eNeuro

Research Article: New Research | Sensory and Motor Systems

Reliability-Weighted Integration of Audiovisual Signals Can Be Modulated by Top-down Control

Tim Rohe[1,2] and Uta Noppeney[1,3]

[1]*Max Planck Institute for Biological Cybernetics, Spemannstr. 38, Tübingen, 72076, Germany*
[2]*Department of Psychiatry and Psychotherapy, University of Tübingen, Tübingen, Germany*
[3]*Computational Neuroscience and Cognitive Robotics Centre, University of Birmingham, Birmingham, United Kingdom*

DOI: 10.1523/ENEURO.0315-17.2018

Received: 11 September 2017

Revised: 18 January 2018

Accepted: 26 January 2018

Published: 16 February 2018

Author contributions: T.R. and U.N. designed research; T.R. performed research; T.R. contributed unpublished reagents/analytic tools; T.R. and U.N. analyzed data; T.R. and U.N. wrote the paper.

Funding: http://doi.org/10.13039/501100000781EC | European Research Council (ERC)
ERC-2012-StG_20111109

Funding: Max Planck Society

Funding: University of Tuebingen
Fortüne 2292-0-0

Conflict of Interest: The authors report no conflict of interest.

This study was funded by the European Research Council (ERC-2012-StG_20111109), the Max Planck Society and a Fortüne grant (2292-0-0) of the University of Tübingen.

Corresponding author: Tim Rohe, Department of Psychiatry and Psychotherapy, Calwerstr. 14, University of Tübingen, 72076 Tübingen, Germany. email: Tim.Rohe@med.uni-tuebingen.de

Cite as: eNeuro 2018; 10.1523/ENEURO.0315-17.2018

Reliability-weighted integration of audiovisual signals can be modulated by top-down control

Tim Rohe[1,2*], Uta Noppeney[1,3]

[1] Max Planck Institute for Biological Cybernetics, Spemannstr. 38, 72076 Tübingen, Germany

[2] Department of Psychiatry and Psychotherapy, University of Tübingen, Tübingen, Germany

[3] Computational Neuroscience and Cognitive Robotics Centre, University of Birmingham, Birmingham, United Kingdom

* Corresponding author: Tim Rohe, email: Tim.Rohe@med.uni-tuebingen.de

Department of Psychiatry and Psychotherapy, Calwerstr. 14, University of Tübingen, 72076 Tübingen, Germany

Abbreviated title: Reliability-weighted audiovisual integration

Number of pages: 72; Number of words: Abstract = 238; Significance Statement = 100; Introduction = 1212; Discussion = 1994; Number of figures: 5; Number of tables: 3; Number of multimedia: 0

Acknowledgements: This study was funded by the European Research Council (ERC-2012-StG_20111109), the Max Planck Society and a Fortüne grant (2292-0-0) of the University of Tübingen. We thank Phillip Ehses for help with the MR parallel imaging sequence. The authors report no conflict of interest.

Abstract

Behaviorally, it is well-established that human observers integrate signals near-optimally weighted in proportion to their reliabilities as predicted by maximum likelihood estimation. Yet, despite abundant behavioral evidence, it is unclear how the human brain accomplishes this feat. In a spatial ventriloquist paradigm, participants were presented with auditory, visual and audiovisual signals and reported the location of the auditory or the visual signal. Combining psychophysics, multivariate fMRI decoding and models of maximum likelihood estimation (MLE), we characterized the computational operations underlying audiovisual integration at distinct cortical levels. We estimated observers' behavioral weights by fitting psychometric functions to participants' localization responses. Likewise, we estimated the neural weights by fitting 'neurometric' functions to spatial locations decoded from regional fMRI activation patterns. Our results demonstrate that low-level auditory and visual areas encode predominantly the spatial location of the signal component of a region's preferred auditory (resp. visual) modality. By contrast, intraparietal sulcus forms spatial representations by integrating auditory and visual signals weighted by their reliabilities. Critically, the neural and behavioral weights and the variance of the spatial representations depended not only on the sensory reliabilities as predicted by the MLE model but also on participants' modality-specific attention and report (i.e., visual vs. auditory). These results suggest that audiovisual integration is not exclusively determined by bottom-up sensory reliabilities. Instead, modality-specific attention and report can flexibly modulate how intraparietal sulcus integrates sensory signals into spatial representations to guide behavioral responses (e.g., localization and orienting).

47 **Significance statement**

48 To obtain an accurate representation of the environment, the brain should integrate noisy sensory
49 signals by weighting them in proportion to their relative reliabilities. This strategy is optimal by
50 providing the most reliable, i.e., least variable percept. The extent to which the brain top-down
51 controls the sensory weights in the integration process remains controversial. The current study
52 shows that the parietal cortex weighs audiovisual signals by their reliabilities. Yet, the sensory
53 weights and the variance of the multisensory representations were also influenced by modality-
54 specific attention and report. These results suggest that audiovisual integration can be flexibly
55 modulated by top-down control.

56

Introduction

In our natural environment our senses are continuously exposed to noisy sensory signals that provide uncertain information about the world. To construct a veridical representation of the environment, the brain is challenged to integrate sensory signals if they pertain to common events. Numerous psychophysics studies have demonstrated that human observers combine signals within and across the senses by weighting them in proportion to their reliabilities with greater weights assigned to the more reliable signal (i.e., the inverse of a signal's variance) (Jacobs, 1999; Ernst and Banks, 2002; Knill and Saunders, 2003; Alais and Burr, 2004). If two signals provide redundant information about the same event (i.e., common-source assumption), this reliability-weighted multisensory integration provides the most precise, i.e., statistically optimal, perceptual estimate (i.e., maximum likelihood estimate, MLE) leading to better performance on a range of tasks such as depth (Ban et al., 2012), shape (Ernst and Banks, 2002), motion (Fetsch et al., 2012) or spatial (Alais and Burr, 2004) discrimination. However, reliability-weighted integration is statistically optimal only for the special case where a single cause elicited the signals, i.e., the common-source assumptions are met. In our natural environment, two signals can arise either from common or separate sources leading to some uncertainty about the causal structure underlying the sensory signals. Mandatory integration of sensory signals would in many instances effectively misattribute information (Roach et al., 2006). In this more natural context, the observer has to infer the causal structure from sensory correspondences such as spatial co-location (Wallace et al., 2004) or temporal correlation (Parise and Ernst, 2016). The observer should then integrate signals in case of a common cause, but segregate them in case of independent causes (Kording et al., 2007). In other words, reliability-

weighted integration is no longer statistically optimal in more general situations where the causal structure of the sensory signals is unknown or the assumption of a common source is violated.

Despite abundant behavioral evidence for near-optimal reliability-weighted integration under experimental conditions which foster the assumption of a common signal cause, the underlying neural mechanisms remain unexplored in the human brain for multisensory signals. For cue combination within a single sensory modality, higher-order visual regions have recently been implicated in reliability-weighted integration of visual-depth cues (Ban et al., 2012). Only recently, elegant neurophysiological studies in non-human primates have started to characterize the neural mechanisms of visual-vestibular integration for heading discrimination. They demonstrated that single neurons (Morgan et al., 2008) and neuronal populations (Fetsch et al., 2012) in the dorsal medial superior temporal area (dMST) integrated visual and vestibular motion near-optimally weighted by their reliabilities. Moreover, the neural weights derived from neural population responses in dMST corresponded closely to the weights governing monkey's behavioral choices.

Over the past decade, accumulating evidence has shown that multisensory integration is not deferred until later processing stages in higher-order association cortices (Beauchamp et al., 2004; Sadaghiani et al., 2009), but starts already at the primary cortical level (Foxe et al., 2000; Ghazanfar and Schroeder, 2006; Kayser et al., 2007; Lakatos et al., 2007; Lewis and Noppeney, 2010; Werner and Noppeney, 2010; Lee and Noppeney, 2014). Previous functional imaging research indicated in a qualitative fashion that sensory reliability modulates regional BOLD responses (Helbig et al., 2012), functional connection strengths (Nath and Beauchamp, 2011) or activation patterns (Ban et al., 2012; Rohe and Noppeney, 2016). For instance, during speech recognition the superior temporal sulcus coupled more strongly with the auditory cortex when

auditory reliability was high but with visual cortex when visual reliability was high (Nath and Beauchamp, 2011). Likewise, using fMRI multivariate pattern decoding a recent study showed that parietal cortices integrated spatial signals depending on their spatial disparity and sensory reliability (Rohe and Noppeney, 2016). However, to our knowledge no previous study has evaluated whether multisensory integration in the human brain follows the quantitative predictions of the MLE model.

Computational models of probabilistic population coding (Ma et al., 2006) suggest that reliability-weighted integration may be obtained by averaging the inputs with fixed weights from upstream populations of neurons that encode the reliability of the sensory input in terms of the sensory gain. By contrast, the recently proposed normalization model of multisensory integration (Ohshiro et al., 2011, 2017) suggests that normalization over a pool of neurons as a canonical computational operation can implement multisensory integration with weights that flexibly adjust to the reliability of the sensory inputs. Critically, in both models reliability-weighted integration depends on a region to have access to inputs from upstream regions that are responsive to auditory and visual inputs. While accumulating evidence suggests that multisensory integration starts already at the primary cortical level (Foxe et al., 2000; Bonath et al., 2007; Kayser et al., 2007; Lakatos et al., 2007; Lewis and Noppeney, 2010; Werner and Noppeney, 2010; Bonath et al., 2014; Lee and Noppeney, 2014), the fraction of multisensory neurons that are influenced by inputs from multiple sensory modalities increases across the cortical hierarchy (Bizley et al., 2007; Dahl et al., 2009). Thus, even if low-level sensory areas are susceptible to limited influence from other sensory modalities, this activity may be less informative (i.e., more unreliable) than that of the preferred sensory modality. As a result,

124 reliability-weighted integration via normalization may be more prominent in higher-order
125 association cortices than in low-level sensory areas.

126 Besides the assumption of a common signal cause, a second assumption of the classical
127 MLE model is that the sensory weights and the variance reduction obtained from multisensory
128 integration depend solely on the bottom-up reliabilities of the sensory inputs irrespective of
129 cognitive influences (i.e., the unisensory reliabilities are not influenced by observers' attentional
130 focus, e.g. selective vs. divided attention). In line with this conjecture, initial psychophysics
131 studies suggested that the sensory weights are immune to attentional influences (Helbig and
132 Ernst, 2008). Yet, more recent psychophysics studies have demonstrated that the sensory weights
133 are modulated by attentional top-down effects (Vercillo and Gori, 2015). Moreover, EEG and
134 fMRI studies revealed profound attentional effects on the neural processes underlying
135 multisensory integration (Talsma et al., 2010; Donohue et al., 2011). The controversial results
136 raise the questions whether the task-relevance of sensory signals influences reliability-weighted
137 integration at the neural level even if the signals' small disparity suggests a common cause.

138 The present study combined psychophysics and fMRI multivariate decoding to
139 characterize the neural processes underlying multisensory integration in a quantitative fashion
140 and to investigate potential top-down effects of modality-specific report and associated
141 attentional effects. We presented participants with auditory, visual and audiovisual signals that
142 were spatially congruent or in a small spatial conflict. On each trial, participants were presented
143 with an auditory and a visual spatial signal from four possible horizontal locations. They located
144 either the visual or the auditory signal by pushing one of four response buttons that corresponded
145 to the four locations. To compute psychometric functions, participants' responses were binarized
146 into left-vs.-right responses. To assess top-down effects of modality-specific report on the

147 behavioral and neural weights, we manipulated whether participants reported the auditory or
148 visual locations. In a model-based analysis, we first investigated whether the sensory weights
149 and variances obtained from psychometric and 'neurometric' functions were in line with the
150 predictions of the MLE model. In a model-free analysis, we next examined whether the sensory
151 weights and variances were influenced by visual signal reliability and/or report of the auditory
152 (or visual) modality.

Materials and Methods

Participants

After giving written informed consent, six healthy volunteers (two females, mean age 28.8 years, range 22-36 years) participated in the fMRI study. All participants had normal or corrected-to normal vision and reported normal hearing. One participant was excluded due to excessive head motion (4.21 / 3.52 STD above the mean of the translational/rotational volume-wise head motion based on the included 5 participants). The study was approved by the human research review committee of the University of Tübingen. A subset of the data (i.e., the audiovisual conditions) have been reported in Rohe and Noppeney (2015a, 2016).

Stimuli

The visual stimulus was a cloud of 20 white dots (diameter: 0.43° visual angle) sampled from a bivariate Gaussian with a vertical standard deviation of 2.5° and a horizontal standard deviation of 2° or 14° (high and low visual reliability). The visual stimulus was presented on a black background (i.e., 100% contrast). The auditory stimulus was a burst of white noise with a 5ms on/off ramp. To create a virtual auditory spatial signal, the noise was convolved with spatially specific head-related transfer functions (HRTFs). The HRTFs were pseudo-individualized by matching participants' head width, heights, depth and circumference to the anthropometry of participants in the CIPIC database (Algazi et al., 2001) and were interpolated to the desired location of the auditory signal.

- Figure 1 about here -

176 *Experimental design and procedure*

177 In the unisensory conditions participants were presented either with auditory or with visual
178 signals of low or high reliability. The signals were sampled from four possible locations along
179 the azimuth (i.e., -10°, -3.3°, 3.3° or 10°). This yielded 4 auditory conditions (i.e., 4 auditory
180 locations) and 4 visual locations x 2 visual reliability (high vs. low) = 8 visual conditions. On
181 each trial participants located either the visual or the auditory signal.

182 In the audiovisual conditions, participants were presented with synchronous auditory and
183 visual signals of high or low visual reliabilities (Fig. 1A). They attended and reported the
184 location either of the visual or auditory signal component. The locations of the auditory and
185 visual signal components were sampled independently from four possible locations. This yielded
186 4 auditory locations x 4 visual locations = 16 audiovisual location combinations that varied in
187 their audiovisual spatial disparities. In the current study, we focused selectively on the
188 audiovisually congruent (A-V = ΔAV = 0°) and slightly conflicting conditions (ΔAV = 6° and =
189 -6°). These small, so-called non-noticeable, spatial conflicts have previously been introduced to
190 test the predictions of the maximum likelihood estimation (MLE) model (e.g., Alais & Burr,
191 2004; Battaglia et al., 2003) as they are assumed to ensure that observers fuse sensory signals
192 into one unified percept. Note that results of the audiovisual conditions with larger disparity
193 (ΔAV > 6°) have been reported in Rohe and Noppeney (2015a, 2016).

194 In total, this MLE study included 52 conditions (Fig. 1B): 4 unisensory auditory
195 conditions, 4 unisensory visual conditions of high visual reliability, 4 unisensory visual
196 conditions of low visual reliability and 40 audiovisual conditions: i.e., (4 audiovisually congruent
197 + 6 audiovisually incongruent conditions with a small spatial disparity) x 2 visual reliability
198 levels (high vs. low) x 2 modality-specific reports (i.e., visual vs. auditory). For the latter model-

free analysis, we obtained variances and sensory weights by fitting psychometric and neurometric functions separately to the perceived and decoded spatial locations (i.e., % perceived right as a function of spatial location) separately for the four conditions in a 2 (visual reliability: high vs. low) x 2 (modality-specific report: auditory vs. visual) factorial design.

On each trial, audiovisual signals were presented for 50 ms duration with a variable inter-stimulus fixation interval of 1.75-2.75 s (Fig. 1A). Participants reported their auditory perceived location in the unisensory auditory and the audiovisual sessions with auditory report. They reported their visual perceived location in the unisensory visual and the audiovisual sessions with visual report. Participants indicated their perceived location by pushing one of four buttons that spatially corresponded to the four signal locations (i.e., -10°, -3.3°, 3.3° or 10° along the azimuth) using their right hand. To compute psychometric functions, participants' responses were binarized into left-vs.-right responses for all analyses. Throughout the experiment, participants fixated a central cross (1.6° diameter).

Unisensory and audiovisual stimuli were presented in separate sessions. Subjects participated in 3-4 unisensory auditory, 3-4 unisensory visual and 20 audiovisual sessions (10 auditory and 10 visual report; apart from one participant who performed 9 auditory and 11 visual report sessions). In the respective sessions we presented the 4 unisensory auditory conditions in 88 trials each, the 8 unisensory visual conditions in 44 trials each and the 32 audiovisual conditions (4 visual stimulus locations x 4 auditory stimulus locations x 2 visual reliability levels) in 11 trials each. Further, 5.9 % null-events (i.e., 'pseudo-events' without a stimulation) were interspersed in the sequence of 352 stimuli per session to estimate stimulus-evoked responses relative to the fixation baseline. To maximize design efficiency, trial types were presented in a pseudorandomized order. We manipulated the modality-specific report (visual vs.

auditory) over sessions in a counterbalanced order within each participant and we presented unisensory and audiovisual runs in a counterbalanced order across participants.

Experimental setup

Audiovisual signals were presented using Psychtoolbox 3.09 (www.psychtoolbox.org) (Brainard, 1997; Kleiner et al., 2007) running under MATLAB R2010a (MathWorks). Auditory stimuli were presented at ~75 dB SPL using MR-compatible headphones (MR Confon). Visual stimuli were back-projected onto a Plexiglas screen using an LCoS projector (JVC DLA-SX21). Participants viewed the screen through an extra-wide mirror mounted on the MR head coil resulting in a horizontal visual field of approx. 76° at a viewing distance of 26 cm. Participants indicated their response using an MR-compatible custom-built button device. Participants' eye movements and fixation were monitored by recording participants' pupil location using an MR-compatible custom-built infrared camera (sampling rate 50 Hz) mounted in front of the participants' right eye and iView software 2.2.4 (SensoMotoric Instruments).

Key predictions of the Maximum Likelihood Estimation model

The majority of multisensory research today has focused on the so-called 'forced fusion case', where observers a priori assume that two signals come from a common source and should hence be integrated. These 'forced fusion criteria' are generally assumed to be met when observers are instructed to locate a single source that emits audiovisual signals (i.e., bi-sensory attention) and the two signals are presented without any conflict or with a small cue conflict such as a spatial disparity of 6° visual angle as employed in our experiment (e.g., Alais and Burr, 2004). Under these classical forced fusion assumptions, the Maximum Likelihood Estimation model makes

two key quantitative predictions for participants' estimates (e.g., spatial estimates) that are formed by integrating auditory and visual signals. The first prediction pertains to the sensory weights applied during the integration process and the second prediction to the variance of the integrated perceived signal location:

Sensory weights: The most reliable unbiased estimate of an object's location ($\hat{S}_{AV}$) is obtained by combining the auditory ($\hat{S}_A$) and visual ($\hat{S}_V$) perceived locations in proportion to their relative reliabilities (r_A, r_V; i.e., the inverse of the variance, $r = 1/\sigma^2$).

$$(1) \quad \hat{S}_{AV} = w_A\hat{S}_A + w_V\hat{S}_V \quad \text{with } w_A = \frac{r_A}{r_A + r_V} = \frac{\frac{1}{\sigma_A 2}}{\frac{1}{\sigma_A 2} + \frac{1}{\sigma_V 2}} \text{ and } w_V = \frac{r_V}{r_A + r_V}$$

$$= \frac{\frac{1}{\sigma_V 2}}{\frac{1}{\sigma_A 2} + \frac{1}{\sigma_V 2}}$$

The variances obtained from the cumulative Gaussians that were fitted to the unisensory visual and auditory conditions were used to determine the 'optimal' weights that participants should apply to the visual and auditory signals in the audiovisual conditions as predicted by the MLE model (equation 1). The empirical weights were computed from the point of subjective equality (PSE) of the psychometric functions of the audiovisual conditions where a small audiovisual spatial disparity of 6° was introduced according to the following equation (Helbig and Ernst, 2008; Fetsch et al., 2012):

$$(2) \quad w_{V,\,emp} = \frac{PSE_{\triangle AV = +6°} - PSE_{\triangle AV = -6°}}{2\ \triangle AV} + \frac{1}{2}$$

Note that the equation assumes that the psychometric functions plot '% perceived right' as a function of the average of the true auditory and visual locations (Fig. 2 and 3).

Variance of the integrated perceived signal location: Multisensory integration reduces the variance of the audiovisual estimate (σ_{AV}^2) in particular for congruent audiovisual trials as compared to the unisensory variances (σ_A^2, σ_V^2):

$$(3)\quad \sigma_{AV}^2 = \frac{\sigma_A^2 \sigma_V^2}{\sigma_A^2 + \sigma_V^2}$$

To generate MLE predictions for the audiovisual variance, the unisensory variances were obtained from the psychometric functions (i.e., cumulative Gaussians) for the auditory and visual signals. The empirical variance of the combined audiovisual estimate was obtained from the psychometric function for the audiovisual conditions.

Behavioral data

Participants' spatial location responses (i.e., four buttons) were categorized as left or right responses. For the unisensory auditory and visual conditions, we plotted the fraction of right responses as a function of the unisensory signal location (Fig. 2E). For the audiovisual spatially congruent and conflicting conditions we plotted the fraction of right responses as a function of the mean signal location of the true auditory and true visual signal locations (separately for the four conditions in our 2 (auditory vs. visual report) x 2 (high vs. low visual reliability) factorial design, Fig. 2A-D).

For the behavioral analysis, we fitted cumulative Gaussian functions individually to the data of each participant (again separately for the four conditions in our 2 (auditory vs. visual report) x 2 (high vs. low visual reliability) factorial design using maximum likelihood estimation

methods as implemented in Palamedes toolbox 1.5.0 (Prins and Kingdom, 2009). To enable reliable parameter estimation for each participant, we employed the following constraints: i) The Gaussians' means (i.e., point of subjective equality, PSE) were constrained to be equal across unisensory and audiovisual congruent conditions (i.e., identical spatial biases were assumed across unisensory and audiovisual congruent conditions). ii) The Gaussians variances (i.e., perceptual thresholds or slopes of the psychometric functions) were constrained to be equal for the congruent and the two conflicting conditions within each combination of visual reliability and modality-specific report. Please note that this is based on the fundamental forced-fusion assumption implicitly adopted in previous research (Ernst and Banks, 2002; Alais and Burr, 2004) whereby the conditions with small non-noticeable cue conflict are considered to be equivalent to congruent conditions. iii) Guess and lapse rate parameters were set to be equal (i.e., guess = lapse rate) and constrained to be equal across all conditions. In other words, we assumed that observers possibly made false responses (e.g., a 'right' response for a signal at -10º) for non-specific reasons such as blinking, inattention etc. with equal probability in their outer left and right hemifields. Based on those constraints we fitted 17 parameters to the 52 data points individually for each participant. More specifically, we fitted one PSE parameter commonly for the unisensory visual, auditory and audiovisual congruent conditions, one PSE parameter each for the eight conflict conditions (i.e., 2 visual reliability X 2 modality-specific report X 2 spatial conflict, ΔAV = -6 or +6; i.e., in total: 9 parameters for PSE). Further, we fitted one slope parameter each for i. the unisensory auditory, ii. low reliable visual, iii. high reliable visual conditions and iv. for each audiovisual condition of the 2 visual reliability X 2 modality-specific report (i.e., 7 slope parameters). Finally, as the conditions were presented in a randomized order

306 we fitted one single guess = lapse rate parameter across all conditions (i.e., one single
307 parameter).

308 The Gaussians' means and variances (σ^2) of the unisensory conditions were used to
309 compute the maximum likelihood predictions for the visual weights (w_V in equation (1)) and the
310 variance of the perceived signal location (σ_{AV}^2 in equation (3)). The empirical visual weights
311 ($w_{V,emp}$ in equation (2)) were computed from the audiovisual conditions with a small spatial cue
312 conflict (i.e., ΔAV = 6° and = -6°). In the main analysis the empirical audiovisual variances were
313 computed jointly from the small cue conflict and congruent audiovisual conditions (cf. modeling
314 constraints above).

315 In a follow-up analysis, we also obtained audiovisual variances selectively for the
316 audiovisual congruent conditions by adding four independent slope parameters for the
317 audiovisual congruent conditions (i.e., 21 parameters in total). As the small disparity trials were
318 not included in the estimation of variance, this follow-up analysis allowed us to investigate
319 whether modality-specific report can influence the integration process even for audiovisual
320 congruent trials. In particular, we asked whether the audiovisual variance for the congruent
321 conditions was immune to modality-specific report as predicted by the classical MLE model or
322 depended on modality-specific report.

323 We evaluated the MLE predictions using classical statistics and a Bayesian model
324 comparison:

325 *Classical statistics:*

326 In a model-based analysis we compared the empirical visual weights and audiovisual
327 variances with the MLE predictions and unisensory auditory and unisensory visual variances at
328 the second (i.e., between subject) random-effects level (Tab. 1). We used non-parametric

Wilcoxon signed rank test to account for the small sample size (n = 5) and potential violations of normality assumptions.

In a model-free analysis, participant-specific visual weights and audiovisual variances were entered into second (i.e., between-subject) level analyses. At the random-effects level, we tested for the effects of visual reliability (high vs. low) and modality-specific report (visual vs. auditory) on the empirical visual weights and audiovisual variances using 2 x 2 repeated measures ANOVAs (Tab. 2). To account for the small sample size we used a non-parametric procedure by computing the ANOVAs on rank-transformed empirical weights and variances (Conover and Iman, 1981). Further, we analyzed whether auditory signals biased visual reports and whether visual signals biased auditory reports by testing whether the visual weight was smaller than one or larger than zero, respectively, while pooling over visual reliability. For these comparisons we used one-sided Wilcoxon signed rank tests.

As we employed a fixed-effects approach for the fMRI data to increase signal to noise ratio, in a follow-up analysis we applied the same fixed-effects approach to the behavioral data to ensure that differences between behavioral and fMRI results did not result from methodological differences.

Unless otherwise stated, results are reported at $p < 0.05$.

Bayesian model comparison:

Using Bayesian model comparison analysis, we compared four models that manipulated whether visual reliability and modality-specific report could affect the PSEs and slopes of the audiovisual psychometric functions and whether their influence was predicted by the MLE model (Tab. 3):

i) Model 1 - Null-model: Visual reliability and modality-specific report were not able to alter PSEs or slopes (i.e., integration of audiovisual signals with constant sensory weights irrespective of modality-specific report or reliability).

ii) Model 2 - MLE model: Visual reliability affected PSEs and slopes as predicted by MLE. Modality-specific report did not influence PSEs or slopes (again as predicted by MLE). Hence, we set the audiovisual PSEs and slopes to the MLE predictions based on the unisensory conditions as described in equation (1) and (3).

iii) Model 3 – Reliability-weighted integration model: Visual reliability influenced PSEs and slopes of the audiovisual conditions, yet not according to the MLE predictions. Hence, we allowed the PSEs and the slopes of the audiovisual conditions to differ across different reliability levels unconstrained by the MLE predictions. Yet, we did not allow top-down influences of modality-specific report to influence audiovisual PSEs or slopes.

iv) Model 4 - Full model: Visual reliability and modality-specific report influenced both PSEs and slopes (i.e., the full model comparable to the analyses using classical statistics above).

For all four models, psychometric functions were individually fitted to participants' behavioral responses as described above. From the models' log likelihood we computed the Bayesian Information Criterion (BICs) as an approximation to the model evidence (Raftery, 1995). Bayesian model comparison (Stephan et al., 2009; Rigoux et al., 2014) was performed at the group level as implemented in SPM12 (Friston et al., 1994) based on the expected posterior probability (i.e., the probability that a given model generated the data for a randomly selected subject), the exceedance probability (i.e., the probability that a given model is more likely than any other model) (Stephan et al., 2009) and the protected exceedance probability (additionally accounting for differences in model frequencies due to chance) (Rigoux et al., 2014).

MRI data acquisition

A 3T Siemens Magnetom Trio MR scanner was used to acquire both T1-weighted anatomical images and T2*-weighted axial echoplanar images (EPI) with BOLD contrast (gradient echo, parallel imaging using GRAPPA with an acceleration factor of 2, TR = 2480ms, TE = 40ms, flip angle=90°, FOV=192 mm×192 mm, image matrix 78×78, 42 transversal slices acquired interleaved in ascending direction, voxel size=2.5×2.5×2.5 mm + 0.25 mm inter-slice gap). In total, we acquired 353 volumes times 20 sessions for the audiovisual conditions, 353 volumes times 6-8 sessions for the unisensory conditions, 161 volumes times 2-4 sessions for the auditory localizer and 159 volumes times 10-16 sessions for the visual retinotopic localizer (see below). This resulted in approximately 18 hours of scanning per participant assigned over 7-11 days. The first three volumes of each session were discarded to allow for T1 equilibration effects.

fMRI data analysis

Spatial ventriloquist paradigm

The fMRI data were analyzed with SPM8 (www.fil.ion.ucl.ac.uk/spm) (Friston et al., 1994). Scans from each participant were corrected for slice timing, realigned and unwarped to correct for head motion and spatially smoothed with a Gaussian kernel of 3 mm FWHM (de Beeck, 2010). The time series in each voxel was high-pass filtered to 1/128 Hz. All data were analyzed in native subject space. The fMRI experiment was modeled in an event-related fashion with regressors entered into the design matrix after convolving each event-related unit impulse with a canonical hemodynamic response function and its first temporal derivative. In addition to modeling the 4 unisensory auditory, the 8 unisensory visual or the 32 audiovisual conditions in a

session, the general linear models (GLM) included the realignment parameters as nuisance covariates to account for residual motion artefacts. The factor modality-specific report (visual vs. auditory) was modeled across sessions. The session-specific parameter estimates pertaining to the canonical hemodynamic response function (HRF) defined the magnitude of the BOLD response to the unisensory or the audiovisual stimuli in each voxel.

To apply the MLE analysis approach to spatial representations at the neural level, we first extracted the parameter estimates pertaining to the HRF magnitude for each condition and session from voxels of regions defined in separate auditory and retinotopic localizer experiments (see below). This yielded activation patterns from the unisensory auditory and visual conditions and the audiovisual congruent ($\Delta AV = 0°$) and small spatial cue conflict ($\Delta AV \pm 6°$) conditions. All activation patterns (i.e., from each condition in each session) were z normalized across all voxels of a region of interest to avoid the effects of region-wide activation differences between conditions. We then trained a linear support vector classification model (as implemented in LIBSVM 3.14 (Chang and Lin, 2011)) to learn the mapping from activation patterns from the audiovisual *congruent* conditions to the categorical left vs. right location of the audiovisual signal in a subject-specific fashion. Importantly, we selectively used activation patterns from audiovisual *congruent* conditions from all but one audiovisual session for support vector classification training (i.e., training was done across sessions of auditory and visual report). The trained support vector classification model was then used to decode the signal location (left vs. right) from the activation patterns of the spatially congruent and *conflicting* audiovisual conditions of the remaining audiovisual session. Hence, given the learnt mapping from audiovisual activation patterns of the congruent conditions to true left vs. right stimulus location class the support vector classifier decoded the stimulus location for activation patterns elicited by

the audiovisual spatially small conflict trials. In a leave-one-out cross-validation scheme, the training-test procedure was repeated for all audiovisual sessions. Finally, the support vector classification model was trained on audiovisual congruent conditions from all audiovisual sessions and then decoded the categorical signal location 'left vs. right' from activation patterns of the separate unisensory auditory and visual sessions.

In line with our behavioral analysis, we plotted the fraction of decoded 'right' as a function of the unisensory signal location for the unisensory auditory and visual conditions (Fig. 3E). For the audiovisual spatially congruent and small cue conflict conditions we plotted the fraction of decoded 'right' as a function of the mean signal location of the true auditory and visual signal locations (separately for auditory/visual report x visual reliability levels; Fig. 3A-D). Because of the lower signal-to-noise ratio of fMRI data, we fitted cumulative Gaussians as neurometric functions to the fraction decoded 'right' pooled (i.e., averaged) across all participants (i.e., fixed-effects analysis). To obtain empirical and MLE predicted weights and variances we employed the same procedure and equations as explained in the section of the behavioral analysis. Confidence intervals for empirical and predicted weights and variances were computed using Palamedes' parametric bootstrap procedure (1000 bootstraps).

In the model-based analysis, we used two-tailed bootstrap tests (5000 bootstrap samples) (Efron and Tibshirani, 1994) to investigate whether empirical sensory weights and variances for audiovisual conditions were significantly different from the MLE predictions. Further, we assessed whether variances for audiovisual conditions were significantly different from variances for unisensory conditions (Tab. 1). For these model-based analyses we parametrically bootstrapped the fraction of decoded 'right' and in turn fitted neurometric functions to the bootstrapped data. From the bootstrapped auditory, visual and audiovisual psychometric

functions we generated bootstrap distributions of MLE predictions for the sensory weights and variances and their empirical counterparts. Bootstrapped null-distributions for a specific parameter comparison (e.g., predicted weight vs. empirical weight) were generated by computing the difference between predicted and empirical parameters (e.g., predicted weight vs. empirical weight) for each bootstrap and subtracting the observed original difference (Efron and Tibshirani, 1994). From this bootstrapped null-distribution the two-tailed significance of a parameter comparison was computed as the fraction of bootstrapped absolute values that were greater or equal to the observed original absolute difference (e.g., violation of MLE prediction: $abs[w_{V,predicted,original} - w_{V,empirical,original}]$). Absolute values were used to implement a two-tailed test (Efron and Tibshirani, 1994). Violations of MLE predictions were tested across modality-specific report because the MLE model does not predict a report modulation (i.e., mean and variance parameters of the psychometric functions were held constant across levels of modality-specific report).

Similarly, in the model-free analysis we used two-tailed bootstrap tests (5000 bootstrap samples) to analyze the effects of visual reliability (high vs. low), modality-specific report (visual vs. auditory) and their interaction on the empirical visual weights and audiovisual variances (Tab. 2). Bootstrapped null-distributions of weights and audiovisual variances for each of the four conditions in our modality-specific report (visual vs. auditory) x visual reliability (high vs. low) design were generated by computing the contrast value of interest (e.g., high minus low visual reliability) for the sensory weights or variances for each bootstrap and subtracting the corresponding contrast value obtained from the original data (Efron and Tibshirani, 1994). From this bootstrapped null-distribution the two-tailed significance (against zero) of the effects of interest (e.g., high vs. low reliability) was computed as the fraction of

bootstrapped absolute contrast values that were greater or equal to the observed original absolute contrast value. Mean and variance parameters of the psychometric functions were set to be equal across the levels of modality-specific report in order to test selectively for the main effect of visual reliability. Conversely, mean and variance parameters of the neurometric functions were set to be equal across levels of visual reliability in order to test selectively for the main effect of modality-specific report. By contrast, mean and variance parameters of the neurometric functions varied across levels of visual reliability and modality-specific report in order to test for the interaction effect of modality-specific report and visual reliability. For all analyses reported in Table 1 and 2 we report p values corrected for multiple comparisons across the three regions of interest using a Bonferroni correction.

Finally, we investigated whether multisensory influences can be observed already at the primary cortical level during (i) audiovisual or (ii) even unisensory (i.e., auditory or visual) stimulation. (i) To assess crossmodal influences during audiovisual stimulation, we computed a one-sided bootstrap test (5000 bootstrap samples) by fitting neurometric functions to bootstrapped data (see above) averaged across visual reliability and modality-specific report. Specifically, we tested whether the empirical weight pertaining to the visual signal was smaller than one (i.e., indicating auditory influence) in visual regions and whether it was larger than zero (i.e., indicating visual influence) in auditory regions. (ii) To assess cross-modal influences during unisensory stimulation, we tested whether the slope (i.e., the perceptual threshold $1/\sigma$) of the neurometric functions was significantly greater than zero in unisensory conditions. As we were only interested in whether the slope was significantly greater than zero (rather than the exact size), we used a constrained approach by fitting neurometric functions to auditory stimulation data in visual cortex and to visual stimulation data (pooled over visual reliability levels) in

auditory cortex with lapse and guess rates set to zero. We determined whether a slope parameter was significantly larger than zero using a one-tailed bootstrap test (5000 bootstrap samples).

Across all analyses we confirmed the validity of the bootstrap tests in simulations showing that simulated p values converged to a nominal alpha-level of 0.05 under the null hypothesis.

Control analyses to account for motor preparation and global activation differences between hemispheres

To account for activations related to motor planning (Andersen and Buneo, 2002), a first control analysis included the trial-wise button responses as a nuisance covariate into the first-level GLM (i.e., one regressor for each of the four response buttons). We then repeated the multivariate decoding analysis using activation patterns from intraparietal sulcus (IPS0-4, see below for the definition) where motor responses were explicitly controlled (Fig. 3-1).

Given the contralateral encoding of space in visual (Wandell et al., 2007) and auditory regions (Ortiz-Rios et al., 2017), a second control analysis evaluated the impact of global activation differences between hemispheres on the classifier's performance. In this control analysis, we z normalized the activation patterns separately for voxels of the left and right hemisphere in each condition prior to multivariate decoding (Fig. 3-2, Fig. 4-1). In other words, multivariate decoding was applied to activation patterns where global activation differences between hemispheres were removed.

Effective Connectivity Analyses

Using Dynamic Causal Modelling (DCM) we investigated the modulatory effects of visual reliability on the effective connectivity from early visual regions to IPS and modality-specific report on the connectivity from prefrontal cortex (PFC) to IPS. For each subject we constructed four bilinear DCMs (Friston et al., 2003). Each DCM included four regions: low-level visual regions (V1-3), low-level auditory regions, IPS0-4 and PFC. Low-level visual and auditory regions and IPS0-4 were defined functionally as described in the section 'auditory and visual retinotopic localizer'. PFC was defined anatomically for each individual as the middle frontal gyrus based on the anatomical cortical parcellation of the Desikan-Killiany atlas (Desikan et al., 2006) implemented in Freesurfer 5.1.0 (Dale et al., 1999). Region-specific time series comprised the first eigenvariate of activations across all voxels within each region that were significant at $p < 0.001$ in the effects-of-interest contrast across all conditions in the first-level within-subject GLMs (F test, uncorrected).

In all DCM models, V1-3, IPS0-4 and low-level auditory regions were bidirectionally connected and PFC was bidirectionally connected to IPS0-4 (i.e., intrinsic connectivity structure; Fig. 5). Synchronous audiovisual signals entered as extrinsic input into V1-3 and low-level auditory regions. Holding intrinsic and extrinsic connectivity structure constant, the 2 x 2 candidate DCMs factorially manipulated the presence/absence of the following modulatory effects: a) visual reliability on V1-3 → IPS0-4 (on vs. off) and b) modality-specific report on PFC → IPS0-4 (on vs. off). After fitting the full model, which included both modulatory effects, to the fMRI data of each subject, we used Bayesian model reduction to estimate the model evidences and parameters of the reduced models (Friston et al., 2016). To determine the most likely of the 4 DCMs given the observed data from all subjects, we implemented a fixed- (Penny et al., 2004) and a random-effects group analysis (Stephan et al., 2009). The fixed-effects group

534 analysis was implemented by taking the product of the subject-specific Bayes factors over
535 subjects (this is equivalent to the exponentiated sum of the log model evidences of each subject-
536 specific DCM) (Penny et al., 2004). The model evidence as approximated by the free energy
537 does not only depend on model fit but also model complexity. Because the fixed-effects group
538 analysis can be distorted by outlier subjects, Bayesian model comparison was also implemented
539 in a random-effects group analysis. At the random-effects level, we report the expected posterior
540 probability, the exceedance probability and the protected exceedance probability (Stephan et al.,
541 2009; Rigoux et al., 2014) (Tab. 4).
542
543 *Auditory and visual retinotopic localizer*
544 Regions of interest along the auditory and visual processing hierarchies were defined in a
545 subject-specific fashion based on auditory and visual retinotopic localizers. In the auditory
546 localizer, participants were presented with brief bursts of white noise at -10° or 10° angle
547 (duration 500 ms, stimulus onset asynchrony 1 s). In a one-back task, participants indicated via a
548 key press when the spatial location of the current trial was different from the previous trial. 20 s
549 blocks of auditory stimulation (i.e., 20 trials) alternated with 13 s of fixation periods. The
550 auditory locations were presented in a pseudorandomized fashion to optimize design efficiency.
551 Similar to the main experiment, the auditory localizer sessions were modeled in an event-related
552 fashion. Auditory-responsive regions were defined as voxels in superior temporal and Heschl's
553 gyrus showing significant activations for auditory stimulation relative to fixation (t test, $p < 0.05$,
554 family-wise error corrected). Within these regions, we defined primary auditory cortex (A1)
555 based on cytoarchitectonic probability maps (Eickhoff et al., 2005) and referred to the remainder

556 (i.e., planum temporale and posterior superior temporal gyrus) as higher order auditory cortex
557 (hA).

558 Visual regions of interest were defined using standard phase-encoded retinotopic
559 mapping (Sereno et al., 1995). Participants viewed a checkerboard background flickering at 7.5
560 Hz through a rotating wedge aperture of 70° width (polar angle mapping) or an
561 expanding/contracting ring (eccentricity mapping). The periodicity of the apertures was 42 s.
562 Visual responses were modeled by entering a sine and cosine convolved with the hemodynamic
563 response function as regressors into the design matrix of the general linear model. The preferred
564 polar angle (or eccentricity, respectively) was determined as the phase lag for each voxel by
565 computing the angle between the parameter estimates for the sine and the cosine. The phase lags
566 for each voxel were projected on the reconstructed, inflated cortical surface using Freesurfer
567 5.1.0 (Dale et al., 1999). Visual regions V1-V3 and IPS0-4 were defined as phase reversal in
568 angular retinotopic maps. IPS0-4 were defined as phase reversal along the anatomical IPS
569 resulting in contiguous, approximately rectangular regions (Swisher et al., 2007).

570 For the decoding analyses, the auditory and visual regions were combined from the left
571 and right hemisphere. Support vector classification training was then applied separately to
572 activation patterns from each region. To improve the signal-to-noise ratio when fitting
573 neurometric functions (cf. Fig. 3 and 4), the decoded signal sides (‘right’ vs. ‘left’) from low-
574 level visual regions (V1-3), intraparietal sulcus (IPS0-4) and low-level auditory regions (A1, hA)
575 regions were pooled. Additional analyses showed similar audiovisual spatial integration within
576 these three regions (Rohe and Noppeney, 2015a, 2016).
577

Results

Spatial ventriloquist paradigm

In the fMRI study, participants were presented with auditory, visual and audiovisual signals sampled randomly from four possible spatial locations along the azimuth (i.e., -10°, -3.3°, 3.3° or 10°) (Fig. 1). Audiovisual signals included in this study were either spatially congruent ($\Delta AV = 0°$) or incongruent with a small spatial conflict ($\Delta AV = \pm 6°$). The reliability of the visual signal was either high or low. Modality-specific report was manipulated by instructing participants to report the location either of the visual or the auditory signal component during the audiovisual conditions.

Figures 2 and 3 present the psychometric functions estimated from the behavioral button responses after categorization into 'left' or 'right' responses and the 'neurometric' functions estimated from spatial locations ('left' vs. 'right') decoded from fMRI responses. The psychometric (resp. neurometric) functions show the fraction of 'right' responses as a function of the mean signal location for each condition. If the visual reliability is greater than the auditory reliability (i.e., visual weight > 0.5), we would expect the function to be shifted toward the right for a positive spatial conflict (A-V = $\Delta AV = +6°$, i.e., the visual signal is presented 6° to the left of the auditory signal) and to the left for a negative spatial conflict ($\Delta AV = -6°$, i.e., the visual signal is presented 6° to the right of the auditory signal). As a consequence, the point of subjective equality (PSE, defined by the abscissa's value for 50% proportion 'right' responses) of the psychometric functions for the spatial conflict conditions can be employed to compute the empirical sensory weights for the different conditions (for further details see (Ernst and Banks, 2002; Fetsch et al., 2012)).

In short, we (i) fitted psychometric or neurometric functions to unisensory, audiovisual congruent and small spatial conflict conditions, (ii) derived the sensory weights from the psychometric/neurometric functions (i.e., shift in PSE) of the conflict conditions and derived the variances from the psychometric/neurometric functions of the spatial conflict and congruent conditions (Fig. 2A-D; 3A-D). In the model-based analysis we compared the sensory weights (Fig. 2F, 3F, 4 A/C) and variances (Fig. 2G, 3G, 4B/D) with MLE predictions that were derived from the unisensory conditions (Fig. 2E, 3E). Because MLE predictions do not depend on modality-specific report, we compared the MLE predictions with the empirical sensory weights and variances while pooling over visual and auditory report.

For both behavioral and neural data, we addressed two questions: First, in a model-based analysis using classical statistics, we investigated whether the MLE predictions that were derived from the unisensory conditions were in line with the empirical sensory weights computed from the audiovisual spatial conflict conditions and the variances computed either from the audiovisual conflict and congruent conditions or from the congruent conditions alone. Second, in a model-free analysis using classical statistics we investigated whether the empirical sensory weights and variances were influenced by visual reliability or modality-specific report. For the psychophysics data we also addressed these two questions using Bayesian model comparison to formally compare the MLE model to alternative models that do or do not allow visual reliability and modality-specific report to influence the PSEs (i.e., Gaussian means) and/or slopes (i.e., Gaussian variances) of the audiovisual conditions.

- Figure 2 about here –

Psychophysics results – Classical statistics

Model-based MLE analysis: The slopes of the psychometric functions for the unisensory conditions indicated that for high visual reliability the visual representations were more reliable than the auditory representations (Fig. 2E). By contrast, for low visual reliability conditions, the variances obtained from the auditory and visual psychometric functions were comparable.

The visual weights obtained from the audiovisual conflict conditions were approximately in line with the MLE predictions derived from those unisensory psychometric functions - though there was a non-significant difference between predicted and empirical visual weights for high visual reliability (Fig. 2F; Tab. 1). Moreover, even though the variance of the perceived signal location was significantly reduced relative to the unisensory auditory condition in case of high visual reliability (Fig. 2G; Tab. 1), it was not reduced relative to the variances obtained for the most reliable unisensory condition. In particular for the low visual reliability conditions where auditory and visual reliabilities were approximately matched, we did not observe a substantial variance reduction as predicted by MLE (Fig. 2G, i.e., a marginally significant difference between MLE predictions and empirical audiovisual variances for low visual reliability, Tab. 1).

Model-free analysis: The visual weights were marginally greater for high relative to low visual reliability (Tab. 2). Yet, contrary to the MLE predictions we also observed a significant effect of modality-specific report on the visual weights. Visual weights were greater for visual relative to auditory report. For visual report, the visual weight was not significantly smaller than one ($p = 0.219$, one-sided Wilcoxon signed rank test pooling across visual reliability) indicating that the auditory signal did not significantly influence visual location reports. For auditory report, the visual weight was significantly larger than zero ($p = 0.032$) indicating that the visual signal

'attracted' auditory location reports, known as ventriloquist effect (Radeau and Bertelson, 1977). Most importantly, we observed a significant interaction between reliability and modality-specific report (Tab. 2). The interaction arose from the fact that the top-down influences of modality-specific report were more pronounced for the low visual reliability conditions when the auditory and visual reliabilities were approximately matched. Indeed, for low visual reliability conditions the psychometric functions of the cue conflict conditions are shifted towards the true visual location for visual report (Fig. 2 D) but towards the true auditory location during auditory report (Fig. 2 B). By contrast, for high visual reliability conditions, the psychometric functions of the cue conflict conditions are shifted towards the true visual location for both auditory and visual report (Fig. 2 A, C).

The variance of the perceived signal location was significantly influenced by visual reliability (Tab. 2), but not by modality-specific report. Critically, we observed a significant interaction between both factors. The significant interaction resulted from the fact that the effect of modality-specific report was revealed predominantly for high visual reliability, but not for low visual reliability, when the auditory and visual reliabilities were approximately matched. The results suggest that the variance of the perceived signal location was influenced predominantly by the sensory modality that needed to be attended and reported. In other words, participants did not fuse sensory signals into one unified percept. Instead, modality-specific report increased the influence of the reported signal in the final percept. Importantly, the interaction effect was also observed when we estimated the audiovisual variance selectively from the audiovisual congruent conditions (interaction of visual reliability and modality-specific report: $F_{1,4} = 34.507$, $p = 0.004$; effect of visual reliability: $F_{1,4} = 23.721$, $p = 0.008$). The results confirm that modality-specific report can selectively increase the influence of the reported sensory signal on the perceived

signal location under classical forced-fusion conditions where sensory signals co-occur in space and time. If observers report the auditory location, the variance is determined predominantly by the variance of the auditory signals (and vice versa for visual report).

Psychophysics results – Bayesian model comparison

In line with the results from classical statistics, the formal Bayesian model comparison demonstrated that the MLE model was not the best model of our data. Instead, the strongest model evidence was observed for a model where visual reliability and modality-specific report influenced the PSE and slope parameters unconstrained by MLE predictions (i.e., protected exceedance probability = 0.916; Tab. 3). Critically, the model evidence combines an accuracy (i.e., model fit) and a complexity term that penalizes complex models with more free parameters. For instance, the MLE model is very parsimonious with only 5 parameters, while the winning model includes 17 free parameters. Our results thus suggest that modeling effects of reliability and modality-specific report are critical to account for observer's localization responses.

Psychophysics results - summary

Collectively, our psychophysics results suggest that auditory and visual signals were integrated approximately weighted by their relative reliabilities. However, the weights were not assigned solely in proportion to the relative bottom-up sensory reliabilities as predicted by the MLE model but were also modulated by modality-specific report and potentially associated attentional processes. The visual weight was greater when the location of the visual signal was attended and reported. Likewise, the variance of the perceived signal location depended on modality-specific report. Hence, irrespective of whether the audiovisual signals were congruent

or in small spatial conflict, participants did not integrate them into one unified percept as predicted by MLE. Instead, they were able to selectively control the influence of auditory or visual signal components depending on task instructions. As a result, observers did not significantly benefit from audiovisual stimulation: there was no reduction in variance of the perceived signal location relative to the most reliable unisensory percept as predicted by MLE optimal integration.

fMRI results

To investigate the neural processes by which human observers integrate sensory signals into spatial representations, we decoded spatial information from fMRI activation patterns. The patterns were extracted from low-level visual regions (V1-V3), low-level auditory regions (primary auditory cortex and planum temporale) and intraparietal sulcus (IPS0-4). We trained a support-vector classification model on fMRI activation patterns selectively from audiovisual congruent conditions ($\Delta AV = 0°$) to learn the mapping from activation patterns to the signal location label (i.e., left vs. right). The trained model then decoded the signal location class (i.e., left vs. right) from activation patterns in audiovisual spatial conflict conditions (i.e., $\Delta AV = \pm 6°$) as well as unisensory auditory and visual conditions. The decoded signal location class, i.e., the 'left/right location response' given by a particular brain area, was then analyzed using the same procedures that were applied to the categorized (i.e., left vs. right) behavioral location responses (see above).

- Figure 3 about here -

716 Visual regions

717 Auditory influences under unisensory auditory stimulation: In line with previous reports of
718 multisensory influences at the primary cortical level (Meyer et al., 2010; Liang et al., 2013;
719 Vetter et al., 2014) we observed a significant positive slope of the psychometric function
720 estimated for the unisensory auditory conditions in low-level visual areas ($p < 0.001$, one-sided
721 bootstrap test). These results indicate that auditory signals (when presented in isolation) elicit
722 spatial representations in low-level visual regions (V1-3). Yet, going beyond previous studies
723 (Meyer et al., 2010; Liang et al., 2013; Vetter et al., 2014) our results demonstrate that these
724 auditory influences on visual cortex (in the absence of concurrent visual signals) are rather
725 limited and induce only unreliable representations when compared to the spatial representations
726 decoded under unisensory visual stimulation (cf. visual and auditory variance obtained for the
727 neurometric functions under unisensory stimulation in low-level visual areas, Fig. 4 B).

728 Model-based MLE analysis: Based on those unisensory visual and auditory neurometric
729 functions, MLE predicted negligible auditory influences on spatial representations during
730 audiovisual stimulation (Fig. 4A). Indeed, in line with those MLE predictions, the
731 representations formed from audiovisual signals relied predominantly on visual input as
732 indicated by a visual weight which did not significantly deviate from one ($p = 0.818$, one-sided
733 bootstrap test pooling the visual weight across conditions). Moreover, in line with MLE
734 predictions, the variance of the audiovisual representations was comparable to unisensory visual
735 variances (Fig 4B, Tab. 1).

736 Model-free analysis: The sensory weights were not significantly modulated by visual
737 reliability or modality-specific report (Tab. 2). Yet, the audiovisual variance was smaller for high
738 as compared to low visual reliability indicating that the representations under audiovisual

stimulation are predominantly determined by the visual signals and hence depend solely on the reliability of the visual signal.

- Figure 4 about here –

Auditory regions

Visual influences under unisensory visual stimulation: In parallel to our findings in visual regions, the slope of the neurometric functions estimated from the unisensory visual conditions was again significantly positive indicating that visual signals alone elicit spatial representations in auditory areas ($p = 0.004$, one-sided bootstrap test pooling across visual reliability). Yet, when compared to the spatial representations decoded under unisensory auditory stimulation, these visual influences on auditory cortex (in the absence of concurrent auditory signals) were rather limited and induced only unreliable representations (cf. visual and auditory variance obtained from the neurometric functions under unisensory stimulation in low-level auditory areas, Fig. 4 D).

Model-based MLE analysis: Based on those unisensory variances, the MLE model predicted a visual weight close to zero (Fig. 4C) and an audiovisual variance approximately identical to the auditory variance for the audiovisual conditions irrespective of visual reliability or modality-specific report (Fig. 4D). While we did not observe any significant deviations from the MLE predictions, the empirical visual weight was greater than predicted by MLE. This was particularly pronounced for high visual reliability conditions. Figure 4C reveals that this deviation emerged predominantly for conditions when the visual signal needs to be attended and reported. These findings may be explained by crossmodal attentional top-down effects operating

762 from vision to audition. Indeed, the visual weight was significantly greater than zero ($p = 0.004$;
763 one-sided bootstrap test pooling the visual weight across conditions) indicating that visual
764 signals exerted a stronger influence on auditory areas during audiovisual stimulation than vice
765 versa (see above: the visual weight was not significantly lower than one in visual regions).
766 Model-free analysis: We did not observe an effect of visual signal reliability, modality-
767 specific report or an interaction between the two factors on the visual weight or variance
768 estimated from the audiovisual conditions in auditory regions (Tab. 2).
769
770 Parietal areas
771 Model-based MLE analysis: In IPS0-4 the neurometric functions for the unisensory conditions
772 indicated that the neural representations for unisensory visual signals were more reliable than
773 those for unisensory auditory signals at both levels of signal reliability (Fig. 3E). This greater
774 reliability of visual IPS representations is consistent with the well-established visual dominance
775 of IPS (Swisher et al., 2007; Wandell et al., 2007). Based on these unisensory variances MLE
776 predicted a visual weight that was close to one for high visual reliability and decreased for low
777 visual reliability. Indeed, the visual weights estimated from the audiovisual conditions were
778 approximately in accordance with these MLE predictions (Fig. 3F, Tab. 1). By contrast, the
779 empirical audiovisual variance was only in line with the MLE predictions for low visual
780 reliability conditions, but significantly smaller than MLE predictions for high visual reliability
781 conditions (Fig. 3G; Tab. 1). This surprising result needs to be further investigated and replicated
782 in future studies.
783 Model-free analysis: In IPS0-4, the visual weight and the audiovisual variance were
784 modulated by visual reliability and modality-specific report (Tab. 2). IPS0-4 integrated

audiovisual signals depending on bottom-up visual reliability and top-down effects of modality-specific report approximately in line with the profile of the behavioral weights (Fig. 3F). Likewise, the audiovisual variance was reduced for high relative to low visual reliability conditions. Moreover, modality-specific report also marginally influenced the variance of the spatial representation obtained from the audiovisual conditions. The variance for the audiovisual conditions was smaller for auditory than visual report (n.b., this marginally significant modulation of variance by modality-specific report was also observed when the analysis focused selectively on the audiovisual spatially congruent conditions, $p = 0.096$). The smaller variance for auditory relative to visual report in IPS contrasts with the variance reduction under visual report observed at the behavioral level (n.b., this difference cannot be explained by methodological differences, because we observed comparable results when applying a fixed-effects analysis at the behavioral level). Potentially this neurobehavioral dissociation can be explained by the fact that the auditory report conditions were more difficult and engaged more attentional resources thereby leading to an increase in reliability of BOLD-activation patterns. Most importantly, however, both behavioral and neural data provide convergent evidence that the sensory weights and to some extent the variances -even for audiovisually congruent trials- depend on both bottom-up visual reliability and top-down effects of modality-specific report.

Control analyses: Eye movements, motor planning, interhemispheric activation differences

No significant differences in eye movement indices (% saccades, % eye blinks, post-stimulus mean horizontal eye position) were observed across any audiovisual conditions (see the supplemental results reported in (Rohe and Noppeney, 2016)). For the unisensory visual conditions, we observed only a small significant effect of the visual signal location on the post-

808 stimulus mean horizontal eye position ($F_{3,9} = 4.9$, $p = 0.028$). However, this effect did not depend
809 on the reliability of the visual signal.

810 Further, a control analysis that decoded IPS activation patterns from a GLM that
811 accounted for participants' trial-wise button responses revealed highly similar results for sensory
812 weights and audiovisual variances (Fig. 3-1) as our initial analysis. These results suggest that IPS
813 represents audiovisual spatial representations that cannot be completely attributed to motor
814 planning and response selection.

815 Finally, given the predominantly contralateral representations of the peri-personal space
816 in visual (Wandell et al., 2007) and auditory regions (Ortiz-Rios et al., 2017) we investigated the
817 impact of global activation differences between the left and right hemispheres on classification
818 performance. When we removed interhemispheric activation differences from activation patterns
819 prior to decoding, we found comparable results for sensory weights and audiovisual variances
820 (Fig. 3-2 and 4-1). Thus, audiovisual spatial representations are encoded in hemisphere-specific
821 activation patterns that go beyond differences in global signal across hemispheres in visual and
822 auditory regions.

823

824 - Figure 5 about here –

825

826 *Dynamic Causal Modelling*

827 Our multivariate pattern analysis showed that visual reliability and modality-specific report
828 influenced visual weights and audiovisual variances in IPS0-4. Using DCM and Bayesian model
829 comparison we next investigated whether these influences were mediated by modulatory effects
830 of reliability on effective connectivity from V1-3 to IPS0-4 and modality-specific report on

connectivity from PFC to IPS0-4 (Fig. 5). PFC potentially mediates the effect of modality-specific report because PFC exerts top-down control on sensory processing (Noudoost et al., 2010; Zanto et al., 2011) by changing the connectivity to parietal regions (Buschman and Miller, 2007). Indeed, in the winning model visual reliability modulated the connection from V1-3 to IPS0-4 and modality-specific report modulated the connection from PFC to IPS0-4 (i.e., protected exceedance probability = 0.699; Tab. 4).

Discussion

The classical MLE model assumes that auditory and visual signals that arise from a common source are integrated weighted by their sensory reliabilities into one unified representation. Critically, the sensory weights are thought to be determined solely by the reliabilities of the sensory signals and immune to task-dependent top-down control. Indeed, abundant evidence suggests that human observers can combine signals within and across the senses near-optimally as predicted by the MLE model (Jacobs, 1999; Ernst and Banks, 2002; van Beers et al., 2002; Knill and Saunders, 2003; Alais and Burr, 2004; Hillis et al., 2004; Saunders and Knill, 2004; Rosas et al., 2005); but see Battaglia et al., 2003). While the forced-fusion assumption of a common signal cause usually holds for integration within a sensory modality (Hillis et al., 2002), it is often violated when integrating signals across sensory modalities (Gepshtein et al., 2005; Parise et al., 2012). For example, it remains controversial whether or not multisensory integration and more specifically the sensory weights can be modulated by top-down control (Helbig and Ernst, 2008; Talsma et al., 2010; Vercillo and Gori, 2015).

The present study investigated the extent to which audiovisual spatial signals are integrated in line with the quantitative predictions of the MLE model at the behavioral and neural levels. Importantly, while many previous MLE studies (Battaglia et al., 2003; Alais and Burr, 2004) presented only signals with a small conflict and asked participants to report the location of the 'audiovisual stimulus' thereby encouraging integration of signals into one unified percept, we instructed participants to attend and report either the visual or the auditory signals (cf. (Stein et al., 1989; Wallace et al., 2004; Kording et al., 2007)). Thus, our task-instructions instructed observers to focus selectively on one signal component rather than treating the two signals as originating necessarily from one common object.

862 At the behavioral level, our results demonstrate that observers did not integrate signals
863 into one unified percept as predicted by MLE. The visual weight increased not only when the
864 visual signal was more reliable but also when it needed to be reported. Most importantly, even
865 when auditory and visual signals were spatially congruent, i.e., likely to originate from one
866 single object, observers were able to focus selectively on one sensory modality as indicated by
867 differences in variance for the spatial representations obtained from auditory and visual report
868 conditions. In other words, modality-specific attention and report modulated not only the sensory
869 weights during the spatial conflict conditions but also during the congruent conditions. Yet, the
870 advantage of being able to selectively control the relative sensory contributions to the final
871 percept came at the price of not obtaining the multisensory benefit (i.e., a reduction in variance
872 for the perceived signal location) that is afforded by reliability-weighted integration according to
873 MLE principles (Ernst and Banks, 2002; Alais and Burr, 2004).

874 Next, we investigated how auditory and visual signals were integrated into spatial
875 representations at the neural level focusing on low-level visual areas (V1-3), low-level auditory
876 areas (primary auditory areas and planum temporale) and parietal areas (IPS0-4). Combining
877 fMRI multivariate decoding with classical MLE analysis (as in our psychophysics analysis), we
878 obtained neural weights and variances of spatial representations from neurometric functions that
879 were computed based on spatial locations decoded from regional BOLD-response patterns.

880 Consistent with previous reports of multisensory influences and interactions in primary
881 sensory areas (Foxe et al., 2000; Bonath et al., 2007; Kayser et al., 2007; Lakatos et al., 2007;
882 Lewis and Noppeney, 2010; Werner and Noppeney, 2010; Bonath et al., 2014; Lee and
883 Noppeney, 2014), unisensory auditory signals elicited spatial representations in visual cortex and
884 visual signals in auditory cortex. In other words, unisensory signals from non-preferred sensory

885 modalities can be decoded from low-level sensory areas (Meyer et al., 2010; Liang et al., 2013;
886 Vetter et al., 2014). Yet, the unisensory neurometric functions demonstrated that the spatial
887 representation decoded from low-level visual (resp. auditory) areas were far more reliable for
888 signals from the preferred than non-preferred sensory modality. As a result and in line with MLE
889 predictions, the sensory weights applied during multisensory integration to the signals from the
890 auditory modality were negligibly small in low-level visual areas. In auditory areas the visual
891 weight was also small at least during auditory report but significantly different from zero.
892 Further, neither the sensory weights nor the variance depended significantly on the reported
893 sensory modality. Instead the variance of the spatial representation decoded from audiovisual
894 signals was comparable to the unisensory visual variance in visual regions and comparable to
895 unisensory auditory variance in auditory regions. Hence, our quantitative analysis based on
896 neurometric functions moves significantly beyond previous research that demonstrated better
897 than chance decoding performance for auditory signals from visual areas and vice versa (Meyer
898 et al., 2010; Liang et al., 2013; Vetter et al., 2014). It demonstrates that signals from the non-
899 preferred sensory modality elicit representations that are far less reliable than those evoked by
900 signals from the preferred sensory modality. Likewise, non-preferred signals exert only limited
901 influences on spatial representations in low-level sensory areas during audiovisual stimulation.
902 Surprisingly, visual signals exerted stronger influences on auditory areas than vice versa
903 potentially reflecting the importance of visual inputs for spatial perception (Welch and Warren,
904 1980).

905 In higher-order areas IPS0-4, unisensory auditory and visual signals elicited spatial
906 representations that were more comparable in their reliabilities. Yet, consistent with the well-
907 known visual response properties of IPS0-4 (Swisher et al., 2007; Wandell et al., 2007) visual

908 stimulation elicited more reliable representations. Hence, as predicted by MLE, IPS0-4 gave a
909 stronger weight to the visual signal during multisensory integration. Potentially, IPS could
910 implement reliability-weighted integration via probabilistic population codes (Ma et al., 2006) or
911 normalization over the pool of neurons within a region (Ohshiro et al., 2011, 2017). Because we
912 used a linear SVM classifier as decoder, it remains unclear which encoding scheme IPS used to
913 represent audiovisual space. To investigate the potential neural implementations, future studies
914 may use explicit encoding models (e.g., estimating voxels' tuning function for space using
915 population receptive fields methods) (Dumoulin and Wandell, 2008) to characterize the effects
916 of reliability-weighted multisensory integration on voxel-response tuning functions.

917 However, in contrast to the MLE predictions the sensory weights in IPS were not only
918 modulated by visual reliability, but also by the sensory modality that needed to be reported. The
919 visual signal had a stronger influence on the decoded spatial representation during visual than
920 auditory report thereby reflecting the sensory weight profile observed at the behavioral level.
921 Likewise, the variance of the spatial representation for audiovisual stimuli in IPS0-4 was
922 marginally influenced by the modality of the reported signal suggesting that the formation of
923 audiovisual representations in IPS0-4 may be susceptible to top-down control. Dynamic Causal
924 Modelling and Bayesian model comparison suggested that these changes in audiovisual spatial
925 representations in IPS0-4 were mediated by modulatory effects: Visual reliability modulated the
926 bottom-up connections from V1-3 to IPS0-4 and modality-specific report modulated the top-
927 down connections from PFC to IPS0-4.

928 Our results demonstrate that observers do not fully integrate auditory and visual signals
929 into unified spatial representations at the behavioral level and neural level in higher-order
930 association areas IPS0-4. Even when auditory and visual signals were spatiotemporally

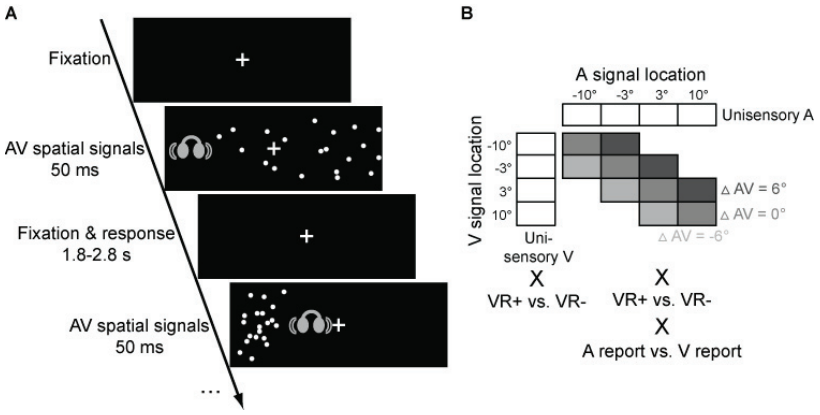
congruent and hence likely to originate from a common source, the sensory signal that needed to be reported had a stronger influence on the spatial representations than the one that was to be ignored. An important aim for future studies is to determine how a change in reported sensory modality modulates audiovisual integration and to dissociate between two main mechanisms: First, modality-specific report may influence the sensory weights via attentional mechanisms. Attention is known to increase the signal-to-noise ratio or reliability of the signal in the attended sensory modality (Desimone and Duncan, 1995; Martinez-Trujillo and Treue, 2004; Briggs et al., 2013; Sprague et al., 2015). Thereby, attention mediates a greater weight in the multisensory integration process (Alsius et al., 2005; Busse et al., 2005; Talsma and Woldorff, 2005; Alsius et al., 2007; Talsma et al., 2007; Talsma et al., 2010; Zimmer et al., 2010a; Zimmer et al., 2010b; Donohue et al., 2011; Vercillo and Gori, 2015; Macaluso et al., 2016); but see Helbig and Ernst, 2008). In this model, auditory and visual signals are integrated weighted by their sensory reliabilities. Yet, in contrast to the MLE model the reliability of each sensory input can be modified prior to audiovisual integration by top-down attention as manipulated by modality-specific report. Second, modality-specific report instructs participants not to fuse signals into one unified percept but to form a spatial estimate selectively for one of the two signals. These instructions may attenuate the integration process even for signals that are collocated in space thereby enabling participants to compute a final spatial estimate that is more strongly based on the reported sensory modality. In this second case, MLE analyses compute a stronger weight for the reported signal because of its task-relevance rather than attentionally increased sensory reliability. Yet, human behavior in this second case is better accommodated by recent Bayesian causal inference models that explicitly model the potential causal structures of the multisensory signals, that is whether they have been caused by common or independent causes (Kording et

al., 2007; Shams and Beierholm, 2010; Wozny et al., 2010; Rohe and Noppeney, 2015a, 2016). In Bayesian causal inference, a final estimate of the spatial location under auditory or visual report is obtained by combining the estimates under the two causal structure, i.e., the MLE reliability-weighted estimate under the assumption of a common source and the estimate of the sensory signals that needs to be reported under the assumption of independent causes. Because the underlying causal structure is uncertain and modality-specific report instructions may further lower the observers' belief that signals are caused by a common source, the reported spatial estimates differ for auditory and visual reports, thereby modelling effects of modality-specific report. Further, because in the course of our experiment the audiovisual signals were spatially uncorrelated across all conditions (i.e., the auditory and the visual signal locations were independently sampled from the four locations, see Fig. 1B), participants might have implicitly learnt a low prior probability of a common cause. Thus, even in conditions in which the audiovisual signals only had a small spatial disparity (which we selectively used in our analyses), participants might have computed a low posterior belief that signals arose from a common cause. In general, previous research has shown that Bayesian causal inference outperforms the MLE model under conditions in which a common cause is unlikely, for example a large spatial discrepancy between the audiovisual signals (Kording et al., 2007; Rohe and Noppeney, 2015a, b). To dissociate the effects of modality-specific attention and report, future studies may use attentional cuing paradigms that pre-cue participants prior to stimulus presentation to attend to the visual (resp. auditory) signal and post-cue them after stimulus presentation to report the location of the auditory (resp. visual) signal.

To conclude, the present study characterized how the brain integrates auditory and visual signals into spatial representations and how these integration processes are modulated by modality-specific report or attention. Combining psychophysics and multivariate fMRI-decoding we demonstrated that classical MLE models cannot fully account for participants' behavioral and neural responses if the experimental context (i.e., modality-specific report and overall uncorrelated audiovisual signals) undermines observers' perception of a common signal cause, thus violating the MLE model's core assumption. While the behavioral and neural weights in parietal cortex depended on the relative sensory reliabilities in line with the quantitative predictions of the MLE model, they were also modulated by whether participants attended and reported the visual or the auditory signal location. Likewise, the variance of the spatial representations depended on task-context to some extent even for collocated audiovisual signals both at the neural and behavioral level. These results suggest that audiovisual integration can be modulated by top-down control. Even when the auditory and visual signals were spatially close (or collocated) and temporally synchronous, modality-specific report influenced how they were weighted and integrated into spatial representations.

994 **Figures**

995 Figure 1



996

997

998

999

1000

1001

Figure 2

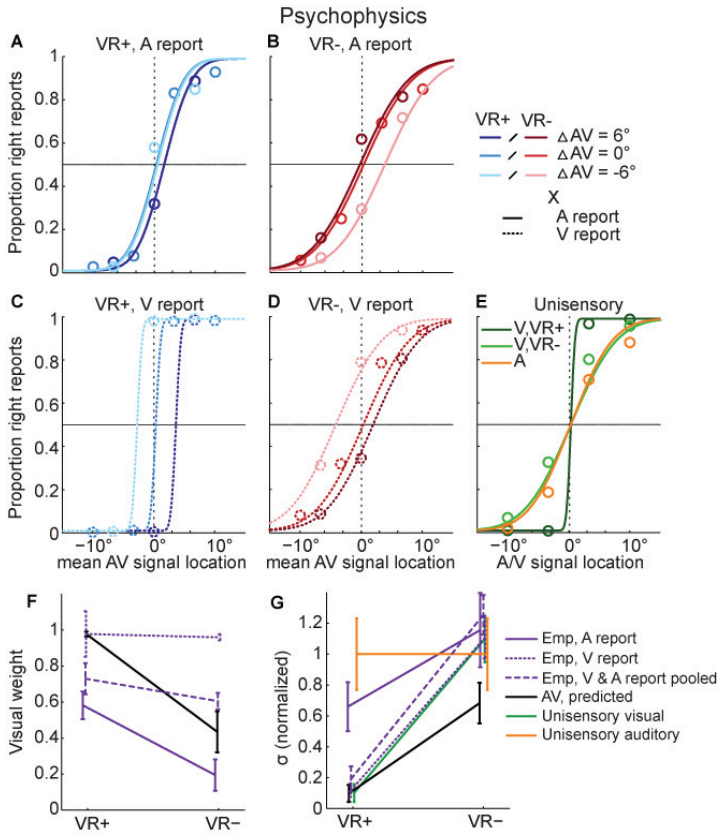
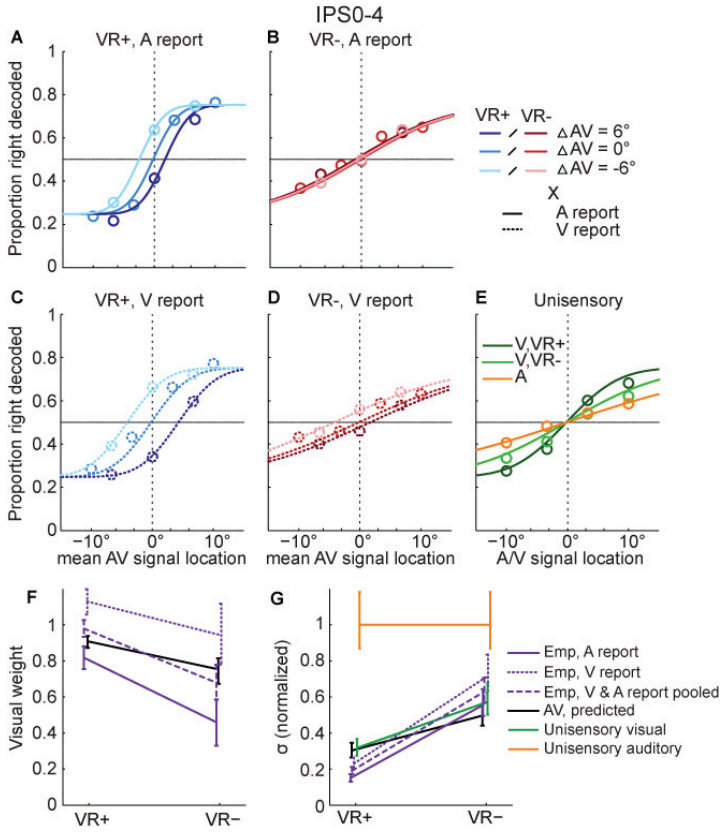
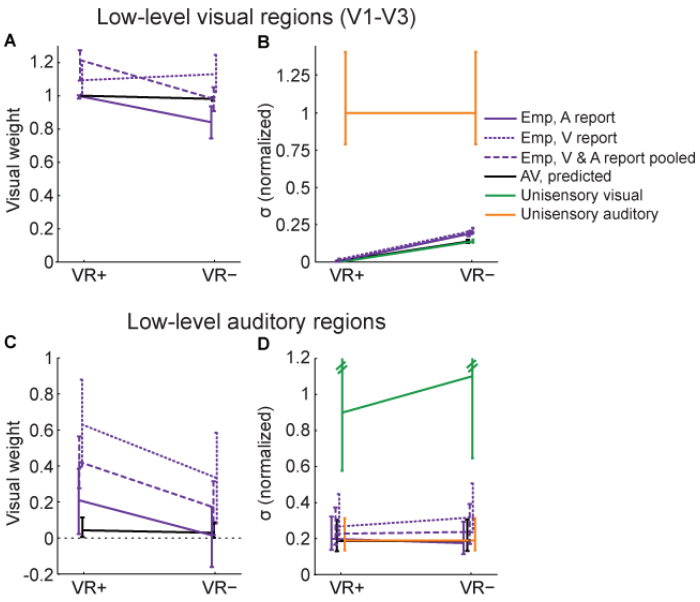


Figure 3



1022 Figure 4



1023

1024

1025

1026

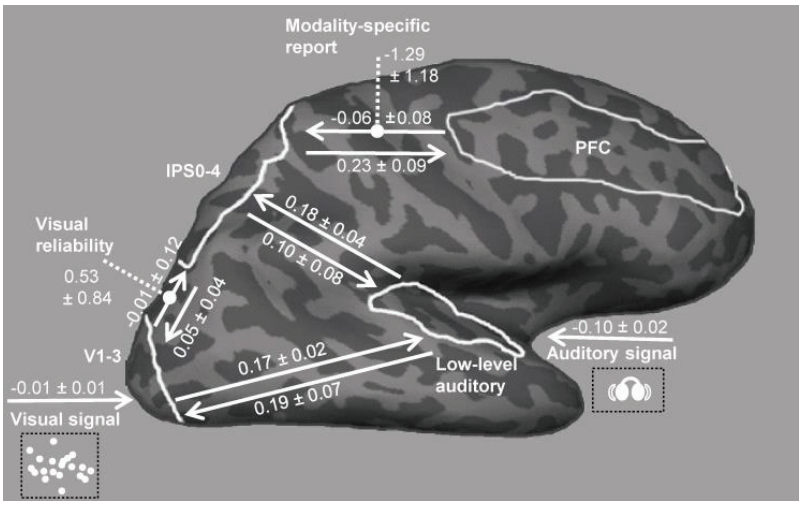
1027

1028

1029

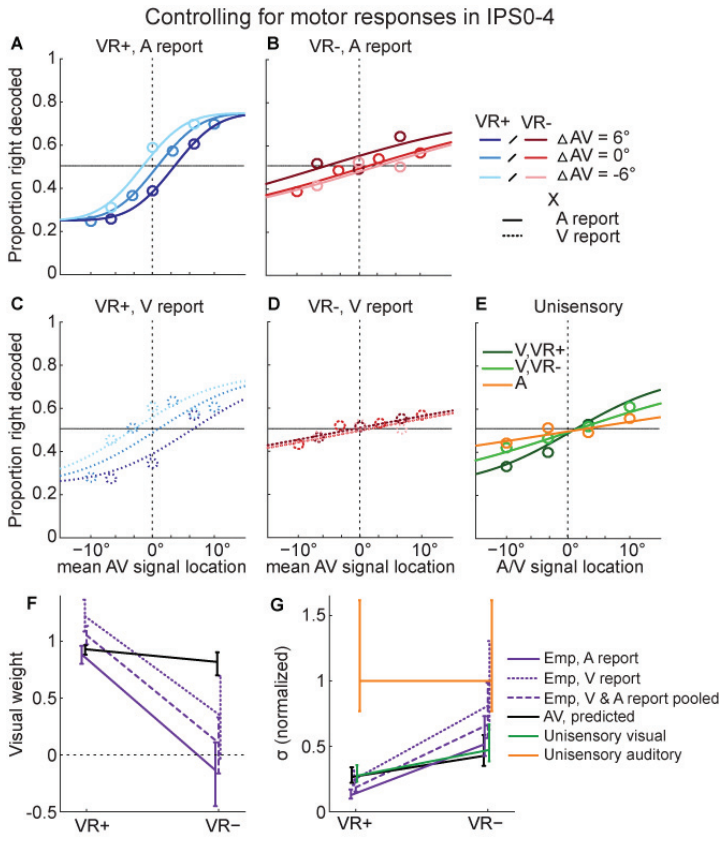
1030

1031 Figure 5



1032

1033 Figure 3-1



1034

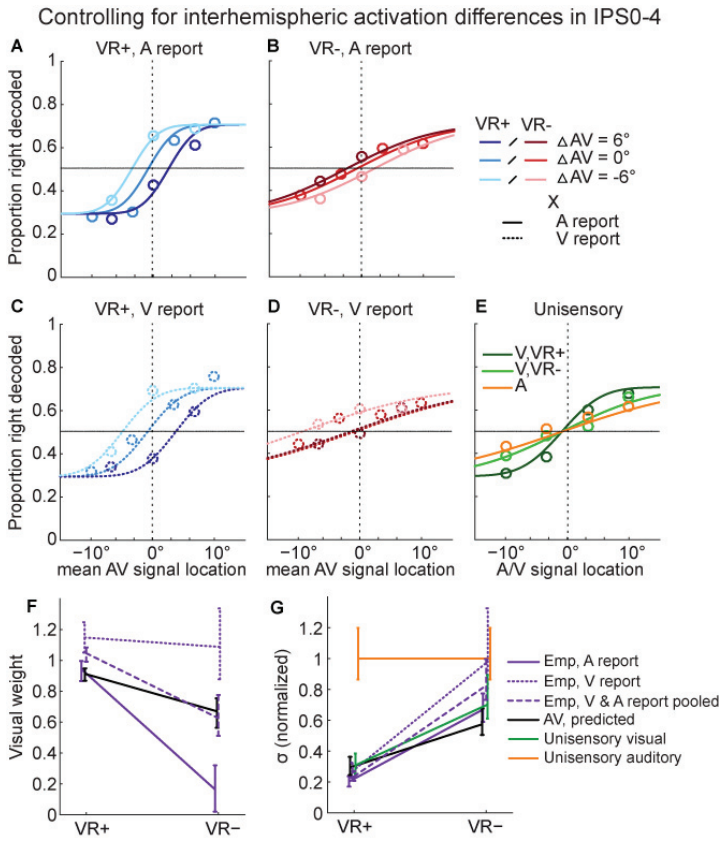
1035

1036

1037

1038

1039 Figure 3-2



1040

1041

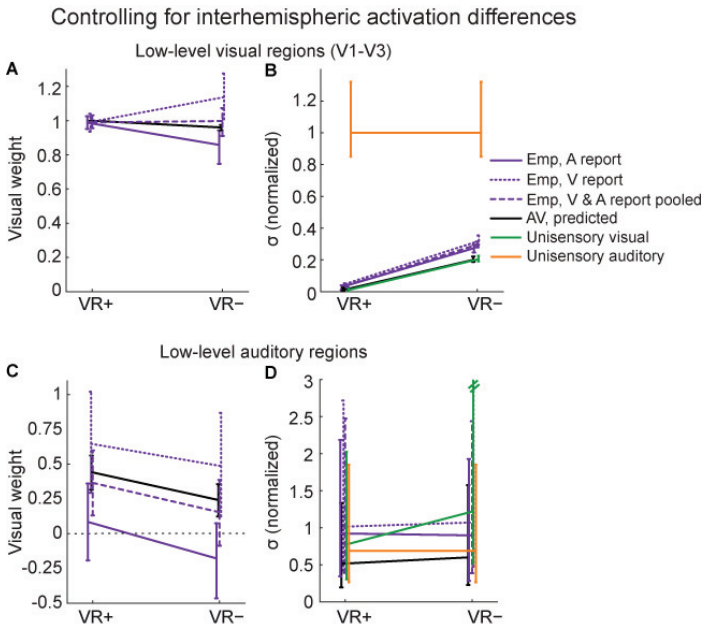
1042

1043

1044

1045

1046 Figure 4-1



1047

1048

1049

1050

1051

1052

Tables

Table 1. Statistical comparison of empirical weights ($w_{V,emp}$) and standard deviations ($\sigma_{AV,emp}$) obtained from the psychometric (behavior) and neurometric (fMRI) functions pertaining to the audiovisual conditions of high and low visual reliability with the MLE predictions ($\sigma_{AV,pred}$, $w_{V,pred}$) and unisensory standard deviations (σ_{uniV}, σ_{uniA}).

		VR+	VR-
		$w_{V,emp} - w_{V,pred}$	
Psychophysics	p	0.0629	0.129
V1-V3	p	0.230	1
IPS0-4	p	0.631	1
Low-level auditory	p	0.064	1
		$\sigma_{AV,emp} - \sigma_{AV,pred}$	
Psychophysics	p	0.188	0.0629
V1-V3	p	0.241	**0.001**
IPS0-4	p	**0.020**	0.275
Low-level auditory	p	1	1
		$\sigma_{AV,emp} - \sigma_{uniV}$	

Psychophysics	p	0.188	0.438
V1-V3	p	0.241	**0.001**
IPS0-4	p	**0.022**	1
Low-level auditory	p	0.883	0.963
		$\sigma_{AV,emp} - \sigma_{uniA}$	
Psychophysics	p	0.063	0.313
V1-V3	p	0.104	0.169
IPS0-4	p	**0.002**	0.086
Low-level auditory	p	1	1

1053

1054

Table 2. Effects of visual reliability (VR), modality-specific report (MR) and their interaction (VRxMR) on empirical weights ($w_{V,emp}$) and standard deviations ($\sigma_{AV,emp}$) obtained from the psychometric (behavior) and neurometric (fMRI) functions.

		$w_{V,emp}$			$\sigma_{AV,emp}$		
		VR	MR	VRxMA	VR	MR	VRxMR
Psychophysics	[F, p]	5.149,0.086	16.308, **0.016**	8.605, **0.043**	19.129, **0.012**	2.172, 0.215	18.892, **0.012**
V1-V3	p	0.346	0.131	0.957	**< 0.001**	0.142	1
IPS0-4	p	**0.022**	**0.001**	1	**<0.001**	0.051	1
Low-level auditory	p	0.693	0.217	1	1	0.419	1

1055

1056

Table 3. Results of the Bayesian model comparison between five competing models of the psychometric data.

Model:	I: Null model	II: MLE model	III: Reliability-weighting	IV: Full model
# parameters	8	5	11	17
R^2 (mean)	0.656	0.658	0.687	0.722
Relative BIC (sum)	0	59.662	1501.077	3202.034
Exp. post. p.	0.111	0.111	0.111	0.667
Exceed. p.	0.014	0.014	0.014	0.957
Prot. exceed. p.	0.026	0.026	0.026	0.921

1057

Table 4. Results of the model comparison of the 2 x 2 Dynamical Causal Models in which visual reliability (VR) modulated the connection from V1-3 to IPS0-4 and modality-specific report (MR) modulated the connection from PFC to IPS0-4.

	Modulation VR & MR	Modulation VR	Modulation MR	No Modulation
Model evidence (FFX)	0	-52.947	-54.033	-90.45
Posterior p. (FFX)	1	0	0	0
Exp. posterior p. (RFX)	0.587	0.136	0.139	0.139
Exceed. p. (RFX)	0.902	0.032	0.033	0.033
Prot. exceed. p. (RFX)	0.699	0.1	0.101	0.101

References

Alais D, Burr D (2004) The ventriloquist effect results from near-optimal bimodal integration. Curr Biol 14:257-262.

Algazi VR, Duda RO, Thompson DM, Avendano C (2001) The cipic hrtf database. In: Applications of Signal Processing to Audio and Acoustics, 2001 IEEE Workshop on the, pp 99-102: IEEE.

Alsius A, Navarra J, Soto-Faraco S (2007) Attention to touch weakens audiovisual speech integration. Exp Brain Res 183:399-404.

Alsius A, Navarra J, Campbell R, Soto-Faraco S (2005) Audiovisual integration of speech falters under high attention demands. Curr Biol 15:839-843.

Andersen RA, Buneo CA (2002) Intentional maps in posterior parietal cortex. Annual review of neuroscience 25:189-220.

Ban H, Preston TJ, Meeson A, Welchman AE (2012) The integration of motion and disparity cues to depth in dorsal visual cortex. Nat Neurosci 15:636-643.

Battaglia PW, Jacobs RA, Aslin RN (2003) Bayesian integration of visual and auditory signals for spatial localization. J Opt Soc Am A Opt Image Sci Vis 20:1391-1397.

Beauchamp MS, Argall BD, Bodurka J, Duyn JH, Martin A (2004) Unraveling multisensory integration: patchy organization within human STS multisensory cortex. Nat Neurosci 7:1190-1192.

Bizley JK, Nodal FR, Bajo VM, Nelken I, King AJ (2007) Physiological and anatomical evidence for multisensory interactions in auditory cortex. Cereb Cortex 17:2172-2189.

Bonath B, Noesselt T, Krauel K, Tyll S, Tempelmann C, Hillyard SA (2014) Audio-visual synchrony modulates the ventriloquist illusion and its neural/spatial representation in the auditory cortex. NeuroImage 98:425-434.

Bonath B, Noesselt T, Martinez A, Mishra J, Schwiecker K, Heinze HJ, Hillyard SA (2007) Neural basis of the ventriloquist illusion. Curr Biol 17:1697-1703.

Brainard DH (1997) The psychophysics toolbox. Spatial vision 10:433-436.

Briggs F, Mangun GR, Usrey WM (2013) Attention enhances synaptic efficacy and the signal-to-noise ratio in neural circuits. Nature 499:476-480.

Buschman TJ, Miller EK (2007) Top-down versus bottom-up control of attention in the prefrontal and posterior parietal cortices. Science 315:1860-1862.

Busse L, Roberts KC, Crist RE, Weissman DH, Woldorff MG (2005) The spread of attention across modalities and space in a multisensory object. Proc Natl Acad Sci U S A 102:18751-18756.

Chang CC, Lin CJ (2011) LIBSVM: a library for support vector machines. ACM Transactions on Intelligent Systems and Technology (TIST) 2:27.

Conover WJ, Iman RL (1981) Rank transformations as a bridge between parametric and nonparametric statistics. The American Statistician 35:124-129.

Dahl CD, Logothetis NK, Kayser C (2009) Spatial organization of multisensory responses in temporal association cortex. J Neurosci 29:11924-11932.

Dale AM, Fischl B, Sereno MI (1999) Cortical surface-based analysis. I. Segmentation and surface reconstruction. NeuroImage 9:179-194.

de Beeck HPO (2010) Against hyperacuity in brain reading: spatial smoothing does not hurt multivariate fMRI analyses? NeuroImage 49:1943-1948.

Desikan RS, Ségonne F, Fischl B, Quinn BT, Dickerson BC, Blacker D, Buckner RL, Dale AM, Maguire RP, Hyman BT (2006) An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. NeuroImage 31:968-980.

Desimone R, Duncan J (1995) Neural mechanisms of selective visual attention. Annual review of neuroscience 18:193-222.

1107 Donohue SE, Roberts KC, Grent-'t-Jong T, Woldorff MG (2011) The cross-modal spread of attention
1108 reveals differential constraints for the temporal and spatial linking of visual and auditory
1109 stimulus events. J Neurosci 31:7982-7990.
1110 Dumoulin SO, Wandell BA (2008) Population receptive field estimates in human visual cortex.
1111 NeuroImage 39:647-660.
1112 Efron B, Tibshirani RJ (1994) An introduction to the bootstrap. London: Chapmann and Hall.
1113 Eickhoff SB, Stephan KE, Mohlberg H, Grefkes C, Fink GR, Amunts K, Zilles K (2005) A new SPM toolbox
1114 for combining probabilistic cytoarchitectonic maps and functional imaging data. NeuroImage
1115 25:1325-1335.
1116 Ernst MO, Banks MS (2002) Humans integrate visual and haptic information in a statistically optimal
1117 fashion. Nature 415:429-433.
1118 Fetsch CR, Pouget A, DeAngelis GC, Angelaki DE (2012) Neural correlates of reliability-based cue
1119 weighting during multisensory integration. Nat Neurosci 15:146-154.
1120 Foxe JJ, Morocz IA, Murray MM, Higgins BA, Javitt DC, Schroeder CE (2000) Multisensory auditory-
1121 somatosensory interactions in early cortical processing revealed by high-density electrical
1122 mapping. Brain Res Cogn Brain Res 10:77-83.
1123 Friston KJ, Harrison L, Penny W (2003) Dynamic causal modelling. NeuroImage 19:1273-1302.
1124 Friston KJ, Holmes AP, Worsley KJ, Poline JP, Frith CD, Frackowiak RSJ (1994) Statistical parametric maps
1125 in functional imaging: a general linear approach. Human brain mapping 2:189-210.
1126 Friston KJ, Litvak V, Oswal A, Razi A, Stephan KE, van Wijk BC, Ziegler G, Zeidman P (2016) Bayesian
1127 model reduction and empirical Bayes for group (DCM) studies. NeuroImage 128:413-431.
1128 Gepshtein S, Burge J, Ernst MO, Banks MS (2005) The combination of vision and touch depends on
1129 spatial proximity. J Vis 5:1013-1023.
1130 Ghazanfar AA, Schroeder CE (2006) Is neocortex essentially multisensory? Trends Cogn Sci 10:278-285.
1131 Helbig HB, Ernst MO (2008) Visual-haptic cue weighting is independent of modality-specific attention. J
1132 Vis 8:21 21-16.
1133 Helbig HB, Ernst MO, Ricciardi E, Pietrini P, Thielscher A, Mayer KM, Schultz J, Noppeney U (2012) The
1134 neural mechanisms of reliability weighted integration of shape information from vision and
1135 touch. NeuroImage 60:1063-1072.
1136 Hillis JM, Ernst MO, Banks MS, Landy MS (2002) Combining sensory information: mandatory fusion
1137 within, but not between, senses. Science 298:1627-1630.
1138 Hillis JM, Watt SJ, Landy MS, Banks MS (2004) Slant from texture and disparity cues: optimal cue
1139 combination. J Vis 4:967-992.
1140 Jacobs RA (1999) Optimal integration of texture and motion cues to depth. Vision Res 39:3621-3629.
1141 Kayser C, Petkov CI, Augath M, Logothetis NK (2007) Functional imaging reveals visual modulation of
1142 specific fields in auditory cortex. J Neurosci 27:1824-1835.
1143 Kleiner M, Brainard D, Pelli D, Ingling A, Murray R, Broussard C (2007) What's new in Psychtoolbox-3.
1144 Perception 36:1.1-16.
1145 Knill DC, Saunders JA (2003) Do humans optimally integrate stereo and texture information for
1146 judgments of surface slant? Vision Res 43:2539-2558.
1147 Kording KP, Beierholm U, Ma WJ, Quartz S, Tenenbaum JB, Shams L (2007) Causal inference in
1148 multisensory perception. PloS one 2:e943.
1149 Lakatos P, Chen CM, O'Connell MN, Mills A, Schroeder CE (2007) Neuronal oscillations and multisensory
1150 interaction in primary auditory cortex. Neuron 53:279-292.
1151 Lee H, Noppeney U (2014) Temporal prediction errors in visual and auditory cortices. Curr Biol 24:R309-
1152 310.
1153 Lewis R, Noppeney U (2010) Audiovisual synchrony improves motion discrimination via enhanced
1154 connectivity between early visual and auditory areas. J Neurosci 30:12329-12339.

1155 Liang M, Mouraux A, Hu L, Iannetti G (2013) Primary sensory cortices contain distinguishable spatial
1156 patterns of activity for each sense. Nature Communications 4:1979.
1157 Ma WJ, Beck JM, Latham PE, Pouget A (2006) Bayesian inference with probabilistic population codes.
1158 Nat Neurosci 9:1432-1438.
1159 Macaluso E, Noppeney U, Talsma D, Vercillo T, Hartcher-O'Brien J, Adam R (2016) The Curious Incident
1160 of Attention in Multisensory Integration: Bottom-up vs. Top-down. Multisensory Research
1161 29:557-583.
1162 Martinez-Trujillo JC, Treue S (2004) Feature-based attention increases the selectivity of population
1163 responses in primate visual cortex. Curr Biol 14:744-751.
1164 Meyer K, Kaplan JT, Essex R, Webber C, Damasio H, Damasio A (2010) Predicting visual stimuli on the
1165 basis of activity in auditory cortices. Nature Neuroscience 13:667-668.
1166 Morgan ML, Deangelis GC, Angelaki DE (2008) Multisensory integration in macaque visual cortex
1167 depends on cue reliability. Neuron 59:662-673.
1168 Nagelkerke NJ (1991) A note on a general definition of the coefficient of determination. Biometrika
1169 78:691-692.
1170 Nath AR, Beauchamp MS (2011) Dynamic changes in superior temporal sulcus connectivity during
1171 perception of noisy audiovisual speech. J Neurosci 31:1704-1714.
1172 Noudoost B, Chang MH, Steinmetz NA, Moore T (2010) Top-down control of visual attention. Current
1173 opinion in neurobiology 20:183-190.
1174 Ohshiro T, Angelaki DE, DeAngelis GC (2011) A normalization model of multisensory integration. Nature
1175 Neuroscience 14:775-782.
1176 Ohshiro T, Angelaki DE, DeAngelis GC (2017) A Neural Signature of Divisive Normalization at the Level of
1177 Multisensory Integration in Primate Cortex. Neuron 95:399-411. e398.
1178 Ortiz-Rios M, Azevedo FA, Kuśmierek P, Balla DZ, Munk MH, Keliris GA, Logothetis NK, Rauschecker JP
1179 (2017) Widespread and Opponent fMRI Signals Represent Sound Location in Macaque Auditory
1180 Cortex. Neuron 93:971-983. e974.
1181 Parise CV, Ernst MO (2016) Correlation detection as a general mechanism for multisensory integration.
1182 Nature Communications 7.
1183 Parise CV, Spence C, Ernst MO (2012) When correlation implies causation in multisensory integration.
1184 Curr Biol 22:46-49.
1185 Penny WD, Stephan KE, Mechelli A, Friston KJ (2004) Comparing dynamic causal models. NeuroImage
1186 22:1157-1172.
1187 Prins N, Kingdom FAA (2009) Palamedes: Matlab Routines for Analyzing Psychophysical Data. In.
1188 Radeau M, Bertelson P (1977) Adaptation to auditory-visual discordance and ventriloquism in
1189 semirealistic situations. Perception & Psychophysics 22:137-146.
1190 Raftery AE (1995) Bayesian model selection in social research. Sociological Methodology 1995, Vol 25
1191 25:111-163.
1192 Rigoux L, Stephan KE, Friston KJ, Daunizeau J (2014) Bayesian model selection for group studies—
1193 revisited. NeuroImage 84:971-985.
1194 Roach NW, Heron J, McGraw PV (2006) Resolving multisensory conflict: a strategy for balancing the
1195 costs and benefits of audio-visual integration. Proc Biol Sci 273:2159-2168.
1196 Rohe T, Noppeney U (2015a) Cortical hierarchies perform Bayesian causal inference in multisensory
1197 perception. PLoS Biol 13:e1002073.
1198 Rohe T, Noppeney U (2015b) Sensory reliability shapes perceptual inference via two mechanisms.
1199 Journal of Vision 15:1-16.
1200 Rohe T, Noppeney U (2016) Distinct computational principles govern multisensory integration in primary
1201 sensory and association cortices. Current Biology 26:509-514.

1202 Rosas P, Wagemans J, Ernst MO, Wichmann FA (2005) Texture and haptic cues in slant discrimination:
1203 reliability-based cue weighting without statistically optimal cue combination. J Opt Soc Am A
1204 Opt Image Sci Vis 22:801-809.
1205 Sadaghiani S, Maier JX, Noppeney U (2009) Natural, metaphoric, and linguistic auditory direction signals
1206 have distinct influences on visual motion processing. J Neurosci 29:6490-6499.
1207 Saunders JA, Knill DC (2004) Visual feedback control of hand movements. J Neurosci 24:3223-3234.
1208 Sereno MI, Dale AM, Reppas JB, Kwong KK, Belliveau JW, Brady TJ, Rosen BR, Tootell RB (1995) Borders
1209 of multiple visual areas in humans revealed by functional magnetic resonance imaging. Science
1210 268:889-893.
1211 Shams L, Beierholm UR (2010) Causal inference in perception. Trends Cogn Sci 14:425-432.
1212 Sprague TC, Saproo S, Serences JT (2015) Visual attention mitigates information loss in small-and large-
1213 scale neural codes. Trends in cognitive sciences 19:215-226.
1214 Stein BE, Meredith MA, Huneycutt WS, McDade L (1989) Behavioral Indices of Multisensory Integration:
1215 Orientation to Visual Cues is Affected by Auditory Stimuli. J Cogn Neurosci 1:12-24.
1216 Stephan KE, Penny WD, Daunizeau J, Moran RJ, Friston KJ (2009) Bayesian model selection for group
1217 studies. NeuroImage 46:1004-1017.
1218 Swisher JD, Halko MA, Merabet LB, McMains SA, Somers DC (2007) Visual topography of human
1219 intraparietal sulcus. J Neurosci 27:5326-5337.
1220 Talsma D, Woldorff MG (2005) Selective attention and multisensory integration: multiple phases of
1221 effects on the evoked brain activity. Journal of cognitive neuroscience 17:1098-1114.
1222 Talsma D, Doty TJ, Woldorff MG (2007) Selective attention and audiovisual integration: is attending to
1223 both modalities a prerequisite for early integration? Cerebral cortex 17:679-690.
1224 Talsma D, Senkowski D, Soto-Faraco S, Woldorff MG (2010) The multifaceted interplay between
1225 attention and multisensory integration. Trends Cogn Sci 14:400-410.
1226 van Beers RJ, Wolpert DM, Haggard P (2002) When feeling is more important than seeing in
1227 sensorimotor adaptation. Curr Biol 12:834-837.
1228 Vercillo T, Gori M (2015) Attention to sound improves auditory reliability in audio-tactile spatial optimal
1229 integration. Frontiers in integrative neuroscience 9:34.
1230 Vetter P, Smith FW, Muckli L (2014) Decoding sound and imagery content in early visual cortex. Current
1231 Biology 24:1256-1262.
1232 Wallace MT, Roberson GE, Hairston WD, Stein BE, Vaughan JW, Schirillo JA (2004) Unifying multisensory
1233 signals across time and space. Exp Brain Res 158:252-258.
1234 Wandell BA, Dumoulin SO, Brewer AA (2007) Visual field maps in human cortex. Neuron 56:366-383.
1235 Welch RB, Warren DH (1980) Immediate perceptual response to intersensory discrepancy. Psychol Bull
1236 88:638-667.
1237 Werner S, Noppeney U (2010) Distinct functional contributions of primary sensory and association areas
1238 to audiovisual integration in object categorization. J Neurosci 30:2662-2675.
1239 Wozny DR, Beierholm UR, Shams L (2010) Probability matching as a computational strategy used in
1240 perception. PLoS Comput Biol 6.
1241 Zanto TP, Rubens MT, Thangavel A, Gazzaley A (2011) Causal role of the prefrontal cortex in top-down
1242 modulation of visual processing and working memory. Nature Neuroscience 14:656-661.
1243 Zimmer U, Itthipanyanan S, Woldorff M (2010a) The electrophysiological time course of the interaction
1244 of stimulus conflict and the multisensory spread of attention. European Journal of Neuroscience
1245 31:1744-1754.
1246 Zimmer U, Roberts KC, Harshbarger TB, Woldorff MG (2010b) Multisensory conflict modulates the
1247 spread of visual attention across a multisensory object. NeuroImage 52:606-616.

1248

1249

Figure legends

Figure 1. Example trial and experimental design. (A) Participants were presented with unisensory auditory, unisensory visual and synchronous audiovisual signals originating from four possible locations along the azimuth. The visual signal was a cloud of white dots. The auditory signal was a brief burst of white noise presented via headphones. Participants localized either the auditory or the visual signal (n.b., for illustrational purposes the visual angles of the cloud have been scaled in a non-uniform fashion in this scheme). (B) In the audiovisual conditions, the experimental design manipulated (1) the location of the visual (V) signal (−10°, −3.3°, 3.3°, 10°) (2) the location of the auditory (A) signal (−10°, −3.3°, 3.3°, 10°), (3) the reliability of the visual signal (low versus high standard deviation of the visual cloud; VR+ vs. VR-), and (4) modality-specific report (auditory versus visual). Only congruent (ΔAV = 0°; ΔAV = A - V) and slightly disparate conditions (ΔAV = ±6°) were used in this study. In unisensory conditions, the experimental design manipulated the location of the auditory signal in auditory conditions and the locations of the visual signals as well as visual reliability in visual conditions.

Figure 2. Psychophysics results: psychometric functions, visual weights and audiovisual variances. In audiovisual (AV) conditions, psychometric functions were fitted to the fraction of 'right' location responses plotted as a function of the mean AV location. Data were fitted separately for audiovisual spatially congruent (ΔAV = 0°) and slightly conflicting conditions (ΔAV = ±6° with ΔAV = A - V). The empirical visual weight is computed from PSE locations of the audiovisual spatially conflicting psychometric functions (see equation 2). If the visual weight is greater than 0.5, the PSE for ΔAV = -6° is left of the PSE for ΔAV = 6°. If the visual

1273 weight is smaller than 0.5, the PSE for ΔAV = -6° is right of the PSE for ΔAV = 6°. If the visual
1274 weight is equal to 0.5, the PSEs for ΔAV = -6° and ΔAV = 6° are identical. **(A-D)** Psychometric
1275 functions for audiovisual spatially congruent and conflicting trials are plotted separately for the
1276 four conditions in our 2 (visual reliability: high, VR+ vs. low, VR-) x 2 (modality-specific
1277 report: auditory vs. visual) factorial design. **(E)** In unisensory conditions, psychometric functions
1278 were fitted to the fraction of right location responses plotted as a function of the signal location
1279 from unisensory auditory (A) and visual conditions of high (V, VR+) and low (V, VR-) visual
1280 reliability. **(F)** Visual weights (mean ± SEM across participants): MLE predicted and empirical
1281 weights for the four conditions in our 2 (visual reliability: high, VR+ vs. low, VR-) x 2
1282 (modality-specific report: auditory vs. visual) factorial design. To facilitate the comparison with
1283 the MLE predictions that do not depend on modality-specific report, the visual weights are also
1284 plotted after pooling the data across both report conditions and re-fitting the neurometric
1285 functions. **(G)** Standard deviations (σ, mean ± SEM across participants): Unisensory and
1286 audiovisual MLE predicted and empirical standard deviations of the perceived spatial locations
1287 for the same combination of conditions as in (F). For illustrational purposes standard deviations
1288 were normalized by the auditory standard deviation (original auditory standard deviation = 5.39
1289 ± 1.25 (mean ± SEM)).
1290
1291 **Figure 3. fMRI results in the intraparietal sulcus: neurometric functions, visual weights**
1292 **and audiovisual variances.** In intraparietal sulcus (IPS0-4), neurometric functions were fitted to
1293 the fraction of decoded 'right' location responses plotted as a function of the mean audiovisual
1294 (AV) location (see figure 2 legend for additional information). **(A-D)** Neurometric functions are
1295 plotted separately for the four conditions in our 2 (visual reliability: high, VR+ vs. low, VR-) x 2

(modality-specific report: auditory vs. visual) factorial design. **(E)** In unisensory conditions, psychometric functions were fitted to the fraction of right location responses plotted as a function of the signal location from unisensory auditory (A) and visual conditions of high (V, VR+) and low (V, VR-) visual reliability. **(F)** Visual weights (mean and 68% bootstrapped confidence interval): MLE predicted and empirical visual weights for 2 (visual reliability: high, VR+ vs. low, VR-) x 2 (modality-specific report: auditory vs. visual) AV conditions. To facilitate the comparison with the MLE predictions that do not depend on modality-specific report, the visual weights are also plotted after pooling the data across both report conditions and re-fitting the neurometric functions. **(G)** Standard deviations (σ, mean and 68% bootstrapped confidence interval): Unisensory and audiovisual MLE predicted and empirical standard deviations for the same combination of conditions as in **(F)**. For illustrational purposes standard deviations were normalized by the auditory standard deviation (original auditory standard deviation = 21.54). For extended analyses controlling motor responses and global interhemispheric activation differences in IPS0-4 see Fig. 3-1 and 3-2.

Figure 4. fMRI results in low-level visual and auditory regions: Visual weights and audiovisual variances. (A) Visual weights (mean and 68% bootstrapped confidence interval): MLE predicted and empirical visual weights for 2 (visual reliability: high, VR+ vs. low, VR-) x 2 (modality-specific report: auditory vs. visual) audiovisual conditions in low-level visual regions (V1-3). To facilitate the comparison with the MLE predictions that do not depend on modality-specific report, the visual weights are also plotted after pooling the data across both report conditions and re-fitting the neurometric functions. **(B)** Standard deviations (σ, mean and 68% bootstrapped confidence interval): Unisensory and audiovisual MLE predicted and

empirical standard deviations for the same combination of conditions as in (A). For illustrational purposes standard deviations were normalized by the auditory standard deviation (original auditory standard deviation = 61.68). **(C)** Visual weights (mean and 68% bootstrapped confidence interval): MLE predicted and empirical visual weights in low-level auditory regions (hA) as shown in (A). **(D)** Standard deviations (σ, mean and 68% bootstrapped confidence interval): Unisensory and audiovisual MLE predicted and empirical standard deviations of spatial representations in low-level auditory regions (hA) as shown in B; note that the upper confidence interval for the visual variance is truncated for illustrational purposes. For illustrational purposes, standard deviations were normalized by a combined visual standard deviation for low and high visual reliability (original visual standard deviation = 38.75, averaged across levels of visual reliability). For extended analyses controlling for global interhemispheric activation differences in low-level visual and auditory regions see Fig. 4-1.

Figure 5. Dynamic causal modelling. In the optimal model (i.e., the model with the highest exceedance probability), visual reliability modulated the connection from V1-3 to IPS0-4 and modality-specific report modulated the connection from PFC to IPS0-4. Values are across-subjects means (± SEM) indicating the strength of extrinsic, intrinsic and modulatory connections. The modulatory effects quantify how visual reliability and modality-specific report change the values of intrinsic connections.

Figure 3-1. fMRI results in the intraparietal sulcus when controlling for motor responses: neurometric functions, visual weights and audiovisual variances. In intraparietal sulcus (IPS0-4), neurometric functions were fitted to the fraction of decoded 'right' location responses

plotted as a function of the mean audiovisual (AV) location (see figure 2 legend for additional information). To control for motor planning in IPS0-4, activation patterns were obtained from a general linear model that modelled participants' trial-wise button responses as a nuisance variable **(A-D)** Neurometric functions are plotted separately for the four conditions in our 2 (visual reliability: high, VR+ vs. low, VR-) x 2 (modality-specific report: auditory vs. visual) factorial design. **(E)** In unisensory conditions, psychometric functions were fitted to the fraction of right location responses plotted as a function of the signal location from unisensory auditory (A) and visual conditions of high (V, VR+) and low (V, VR-) visual reliability. **(F)** Visual weights (mean and 68% bootstrapped confidence interval): MLE predicted and empirical visual weights for 2 (visual reliability: high, VR+ vs. low, VR-) x 2 (modality-specific report: auditory vs. visual) AV conditions. To facilitate the comparison with the MLE predictions that do not depend on modality-specific report, the visual weights are also plotted after pooling the data across both report conditions and re-fitting the neurometric functions. **(G)** Standard deviations (σ, mean and 68% bootstrapped confidence interval): Unisensory and audiovisual MLE predicted and empirical standard deviations for the same combination of conditions as in (F). For illustrational purposes standard deviations were normalized by the auditory standard deviation.

Figure 3-2. fMRI results in the intraparietal sulcus when controlling for global interhemispheric activation differences: neurometric functions, visual weights and audiovisual variances. In intraparietal sulcus (IPS0-4), neurometric functions were fitted to the fraction of decoded 'right' location responses plotted as a function of the mean audiovisual (AV) location (see figure 2 legend for additional information). To control for global interhemispheric activation differences, activation patterns were z normalized separately for the left and right

hemisphere within each condition prior to multivariate decoding. **(A-D)** Neurometric functions are plotted separately for the four conditions in our 2 (visual reliability: high, VR+ vs. low, VR-) x 2 (modality-specific report: auditory vs. visual) factorial design. **(E)** In unisensory conditions, psychometric functions were fitted to the fraction of right location responses plotted as a function of the signal location from unisensory auditory (A) and visual conditions of high (V, VR+) and low (V, VR-) visual reliability. **(F)** Visual weights (mean and 68% bootstrapped confidence interval): MLE predicted and empirical visual weights for 2 (visual reliability: high, VR+ vs. low, VR-) x 2 (modality-specific report: auditory vs. visual) AV conditions. To facilitate the comparison with the MLE predictions that do not depend on modality-specific report, the visual weights are also plotted after pooling the data across both report conditions and re-fitting the neurometric functions. **(G)** Standard deviations (σ, mean and 68% bootstrapped confidence interval): Unisensory and audiovisual MLE predicted and empirical standard deviations for the same combination of conditions as in (F). For illustrational purposes standard deviations were normalized by the auditory standard deviation.

Figure 4-1. fMRI results in low-level visual and auditory regions when controlling for interhemispheric activation differences: Visual weights and audiovisual variances. To control for interhemispheric activation differences, activation patterns were z normalized separately in the left and right hemisphere within each condition prior to multivariate pattern decoding. **(A)** Visual weights (mean and 68% bootstrapped confidence interval): MLE predicted and empirical visual weights for 2 (visual reliability: high, VR+ vs. low, VR-) x 2 (modality-

specific report: auditory vs. visual) audiovisual conditions in low-level visual regions (V1-3). To facilitate the comparison with the MLE predictions that do not depend on modality-specific report, the visual weights are also plotted after pooling the data across both report conditions and re-fitting the neurometric functions. **(B)** Standard deviations (σ, mean and 68% bootstrapped confidence interval): Unisensory and audiovisual MLE predicted and empirical standard deviations for the same combination of conditions as in (A). For illustrational purposes standard deviations were normalized by the auditory standard deviation. **(C)** Visual weights (mean and 68% bootstrapped confidence interval): MLE predicted and empirical visual weights in low-level auditory regions (hA) as shown in (A). **(D)** Standard deviations (σ, mean and 68% bootstrapped confidence interval): Unisensory and audiovisual MLE predicted and empirical standard deviations of spatial representations in low-level auditory regions (hA) as shown in B; note that the upper confidence interval for the visual variance is truncated for illustrational purposes. For illustrational purposes, standard deviations were normalized by a combined visual standard deviation for low and high visual reliability.

Table legends

Table 1

Note: Numbers denote t and p values for psychophysics parameters and p values for neurometric parameters. Psychophysics parameters were compared using two-tailed Wilcoxon signed rank tests on individual parameters (random-effects analysis, df = 4). Neurometric parameters from V1-V3, IPS0-4 and low-level auditory regions were compared using a two-tailed bootstrap test (5000 bootstraps) on parameters computed across the sample (fixed-effects analysis). All comparisons of neurometric parameters were Bonferroni corrected across the three regions of interest. A = auditory, V = visual, VR+/- = High / low visual reliability.

Table 2

Note: Numbers denote F and p values for psychophysics parameters and p values for neurometric parameters. Effects on psychophysics parameters were computed using a repeated measures ANOVA on rank-transformed weights and standard deviations (random-effects analysis, n = 5, df1 = 1, df2 = 4). Effects on neurometric parameters were computed using two-tailed bootstrap test (5000 bootstraps) on parameters computed across the sample (fixed-effects analysis). The analyses for neurometric weights and standard deviations were Bonferroni corrected across the three regions of interest.

Table 3

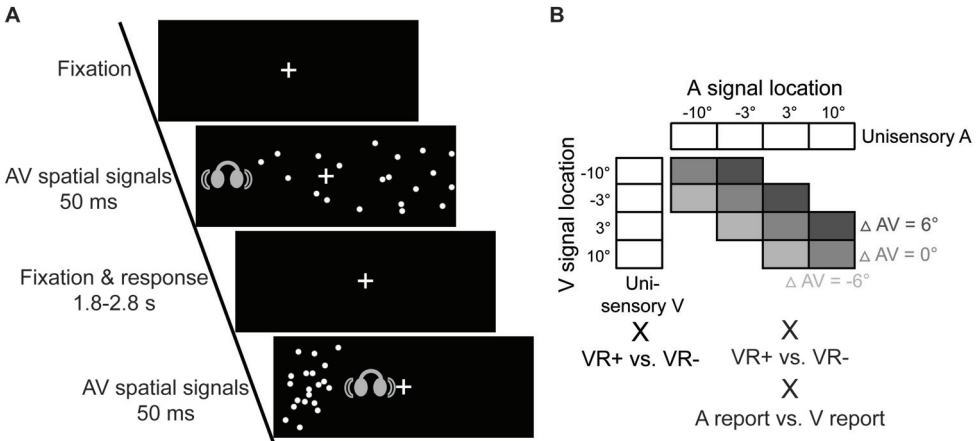
Note: Model I: In the null model, neither PSEs nor slopes depended on visual reliability or modality-specific report. II: In the MLE model, audiovisual PSEs and slopes were predicted based on unisensory variances as described in equation (1) and (3). III: In the reliability-weighted integration model, PSEs and slopes depended on visual reliability unconstrained by MLE predictions. IV: In the full model, PSEs and

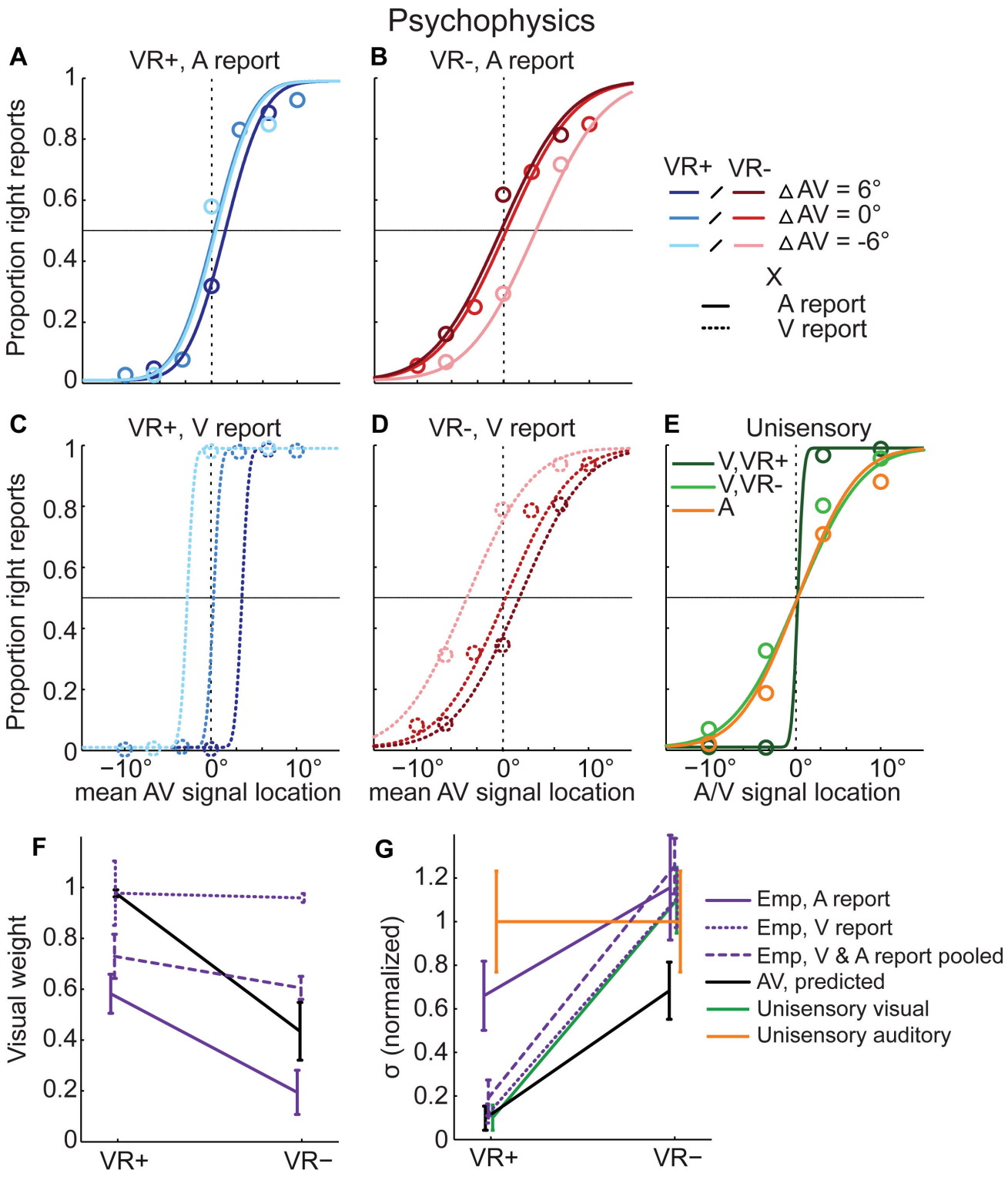
slopes depended on visual reliability unconstrained by MLE predictions and modality-specific report (MR). R^2, coefficient of determination, corrected for the binary response option (Nagelkerke, 1991). Relative BIC = Bayesian Information Criterion (i.e., an approximation to the model evidence) at the group level, i.e., subject-specific BICs summed over all subjects (BIC = LL − 0.5 M ln(N), LL = log likelihood, M = number of parameters, N = number of data points) of a model relative to the null model (n.b. a greater relative BIC indicates that a model provides a better explanation of our data). Exp. Post. p = Expected posterior p. = probability that a given model generated the data for a randomly selected subject. Exceed. p. = Exceedance p. = probability that a given model is more likely than any other model. Prot. Exceed. p. = exceedance p. controlled for the fact that the observed variability in model evidences occurred by chance, i.e., it quantifies the probability that one model is more frequent than any others beyond chance.

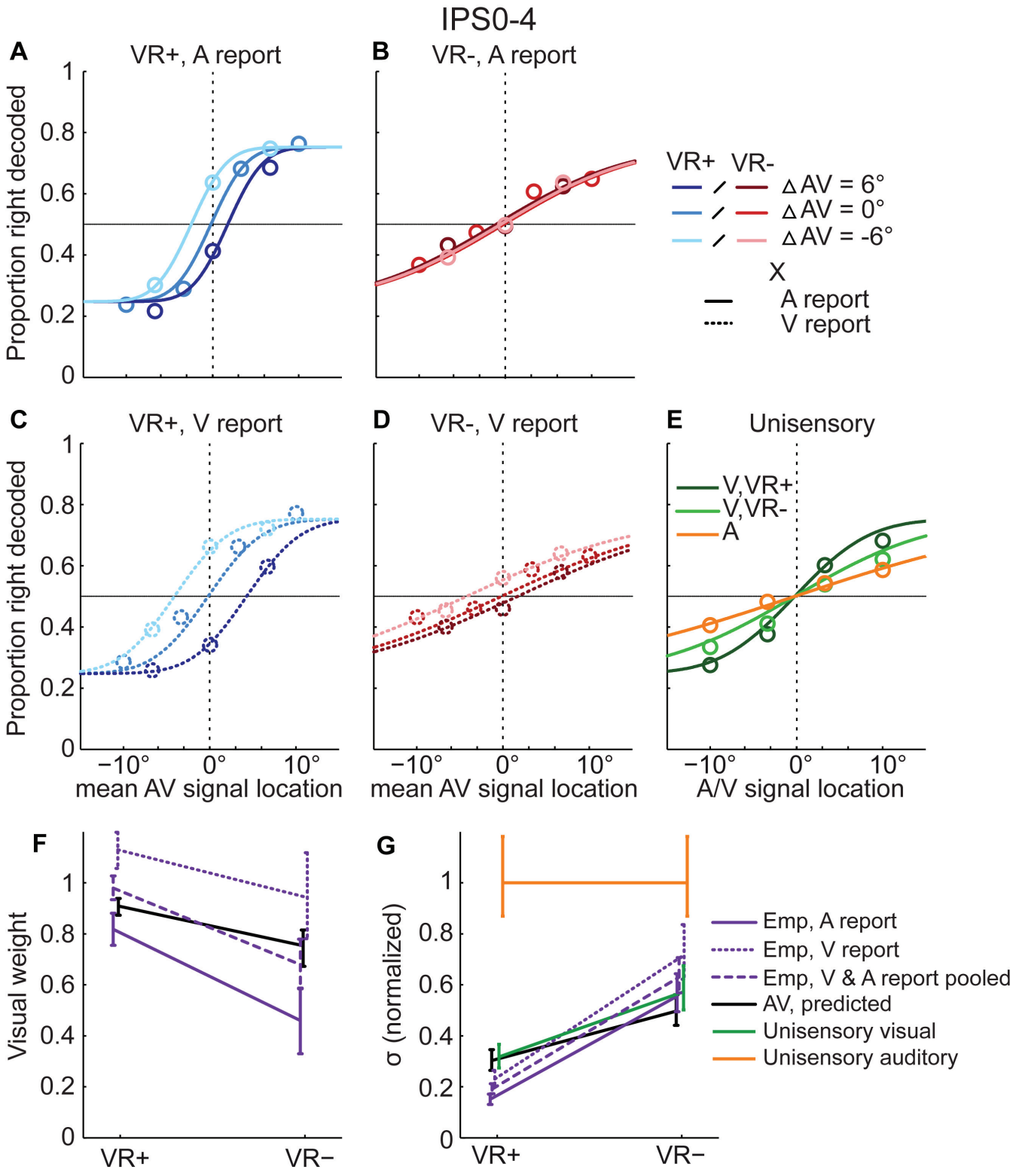
Table 4

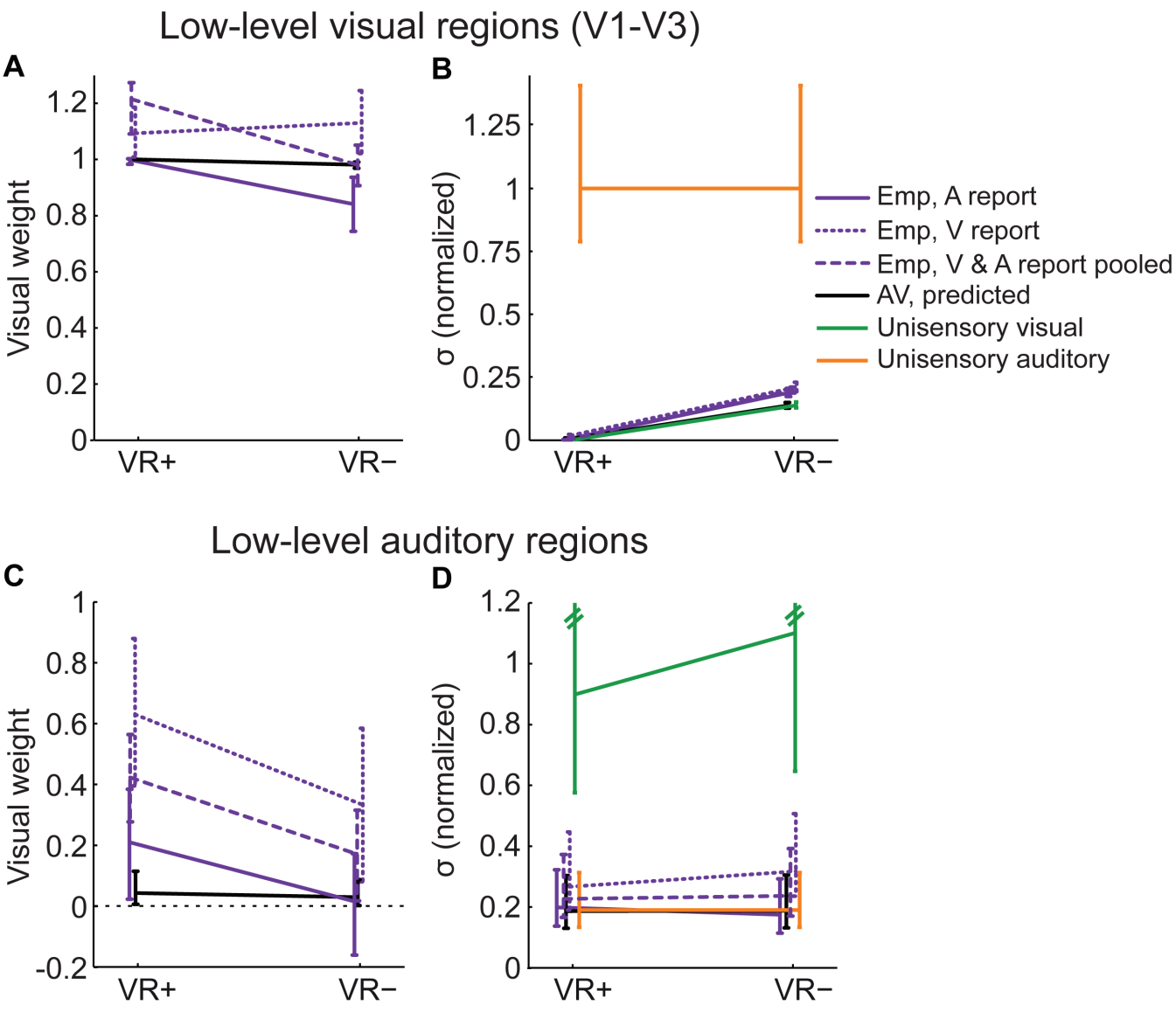
Note: FFX = fixed-effects analysis. RFX = random-effects analysis. p. = probability. Model evidence = Free energy (relative to full model) summed over participants (i.e., larger is better). Exp. Post. p = Expected posterior p. = probability that a given model generated the data for a randomly selected subject. Exceed. p. = Exceedance p. = probability that a given model is more likely than any other model. Prot. exceed. p. = exceedance p. controlled for the fact that the observed variability in model evidences occurred by chance, i.e., it quantifies the probability that one model is more frequent than any others beyond chance.

1451









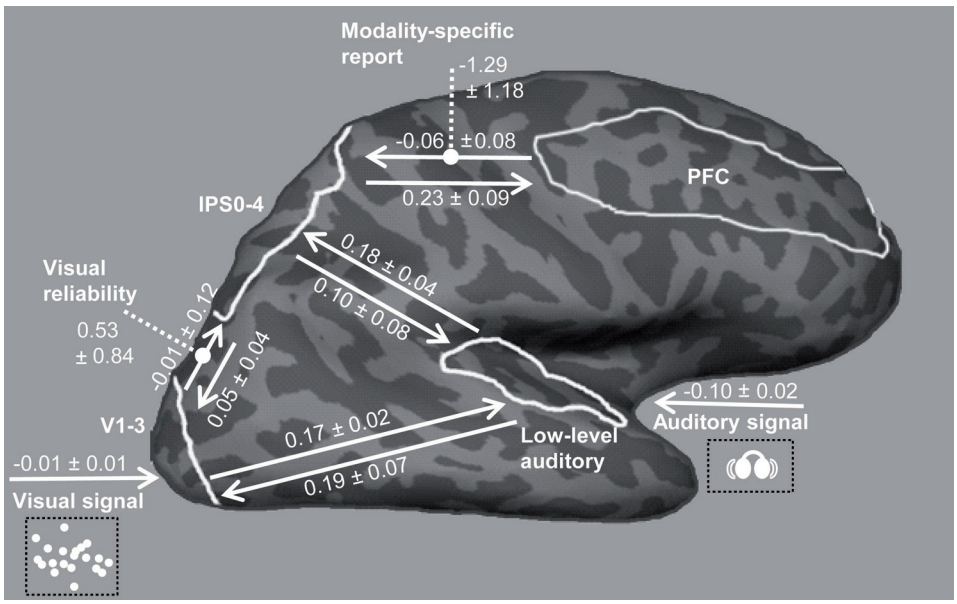


Table 1. Statistical comparison of empirical weights ($w_{V,emp}$) and standard deviations ($\sigma_{AV,emp}$) obtained from the psychometric (behavior) and neurometric (fMRI) functions pertaining to the audiovisual conditions of high and low visual reliability with the MLE predictions ($\sigma_{AV,pred}$, $w_{V,pred}$) and unisensory standard deviations (σ_{uniV}, σ_{uniA}).

		VR+	VR-
		$w_{V,emp} - w_{V,pred}$	
Psychophysics	p	0.0629	0.129
V1-V3	p	0.230	1
IPS0-4	p	0.631	1
Low-level auditory	p	0.064	1
		$\sigma_{AV,emp} - \sigma_{AV,pred}$	
Psychophysics	p	0.188	0.0629
V1-V3	p	0.241	**0.001**
IPS0-4	p	**0.020**	0.275
Low-level auditory	p	1	1
		$\sigma_{AV,emp} - \sigma_{uniV}$	
Psychophysics	p	0.188	0.438
V1-V3	p	0.241	**0.001**
IPS0-4	p	**0.022**	1

Low-level auditory	p	0.883	0.963
		$\sigma_{AV,emp} - \sigma_{uniA}$	
Psychophysics	p	0.063	0.313
V1-V3	p	0.104	0.169
IPS0-4	p	**0.002**	0.086
Low-level auditory	p	1	1

Table 2. Effects of visual reliability (VR), modality-specific report (MR) and their interaction (VRxMR) on empirical weights ($w_{V,emp}$) and standard deviations ($\sigma_{AV,emp}$) obtained from the psychometric (behavior) and neurometric (fMRI) functions.

		$w_{V,emp}$			$\sigma_{AV,emp}$		
		VR	MR	VRxMA	VR	MR	VRxMR
Psychophysics	[F, p]	5.149,0.086	16.308, **0.016**	8.605, **0.043**	19.129, **0.012**	2.172, 0.215	18.892, **0.012**
V1-V3	p	0.346	0.131	0.957	**< 0.001**	0.142	1
IPS0-4	p	**0.022**	**0.001**	1	**<0.001**	0.051	1
Low-level auditory	p	0.693	0.217	1	1	0.419	1

Table 3. Results of the Bayesian model comparison between five competing models of the psychometric data.

Model:	I: Null model	II: MLE model	III: Reliability-weighting	IV: Full model
# parameters	8	5	11	17
R^2 (mean)	0.656	0.658	0.687	0.722
Relative BIC (sum)	0	59.662	1501.077	3202.034
Exp. post. p.	0.111	0.111	0.111	0.667
Exceed. p.	0.014	0.014	0.014	0.957
Prot. exceed. p.	0.026	0.026	0.026	0.921

Table 4. Results of the model comparison of the 2 x 2 Dynamical Causal Models in which visual reliability (VR) modulated the connection from V1-3 to IPS0-4 and modality-specific report (MR) modulated the connection from PFC to IPS0-4.

	Modulation VR & MR	Modulation VR	Modulation MR	No Modulation
Model evidence (FFX)	0	-52.947	-54.033	-90.45
Posterior p. (FFX)	1	0	0	0
Exp. posterior p. (RFX)	0.587	0.136	0.139	0.139
Exceed. p. (RFX)	0.902	0.032	0.033	0.033
Prot. exceed. p. (RFX)	0.699	0.1	0.101	0.101