time) indicated an acceleration of both sensory encoding and decision formation for multisensory-cue condition as compared to unisensory-cue conditions. Lastly, direct decoding procedure between unisensory trials and multisensory trials demonstrated two periods of multisensory integration intimately linked to sensory encoding and decision formation. All the results were reproducible in both detection and categorization tasks. Together, these findings demonstrate that early multisensory integration accelerates sensory encoding, while later multisensory integration accelerates decision formation.

2. Material and method

2.1. Participants

Data were collected from 12 subjects (4 females, age range 23–35 year old, mean = 26.7, mean absolute deviation (m.a.d.) = 2.6) in the EEG experiments. All participants had normal hearing and normal or corrected-to-normal vision. The study was conducted in accordance with the Declaration of Helsinki and approved by the Inserm (Institut National

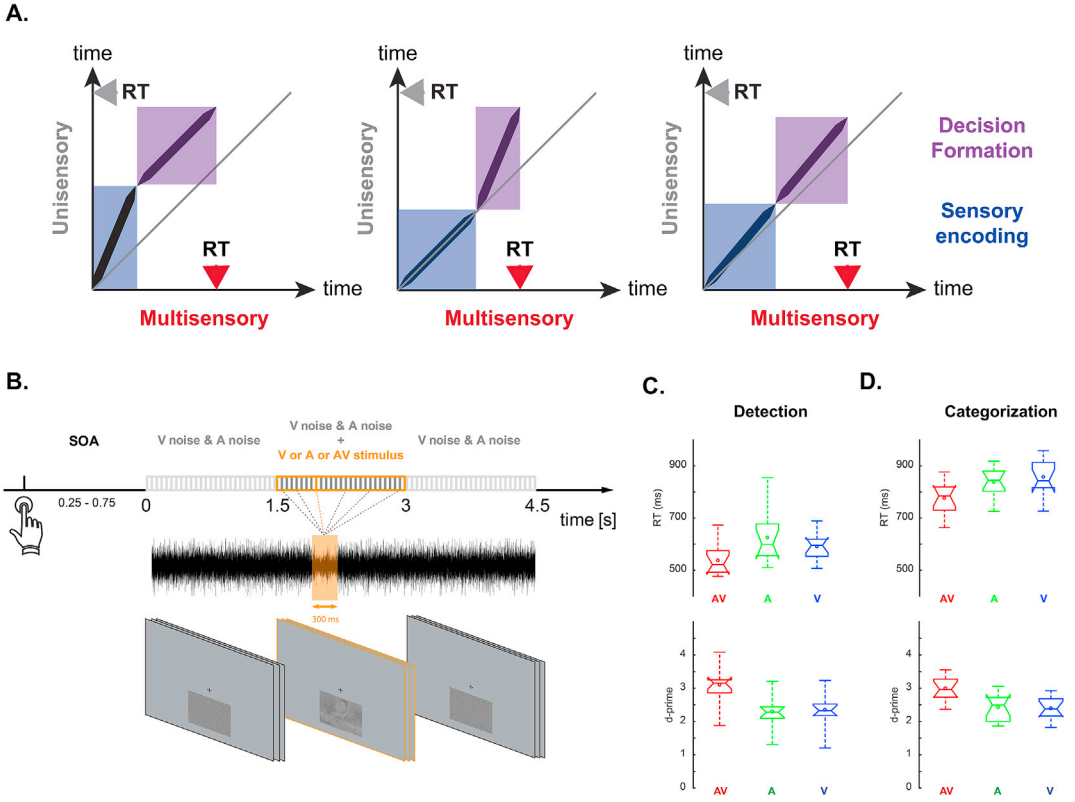

Fig. 1. (A) Hypothesized difference in perceptual decision making dynamics between unisensory and multisensory signals during sensory encoding (in blue) and during decision formation (in purple). Left: Once received, sensory inputs are fully integrated to reduce uncertainty and accelerate sensory processing, which in turn fuels a supramodal decision process. Middle: Information from the different senses is processed in parallel during sensory encoding and the facilitatory effect of multisensory integration takes place only during the decision formation stage. Right: The different modalities are combined at both processing stages. (B) Experimental design: participants had to detect, or to categorize, an unpredictable cue fading in/off from a continuous stream of audio-visual noise. (C) Behavioral performance: Response times and d-primes plotted for audiovisual (red, AV), auditory (green, A) and visual (blue, V) cue conditions during detection and categorization tasks.

de la Santé et de la Recherche médicale) ethical committee (Institutional Review Board IRB00003888 - agreement n°14–156). Written informed consent was obtained from all volunteers before the experiment.

2.2. Experimental setup

EEG was acquired from 128 electrodes and 4 additional electrodes, placed above and below the dominant eye and on the left and the right sides of the eyes, were used to record EOG. Signals were digitized at 2048 Hz, with a 24-bit A/D conversion BioSemi ActiveTwo system (Netherlands). During the experiment, participants were seated comfortably in a quiet dark room. A chin-rest (SR Research) was used to maintain stable head position while fixating on a black cross continuously displayed on a gray background. The screen, on which instructions and visual stimuli were presented, was located 80 cm from the participant (BenQ XL2411, refresh rate: 100 Hz, resolution: 1920x1080). The dynamic visual sequence of stimulation was presented below the central fixation, with the image centered at 1/4 of the display height. Sounds were delivered through earphones (Etymotic ER.4), using a dedicated audio card (Sound Blaster Audigy 5/Rx). Responses were collected through a numeric keypad. The experiment was programmed and controlled using Presentation software (version 18.1 from Neuro-Behavioral Systems, USA).

2.3. Stimuli

Stimuli were primarily from the CerCo databases (the Brain and Cognition research Center, CerCo, UMR 5549 CNRS). They were typical sounds and images of living or non-living categories (respectively birds, dogs, monkeys and cars, guitars, phones). All stimuli were equated using the following procedure. Sounds were first calibrated (11025Hz, 16 bits, mono) and then rms-normalized. Auditory noise was added by permuting samples with a morphed average of all sounds. Images were first cropped (400x300 pixels) and converted to black and white. Intensity range was normalized and image histograms equalized to the mean image. Visual noise was created by shuffling pixels. For both sounds and images, the signal-to-noise ratio (SNR) ranged from 0% to 100% in 5% incremental steps.

The stimuli used for the main experiment were selected from a larger pool of stimuli after behavioral pilot on a separate group of participants ($n = 8$). Every selected item reached at least 50% of correct recognition at 50% of SNR in the pretests to ensure representativeness and homogeneity of the final set of stimuli. Following the piloting, the selected set of stimuli contained 60 sounds and 240 images with an equal number of items per category.

During the main experiment, unisensory performance accuracy was kept at 70% using a continuous adaptive staircase procedure (adaptive up-down method (Levitt, 1971), with separate staircase for each category in each unisensory condition).

2.4. Experimental design

At the beginning of each trial, the participant was prompted to press a button (see Fig. 1 B). Following a random SOA (250–750 ms), an audio-visual sequence started. Each sequence lasted 4.5 s and contained dynamic audio-visual noise. At any random time between 1.5 s and 3 s, an unpredictable signal cue was presented in any modality (auditory, visual or both at the same time) on 85% of trials. The remaining 15% were trials with audio-visual noise only (catch trials).

During the audio-visual sequence, smooth transitions between noise and signal cue were achieved by a faint gradual increase/decrease of signal-to-noise in the course of target presentation (300 ms cycle for signal fading-in/off). The visual stream was constructed by displaying at every screen refresh a random picture with a given signal-to-noise ratio. For each visual sequence, the pictures were chosen from pre-selected image set (with 500 versions at 0% signal-to-noise and 100 versions for each signal-to-noise above 0%). Each auditory stream was constructed by concatenating the different noise versions of a selected sound (either at 0% signal-to-noise or at a given signal-to-noise ratio for the signal cue). Each participant were exposed to the entire set of stimuli, each stimulus was used the same amount of time and audio-visual sequences were unique (*i.e.* different sequences were produced for each stimulus and for each participant).

Every participant performed two tasks. In the detection task, participants had to indicate whether a signal cue was presented in either or both modalities by pressing the appropriate response button. In the categorization task, participants had to indicate if the signal cue was an animal or an inanimate object by pressing the corresponding response button (counterbalanced across participants). In case of audio-visual cues, the two modalities were congruent (*e.g.,* the image of a bird was presented with the sound of a bird). No specific instruction was given regarding speed-accuracy trade-off. Participants were told that difficulty was adjusted depending on their accuracy (online staircase procedure, see above) and that late responses given at the end of a stimulus sequence were not taken into account (to avoid mixing the brain signals related to motor response with those elicited by the offset of the audiovisual sequence). Task order was counterbalanced between participants. Each task was divided into five consecutive blocks, during which conditions were randomized (four blocks contained 165 sequences and one block contained 168 sequences). To maintain vigilance, participants were encouraged to make self-paced breaks.

2.5. Analysis of behavior

Participants' responses were analyzed off-line to investigate performance accuracy and response times. To take into account responses given when no signal cue was presented (catch trials), perceptual sensitivity was computed (d-prime). To do so we followed a conservative criterion. First, hit rates were calculated by including hits, misses and trials with undefined response (*e.g.* double button press) in the total number of trials (plus incorrect responses in the categorization task). Second, false alarm rates were calculated by summing response rate for catch trials (mean $\pm$ m.a.d.; detection task: 3.8% $\pm$2.7; categorization task: 3.3% $\pm$2.0) and response rate for responses occurring before the target presentation in any condition (detection task: 1.0% $\pm$0.8; categorization task: 0.3% $\pm$0.3). Once computed, d-prime values were subjected to a paired random permutation test (10 000 iterations), comparing performance between multisensory-cue and unisensory-cue conditions. Effect sizes were estimated using Cohen's *d* for paired samples.

Mean participant's performance accuracy and response times (RT) were computed and subjected to the same paired random permutation test. Performance accuracy was used to compute multisensory gain: The difference between performance in the multisensory condition and the best performance in one of the two unisensory conditions. Each participants RT distributions were further analyzed using the most appropriate model for each task: (i) the race model for the detection task and (ii) the diffusion decision model (DDM) for the categorization task.

Behavioral analysis of the Detection task. The detection task used here required participants to give the same response despite the differences in cue type (*i.e.* auditory, visual or audio-visual cue). This type of paradigm leads to the typical redundant signal effect (RSE) characterized by RT acceleration when both cues are presented together (*i.e.* audio-visual condition) as compared to conditions with the single cue (auditory or visual). Typically, this RSE can be explained by a simple statistical facilitation: in the redundant condition (multisensory cue condition), the faster of the two signals would trigger the response (Raab, 1962). To test whether the observed RSE is merely a result of such statistical facilitation, we tested the *race model inequality* (or *Miller's bound,* (Miller, 1982), see also for review (Gondan and Minakata, 2016)):

$$P_{AV}(t) \leq P_A(t) + P_V(t)$$

With P being the cumulative probability that a response has been triggered at time t following the presentation of audio-visual signal (AV), auditory signal (A) or visual signal (V).

Under the *race model inequality*, the cumulative RT distribution of the audio-visual condition cannot exceed the sum of the distributions in unisensory conditions (auditory and visual conditions), and therefore assumes that there are no interactions between the sensory signals. For each participant, the cumulative distribution function (CDF) for each experimental condition was calculated using 50 quantiles. To down-sample the collected RTs to a common sample size (*i.e.* 50 bins) a linear interpolation was performed. Then, we tested model violation by comparing the area between the empirical CDF in the multisensory cue condition and the *Miller's bound* (model predicted CDF) using a one-sample *t*-test, with participants as repeated measures and the area as the dependant measure (Colonius and Diederich, 2006; Innes and Otto, 2019). If the empirically observed and the theoretically derived cumulative audio-visual RT distributions are significantly different, it can be concluded that faster RTs in the multisensory condition cannot be explained by a simple statistical facilitation due to the presence of two signal cues.

The *race model inequality* assumes absence of correlation between unisensory responses (*i.e.* the join probability is supposed to be negligible) and context invariance (*i.e.* statistical independence of unisensory processes, see Otto et al., 2017 for an extended discussion). Therefore, next we applied the *context variant race model* which includes two free parameters to account for the aforementioned assumptions inherent to the *race model inequality*. The *context variant race model* estimates RT distribution of the multisensory condition using RT distributions from the unisensory conditions by means of the LATER model (Linear Approach to Threshold with Ergodic Rate), further complemented with the two free parameters. The LATER model posits that the behavioral response is initiated when a decision variable reaches a threshold, and that this decision variable is defined as a ramp with a slope drawn from a Gaussian distribution (note here that 1/RTs distribution follows a Gaussian distribution, also all modeling described here are performed in the reciprocal space). The LATER model has been shown to be effective in modeling decisions and RTs in one-alternative tasks (*e.g.* signal detection task), especially when the stimulus is unpredictable in time like in the detection task we used (Carpenter and Williams, 1995; Noorani and Carpenter, 2016). As a first step, the LATER model was fitted to each unisensory RT distribution using the minimum variance unbiased estimator (computed using the function *normfit* in Matlab). That is, RTs were described by a reci-normal distribution defined by two parameters: *mu* and *sigma*. Next, these best fitting estimates of unisensory RT distribution were used to fit the *context variant race model* to the multisensory RT distribution, assuming that the response in the multisensory condition is triggered by a unisensory signal with the higher drift diffusion rate (*i.e.* race logic). In addition, two free parameters were added to the model: *Rho*, a measure of correlation between unisensory signals, and *Eta*, the variability of the decision variable rate (Otto and Mamassian, 2012). These free parameters account for the context invariance assumption (*i.e.* statistical independence of unisensory processes, see above) assumed by the standard *race model inequality* and therefore permit the *context variant race model* to violate *Miller's bound*. Following the fitting procedure (performed using the *Maximum Likelihood Estimates* function from Matlab), the free parameters were obtained for each individual and subjected to a one sample *t*-test. Last, to assess how much the best-fitting model explains RT acceleration, we compared the model-fit multisensory benefit to the empirically observed multisensory benefit. That is, for both RT CDFs (*i.e.* modelled and empirically observed), we estimated the area (using the mean of CDFs difference) delimited by the faster of the unisensory CDFs. Comparison between multisensory benefit estimated by the model and empirical multisensory benefit was then performed across participants using one-tail Pearson's linear correlation to reflect attended positive relationship.

Behavioral analysis of the Categorization task. The categorization task used here was a two-alternative forced choice task, in which participants had to categorize the stimuli as being living/animal or inanimate/object. The Diffusion Decision Model (DDM) was built and extensively applied to modeling perceptual decision making for similar binary choice tasks (Ratcliff, 1978; Ratcliff and McKoon, 2008). This approach proposes to map different cognitive processes to different psychologically meaningful parameters (Ratcliff et al., 2016; Voss et al., 2013):

- Drift rate (v): the mean rate with which decision process approaches decision threshold triggering the behavioral response, which can be interpreted as information uptake and is linked to task difficulty.
- Threshold separation (a): the distance between decision thresholds, which reflects the amount of information considered for making a decision (a measure of response criterion: conservative *vs.* liberal).
- Starting point (z): bias toward one of the two possible outcomes (asymmetry in response criterion). Here a relative starting point (zr) was used, such that threshold separation a can be estimated (relative bias allow $a \neq 1$).
- Duration of non-decisional processes ($t0$): which includes the encoding processes preceding the decisional period, memory access and motor processes of the response system.
- A difference parameter (d) allows adjusting the difference in duration of non-decisional processes between the two choices.
- Inter-trial variability of the different parameters: drift rate (sv), starting point (sz) and time constant ($st0$). These parameters permit to take into account trial-to-trial fluctuations in processing of different stimuli. Specifically, the drift rate follows a normal distribution with the mean v and the standard deviation sv, while starting point and non-decision constants follow uniform distributions with the means z and $t0$, and sz and $st0$, respectively.

We hypothesized that multisensory RT acceleration can be linked to four DDM parameters: (1) the drift rate variability across conditions v and (2) across trials sv; (3) the variability of non-decisional constant across conditions $t0$; (4) and across trials $st0$. Threshold separation, a, was allowed to vary freely, but was kept fixed between conditions, to account for the differences in response caution across participants. Likewise, other parameters were kept constant across conditions. We assumed that there was no a priori bias toward (or against) any response (otherwise, it would imply that participants had prior knowledge of the upcoming event) and therefore we fixed the relative starting point of decision trajectory at mid-point between the two decision thresholds ($zr = 0.5$), and set it constant across trials ($szr = 0$). Similarly we assumed that there was no reason to expect differences in speed of response execution between the two categories ($d = 0$). For completeness, we also tested several additional models, in which we varied the amount of free parameters and verified the relevance of each parameter in predicting the RT distributions and multisensory benefit (see below).

We computed the DDM for each participants using fast-dm-30 (Voss et al., 2015). Fast-dm-30 allows to estimate DDM parameters and to calculate values of predicted RT CDF for a given parameter set. To optimize the fit of all free parameter estimates, we used the Kolmogorov-Smirnov test, which utilizes the entire RT distribution. As a consequence, it is a robust and stable criterion that does not imply data reduction (such as binning procedure) and is less influenced by the outliers (Voss et al., 2015, 2013).

Model was evaluated by computing a goodness of fit measure at the population level by comparing the DDM predicted and the empirical RT distributions. That is, we computed the integral of the absolute difference between best-fitting CDF from the DDM (averaged for the two-categories) and the corresponding empirical CDF (RT distributions were equated in number of trials by down-sampling to 50 bins using linear interpolation). The area between the curves was computed for each condition and then averaged across subjects. The lower this goodness of fit, the better DDM predicts empirical data. To validate the best model fit in a graphical way, empirical and predicted RT quantiles (25, 50 and 75%) for the different conditions were plotted for all subjects. In so-called quantile-probability plots, a good fit can be assumed when all data points are positioned along

the diagonal. Last, the estimated free parameters that we allowed to vary between conditions were subjected to Kruskal-Wallis test to assess the differences between conditions.

While the goodness of fit allows making inferences regarding the shape of RT distribution of different experimental conditions taken separately, it does not provide information regarding the predicted multisensory benefit (modelled CDFs of the different conditions can be shifted in any direction with respect to each other). Thus, to take into account the relative position of the modelled CDFs between conditions, we further estimated the multisensory benefit predicted by the DDM. This procedure is similar to the one used in the detection task and consists in comparing predicted and observed multisensory benefit using a correlation analysis (see previous section).

2.6. Signal preprocessing

Electrophysiological data (EEG, EOG) were scanned using a semi-automatic artifact detection procedure based on signal characteristic (epochs with utmost signal variance, maximum absolute amplitude and z-score - computed across trials and channels separately – where rejected). To perform a sanity check and exclude remaining artifacts, continuous data were visually inspected leading to an average of 128 ± 19 hits trials per condition and 77 ± 11 catch trials (detection task: AV = 153 ± 20, A = 116 ± 18, V = 117 ± 16, AVnoise = 79 ± 11; categorization task: AV = 150 ± 21, A = 117 ± 19, V = 114 ± 19, AVnoise = 75 ± 12). Artifacted channels were interpolated using spline method, then average reference was applied. Ambient noise (50 Hz and 100Hz–150Hz harmonics) was removed by fitting a sine and cosine at the specified frequency to the single trial data and subsequently subtracting the estimated components. Trials were defined from $\pm$ 1.625 from stimulus onset fade-in. For audiovisual noise trials (*i.e.* catch trials, no cue target), a sham onset time point was randomly picked between the 1.5–3s period (*i.e.* when signal cue can occur in the visual, auditory and audio-visual conditions). Trials were demeaned and high-pass filtered at 0.5Hz using a windowed Sinc FIR filter. To keep the variance of signal cue SNR across participants in a reasonable range, trials with extreme SNR were discarded on the basis of two signal-to-noise steps above and below the median in each category (in either the visual or the auditory condition). This threshold was chosen after scanning all the data and led to rejection of only a few trials (2.9 ± 1.7 per condition and per subjects). Finally, to increase the spatial resolution/reduce volume condition, we computed an estimate of the Scalp Current Density (*i.e.* surface Laplacian). This spatial filter is based on a second-order derivative of the scalp topography using 3D-spline method (Perrin et al., 1989). The data were processed offline by using custom-written scripts in MATLAB (MathWorks, MA, USA), the FieldTrip Toolbox (Oostenveld et al., 2011) and the LIBLINEAR library for large linear classification (http://www.csie.ntu.edu.tw/~cjlin/liblinear).

2.7. Classical ERP-analysis

For each participant, after down sampling the EEG data to 512 Hz, evoked-related potential (ERPs) were computed and baseline-corrected (−100 msec to onset of signal cue fade-in). Multisensory effects were assessed by mean of the additive model (Besle et al., 2004) through cluster-based permutation tests (Maris and Oostenveld, 2007). In short, ERPs from the multisensory condition were summed with ERPs from catch trials and then compared to the sum of ERPs from the unisensory conditions using non-parametric statistics and corrected for multiple comparison by taking the maximum sum of the cluster-level statistics (alpha level = 0.05, number of randomization = 1000, minimum number of neighboring channels per cluster = 2).

ERPs were further analyzed to highlight the topographical similarity between conditions by calculating the spatial correlation between every scalp maps of the multisensory condition and of the sum of unisensory conditions. This was performed by computing the Dissimilarity index as followed (Murray et al., 2008):

$$DISS_{u,v} = \sqrt{\frac{1}{n} \times \sum_{i=1}^{n} \left(\frac{u_i}{GFP_u} - \frac{v_i}{GFP_v}\right)^2}$$

where: n is the number of electrodes, u_i and v_i are the measured potentials at the ith electrode for conditions u and v, and the Global Field Power is the standard deviation of all electrodes.

The Dissimilarity index was then normalized to obtain the Spatial Correlation as followed:

$$SpCorr_{u,v} = 1 - \frac{Diss^2_{u,v}}{2}$$

Computation of this measure for every pair of topographies over time produces a matrix of topographical similarity for every time points. Topographical similarity range from −1, when two topographies are perfectly reversed in polarity, to +1, when they are identical. For each participant, the topographical similarity matrix was computed during the baseline period, which served as a control, and over the post fade-in onset period (from 0 to $\pm$ 800 ms and from 0 to $\pm$ 1000m for the detection and the categorization task respectively). The values were then subjected to a cluster based permutation paired-test, similar to the procedure described above (here the clusters correspond to the neighboring pixels in the topographical similarity matrix). To assess the asymmetry of the topographical similarity matrix across participants, values above and below the diagonal were summed and then contrasted using a paired-sample *t*-test.

2.8. Multivariate pattern analysis

Principle: The aim of MVPA applied to EEG is to isolate activations that are specific to a given brain operation by optimally combining the information from all electrodes at a given time point. In the present study, to obtain the topographical weights that maximally discriminate experimental conditions, we used the LIBLINEAR library (Fan et al., 2008) in a *Monte-Carlo* stratified cross-validation procedure (hold-out method with 200 iterations). Importantly, we applied different decoder types: the first one was used to classify conditions (binary classifier, (L2-regularized logistic regression mode), and the second one was a RT-based classifier (non-binary classifier, L2-regularized L2-loss support vector regression mode). Whereas the first decoder type characterizes the full dynamics of activation patterns, the second decoder type isolates brain activity that predicts RT. Thus, this double decoding procedure permits to study processes, which may partly overlap in time: sensory encoding and decision formation.

Decoding procedure: Data were down-sampled (128 Hz) to reduce computation time while maintaining sufficient temporal resolution (Grootswagers et al., 2017). Trials were defined either relative to the cue fade-in onset (−200 to 1300 msec) or relative to the response (−1500 to 200 msec). Each trial was baselined using -200-0 msec interval (cue--locked analysis) or the entire epoch (RT-locked analysis). On each cross-validation iteration (CV), the data-set was randomly split into a training set (90% of the trials) and a testing set (the remaining 10% of the trials), each condition being equally represented by the same amount of trials (stratified cross-validation). Last, the signal at each electrode was normalized across trials using the estimates from the training set (Crouzet et al., 2015; Edwards et al., 2018).

Weights projection: For each time point, a weight was assigned to each electrode which reflects how this feature contributed to maximizing the decoding. After the decoding procedure, and to further exploit weighting information, the electrode weights were used in two ways. First, to characterize brain activation that best differentiate experimental conditions, topographical weights were transformed back into activation patterns by multiplying them with the covariance in the data. Second, to track the temporal course of the cognitive operation isolated by a given classifier, topographical weights were applied to the single-trial time series and used for additional statistical analyses (see below *Link between classifiers* section).

Time generalization: Classifiers that best differentiate conditions at a

given time point were tested on every other time point, leading to a "temporal generalization" matrix. This step was performed within the cross-validation iteration which implies that trials used for training and testing (at the same/different time point) were from different trial sets. Such temporal generalization permits to draw the blueprint of brain processes by distinguishing canonical motifs that are not accessible otherwise. Actually, similar variation in decoding performance over time can originate from different scenarios that can be revealed through temporal generalization. For instance the reactivation of a unique process can have the same profile of decoding performance as a chain of different processes. In the former case the temporal generalization would reveal a checkerboard like motif where the process is repeated at different time moment, while in the second case a diagonal motif would be seen in the temporal generalization matrices (for further instantiations see (King and Dehaene, 2014)).

Generalization across conditions: To assess the similarities and differences between brain responses across conditions, the classifiers trained in one condition (*e.g.*, response time decoding of unisensory-cue trials) were tested for generalization on other conditions (*e.g.,* response time decoding of multisensory-cues trials). This step was performed at each iteration on normalized data. Here, we also tested for the temporal generalization across conditions, following the same procedure as for within-condition temporal generalization described above. Generalization across time and experimental conditions captures the neural architecture of brain operations and reveals how processing stages change between experimental conditions (King et al., 2016; Marti et al., 2015; Myers et al., 2015; Salti et al., 2015).

Statistical validation: For each iteration, the classifier generated a measure of the decoding performance (either the Area Under the Curve or the correlation coefficient depending on the type of decoding performed: between conditions or RT prediction). This measure relates to the proportion of test trials that were correctly classified using the topographical weights computed from the training set. Chance levels were obtained by running the same procedure while shuffling the labels between conditions. As within-subjects analysis, classifier performance from each iteration were used to compute confidence intervals of the mean using a bootstrapping procedure (95% CI from 1000 boostraps) which was then compared with the results from shuffled data to estimate statistically significant decoding performance. Between-subject analysis was performed using a one-tail paired random permutation test on the mean performance across CV iterations for real and shuffled conditions (10 000 iterations). To control for multiple comparisons, a cluster-based correction was applied with false alarm rate set at 0.005% cut-off (Maris and Oostenveld, 2007).

Relationship between classifiers: We performed several different decoding procedures to characterize distinct cognitive processes. Trials with unisensory signal cues were classified either against audio-visual noise or were used to predict RT (using different decoder type, see above). These different classifiers permitted to identify activation patterns corresponding to early sensory encoding and later decision formation, which were next tested for generalization on multisensory trials (generalization procedure). Direct decoding between unisensory trials and multisensory trials revealed an early and a late period of multisensory integration and corresponding activation patterns (see results section). To relate these different cognitive processes at the single-trial level, the co-occurrence of the activation patterns that were derived from different decoding procedures were assessed as follows. First, for each subject, and for each type of decoding, the classifiers showing the highest decoding performance during the early and the later processing stage were identified. This allowed to select the classifiers which best discriminated conditions at the subject level. Second, the corresponding weights were applied onto multisensory-cue trials (hits only) to obtain a single-trial time series for each decoder type (*i.e.,* sensory encoding, decision formation and early/late multisensory integration). Third, from each classifier-weighted single-trial time series the maximum absolute amplitude was extracted (4 numbers per trial). This provided an estimate of respective cognitive process strength (early sensory encoding, later decision formation, early and late multisensory integration) at the single trial level. Fourth, single-trial maximum absolute amplitudes were correlated to evaluate the relationship between, respectively, early/late multisensory integration and sensory encoding/decision formation. Given the hypothesized positive relationship, one-tail Pearson correlation was computed. Finally, individual correlations were subjected to a random effects test. That is, the correlation coefficients obtained for each participant (Fisher z-transformed) were tested at the group level using one sample *t*-test (*i.e.* against zero).

3. Results

3.1. Behavior

Behavioral benefits of multisensory integration were investigated in two tasks, where participants had to detect, or to categorize, a target-cue embedded in a stream of audio-visual noise (Fig. 1 B). This signal cue was either auditory (A), visual (V), or audio-visual (AV), and could be presented at any moment in the stream of noise or be absent (catch trials). Signal cue consisted of an unpredictable faint gradual increase-decrease of signal-to-noise-ratio (SNR). The amount of SNR was titrated to maintain performance at an accuracy level of 70% in either unisensory conditions (Detection task: A = 67.4 ± 1.8, V = 69.2 ± 2.5, Categorization task: A = 70.3 ± 2.9, V = 69.0 ± 3.4; mean % +/− m.a.d.). As compared to the unisensory conditions, accuracy in the multisensory condition was higher in both tasks (89.4 ± 1.5 and 86.2 ± 2.4; mean % +/− m.a.d in the Detection and the Categorization task respectively). Which leads to significant multisensory gain in both tasks (Detection: 0.2 ± 0.03 m.a.d, p = 0.0006, Cohen's d = 5.2584; Categorization: 0.15 ± 0.04 m.a.d, p = 0.0003; Cohen's d = 4.0776). Perceptual sensitivity, depicted in Fig. 1 C, was also found to be higher in the multisensory condition as compared to unisensory conditions (Detection task: for AV vs. A p = 0.0002 & Cohen's d = 5.9588, for AV vs. V p = 0.0002 & Cohen's d = 5.2498; Categorization task: for AV vs. A p = 0.0003 & Cohen's d = 3.2181, for AV vs. V p = 0.0003 & Cohen's d = 7.3189); with false alarm rate being relatively low (see Materials and Methods).

In both tasks, more accurate performance in multisensory condition was accompanied by significantly faster response times (Detection task: AV = 537 ms ± 46 m.a.d., A = 624 ± 73, V = 591 ± 39, for AV vs. A p = 0.0002 & Cohen's d = 1.7408, for AV vs. V p = 0.0004 & Cohen's d = −2.1875; Categorization task: AV = 776 ± 52, A = 838 ± 45, V = 857 ± 55, for AV vs. A p = 0.0004 & Cohen's d = 2.334, for AV vs. V p = 0.0003 & Cohen's d = −3.171). To further examine the observed RT acceleration, we analyzed behavioral data taking into account the entire RT distribution. For this, we used computational models that are appropriate for each task: the race models for the detection task and the Diffusion Decision Model (DDM) for the categorization task.

3.2. Detection task: the race models

To assess whether RT acceleration in the multisensory condition can be explained by statistical facilitation due to signal redundancy (*i.e.* both cues providing the same information), we applied the race model inequality (see Materials and Methods section for details). This model relies on the assumption that if the two components of a redundant signal are processed in separate channels, the faster channel determines the response time (Gondan and Minakata, 2016; Miller, 1982; Otto and Mamassian, 2017). The comparison of this so-called Miller's bound to the observed cumulative distribution function (CDF) of multisensory RT across participants demonstrates that RT distribution in the multisensory condition cannot be fully explained by a simple statistical facilitation (mean = $4.2 \times 10^{-3} \pm 0.3 \times 10^{-3}$ s.e.; t(11) = 4.22, p = 0.0014, Miller's bound violation is illustrated at the population level in Fig. 2 A with a dark shaded area).

We next applied the *context variant race model*, a refined race model recently developed by Otto and colleagues, which allows for possible interactions between unisensory channels to explain multisensory benefit

(Otto et al., 2017, 2011). This model was derived in two steps. First, we fitted the LATER model (Noorani et al., 2016; Carpenter et al., 1995) to the unisensory CDFs and estimated for each distribution *mu* and *sigma* parameters (respectively: A: 1.76 ± 0.20; V: 1.78 ± 0.12 and A: 0.45 ± 0.07; V: 0.33 ± 0.06; mean ± m.a.d.). Second, we combined these parameter estimates to model multisensory RTs distribution (mean log-likelihood of the MLEs = −24.7 ± 7.1 m.a.d.). For illustration purposes, Fig. 2 A depicts the subject-average best-fitting *context variant race model* and the corresponding empirically observed multisensory CDF. We compared modelled and empirical data at the subject level by computing corresponding multisensory benefit (*i.e.* the difference between the multisensory CDF, modelled or empirical, and the faster of the two unisensory CDFs; light shaded area in Fig. 2 A). Correlation between model-fit benefits and empirical benefits shows that the best-fitting model almost fully explains multisensory RT acceleration across participants ($r = 0.99$, $p = 4.1 \times 10^{-9}$, see inset Fig. 2 A). Following the validation step of the best-fitting model, we inspected the free parameters and found that both Rho (0.44 ± 0.25; mean ± m.a.d.) and Eta (0.05 ± 0.04) were significantly different from zero across participants (respectively: $t(11) = 4.96$, $p = 0.0004$ and $t(11) = 4.24$, $p = 0.0014$). Note that the values reported in Fig. 2 legend correspond to the application of the model at the population level (*i.e.* RT averaged across participants, see above). These results demonstrate that inclusion of interactions between the unisensory features in the *context variant race model* resulted in increase of explained variance as compared to simple statistical facilitation (Miller's bound).

3.3. Categorization task: the diffusion model

The Diffusion Decision Model (DDM) is used to model performance in two-alternative forced choice tasks, like the categorization task employed here. The DDM models the accumulation of evidence between two threshold criterions - one for each response alternative. The DDM allows modeling RT distributions using up to eight free parameters, each of which is associated with psychologically meaningful process (Ratcliff et al., 2016; Voss et al., 2013). Here, we reasoned that multisensory RT acceleration could be linked to four DDM parameters that we allowed to vary between conditions: the rate of decision formation (drift rate, *v*), the duration of non-decisional processes (non-decisional constant, *t0*) and inter-trial variability of these two parameters (*sv* and *st0* respectively). To account for the differences in response caution across participants, the distance between decisional bounds (threshold separation *a*) was allowed to freely vary across participants, but was kept fixed between conditions. The last three parameters were kept fixed.

After fitting the DDM to the data separately for each participant, statistical analyses of the free parameters across participants revealed significant differences between the multisensory condition and the unisensory conditions (see Table 1), for *t0* (*i.e.* non-decisional time constant), indicating that the duration of non-decisional processes is reduced in the multisensory trials as compared to unisensory trials. Critically, there was no difference for *v* (*i.e.* drift rate). Last, for multisensory trials, *sv* (*i.e.* inter-trial variability of the drift rate) was significantly greater, while *st0* (*i.e.* inter-trial variability of non-decisional time constant) was significantly reduced. Empirical CDFs averaged across participants and the corresponding best-fitting DDM estimates are represented in Fig. 2 B. Inset depicts quantile probability plot, which endorses high model fit at the level of single subjects for all three conditions.

The appropriateness of chosen free parameter was verified through the assessment of alternative models for which we increased the number of free parameters and/or combined them (see Supplementary Table 1). Overall, all tested models revealed the importance of *t0* for predicting RT distribution and multisensory benefit. The predictions were further improved when *t0* was combined with *sv* and *st0* (*i.e.:* inter-trial variability drift rate and non-decisional constant). Critically, in all best-fitting models parameter estimates significantly differed between conditions: with *t0* and *st0* being systematically smaller in the Audio-Visual condition, while *sv* being systematically greater. These analyses confirmed that in the categorization task, faster RTs in multisensory cue trials are

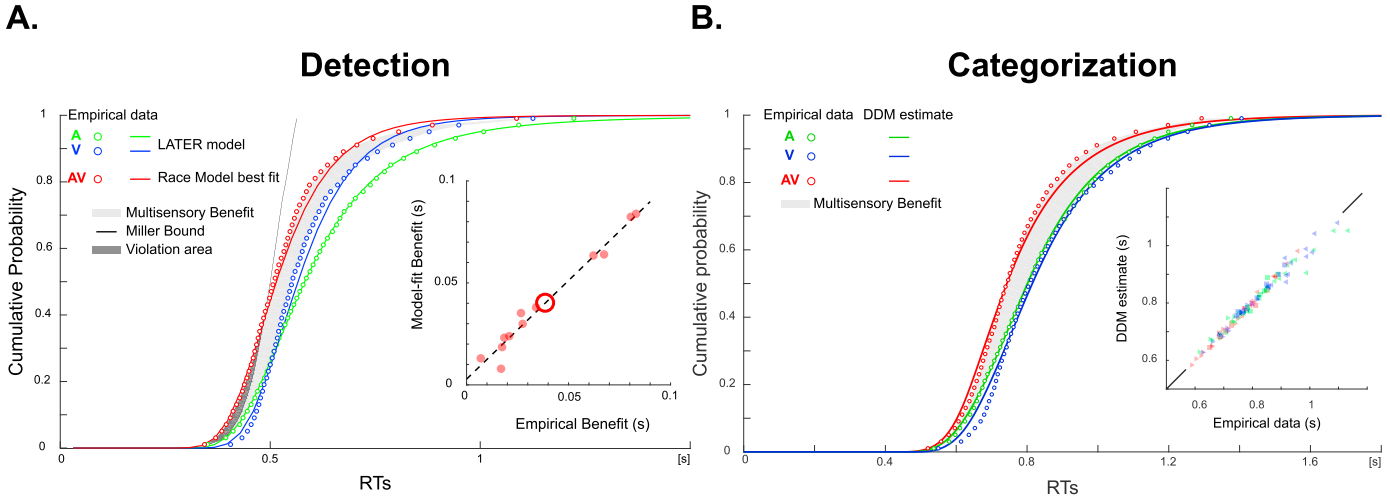

Fig. 2. Response time modeling. (A) Detection task: race models. Circles represent the cumulative distribution function of RT averaged across participants for the three conditions (Auditory in green, Visual in blue and Audio-Visual in red). Light gray shaded area represents the empirical multisensory benefit that is the difference between the multisensory condition and the faster of the two unisensory conditions. The black line represents Miller's bound as defined by *race model inequality*, with the dark gray shaded area highlighting violation. The red line depicts the best-fitting *context variant race model*. This model is constrained by the fits of the LATER model from the unisensory conditions (green and blue line for the auditory and the visual condition respectively) and two free parameters: the correlation (Rho = 0.35) and the variance (Eta = 0.065). Note that for illustration purpose the model was here computed on RT data averaged across participants. Inset: goodness of fit at the single subject level depicted as the correlation between empirical multisensory benefit and prediction from the best-fitting *context variant race model* (each dot represent a participant, the big red circle represents the population mean). The black dotted line represents the least-square regression. (B) Categorization task: diffusion models. Circles represent the cumulative distribution function of RT for the three conditions. Light gray shaded area represents the empirical multisensory benefit. The lines represent the best-fitting diffusion models (DDM) for each condition. Inset depicts quantile-probability plots that is the link between empirical and predicted RT quantiles (25, 50 and 75% respectively depicted by triangle facing right, square and triangle facing left). Each symbol represents the quantile of a participant. The black lines at the extremities indicate the diagonal (*i.e.* perfect correlation).

Table 1
Comparison of the free parameters between conditions for the main DDM (mean ± mean absolute deviation). Statistical analyses show significant differences between the multisensory condition and the unisensory conditions for *t0* (*i.e.* non-decisional time constant), indicating that the duration of non-decisional processes is reduced in the multisensory trials as compared to unisensory trials. While *st0* (*i.e.* inter-trial variability of non-decisional time constant) was significantly reduced for multisensory trials, *sv* (*i.e.* drift rate inter-trial variability) was significantly greater.

Parameters	AV	A	V	statistics
t0	578.5 ± 40.9	627.7 ± 31.2	638.2 ± 49.5	$\chi^2(2) = 7.66$ $p = 0.0217$
v	0.12 ± 0.09	0.15 ± 0.07	0.16 ± 0.11	$\chi^2(2) = 0.42$ $p = 0.8122$
st0	0.96 ± 0.32	0.68 ± 0.34	0.38 ± 0.21	$\chi^2(2) = 11.37$ $p = 0.0034$
sv	0.15 ± 0.04	0.29 ± 0.06	0.22 ± 0.08	$\chi^2(2) = 13.41$ $p = 0.0012$

explained in the DDM by shorter and less variable non-decision time; and by an increased variance of the drift rate.

3.4. Classical ERP-analysis

EEG data were first analyzed by means of a traditional approach using ERPs. For each condition and each task, evoked activity was marked by a first focal negative topography followed by the emergence of a strong positive bilateral component extending over centro-parietal electrodes (Fig. 3). Interestingly the location of the early negative component was contingent on the condition. In the auditory condition, two peaks occurred over left and right central electrodes, while in the visual condition, the peak was located over posterior midline electrodes. In the multisensory condition, the early negative topography was a combination of visual and auditory topographies. To assess if this multisensory response was simply equivalent to the sum of unisensory responses, we tested the additive model. Statistical comparison of the multisensory condition to the sum of unisensory conditions revealed significant differences already at 150 ms after the cue fading-onset, which coincided with the early negative component. This multisensory effect lasted until response was made. However, comparing the same time point across conditions does not permit to determine if the later significant difference (*i.e.* after 300 ms) results from a genuine amplitude difference across conditions or if it reflects faster overall brain dynamics in the multisensory condition (*i.e.* temporal shift related to early multisensory integration). To dissociate between these two possibilities, we computed the topographical spatial correlation between unisensory and multisensory conditions over time. Spatial correlation matrix revealed higher correlation above the diagonal indicating that similar topographies are elicited at earlier latencies in the multisensory condition as compared to unisensory conditions (Fig. 4). This asymmetry was tested statistically by comparing the summed spatial correlation values above and below the diagonal. For both tasks, this analysis confirmed higher correlation above the diagonal (Detection: $t(11) = 2.68$, $p = 0.021$; Categorization: $t(11) = 5.36$, $p = 0.0002$). Topographical spatial correlation analysis highlights the difficulty to investigate later stages of multisensory integration and therefore to establish if multisensory integration occurs during decision formation. Therefore we turned to supervised machine learning analysis approach which enabled us to characterize brain operations by optimally combining information from every electrode and to compare activation pattern across time points and conditions.

3.5. EEG decoding analysis of sensory encoding and decision formation following unisensory signal cue

To characterize the functional processing stages at play following unisensory cue we used a series of supervised machine learning analysis. Firstly, a linear multivariate classifier was trained to distinguish trials containing unisensory signal cues (auditory/visual) embedded in audio-visual noise from trials containing only audio-visual noise. Time resolved decoding performance gradually increased above chance level following unisensory cue onset, peaked and returned to chance level (blue curves in Fig. 5 A and B). Secondly, we performed a non-binary classification to decode correct unisensory trials as a function of response times (RT) and thus evaluate the formation of decision over time. Compared to the first classifier, classification performance of this RT-based classifier rose later above chance but peaked at the same latency that is before the behavioral response was made (purple curve in Fig. 5 A and B). To facilitate the comparison between the results from the two classifications and further delineate the temporal extent of brain processes related to perceptual decision making, we computed the topographical representation of classifier weights (activation patterns) and performed temporal generalization. Temporal generalization matrix is obtained by testing across all time points a decoder trained at a given time point and thereby characterizes canonical motif of neural operation (*e.g.* sustained, chained or reactivated see (King and Dehaene, 2014) for a comprehensive review).

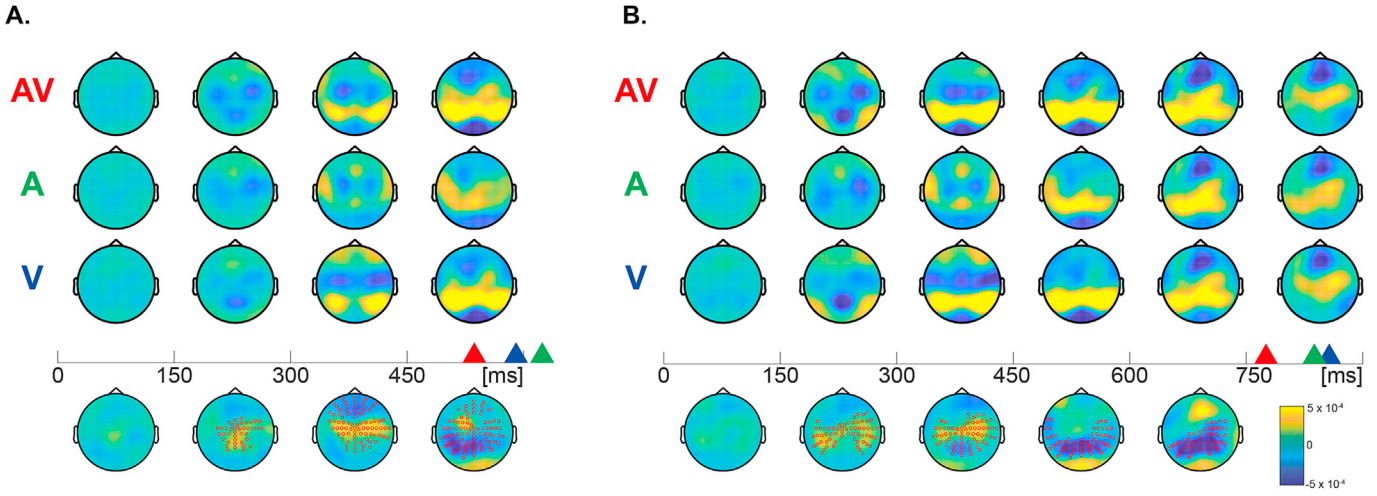

Fig. 3. Classical ERP-analysis. Evoked responses for Audio-Visual, Auditory and Visual signal cue trials (from top to bottom) in the detection task (A) and the categorization task (B). Mean topographies are depicted every 150 ms (averaged over the corresponding window). Corresponding subject-average mean RTs are represented by triangle on the time axis (AV in red, A in green and V in blue). The lower row represents topographical difference between (AV + catch) and (A + V). Red circles correspond to electrodes showing significant amplitude difference following statistical cluster analysis.

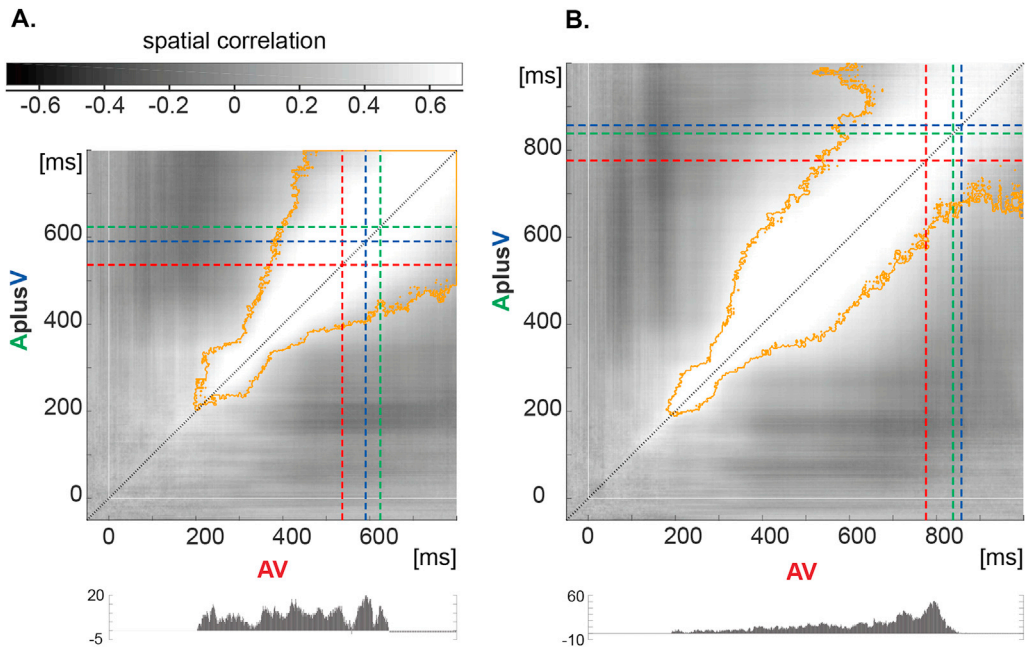

Fig. 4. Topographical spatial correlation. For both the detection task (A) and the categorization task (B), the matrices depict the spatial correlation computed for every time point between multisensory (x-axis) and unisensory topographies (y-axis), that is Audio-Visual versus Auditory plus Visual signal cue trials. Dotted lines represent mean RTs for Audio-Visual (in red), Auditory (in green) and Visual (in blue) conditions. Orange contours represent significant topographical correlation following statistical cluster analysis.

Below the matrices is depicted the number of significant voxels above the diagonal minus the number of voxels below the diagonal at the same latency (*i.e.* axis perpendicular to the diagonal). The overall higher correlation above the diagonal indicates that similar topographies are elicited earlier in the multisensory condition.

In both tasks, temporal generalization matrix captured by the unisensory signal cue vs. noise classifier revealed two *islands* of significant decoding, an early and a late, each being associated with a specific activation pattern (Supplementary Figure. 1 A and C). The first one consisted in three prominent negativities, similar to the modality specific components found in the classical-ERP analyses (see Fig. 3). The second stage consisted in a positive centro-parietal topography similar to the second component rising before response time in the classical-ERP analyses. This second stage was also isolated by the RT-based classifier, where only one late *island* of significant decoding was found and peaked about hundreds of milliseconds before behavioral response indicating its role in the preceding decision process (Supplementary Figure. 1 B and D). In summary, our dual decoding procedure was employed: firstly, to characterize the full dynamics of activation patterns; and secondly, to highlight the brain activity predicting RT. This twofold approach permitted to draw a chain of two processes, partly overlapping, from sensory encoding to decision formation.

3.6. Decoding generalization to multisensory signal cues reveals acceleration at both sensory and decision stages

Next, we sought to uncover the neural mechanisms underlying behavioral performance benefit following multisensory cues by applying a cross-condition decoding approach. Classifiers derived from unisensory trials were used to discriminate trials containing multisensory signal cues versus trials containing only audio-visual noise. Moreover, to account for possible temporal differences across conditions and thus accommodate the decoding of brain operations happening at different latencies we

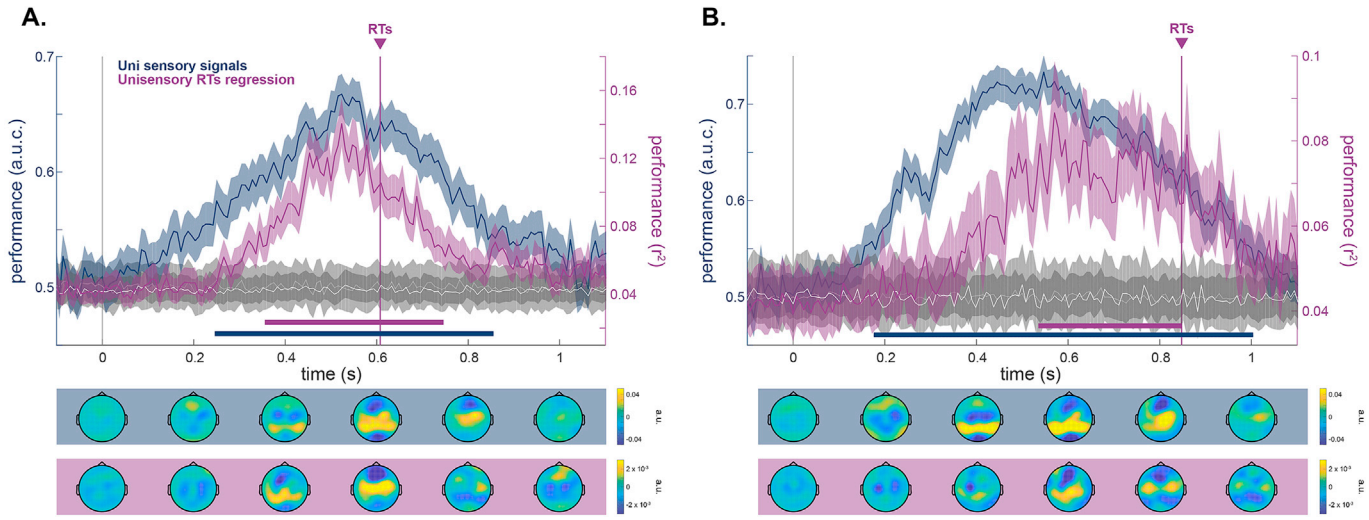

Fig. 5. Time-resolved pattern classification of unisensory signal cue encoding and decision formation for the detection task (A) and the categorization task (B). Mean decoding performance as a function of time relative to cue onset for the two different decoders (±95% mean confidence intervals). In dark blue: decoding of unisensory signal cue trials against noise trials. In purple: decoding based on response times. For both, chance levels are depicted in gray. Vertical line represents subject-average response time. Color-coded shapes below the x-axis signify statistically significant decoding performance compared to the chance level (corrected for multiple comparisons across time, $p < .005$). Topographical plots underneath depict activation patterns corresponding to the classifier weights (mean over 200 ms time window).

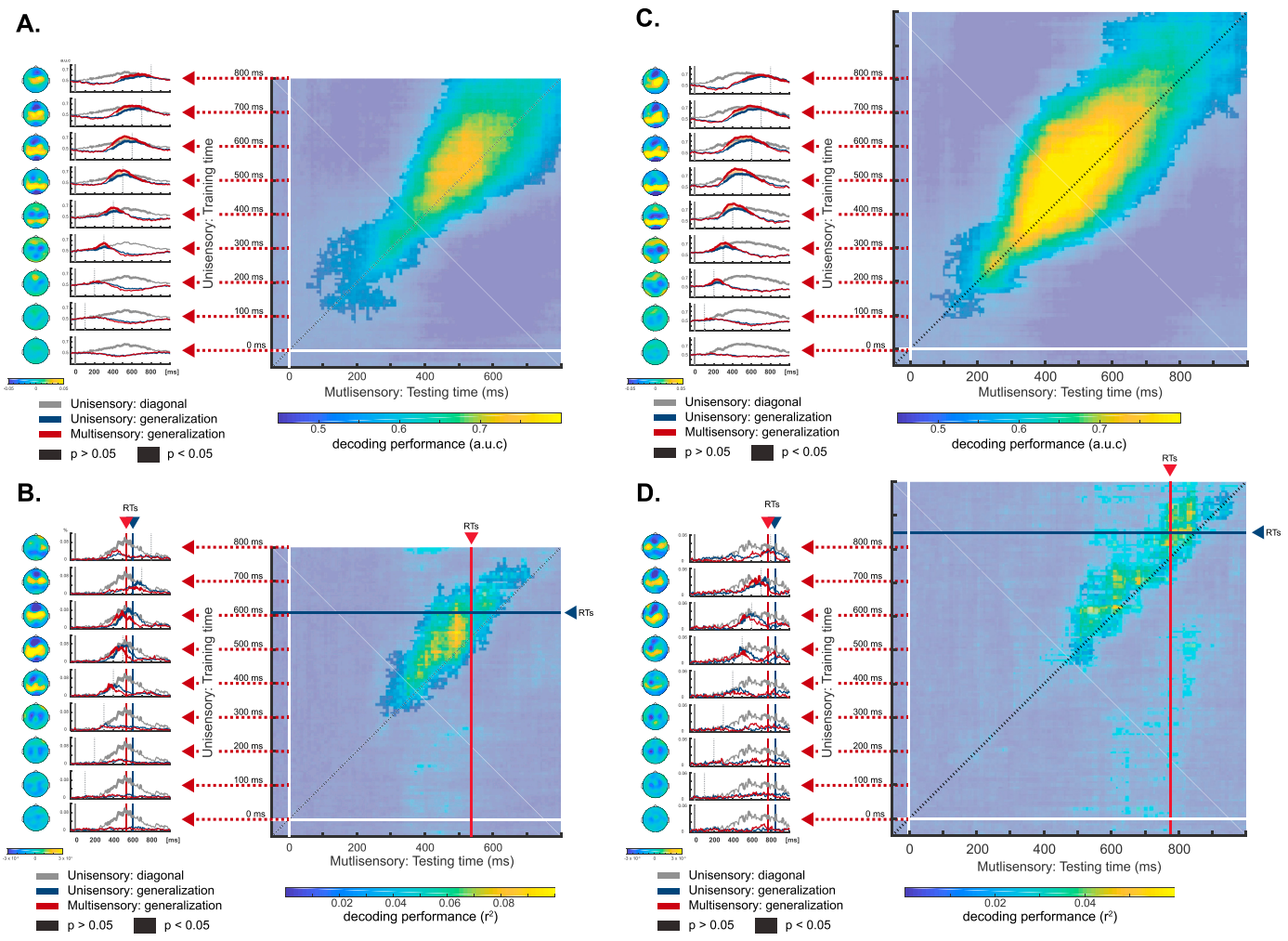

Fig. 6. Generalization from unisensory to multisensory condition for the detection task (left panels) and the categorization task (right panels). Classifiers trained on unisensory trials were tested either to distinguish multisensory signal cue trials vs. trials containing audio-visual noise only (A and C), or to decode response times of multisensory-cue trials (B and D). For each panel the right-side plot shows the temporal generalization in time across conditions (i.e. a classifier trained at a given time (y-axis) is tested at every other time moment (x-axis)). Blue horizontal and red vertical lines represent mean response times for unisensory trials and multisensory trials respectively. The left-side plot of each panel shows at every 100 ms of the training time (with 0 ms on the bottom) the activation pattern and corresponding decoding performance of the classifier trained-tested on unisensory trials (in blue) and then tested on multisensory trials (in red). Thicker line corresponds to decoding performance higher than chance. Orange areas indicate periods of significant process acceleration in the multisensory condition as to compare to unisensory condition. The gray lines represent decoding performance for unisensory trials along the diagonal of the unisensory temporal generalization matrix (dotted line along the diagonal in Supplementary Figure 1).

further tested the capacity of classifiers to generalize across time. This approach revealed that classifiers trained on unisensory trials were able to decode multisensory trials effectively, indicating that unisensory and multisensory decision making follows a similar chain of events. Furthermore, matrix generalization unveiled an off-diagonal pattern: unisensory classifiers led to higher decoding performance at earlier latencies for multisensory trials (Fig. 6, right column of each panel). As evidenced by classification against audio-visual noise (Fig. 6 A and C) and RT-based decoding (Fig. 6 B and D), this acceleration pattern occurred during both sensory encoding and decision formation stages. In order to verify that the acceleration significantly increased with time we calculated the distance between the multisensory and unisensory decoding time courses every 100 ms (see orange areas in Fig. 6, middle column of each panel). This integral between significant decoding performances was found to increase linearly with time (Pearson correlation: detection task: $p = 0.027$, $rho = 0.74$; categorization task: $p = 0.023$, $rho = 0.72$). This measure is equivalent to estimating the difference between decoding performance of multisensory versus unisensory generalization matrix and to show that this difference increases over time (*i.e.* distance to the diagonal increases along time). Thus, the present results demonstrate that combination of multisensory-cues speeds-up neural dynamic all along the course of processing and thereby strongly supports the view that multisensory integration processes are at play during sensory encoding and during decision formation (third hypothesis depicted in Fig. 1 A).

To verify that pooling unisensory trials in the decoding procedure does not undermine early sensory processing, we performed an alternative decoding procedure, in which auditory and visual trials were decoded separately (either against catch trials or to decode RT). Then the weights obtained from the two unisensory conditions were summed to decode multisensory trials (time generalization across conditions). Combining unisensory decoding led to equivalent results (see Supplementary Figure 3).

Temporal generalization across conditions illustrates the chain of processes common to unisensory and multisensory conditions. However, neural activity specific to multisensory processing (*i.e.,* multisensory integration) was not targeted using this approach. To examine if the acceleration of brain network activation in multisensory-cues

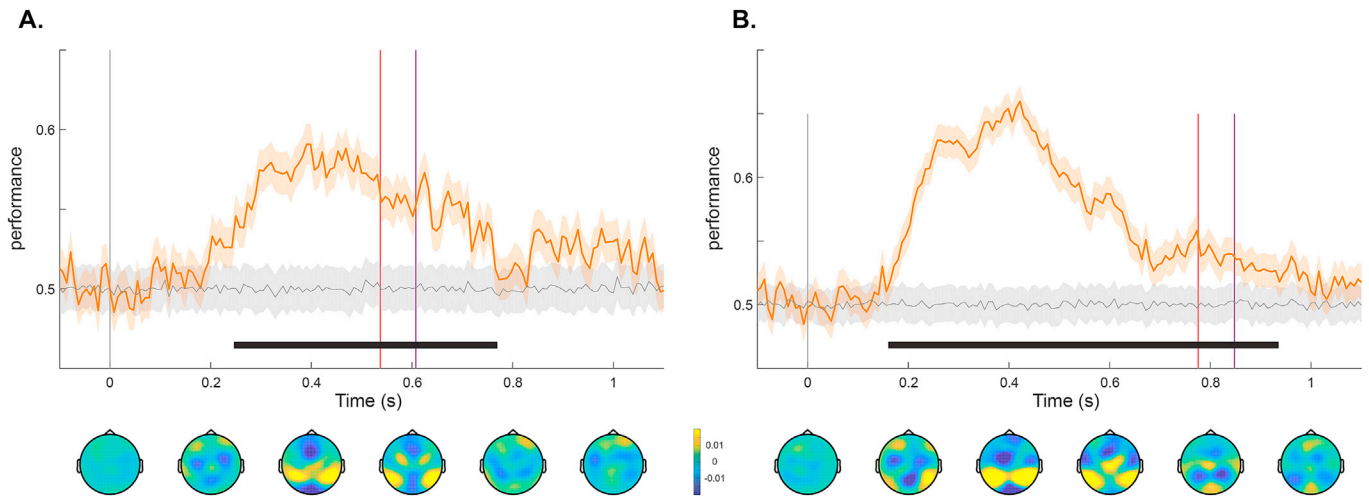

Fig. 7. Time resolved decoding of multisensory integration for the detection task (left panels) and the categorization task (right panels). Mean classification performance (orange line) as a function of time (+/− CI mean; chance levels in gray), with the black shape under the x-axis indicating statistically significant decoding performance compared to the chance level (corrected for multiple comparisons across time, $p < .005$). Topographical maps underneath depict the activation patterns from time-resolved classification procedure (mean over 200 ms time window). Vertical lines represent mean response times for multisensory trials (in red), and unisensory trials (in purple).

condition is the only difference with unisensory-cue condition, we compared the temporal generalization across conditions with direct decoding of multisensory-cues condition. That is, we trained classifiers to distinguish multisensory signal cues trials from audio-visual noise only (multisensory classifiers). Thereafter, we compared decoding performance of multisensory classifiers with the highest off-diagonal performance from the generalization matrix across time and condition (depicted in Fig. 6). Comparison of indirect and direct decoding showed a similar time-course (Supplementary Figure 2). However, during an early and a later time period, decoding performance of multisensory classifiers (direct approach) was significantly higher than that obtained from temporal generalization across condition (indirect approach); even though temporal differences between conditions were accounted for. This result indicates that decoders built from multisensory trials extracted neural activity which could not be captured by the decoders build from unisensory trials alone, neural activity likely related to the integration of multisensory-cues (*i.e.*, related to multisensory integration processes). Conversely, it can be argued that the higher decoding accuracy obtained from the direct multisensory classifiers as compared to the indirect ones (*i.e.* trained on unisensory conditions) provides only partial evidence that the direct classifier captured genuine multisensory integration processes. It could be the case that the direct approach gave better decoding performance even though the EEG patterns in the multisensory condition is purely additive (*i.e.* not integrative). To address this issue we performed a direct comparison between unisensory and multisensory conditions.

3.7. Time-resolved decoding of multisensory integration and links to sensory and decision stages

To directly tackle multisensory integration, we trained classifiers to differentiate between correct multisensory-cues trials and correct unisensory-cue trials. Time-resolved pattern classification showed above chance performance starting about 200 ms after the cue fade-in onset, characterized by two main activation patterns (Fig. 7). The earliest period presented three negative peaks while the later period showed a bilateral parietal positivity. To assess if the late period, showing the highest decoding performance, was caused by a time lag between conditions (due to the acceleration revealed by the temporal generalization across conditions, see above), we ran the same analysis using trials time-locked to the response. This second analysis confirmed the significance of the latest period characterized by the bilateral parietal positivity (Supplementary Figure 4). As a control analysis, we tried to decode missed multisensory-cues trials versus missed unisensory-cue trials and did not find above chance performance. Such null result sets a parallel between classifier efficiency in decoding multisensory signal and the possible absence of multisensory integration when signal cues were not integrated effectively.

We then aimed at evaluating how much the late multisensory process depends on the early multisensory process. For each participant, we identified the classifier weights leading to the highest decoding performance between correct multisensory-cues trials and correct unisensory-cue trials, during the early (0–300 ms) and the late period (300 ms to the response) and applied these weights to the multisensory-cue trials. This procedure provided us with single-trial estimates of the early and late multisensory classifier over time, which in turn, allowed us to evaluate the relationship between their maximum strength. This correlation, computed at the single trial level, revealed at the group level a link between the strength of early and late multisensory integration classifiers (detection task: $t(11) = 4.64$, $p = 7.2 \times 10–4$; categorization task: $t(11) = 2.02$, $p = 0.068$). In summary, the first multisensory processing stage fosters the later multisensory processing stage.

After characterizing two main activation patterns related to multisensory integration (Fig. 7), we assessed how they were associated with the identified processing stages of perceptual decision making from decoding unisensory trials: sensory encoding and decision formation stages. For each participant, and each type of decoding (*i.e.,* decoding sensory encoding, decision formation and early/late multisensory integration), we identified the classifier weights leading to the highest decoding performance of the multisensory trials, during an early (0–400 ms) and a late period (200 ms to the response) and applied these weights to the multisensory-cues trials. Note that here we allowed for an overlap between stages to avoid a strict sequential perspective over processing stages. This procedure provided us with single-trial estimates of each decoder type over time, which in turn allowed us to evaluate the relationship between the corresponding cognitive processes (early/late multisensory integration and sensory encoding/decision formation). That is if the signal strength of sensory encoding and the signal strength of decision formation were linked to the strength of multisensory integration. This comparison performed for each subject at the group level revealed a significant linear relationship between early multisensory integration and sensory encoding (detection task: $t(11) = 7.68$, $p = 9.6 \times 10–6$; categorization task: $t(11) = 5.35$, $p = 2.3 \times 10–4$), and late multisensory integration and decision formation

(detection task: $t(11) = 5.61$, $p = 1.6 \times 10–4$; categorization task: $t(11) = 5.12$, $p = 3.3 \times 10–4$). Overall these results revealed a link between the cognitive processes isolated by the different classifiers: the strength of early multisensory integration was related to the strength of sensory encoding, while the strength of later multisensory integration was related the strength of decision formation and to response time variations.

4. Discussion

Sensory uncertainty is resolved faster and more accurately using multisensory cues as compared to unisensory cues. However, the origin of such benefit is still unclear and could arise from multisensory integration during sensory encoding and/or during decision formation. In the present study, we aimed to pinpoint the critical processing stages during which multisensory integration influences perceptual decision making. First, modeling of behavioral data revealed the influence of multisensory integration on decision formation process. Next, using a supervised machine learning approach on EEG data, we identified specific patterns associated with two functional stages of perceptual decision making. Then, we demonstrated that multisensory benefit originates from the acceleration of both processing stages: sensory encoding and decision formation. Finally, we uncovered that the acceleration of these processes was not simply related to the speed-up of the unisensory dynamics but tightly linked to distinct multisensory integration processes, which were at play both during sensory encoding and decision formation.

4.1. Multisensory integration at play during perceptual decisional making, evidences at the behavioral level

The hallmark of multisensory integration is behavioral outcome, which typically translates in higher accuracy and faster response times. In the present study, we observed multisensory benefits in two tasks: when participants were asked to detect or to categorize multisensory cues their accuracy and sensitivity were higher in the context of multisensory as compared to unisensory cues; and multisensory gain was accompanied by faster RTs (Fig. 1). To account for behavioral response variability in unisensory contexts, research on perceptual decision making has converged on an elemental computational architecture (Bogacz, 2007; Smith and Ratcliff, 2004). In this framework, ongoing noisy activity at the sensory level is minimized by the accumulation of evidence over time, which leads to appropriate motor response execution when a response criterion is reached. Within this framework a key question resides in determining the origin of multisensory RT speed-up.

In detection tasks, multisensory integration has been formulated as follows. Either sensory evidence originating from two modalities is accumulated by two 'racing' parallel decision processes: the fastest reaching the threshold criterion triggers the response (*i.e.* race model statistical facilitation). Alternatively, sensory evidence originating from two modalities could be integrated in a single decision process, resulting in a more rapid accumulation of evidence towards the threshold criterion (*i.e.* co-activation model). The former hypothesis, so-called *race model*, has been extensively applied in multisensory research to estimate if multisensory RT acceleration can be explained by a simple statistical facilitation (Miller, 1982, see for review Godan et al., 2016).

On that base, we first compared RT distribution from the multisensory condition to *Miller's bound* and found that RT acceleration cannot be explained by a simple statistical facilitation (see Miller's bound violation in Fig. 2). This result suggested that either there is a unique decision process which integrates the two sensory signals (*i.e.* co-activation model), or the *race model* does not extend well to multisensory contexts and requires amendments to explain multisensory acceleration. Therefore, we next applied the *context variant race model* which introduces two interactions parameters to account for possible change(s) when multisensory cues are present (Otto and Mamassian, 2017). Best-fitting estimates obtained from the *context variant race model* demonstrated the relevance of both interaction parameters for accurate fit of the entire multisensory RT distribution and almost perfectly explained the empirical behavioral benefit (see Fig. 2 A inset). The first interaction parameter Rho, a correlation parameter, was significantly greater than zero, suggesting that a common source of variability contributed to the overall variability of latencies (*e.g.* vigilance fluctuation or motor act). The second interaction parameter, Eta, characterized a change in decision formation following multisensory cues. That is additional variability in the rate of evidence accumulation process was needed to accurately model multisensory RT distribution. This finding directly challenges the context invariance assumption of the classic race model, which states that the decision process related to one sensory signal is not affected by the decision process of the other sensory signal (see (Otto and Mamassian, 2017) for a thorough review). Therefore, when two parallel unisensory decision processes are assumed to take place, some interaction between them must be considered to model multisensory benefit. To summarize, our analyses of RT distribution in the detection task revealed that statistical facilitation alone cannot account for multisensory RT acceleration. Multisensory benefit can be explained by either a unique decision process that integrates the two sensory signals, or by two decision processes that interact. Critically, in either case multisensory integration is at play during decision formation.

Two alternative forced choice tasks have been widely utilized to investigate perceptual decision making in unisensory contexts (Gold and Shadlen, 2007; Ratcliff et al., 2016). The prominent Diffusion Decision Model (DDM) sketches binary choice as the accumulation of evidence evolving between two threshold criterions, one for each choice. The DDM parameters have been proposed to be psychologically meaningful and map onto different cognitive processes (Ratcliff and McKoon, 2008; Voss et al., 2013). We computed the DDM in unisensory and multisensory contexts and compared the derived parameters to investigate which of the cognitive processes are modulated by multisensory cues. We found that non-decisional time and its variability were significantly reduced in the multisensory condition. This result indicates that acceleration of responses originated from acceleration of non-decisional time(s) that is/are sensory encoding and/or motor preparation. Here, we cannot fully exclude that multisensory integration process could occur during motor preparation and this hypothesis merits to be tested in a dedicated paradigm – an important test by itself. Nonetheless, in light of our electrophysiological results (*i.e.* showing that early multisensory integration speeds-up sensory encoding), it is more likely that acceleration of behavioral multisensory responses stems largely from acceleration of sensory encoding. We further found that, while the DDM did not show differences for the mean accumulation rate between conditions, the variance of accumulation rate was greater in the multisensory condition as compared to unisensory conditions, and thus contributed to faster RT when sensory evidences can be gathered from multisensory cues (see below). Lastly, the analysis of RT distribution in the categorization task using the DDM model explained multisensory benefit by multisensory integration effect at different cognitive processes of perceptual decision making: during sensory encoding and during decision formation.

In summary, our investigation of individual RT distribution in both tasks revealed that inter-trial variability is a key component of RT speed-up in the multisensory condition. It appears that in the presence of two sensory cues, decision formation process presents a greater variability, which leads the accumulation of evidence to reach the decisional criterion earlier than when a single signal cue is available. This greater variability can be directly caused by the multiple sources of sensory evidence (from one or two sensory channels), and/or related to additional fluctuations within each sensory channel due to cross-modal modulation (*i.e.* the fact that early sensory cortex is modulated by signals from the non-preferred modality, see for instance: (Lemus et al., 2010)).

4.2. The course of perceptive decision making following unisensory-cue

Pioneering single-unit animal studies revealed that decision-related

activity recorded in the lateral intraparietal cortex and frontal eye fields paralleled temporal accumulation of sensory evidence leading to the behavioral response (Gold and Shadlen, 2007; Heekeren et al., 2008). Neural correlates of perceptual decision making have since been identified in multiple brain areas, with neural code of decision making process spanning across the cortical hierarchy: from sensory areas to parietal and frontal regions (de Lafuente and Romo, 2006; Siegel et al., 2015; Tsunada et al., 2016). In humans, a series of seminal EEG studies identified a brain signal, the centro parietal positivity (CPP), presenting many characteristics of a neural signature of decision formation (O'Connell et al., 2012; Twomey et al., 2016). In our study, the activation pattern and the time-course of decoding performance of the RT-based classifier (Fig. 5) highly resembles the topography and the temporal characteristics of the CPP: positive topography emerging in parietal regions and evolving towards the central regions, and progressive increase in amplitude peaking before the behavioral responses. These analogies further support the view that the RT-based classifier characterized the neural signal coding for decision formation.

Decoding of unisensory signal cues from noise unveiled two processing stages: sensory encoding and decision formation. The latter stage was equivalent to the decision formation stage characterized by the RT-based classifier, with a CPP-like activation pattern. In the early stage, the neural signal peaked about 150–200 ms after cue fade-in onset, when signal-to-noise was the largest in the stimulus sequence (Fig. 5). Its topography was dominated by three negative peaks comparable to the modality specific early activity observed in the classical ERP analysis (Fig. 3). To that regard, it is reminiscent of the well described N2 component typically related to sensory encoding process and modulated by attention when a target selection has to be performed (Gamble and Woldorff, 2015; Loughnane et al., 2016; Luck and Hillyard, 1994). Importantly, an early target selection component has been recently found to modulate the onset and the rise of CPP (Loughnane et al., 2016) and thereby appears as a processing step mediating decision formation in the context of unpredictable cues. It is important to note that while the decoding procedure highlighted the existence of two processing stages, it does not demonstrate that their relationship is purely serial, that is decision formation would not start until sensory encoding is completed. Rather the decoding time-courses from the two decoding approaches shows a transition period between the two stages where activation patterns differ: each classifier (against audio-visual noise and RT-based) captured the co-existent activities related to different cognitive processes. Thus, sensory encoding and decision formation are partly overlapping processes. Altogether, decoding analyses of unisensory-cue trials (*i.e.,* against audio-visual noise and RT-based) are concordant and trace the temporal trajectory of neural processes from sensory encoding to decision-related signal gradually building-up before the response. Finally, our results demonstrate that this chain of processes is similar across modalities, and thus complement the findings from human EEG/MEG studies in the visual domain (Mostert et al., 2016; Ratcliff et al., 2009; Wyart et al., 2012).

The topographical dynamics isolated by the two decoding approaches are complex, and we here focused on the topographies that were strongest and similar across detection and discrimination tasks. However, it is possible that our experimental design elicited additional activity, such as the Contingent Negativity Variation (CNV). This expectancy-based evoked activity is present when a warning signal precedes an imperative stimulus, which would correspond in our task design to the onset of the audio-visual stream and to the unpredictable signal cue respectively. CNV sustained activity is thought to be supramodal (Pasinski et al., 2016; Walter et al., 1964) and localized in anterior regions (Gómez et al., 2003; Hultin et al., 1996; Lamarche et al., 1995; Nagai et al., 2004; Rosahl and Knight, 1995). While it is possible that a CNV evolved alongside the main activities isolated by the classifiers, it is unlikely to drive all the findings reported here because: (1) we applied SCD transformation, which attenuates the volume conduction effects and thus minimizes the possible influence of CNV that is prominent over the frontal electrodes (as in (Kelly and O'Connell, 2013)); (2) we demeaned and baseline-corrected the data, which reduces the CNV activity that could be present before the cue fade-in onset (*i.e.* the fore-period). Most importantly, the effect of CNV, if any, should be most prominent in the target absent condition (*i.e.* catch trial with only audio-visual noise condition), which was not observed in our data.

4.3. Multisensory signal cues accelerate both sensory and decision processes

To portray the dynamic of brain processes following the presentation of multisensory-cues as compared to unisensory-cues, we performed a temporal generalization across conditions. The critical advantage of generalizing in time relies on the fact that unlike classical approach comparing the same time points between conditions, temporal generalization matrix provides comparisons between all time-points allowing to relate brain operations occurring at different latencies. Cross-condition decoding revealed that classifiers trained on unisensory trials were able to decode multisensory trials successfully: They were effective in both processing stages (*i.e.,* sensory encoding and decision formation) and their performance was high for the two types of decoding procedures (*i.e.,* against audio-visual noise trials and based on RT). This cross-condition decoding demonstrated that unisensory and multisensory decision making share the same trajectory. However, time generalization revealed that decoding performance of multisensory trials was more accurate and reached significance earlier than the ones they were trained on (*i.e.,* unisensory trials). Critically, this acceleration was not limited to a given period but increased with time. Thus, our results demonstrate for the first time that multisensory-cues accelerate neural processing dynamics during sensory encoding as well as during decision formation (as depicted in the third hypothesis depicted in Fig. 1 A).

Acceleration during sensory encoding was suggested by a body of work describing how multisensory integration influences early sensory processing (Cappe et al., 2012; Foxe and Schroeder, 2005; Kayser et al., 2008; Lakatos et al., 2007; Mercier et al., 2015, 2013; Romei et al., 2009; van Wassenhove et al., 2005). Based on this research, it could be hypothesized that in the case of congruent multisensory source of information, a speed-up at the sensory stage caused by multisensory integration would pass onto the decision formation stage (first hypothesis depicted in Fig. 1 A). This hypothesis is supported by a recent study that established a link between the N2 and the CPP: in the context of unpredictable source of information, the latency of early target selection signal *(i.e.* N2) preludes to faster RTs through earlier evidence accumulation as measured by the CPP (Loughnane et al., 2016).

Alternatively, RT acceleration in multisensory-cue context could result from a faster decision rate: either by a higher mean rate or a greater variance (see behavioral modeling results and second hypothesis depicted in Fig. 1 A). Acceleration during decision formation was also suggested by works on perceptual decision making, in which decision formation rate, or variability, has been shown to vary with the amount of sensory evidence (Hanks and Summerfield, 2017; Heekeren et al., 2008; O'Connell et al., 2018; Romo and de Lafuente, 2013). Our study reconciles these two non-mutually exclusive hypotheses by showing that multimodal cues speedup sensory encoding (which leads to an earlier decision formation) and decision formation is further accelerated by multisensory evidence (third hypothesis depicted in Fig. 1 A). This acceleration of neural processes during both sensory encoding and decision formation highly suggests that multisensory integration is at play during each processing stage.

4.4. Multisensory integration arises during sensory encoding and decision formation

Multisensory interactions are pervasive in human brain and complete different processes along the cortical hierarchy (Ghazanfar and Schroeder, 2006; Rohe and Noppeney, 2015; Werner and Noppeney, 2010). Sensory regions are the earliest cortical stages of multisensory convergence (Cappe

et al., 2009; Foxe and Schroeder, 2005). In these areas, neural activity is modulated by cross-modal inputs (Kayser et al., 2008; Lakatos et al., 2007; Mercier et al., 2015, 2013). These early modulations mainly relate to low level features of the different sensory inputs, such as their co-occurrence in a short temporal window and/or in a small region of space. Sensory regions also closely interact with higher order areas (*e.g.,* temporal, parietal and frontal associative cortex) which mediate integration processes at a higher level; for instance, the congruency of multisensory signals, their reliability or task relevance (Kayser et al., 2017; Noppeney et al., 2010; Rohe and Noppeney, 2016). Critically, while associative cortices are largely multisensory, at the same time they play an important role in decision making (Bizley et al., 2016; Ghazanfar and Schroeder, 2006; Romo and de Lafuente, 2013).

To examine neural activity that was specific to the integration of multisensory-cues leading to behavioral benefit, we used a linear classifier to decode multisensory-cues trials from unisensory-cue trials. We found that the temporal and spatial neural signatures of multisensory integration were concomitant with sensory encoding and decision formation stages as identified by decoding unisensory signal cues. The earliest period of multisensory integration was characterized by a topography consisting of three negative peaks – a topography comparable to early sensory encoding signal (*i.e.,* N2 see above), while the later multisensory integration period was characterized by a centro-parietal positivity akin to decision formation signal (*i.e.,* CPP see above). Moreover, the strength of early and late activation patterns were correlated respectively with the strength of sensory encoding and decision formation (see Results section 3.7), implying a functional link between these processing steps and multisensory integration.

Our finding of distinct temporally unfolding multisensory integration processes are complementary to two recent studies investigating a causal inference during multisensory perception (Cao et al., 2019; Rohe et al., 2019). These studies dissected the neural dynamics of hierarchical causal inference and uncovered different multisensory computations taking place at multiple timescales within distinctive brain networks. They suggested that, at first, a forced-fused representation is built, which corresponds to early, non-specific, multisensory interaction during the encoding process of sensory signals. Next, this fused representation and the primary segregated representations are weighted, as a function of reliability and/or disparity, to account for context-dependent multisensory integration. This strategy gives more flexibility, as it allows differentiating common and independent sensory sources through co-existing computational solutions: fusion and segregation. In the context of the present study, signal cues are embedded in a continuous flow of audio-visual noise from which they must be segregated. It follows that the early period of multisensory integration (*i.e.* occurring during sensory encoding) may correspond to the aforementioned fused representation, which is then evaluated in a context dependent manner during the later period of multisensory integration that is happening during decision formation. This perspective further implies a causal link between initial representations and their subsequent context-dependent weighting, as the later relies on the former. We verified this implication using single-trial based correlations between the initial multisensory encoding process and the later multisensory decision process; the result shows a degree of dependency, and therefore advocates for the hierarchical nature of multisensory integration.

Accordingly, our result substantiates the existence of distinct multisensory processes shaping distinct computational stages and demonstrates for the first time an interplay between multisensory integration and decision making, not only by accelerating sensory processing and thereby advancing the onset of decision formation, but also by concurrently fostering decision formation. As such, multisensory integration appears as a crucial factor in perceptual decision making which should be taken into account for building a complete understanding of this multifaceted process.

5. Conclusion

In the present study we employed behavioral modeling and supervised machine learning to demonstrate that multisensory integration and perceptual decision making are nested, and not sequential processes. By applying diffusion models appropriate for each task (*i.e.* detection and categorization), we found consistent indications that multisensory interaction occurs during both perceptual decision making stages. Next, using EEG decoding, we tracked the processing stages (*i.e.,* sensory encoding and decision formation) of perceptual decision making following unisensory-cues. From there, we applied cross-condition temporal generalization decoding and demonstrated that both sensory encoding and decision formation stages were accelerated. Finally, we characterized multisensory integration and revealed that early and late periods of multisensory integration were tightly linked to sensory encoding and decision formation, respectively, with the later period relying on the former. These results were reproducible across both detection and discrimination tasks. In conclusion, our study demonstrates that multisensory signals foster decision making by accelerating both sensory encoding stage and decision formation stages.

Declaration of competing interest

The authors declare no competing financial interests.

Credit authorship contribution statement

Manuel R. Mercier: Conceptualization, Data curation, Formal analysis, Funding acquisition, Project administration, Investigation, Resources, Methodology, Software, Validation, Visualization, Writing - original draft, Writing - review & editing. **Celine Cappe:** Funding acquisition, Project administration, Resources, Writing - review & editing.

Declaration of competing interest

The authors declare no competing financial interests.

Acknowledgments

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie-Sklodowska-Curie grant agreement No 798853. European Commission, Research Executive Agency, Individual Fellowships, Reintegration panel, project MIMe (H2020 MSCA - IF 798853 - MIMe to M.R.M.). This work also supported by the Agence Nationale de la Recherche (IBM 12-PDOC-0008 to C.C.). The authors would like to express their gratitude to Rasa Gulbinaite and Sébastien Crouzet for their helpful insights.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.neuroimage.2020.116970.

References

Atilgan, H., Town, S.M., Wood, K.C., Jones, G.P., Maddox, R.K., Lee, A.K.C., Bizley, J.K., 2018. Integration of visual information in auditory cortex promotes auditory scene analysis through multisensory binding. Neuron 97, 640–655. https://doi.org/10.1016/j.neuron.2017.12.034 e4.

Besle, J., Fort, A., Giard, M.-H., 2004. Interest and validity of the additive model in electrophysiological studies of multisensory interactions. Cognit. Process. 5, 189–192. https://doi.org/10.1007/s10339-004-0026-y.

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

Bizley, J.K., Nodal, F.R., Bajo, V.M., Nelken, I., King, A.J., 2007. Physiological and anatomical evidence for multisensory interactions in auditory cortex. Cerebr. Cortex 17, 2172–2189. https://doi.org/10.1093/cercor/bhl128.

Bogacz, R., 2007. Optimal decision-making theories: linking neurobiology with behaviour. Trends Cogn. Sci. (Regul. Ed.) 11, 118–125. https://doi.org/10.1016/j.tics.2006.12.006.
Cao, Y., Summerfield, C., Park, H., Giordano, B.L., Kayser, C., 2019. Causal inference in the multisensory brain. Neuron 102, 1076–1087. https://doi.org/10.1016/j.neuron.2019.03.043 e8.
Cappe, C., Morel, A., Barone, P., Rouiller, E.M., 2009. The thalamocortical projection systems in primate: an anatomical support for multisensory and sensorimotor interplay. Cerebr. Cortex 19, 2025–2037. https://doi.org/10.1093/cercor/bhn228.
Cappe, C., Thelen, A., Romei, V., Thut, G., Murray, M.M., 2012. Looming signals reveal synergistic principles of multisensory integration. J. Neurosci. 32, 1171–1182. https://doi.org/10.1523/JNEUROSCI.5517-11.2012.
Cappe, C., Thut, G., Romei, V., Murray, M.M., 2010. Auditory-visual multisensory interactions in humans: timing, topography, directionality, and sources. J. Neurosci. 30, 12572–12580. https://doi.org/10.1523/JNEUROSCI.1099-10.2010.
Carpenter, R.H., Williams, M.L., 1995. Neural computation of log likelihood in control of saccadic eye movements. Nature 377, 59–62. https://doi.org/10.1038/377059a0.
Colonius, H., Diederich, A., 2006. The race model inequality: interpreting a geometric measure of the amount of violation. Psychol. Rev. 113, 148–154. https://doi.org/10.1037/0033-295X.113.1.148.
Crouzet, S.M., Busch, N.A., Ohla, K., 2015. Taste quality decoding parallels taste sensations. Curr. Biol. 25, 890–896. https://doi.org/10.1016/j.cub.2015.01.057.
de Lafuente, V., Romo, R., 2006. Neural correlate of subjective sensory experience gradually builds up across cortical areas. Proc. Natl. Acad. Sci. Unit. States Am. 103, 14266–14271. https://doi.org/10.1073/pnas.0605826103.
Edwards, G., VanRullen, R., Cavanagh, P., 2018. Decoding trans-saccadic memory. J. Neurosci. 38, 1114–1123. https://doi.org/10.1523/JNEUROSCI.0854-17.2017.
Fan, R.-E., Chang, K.-W., Hsieh, C.-J., Wang, X.-R., Lin, C.-J., 2008. LIBLINEAR: a library for large linear classification. J. Mach. Learn. Res. 9, 1871–1874.
Foxe, J.J., Schroeder, C.E., 2005. The case for feedforward multisensory convergence during early cortical processing. Neuroreport 16, 419–423.
Gamble, M.L., Woldorff, M.G., 2015. Rapid context-based identification of target sounds in an auditory scene. J. Cognit. Neurosci. 27, 1675–1684. https://doi.org/10.1162/jocn_a_00814.
Ghazanfar, A., Schroeder, C., 2006. Is neocortex essentially multisensory? Trends Cognit. Sci. 10, 278–285. https://doi.org/10.1016/j.tics.2006.04.008.
Gold, J.I., Shadlen, M.N., 2007. The neural basis of decision making. Annu. Rev. Neurosci. 30, 535–574. https://doi.org/10.1146/annurev.neuro.29.051605.113038.
Gómez, C.M., Marco, J., Grau, C., 2003. Preparatory visuo-motor cortical network of the contingent negative variation estimated by current density. Neuroimage 20, 216–224.
Gondan, M., Minakata, K., 2016. A tutorial on testing the race model inequality. Atten. Percept. Psychophys. 78, 723–735. https://doi.org/10.3758/s13414-015-1018-y.
Grootswagers, T., Wardle, S.G., Carlson, T.A., 2017. Decoding dynamic brain patterns from evoked responses: a tutorial on multivariate pattern analysis applied to time series neuroimaging data. J. Cognit. Neurosci. 29, 677–697. https://doi.org/10.1162/jocn_a_01068.
Hanks, T.D., Summerfield, C., 2017. Perceptual decision making in rodents, monkeys, and humans. Neuron 93, 15–31. https://doi.org/10.1016/j.neuron.2016.12.003.
Heekeren, H.R., Marrett, S., Ungerleider, L.G., 2008. The neural systems that mediate human perceptual decision making. Nat. Rev. Neurosci. 9, 467–479. https://doi.org/10.1038/nrn2374.
Hultin, L., Rossini, P., Romani, G.L., Högstedt, P., Tecchio, F., Pizzella, V., 1996. Neuromagnetic localization of the late component of the contingent negative variation. Electroencephalogr. Clin. Neurophysiol. 98, 435–448. https://doi.org/10.1016/0013-4694(96)95507-8.
Innes, B.R., Otto, T.U., 2019. A comparative analysis of response times shows that multisensory benefits and interactions are not equivalent. Sci. Rep. 9, 2921. https://doi.org/10.1038/s41598-019-39924-6.
Kayser, C., Petkov, C.I., Logothetis, N.K., 2008. Visual modulation of neurons in auditory cortex. Cerebr. Cortex 18, 1560–1574. https://doi.org/10.1093/cercor/bhm187.
Kayser, S.J., Philiastides, M.G., Kayser, C., 2017. Sounds facilitate visual motion discrimination via the enhancement of late occipital visual representations. Neuroimage 148, 31–41. https://doi.org/10.1016/j.neuroimage.2017.01.010.
Keil, J., Senkowski, D., 2018. Neural oscillations orchestrate multisensory processing. Neuroscientist 24, 609–626. https://doi.org/10.1177/1073858418755352.
Kelly, S.P., O'Connell, R.G., 2013. Internal and external influences on the rate of sensory evidence accumulation in the human brain. J. Neurosci. 33, 19434–19441. https://doi.org/10.1523/JNEUROSCI.3355-13.2013.
King, J.-R., Dehaene, S., 2014. Characterizing the dynamics of mental representations: the temporal generalization method. Trends Cognit. Sci. 18, 203–210. https://doi.org/10.1016/j.tics.2014.01.002.
King, J.-R., Pescetelli, N., Dehaene, S., 2016. Brain mechanisms underlying the brief maintenance of seen and unseen sensory information. Neuron 92, 1122–1134. https://doi.org/10.1016/j.neuron.2016.10.051.
Lakatos, P., Chen, C.-M., O'Connell, M.N., Mills, A., Schroeder, C.E., 2007. Neuronal oscillations and multisensory interaction in primary auditory cortex. Neuron 53, 279–292. https://doi.org/10.1016/j.neuron.2006.12.011.
Lamarche, M., Louvel, J., Buser, P., Rektor, I., 1995. Intracerebral recordings of slow potentials in a contingent negative variation paradigm: an exploration in epileptic patients. Electroencephalogr. Clin. Neurophysiol. 95, 268–276. https://doi.org/10.1016/0013-4694(95)00117-h.
Lemus, L., Hernández, A., Luna, R., Zainos, A., Romo, R., 2010. Do sensory cortices process more than one sensory modality during perceptual judgments? Neuron 67, 335–348. https://doi.org/10.1016/j.neuron.2010.06.015.
Loughnane, G.M., Newman, D.P., Bellgrove, M.A., Lalor, E.C., Kelly, S.P., O'Connell, R.G., 2016. Target selection signals influence perceptual decisions by modulating the onset and rate of evidence accumulation. Curr. Biol. 26, 496–502. https://doi.org/10.1016/j.cub.2015.12.049.
Luck, S.J., Hillyard, S.A., 1994. Electrophysiological correlates of feature analysis during visual search. Psychophysiology 31, 291–308.
Maris, E., Oostenveld, R., 2007. Nonparametric statistical testing of EEG- and MEG-data. J. Neurosci. Methods 164, 177–190. https://doi.org/10.1016/j.jneumeth.2007.03.024.
Marti, S., King, J.-R., Dehaene, S., 2015. Time-resolved decoding of two processing chains during dual-task interference. Neuron 88, 1297–1307. https://doi.org/10.1016/j.neuron.2015.10.040.
Mercier, M.R., Foxe, J.J., Fiebelkorn, I.C., Butler, J.S., Schwartz, T.H., Molholm, S., 2013. Auditory-driven phase reset in visual cortex: human electrocorticography reveals mechanisms of early multisensory integration. Neuroimage 79, 19–29. https://doi.org/10.1016/j.neuroimage.2013.04.060.
Mercier, M.R., Molholm, S., Fiebelkorn, I.C., Butler, J.S., Schwartz, T.H., Foxe, J.J., 2015. Neuro-oscillatory phase Alignment drives speeded multisensory response times: an electro-corticographic investigation. J. Neurosci. 35, 8546–8557. https://doi.org/10.1523/JNEUROSCI.4527-14.2015.
Miller, J., 1982. Divided attention: evidence for coactivation with redundant signals. Cognit. Psychol. 14, 247–279. https://doi.org/10.1016/0010-0285(82)90010-X.
Mostert, P., Kok, P., de Lange, F.P., 2016. Dissociating sensory from decision processes in human perceptual decision making. Sci. Rep. 5 https://doi.org/10.1038/srep18253.
Murray, M.M., Brunet, D., Michel, C.M., 2008. Topographic ERP analyses: a step-by-step tutorial review. Brain Topogr. 20, 249–264. https://doi.org/10.1007/s10548-008-0054-5.
Myers, N.E., Rohenkohl, G., Wyart, V., Woolrich, M.W., Nobre, A.C., Stokes, M.G., 2015. Testing sensory evidence against mnemonic templates. Elife 4, e09000. https://doi.org/10.7554/eLife.09000.
Nagai, Y., Critchley, H.D., Featherstone, E., Fenwick, P.B.C., Trimble, M.R., Dolan, R.J., 2004. Brain activity relating to the contingent negative variation: an fMRI investigation. Neuroimage 21, 1232–1241. https://doi.org/10.1016/j.neuroimage.2003.10.036.
Noorani, I., Carpenter, R.H.S., 2016. The LATER model of reaction time and decision. Neurosci. Biobehav. Rev. 64, 229–251. https://doi.org/10.1016/j.neubiorev.2016.02.018.
Noppeney, U., Ostwald, D., Werner, S., 2010. Perceptual decisions formed by accumulation of audiovisual evidence in prefrontal cortex. J. Neurosci. 30, 7434–7446. https://doi.org/10.1523/JNEUROSCI.0455-10.2010.
O'Connell, R.G., Dockree, P.M., Kelly, S.P., 2012. A supramodal accumulation-to-bound signal that determines perceptual decisions in humans. Nat. Neurosci. 15, 1729–1735. https://doi.org/10.1038/nn.3248.
O'Connell, R.G., Shadlen, M.N., Wong-Lin, K., Kelly, S.P., 2018. Bridging neural and computational viewpoints on perceptual decision-making. Trends Neurosci. 41, 838–852. https://doi.org/10.1016/j.tins.2018.06.005.
Otto, T.U., Mamassian, P., 2017. Multisensory decisions: the test of a race model, its logic, and power. Multisensory Res. 30, 1–24. https://doi.org/10.1163/22134808-00002541.
Otto, T.U., Mamassian, P., 2012. Noise and correlations in parallel perceptual decision making. Curr. Biol. 22, 1391–1396. https://doi.org/10.1016/j.cub.2012.05.031.
Pasinski, A.C., McAuley, J.D., Snyder, J.S., 2016. How modality specific is processing of auditory and visual rhythms? Psychophysiology 53, 198–208. https://doi.org/10.1111/psyp.12559.
Perrin, F., Pernier, J., Bertrand, O., Echallier, J.F., 1989. Spherical splines for scalp potential and current density mapping. Electroencephalogr. Clin. Neurophysiol. 72, 184–187. https://doi.org/10.1016/0013-4694(89)90180-6.
Raab, D.H., 1962. Division OF psychology: statistical facilitation OF simple reaction times*. Trans. N. Y. Acad. Sci. 24, 574–590. https://doi.org/10.1111/j.2164-0947.1962.tb01433.x.
Ratcliff, R., 1978. A theory of memory retrieval. Psychol. Rev. 85, 59–108. https://doi.org/10.1037/0033-295X.85.2.59.
Ratcliff, R., McKoon, G., 2008. The diffusion decision model: theory and data for two-choice decision tasks. Neural Comput. 20, 873–922. https://doi.org/10.1162/neco.2008.12-06-420.
Ratcliff, R., Philiastides, M.G., Sajda, P., 2009. Quality of evidence for perceptual decision making is indexed by trial-to-trial variability of the EEG. Proc. Natl. Acad. Sci. Unit. States Am. 106, 6539–6544. https://doi.org/10.1073/pnas.0812589106.
Ratcliff, R., Smith, P.L., Brown, S.D., McKoon, G., 2016. Diffusion decision model: current issues and history. Trends Cogn. Sci. (Regul. Ed.) 20, 260–281. https://doi.org/10.1016/j.tics.2016.01.007.
Rohe, T., Ehlis, A.-C., Noppeney, U., 2019. The neural dynamics of hierarchical Bayesian causal inference in multisensory perception. Nat. Commun. 10 https://doi.org/10.1038/s41467-019-09664-2.
Rohe, T., Noppeney, U., 2016. Distinct computational principles govern multisensory integration in primary sensory and association cortices. Curr. Biol. 26, 509–514. https://doi.org/10.1016/j.cub.2015.12.056.
Rohe, T., Noppeney, U., 2015. Cortical hierarchies perform bayesian causal inference in multisensory perception. PLoS Biol. 13, e1002073 https://doi.org/10.1371/journal.pbio.1002073.
Romei, V., Murray, M.M., Cappe, C., Thut, G., 2009. Preperceptual and stimulus-selective enhancement of low-level human visual cortex excitability by sounds. Curr. Biol. 19, 1799–1805. https://doi.org/10.1016/j.cub.2009.09.027.
Romo, R., de Lafuente, V., 2013. Conversion of sensory signals into perceptual decisions. Prog. Neurobiol. 103, 41–75. https://doi.org/10.1016/j.pneurobio.2012.03.007.

Rosahl, S.K., Knight, R.T., 1995. Role of prefrontal cortex in generation of the contingent negative variation. Cerebr. Cortex 5, 123–134. https://doi.org/10.1093/cercor/5.2.123.

Salti, M., Monto, S., Charles, L., King, J.-R., Parkkonen, L., Dehaene, S., 2015. Distinct cortical codes and temporal dynamics for conscious and unconscious percepts. eLife 4. https://doi.org/10.7554/eLife.05652.

Schroeder, C.E., Foxe, J., 2005. Multisensory contributions to low-level, 'unisensory' processing. Curr. Opin. Neurobiol. 15, 454–458. https://doi.org/10.1016/j.conb.2005.06.008.

Siegel, M., Buschman, T.J., Miller, E.K., 2015. Cortical information flow during flexible sensorimotor decisions. Science 348, 1352–1355. https://doi.org/10.1126/science.aab0551.

Smith, P.L., Ratcliff, R., 2004. Psychology and neurobiology of simple decisions. Trends Neurosci. 27, 161–168. https://doi.org/10.1016/j.tins.2004.01.006.

Stein, B.E., Meredith, M.A., 1993. The merging of the senses, 2. print. In: Cognitive Neuroscience Series. MIT Press, Cambridge, Mass.

Talsma, D., Senkowski, D., Soto-Faraco, S., Woldorff, M.G., 2010. The multifaceted interplay between attention and multisensory integration. Trends Cognit. Sci. 14, 400–410. https://doi.org/10.1016/j.tics.2010.06.008.

ten Oever, S., Romei, V., van Atteveldt, N., Soto-Faraco, S., Murray, M.M., Matusz, P.J., 2016. The COGs (context, object, and goals) in multisensory processing. Exp. Brain Res. 234, 1307–1323. https://doi.org/10.1007/s00221-016-4590-z.

Tsunada, J., Liu, A.S.K., Gold, J.I., Cohen, Y.E., 2016. Causal contribution of primate auditory cortex to auditory perceptual decision-making. Nat. Neurosci. 19, 135–142. https://doi.org/10.1038/nn.4195.

Twomey, D.M., Kelly, S.P., O'Connell, R.G., 2016. Abstract and effector-selective decision signals exhibit qualitatively distinct dynamics before delayed perceptual reports. J. Neurosci. 36, 7346–7352. https://doi.org/10.1523/JNEUROSCI.4162-15.2016.

van Wassenhove, V., Grant, K.W., Poeppel, D., 2005. Visual speech speeds up the neural processing of auditory speech. Proc. Natl. Acad. Sci. Unit. States Am. 102, 1181–1186. https://doi.org/10.1073/pnas.0408949102.

Voss, A., Nagler, M., Lerche, V., 2013. Diffusion models in experimental psychology: a practical introduction. Exp. Psychol. 60, 385–402. https://doi.org/10.1027/1618-3169/a000218.

Voss, A., Voss, J., Lerche, V., 2015. Assessing cognitive processes with diffusion model analyses: a tutorial based on fast-dm-30. Front. Psychol. 6, 336. https://doi.org/10.3389/fpsyg.2015.00336.

Walter, W.G., Cooper, R., Aldridge, V.J., Mccallum, W.C., Winter, A.L., 1964. Contingent negative variation: an electric sign OF sensorimotor association and expectancy IN the human brain. Nature 203, 380–384. https://doi.org/10.1038/203380a0.

Welch, R.B., Warren, D.H., 1980. Immediate perceptual response to intersensory discrepancy. Psychol. Bull. 88, 638–667.

Werner, S., Noppeney, U., 2010. Distinct functional contributions of primary sensory and association areas to audiovisual integration in object categorization. J. Neurosci. 30, 2662–2675. https://doi.org/10.1523/JNEUROSCI.5091-09.2010.

Wyart, V., de Gardelle, V., Scholl, J., Summerfield, C., 2012. Rhythmic fluctuations in evidence accumulation during decision making in the human brain. Neuron 76, 847–858. https://doi.org/10.1016/j.neuron.2012.09.015.