
Research Article: New Research | Cognition and Behavior

Shared physiological correlates of multisensory and expectation-based facilitation

<https://doi.org/10.1523/ENEURO.0435-19.2019>

Cite as: eNeuro 2020; 10.1523/ENEURO.0435-19.2019

Received: 21 October 2019

Revised: 16 December 2019

Accepted: 23 December 2019

This Early Release article has been peer-reviewed and accepted, but has not been through the composition and copyediting processes. The final version may differ slightly in style or formatting and will contain links to any extended data.

Alerts: Sign up at www.eneuro.org/alerts to receive customized email alerts when the fully formatted version of this article is published.

Copyright © 2020 Kayser and Kayser

This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International license, which permits unrestricted use, distribution and reproduction in any medium provided that the original work is properly attributed.

1 **1. Manuscript Title (50 word maximum)**

2 Shared physiological correlates of multisensory and expectation-based facilitation

3 **2. Abbreviated Title (50 character maximum)**

4 Physiological correlates of cross- and intra-modal combination

5 **3. List all Author Names and Affiliations in order as they would appear in the published article**

6 Stephanie J. Kayser^{1,2}, Christoph Kayser^{1,2}

7 ¹Department for Cognitive Neuroscience, Faculty of Biology, ²Cognitive Interaction Technology,
8 Bielefeld University, 33615 Bielefeld, Germany

9 **4. Author Contributions:**

10 SJK, CK Designed research, SJK Performed research, SJK, CK Analyzed data, SJK, CK Wrote the paper.

11 **5. Correspondence should be addressed to (include email address)**

12 Stephanie Kayser (stephanie.kayser@uni-bielefeld.de)

13 **6. Number of Figures** **3**

14 **7. Number of Tables** **0**

15 **8. Number of Multimedia** **0**

16 **9. Number of words for Abstract** **218**

17 **10. Number of words for Significance Statement** **101**

18 **11. Number of words for Introduction** **555**

19 **12. Number of words for Discussion** **1809**

20 **13. Acknowledgements**

21 We are grateful to Caren Strote for her help with data collection.

22 **14. Conflict of Interest** Authors report no conflict of interest

23 **15. Funding source**

24 This work was supported by the European Research Council (ERC-2014-CoG; grant n°
25 646657).

1 *Shared physiological correlates of multisensory and expectation-based facilitation*

2

3

4

5 *Abstract*

6 Perceptual performance in a visual task can be enhanced by simultaneous multisensory information,
7 but can also be enhanced by a symbolic or amodal cue inducing a specific expectation. That similar
8 benefits can arise from multisensory information and within-modality expectation raises the
9 question of whether the underlying neurophysiological processes are the same or distinct. We
10 investigated this by comparing the influence of three types of auxiliary probabilistic cues on visual
11 motion discrimination in humans: i) acoustic motion, ii) a pre-motion visual symbolic cue, iii) a post-
12 motion symbolic cue. Using multivariate analysis of the EEG data we show that both the multisensory
13 and preceding visual symbolic cue enhance the encoding of visual motion direction as reflected by
14 cerebral activity arising from occipital regions around 200-400ms post-stimulus onset. This suggests a
15 common or overlapping physiological correlate of crossmodal and intramodal auxiliary information,
16 pointing to neural mechanism susceptible to both multisensory and more abstract probabilistic cues.
17 We also asked how pre-stimulus activity shapes the cue-stimulus combination and found a
18 differential influence on crossmodal and intramodal combination: while alpha power modulated the
19 relative weight of visual motion and the acoustic cue, it did not modulate the behavioural influence
20 of a visual symbolic cue, pointing to differences in how pre-stimulus activity shapes the combination
21 of multisensory and abstract cues with task-relevant information

22

23 *Significance Statement*

24 Perception can be enhanced by the combination of multisensory information and by the exploitation
25 of amodal or symbolic cues presented within the same modality as the task-relevant information. We
26 here asked whether the physiological correlates reflecting the behavioural benefits induced by each
27 type of cue are similar or not. Using multivariate analysis of EEG data we show that the perceptual
28 enhancement induced by an acoustic cue and a visual-symbolic cue for the discrimination of visual
29 motion arise from the same physiological source. This suggests that the impact of multisensory
30 information and more abstract sensory expectations on perception arise from shared mechanisms.

31

32 *Introduction*

33 Perception often benefits from additional information besides that made available by the primary
34 and task-relevant features. One example is multisensory integration: often our performance in
35 detecting or discriminating stimuli is enhanced when the same information is presented in more than
36 one sensory modality (Stein and Meredith, 1993; Kayser and Logothetis, 2007; Stein and Stanford,
37 2008). Here, perceptual benefits can arise both from the combination of partially redundant
38 information, as in Bayesian fusion (Ernst and Bulthoff, 2004; Angelaki et al., 2009), or the auxiliary
39 influence of one apparently irrelevant stimulus onto task performance in another modality (Jaekl and
40 Harris, 2009; Caclin et al., 2011; Kim et al., 2012; Gleiss and Kayser, 2014b). As suggested by recent
41 work, the neural mechanisms underlying multisensory benefits comprise a cascade of neural
42 processes that commence with early multisensory influences in low-level sensory regions, continue
43 with the automatic merging of multisensory information in parietal cortex, and culminate in the task-
44 and context-dependent arbitration of different perceptual strategies of exploiting multisensory
45 information in the frontal lobe (Rohe and Noppeney, 2015; Aller and Noppeney, 2019; Cao et al.,
46 2019; Rohe et al., 2019).

47 From a behavioural perspective, similar benefits for perception also emerge from other types of
48 auxiliary information, such as prior knowledge or expectations elicited by cues presented either
49 within the same modality or in an amodal fashion (Summerfield and Egner, 2009; de Lange et al.,
50 2013; Summerfield and de Lange, 2014). For example, priming participants to expect a specific
51 sensory attribute boosts perceptual performance and reduces reaction times in discrimination tasks.
52 One proposed mechanism by which expectation can improve perception is by an enhancement of
53 the encoding of sensory information in low level sensory cortices, in addition to post-sensory
54 decision level effects (Kok et al., 2012; Jiang et al., 2013; Cheadle et al., 2015; Bang and Rahnev,
55 2017; de Lange et al., 2018; Rungratsameetaweemana et al., 2018). The parallels between
56 traditional cueing paradigms and studies investigating multisensory paradigms, in particular those
57 focusing on auxiliary multisensory influences, raise the question in how far the neural mechanisms
58 mediating the behavioural benefits arising from multisensory combination are the same as those
59 mediating within-modality priming effects.

60 We here compare the physiological correlates of multisensory combination and within-modality
61 cueing using EEG while human participants performed a visual motion discrimination task.
62 Specifically, we capitalize on previous work that has established the cerebral correlates of
63 multisensory combination in a task where acoustic motion enhances the discrimination of horizontal
64 visual motion (Kayser et al., 2017; Kayser and Kayser, 2018): the physiological correlates of the
65 perceptual benefit arising from congruent audio-visual stimuli in this task were shown to comprise
66 the enhancement of visual motion encoding in (low-level) visual cortices at latencies of around
67 300ms post-stimulus onset. We here exploit this validated paradigm to ask whether the perceptual
68 enhancement induced by a probabilistic within-modality visual cue arises from the same
69 physiological processes that mediate the multisensory benefit.

70 Furthermore, we investigated the role of pre-stimulus activity in shaping the influence of task-
71 relevant visual motion and the visual cue on behaviour. Previous work has shown that multisensory
72 perception is shaped by the state of oscillatory brain activity prior to a stimulus (Keil et al., 2012;
73 Lange et al., 2013; Gleiss and Kayser, 2014b; Kaiser et al., 2019). For example, the overall tendency to
74 combine visual and acoustic cues during temporal rate judgements was influenced by the level of
75 pre-stimulus alpha power (Rohe et al., 2019). Given this link of pre-stimulus activity and multisensory
76 combination we asked whether between-modality multisensory integration and within-modality
77 combination of a cue and stimulus are shaped in a similar manner by pre-stimulus activity.

78 *Methods*

79 We report data obtained from a planned sample of 20 healthy adult participants (9 males; age $23.4 \pm$
80 3.3 years). These participated following written informed consent and briefing about the purpose of
81 the study. All had self-reported normal hearing and tested normal vision, declared no previous
82 history of neurological disorders, and were right-handed (Oldfield, 1971). The study was conducted in
83 accordance with the Declaration of Helsinki and was approved by the ethics committee of Bielefeld
84 University. The sample size was decided a priori based on general recommendations for behavioural
85 studies (Simmons et al., 2011). Data collection proceeded until the sample of 20 acceptable
86 participants was obtained, whereby some ($n=7$) participants dropped out because they either did not
87 pass the sight test (one) or during an initial screening session exhibited a perceptual threshold for
88 visual motion discrimination that was higher than a *a priori* set criterion of 25% motion coherence (six
89 participants; see below).

90 *Experimental design and stimuli*

91 The task required participants to discriminate the direction (left- or rightwards) of random dot visual
92 motion (**Fig. 1A**). The motion stimulus was presented following the onset of a fixation dot (800 to
93 1200ms uniform delay), lasted 800ms and was presented on a 24" high-performance LCD monitor
94 (Asus PG279Q) at 1920×1080 pixel resolution and a refresh rate of 120 Hz. The random dot pattern
95 consisted of 1300 white, limited-lifetime dots that were presented on a neutral grey screen (16
96 cd/m^2 background luminance) and covered 10° of visual angle (with the centre 0.8° devoid of dots).
97 Individual dots were 0.1° in diameter and moved at $4^\circ/\text{s}$. Each dot had a life-time of 9 frames (at
98 120Hz). A small percentage of dots moved coherently in the same direction (left or right). In an initial
99 screening session, we determined participant's psychometric thresholds (around 71% correct
100 performance) for discriminating motion directions using three interleaved 2-down 1-up staircases. In
101 the actual experiment the motion coherence was additionally manipulated over the 800ms (96
102 frames) of stimulus presentation: every period of 12 consecutive non-overlapping frames was
103 characterized by a different and independent coherence level (drawn from a normal distribution
104 centred on the participant's specific threshold with a SD of 10%, limited to the range 0 to 100%). This
105 was done to allow the quantification of perceptual weights, that is, how much each epoch of 12
106 frames contributed to the participant's responses (see below). Across participants the coherence
107 thresholds were $18.2 \pm 1.9\%$ (mean \pm s.e.m). From 400 to 500ms following stimulus offset a response
108 cue was presented prompting the participants to respond. Given that some of the analyses reported
109 here essentially reproduce a previous study (Kayser et al., 2017), it is important to note that the
110 stimulus used here differed from the previous study in that the motion coherence here changed
111 randomly over the course of a single trial, while it was constant in that previous study. Also, the
112 previous study used a speeded reaction time task, while we here used a fixed stimulus duration and a
113 task emphasizing accuracy (see below).

114 The experiment comprised three conditions (**Fig. 1A**). In the first condition, visual motion was
115 accompanied by a dynamic acoustic stimulus mimicking motion in either the same ('congruent', 66%
116 of trials) or the opposite direction ('incongruent', 33% of trials) as the visual motion. Sounds were
117 composed from white noise (at 44.1 kHz sampling rate) with an amplitude that was linearly
118 modulated from 0 to the maximal level in opposite directions on left and right speakers over the
119 stimulus period, inducing a percept of continuous acoustic motion (Meyer and Wuerger, 2001;
120 Moore, 2003). Sounds were presented with a peak amplitude of 65 dB(A) SPL r.m.s. level using
121 speakers placed to the left and right of the monitor; on- and offsets were cosine ramped (8ms). In
122 the second condition, visual motion was preceded by a symbolic visual cue that indicated the likely
123 (66% correct) direction of the subsequent visual motion stimulus. The cue was presented during the
124 entire fixation period. In the third condition, the visual motion stimulus was followed by a symbolic

125 visual cue that indicated the likely (66% correct) direction of the preceding motion stimulus. The cue
126 was presented starting 400 to 500ms (uniform) following stimulus offset for a period of 500 to
127 600ms. Participants completed 936 trials in total, 156 per motion direction and cue condition.
128 Individual trials were separated by inter-trial intervals of 1200 to 1500ms (uniform) and were
129 grouped into four experimental blocks, with each block comprising all three conditions. Within each
130 block groups of 78 subsequent trials featured the same condition ('mini-blocks'), with the three
131 conditions (mini-blocks of 78 trials) appearing in pseudo-random order within each experimental
132 block. Participants were instructed 'to discriminate the direction of visual motion and to respond as
133 quickly and accurately as possible following the response cue, while making use of the auxiliary
134 information' by pressing a left or right arrow key on a keyboard (Das Keyboard, TX, USA) using the
135 same hand for both keys.

136 During the experiment monocular (right) eye movements were monitored using an EyeLink 1000
137 System (SR Research, Canada) at 250Hz. For six participants the system did not reliably track the eye
138 and the eye tracking data could not be analysed.

139

140 *Analysis of behavioural and eye tracking data*

141 From the behavioural responses we calculated the percentage of correct responses and reaction
142 times aligned to the response cue. For an analysis based on signal detection theory we computed hit
143 and false-alarm rates relative to left-wards motion (Green and Sweets, 1966). To compute the
144 response bias, we analysed trials with the cue pointing to the left and right separately, and combined
145 the bias measures (c), after converting both sides to have the same sign (Bang and Rahnev, 2017).
146 We quantified the perceptual use of the moment by moment visual motion evidence using reverse
147 correlation (Marmarelis, 1978; Eckstein and Ahumada, 2002). Specifically, we computed perceptual
148 response templates (weights) relative to the two response options, and normalized these within
149 participants to z-scores using a shuffling procedure (Neri and Heeger, 2002; Chauvin et al., 2005).
150 These templates were computed based on the amount of motion evidence available in the random
151 dot motion at each moment in time. This level of motion evidence was extracted post-hoc from the
152 single trial motion stimulus based on algorithms previously used to detect horizontal motion using
153 parameters suitable for the human visual system (Kiani et al., 2008; Urai et al., 2017). The resulting
154 perceptual weights indicate how strongly (relative to chance) the visual motion influenced the
155 participant's response. We also modelled the participants single trial choice using logistic regression,
156 entering the visual motion stimulus, the cue, their interaction with pre-stimulus power as predictors.
157 Similarly, we modelled performance (response accuracy) based on pre-stimulus power and its
158 interaction with motion-cue congruency.

159 From the eye tracking data we extracted fixation events detected by the eye tracking system (using
160 the EyeLink 1000 'cognitive' setting) and computed the standard deviation of all fixated positions
161 along horizontal and vertical dimensions during stimulus presentation. In addition, we extracted
162 saccadic eye movements, as detected by the EyeLink 100 system (velocity threshold of 30°/s;
163 acceleration threshold 8000°/s). We then counted the number of saccades exceeding an amplitude
164 of 0.8° executed during the stimulus interval, hence excluding smaller micro-saccades (Rolfs, 2009).

165

166 *Setup and EEG recordings*

167 The experiment took place in a dark and electrically shielded room (E:Box, Desone, Germany).
168 Stimulus presentation was controlled from Matlab (Mathworks) using routines from the

169 Psychophysics toolbox (Brainard, 1997). Sound levels were calibrated using a sound level meter
170 (Model 2250; Bruel&Kjær, Denmark). EEG signals were continuously recorded using an active 128
171 channel BioSemi system (BioSemi, The Netherlands) using Ag-AgCl electrodes mounted on an elastic
172 cap. Four additional electrodes were placed near the outer canthi and below the eyes to obtain the
173 electro-oculogram (EOG). Electrode offsets were kept below 25mV. Data were acquired at a
174 sampling rate of 1000Hz using a low pass filter of 208Hz.

175 *EEG data analysis*

176 Data analysis was carried out offline with MATLAB (R2017a; The MathWorks Inc., Natick, MA), using
177 the FieldTrip toolbox (version fieldtrip-20171001) (Oostenveld et al., 2011) and custom written
178 routines. The data from the different blocks were pre-processed separately by band-pass filtering
179 (0.6-70Hz), re-sampling to 150Hz and de-noising using ICA. We removed ICA components that likely
180 reflect eye movement artefacts, localized muscle activity or poor electrode contacts. These were
181 identified following definitions provided in the literature (O'Beirne and Patuzzi, 1999; Hipp and
182 Siegel, 2013). On average we rejected 16.4 ± 1.6 components (mean \pm s.e.m.). Periods contaminated
183 by eye blinks or movements were identified using horizontal, vertical and radial EOG signals (Keren et
184 al., 2010; Quax et al., 2019). We rejected trials based on a threshold of 3 standard deviations above
185 mean of the high-pass filtered EOGs, or during which the peak amplitude on any electrode exceeded
186 $\pm 175\mu\text{V}$. On average, we retained $93.7 \pm 2.1\%$ of trials. Also, for three participants one channel was
187 deemed bad and interpolated with its neighbours, for one participant two channels were
188 interpolated. For subsequent analysis the EEG signals were referenced to the common average
189 reference. The analysis of eye tracking data showed that the retained trials contained very few
190 saccadic eye movements, suggesting that the EEG data were very little confounded by eye
191 movements (see Results).

192 To extract EEG signatures of brain activity relevant for the visual motion discriminant task, we used a
193 multivariate regularized linear discriminant analysis (LDA). We implemented an LDA to identify a
194 projection of the EEG data in sliding windows of 92ms duration (window step of 46ms) that
195 maximally discriminated between the two directions of motion (across all three cue conditions).
196 Within each window, the EEG activity of each electrode was averaged over time and the multivariate
197 classifier was trained using the activity of all EEG electrodes. The data are presented such that the
198 classification performance, or time points noted, correspond to the first time point within the
199 classification window. Each LDA projection obtained at one time point, i.e. from one 92ms data
200 window, was defined by a projection vector, w , which describes a one dimensional weighted
201 combination of the EEG data (Parra et al., 2005). The regularization parameter was taken from
202 previous work and set to 0.1. The classification performance was quantified using the area under the
203 receiver operator characteristic (AUC) based on 6-fold cross validation. Given potentially unequal
204 trial numbers for each condition, we repeated the discriminant analysis 100-times using a random
205 subset of 80% of the available trials for each condition, averaging the resulting AUC and projection
206 vectors. The scalp topographies for each discriminant component, i.e. time point, were derived using
207 the corresponding forward model defined as the normalized correlation between the discriminant
208 component and time-averaged EEG activity (Parra et al., 2005).

209 The discriminant output provides a measure of the single trial evidence contained in the EEG signal
210 about two conditions of interest. It can serve as a sensitive representation of the cerebral encoding
211 of the task relevant sensory information (Parra et al., 2005; Philiastides et al., 2014; Kayser et al.,
212 2016; Grootswagers et al., 2018; Cichy et al., 2019). Following previous work, we exploited this
213 projection to ask which specific LDA component (i.e. when in time) is affected by each of the cues. To
214 this end we obtained single trial projections of the discriminant activity by applying the weights

215 extracted at specific time points of interest (P1, P2 and P3 in **Fig. 2A**) to all trials and time points. We
216 then compared the amount of sensory evidence in the discriminant components by comparing their
217 magnitude (ignoring the difference in sign arising from the two motion directions) between
218 congruent and incongruent trials. This was done using six-fold cross-validation, computing LDA
219 weights based on one subset of trials and quantifying the congruency effect on the remaining trials.
220 Similarly, we used cross-validation when entering the LDA evidence as a predictor into a linear model
221 of choice (or accuracy), again computing the LDA weights and the neuro-behavioural regression
222 model using distinct sets of trials.

223 Time frequency representations of the pre-stimulus activity were obtained using wavelet analysis in
224 FieldTrip. Frequencies ranged from 2Hz to 50Hz, in steps of 1Hz below 16Hz and steps of 2Hz above.
225 The wavelet width scaled with frequency, from 3 cycles at 2Hz to 9 cycles above 35 Hz. To ensure
226 that power estimates were not contaminated with stimulus evoked responses, we zero-windowed
227 the post-stimulus period. Power estimates were z-scored across trials. To reduce the statistical
228 complexity of testing for significant effects across time, frequency and electrodes we applied this
229 time-frequency analysis to the one-dimensional single-trial LDA projections of interest, hence
230 focusing on time and frequency dimensions only.

231

232 *EEG Source analysis*

233 A confirmatory source analysis was implemented by first obtaining single trial source signals using a
234 linear constrained minimum variance beamformer in Fieldtrip (6% normalization, using the
235 covariance matrix obtained from -0.6 to -0.1s prior to response). A standardized head model based
236 on the average template brain of the Montreal Neurological Institute was used as single participant
237 MRI data were not available. Lead-fields were computed using a 3D grid with 6mm spacing. The
238 activity at individual grid points was correlated with the linear discriminant signal over trials at the
239 single participant level, analogous to obtaining the forward scalp distribution via the correlation of
240 sensor and discriminant activity (Parra et al., 2005; Haufe et al., 2014). Correlation volumes were z-
241 transformed and averaged across participants.

242 *Statistical procedures and effect sizes*

243 The analysis of behavioural data was based on a repeated-measures ANOVA and post-hoc paired t-
244 tests. Effect sizes are reported as Cohen's d for paired t-tests and partial eta-squared (η_p^2) for the
245 ANOVA (Kline, 2004). Measures of signal detection theory were derived relative to left-ward motion
246 as 'to be detected' stimulus. Statistical testing for perceptual weights was based on a percentile
247 bootstrap distribution obtained by randomly resampling (2000 times) participants with replacement.
248 Here the effect sizes are reported as Cohens d for the underlying paired t-test. Significance testing of
249 the single trial discriminant performance (AUC), of congruency effects in discriminant activity, and of
250 the influence of oscillatory power in the regression model of choice were based on group-level
251 cluster-based permutation procedures computed by randomly permuting effect signs (Nichols and
252 Holmes, 2002; Maris and Oostenveld, 2007). The detailed parameters were: 2000 permutations;
253 clustering bins with significant first-level tests (uncorrected at $p < 0.05$; or ROC above the 95%
254 percentile of the distribution across bins); minimal cluster size of at least 3 neighbours; computing
255 the cluster-mass within each cluster; performing a two-sided test at $p < 0.05$ on the clustered data. For
256 these tests we report the cluster-value (t_{sum}) of the significant cluster as effect size in addition to
257 Cohen's d derived from the univariate t-value at the peak location.

258

259

260 *Results*

261

262 *Cueing improves discrimination performance*

263 Participants (n=20) performed a visual motion discrimination task (left vs. right-wards random dot
264 motion, 800ms fixed duration) around their respective individual thresholds, as determined in an
265 initial screening session. The different experimental conditions offered participants three types of
266 additional cues (**Fig. 1A**): i) during the *acoustic cue condition* visual motion was accompanied by
267 acoustic spatial motion that moved, in 66% of trials, in the same direction as the visual motion
268 stimulus. For this task previous work has established the electrophysiological correlates of audio-
269 visual integration, by first extracting visual-motion sensitive EEG components using single trial linear
270 discriminant analysis, and then determining when this component is enhanced by congruent (vs.
271 incongruent) audio-visual information (Kayser et al., 2017; Kayser and Kayser, 2018). During ii) a
272 *visual-pre cue condition* we provided participants with a prior visual symbolic cue that indicated the
273 likely (66% chance of being correct) direction of the subsequent motion stimulus. This type of cue is
274 known to facilitate perceptual performance, possibly by enhancing the encoding of the task-relevant
275 subsequent information (Summerfield and de Lange, 2014). And last, during iii) a *visual-post cue*
276 *condition*, we provided participants with a visual symbolic cue that was presented subsequent to the
277 random dot pattern and indicated the likely (66% correct) direction of the preceding motion
278 stimulus. While such a post-stimulus cue can also facilitate performance (Bang and Rahnev, 2017), it
279 can't do so by influencing the encoding of visual motion information *during* the stimulus, and here
280 served as a control condition for the EEG analysis.

281 As expected, response accuracy was significantly affected by the congruency between cue and
282 motion stimulus. A repeated measures ANOVA revealed no overall effect of cue ($F(2,119)=1.17$,
283 $\eta_p^2=0.008$, $p=0.32$; **Fig. 1B**), but a significant influence of congruency ($F(1,119)=29.5$, $\eta_p^2=0.35$, $p<10^{-4}$)
284 and a cue-congruency interaction ($F(2,119)=11.2$, $\eta_p^2=0.09$, $p<10^{-3}$). Post-hoc tests revealed that for
285 each cue performance was more accurate when cue and visual motion were congruent (paired t-
286 tests, acoustic cue: $t(19) = 6.3$, Cohen's $d = 1.4$, $p<10^{-5}$; pre-cue: $t(19)=2.3$, $d = 0.5$, $p=0.03$; post-cue:
287 $t(19) = 5.1$, $d = 1.1$, $p=0.001$). Also, the congruency benefit was significantly stronger for the acoustic
288 cue compared to both symbolic cues ($t(19)=4.4$ vs. visual-pre and $t(19)=4.9$ vs. visual-post, both $d >$
289 0.9 , $p<10^{-3}$), while the two symbolic cues did not differ ($t(19)=0.7$, $d = 0.15$, $p=0.49$).

290 When quantified using signal detection theory the data revealed that sensory precision (d -prime) was
291 significantly higher for congruent compared to incongruent trials for each cue (acoustic cue: d
292 prime= 1.9 ± 0.11 vs 0.54 ± 0.14 , $t(19)=6.7$, $p<10^{-3}$; pre-cue: 1.44 ± 0.12 vs. 0.92 ± 0.12 , $t(19)=2.6$, $p=0.016$;
293 post-cue: 1.60 ± 0.08 vs. 1.01 ± 0.12 , $t(19)=4.56$, $p=0.0002$). To quantify the influence of the cue on
294 response bias, we extracted the magnitude of this bias (Bang and Rahnev, 2017) (acoustic cue:
295 0.34 ± 0.05 ; pre-cue: 0.13 ± 0.05 ; post-cue: 0.14 ± 0.03). This bias was significantly stronger for the
296 acoustic cue compared to the symbolic cues ($t(19)=4.6$ and 4.3 , Cohen's $d>1.0$, $p<10^{-3}$) and did not
297 differ between the latter two ($t(19)=0.3$, $d = 0.06$, $p=0.75$). Overall these results show that
298 participant's performance was affected by all three cue types, although the acoustic stimulus had the
299 strongest influence both in terms of decision bias and overall accuracy benefit.

300 Reaction times were not affected by the cue-stimulus congruency or cue type. An ANOVA revealed
301 no significant effects (all $p>0.05$, maximal $\eta_p^2=0.01$); as expected, post-hoc t-tests showed only small
302 and insignificant effects (acoustic cue: $t(19)=-1.1$, Cohen's $d=0.25$, $p=0.27$, pre-cue: $t(19)=-1.5$,
303 $d=0.34$, $p=0.13$, post-cue: $t(19)=-0.98$, $d=0.22$, $p=0.33$).

304 The eye tracking data confirmed that participants maintained the required central fixation during the
305 stimulus period: the mean and s.e.m. for the horizontal and vertical eye position during the stimulus
306 were $1.8 \pm 0.5^\circ$ and $1.5 \pm 0.4^\circ$ ($n=14$ available participants). In addition, we found that saccadic eye
307 movements ($> 0.8^\circ$) executed during the stimulus epoch were very rare ($0.9 \pm 1.4\%$ of trials, mean \pm
308 SD; max. 4.1%).

309 *Cueing does not affect temporal sampling of information*

310 To understand whether visual motion discrimination was affected by the entire duration (800ms) of
311 the visual motion stimulus, or only a particular epoch of this, we computed psychophysical response
312 templates (Neri and Heeger, 2002; Chauvin et al., 2005). This was possible as the momentary amount
313 of evidence about the visual motion direction was manipulated randomly over time and trials,
314 allowing us to determine the time course of the perceptual use of this information using reverse
315 correlation (Marmarelis, 1978; Bang and Rahnev, 2017). This revealed that for all cue types motion
316 information was significantly related to participant's choices for most time points throughout the
317 stimulus, except the first 120ms (**Fig. 1C**, top panel; a group-level two-sided percentile bootstrap
318 confidence interval across all conditions, $p < 0.01$ uncorrected). To determine whether the congruency
319 of cue and visual motion affected the perceptual use of motion evidence, we contrasted congruent
320 and incongruent trials for each cue separately (**Fig. 1C**, bottom panel). This revealed no clear
321 difference between congruent and incongruent trials (a group-level two-sided percentile bootstrap
322 confidence intervals was inconclusive, $p > 0.05$ uncorrected), and the overall effect size of any
323 congruency effect was small (the peak Cohen's d for the underlying paired t-test was moderate $d =$
324 0.47 , with the overall mean effect across time and conditions $d = 0.18$). This suggests that any
325 influence of cue-stimulus congruency on response accuracy does not originate from the differential
326 sampling of sensory information at any particular time point. This also indicates that for the analysis
327 of the EEG data all time points are a priori equally relevant for the different conditions.

328

329 *Cueing influences the physiological correlates of visual motion encoding*

330 To understand where and when the cues affected the sensory encoding and decision process, we
331 first extracted signatures of visual motion encoding from the EEG data using single trial classification.
332 Previous work has shown that the audio-visual perceptual benefit correlates with the enhancement
333 of the cerebral encoding of visual motion information in occipital regions around 300ms post-
334 stimulus onset (Kayser et al., 2017; Kayser and Kayser, 2018). The physiological correlates of this can
335 be extracted by first determining EEG components sensitive to visual motion direction using single
336 trial classification, and then determining the influence of cue-stimulus congruency on this
337 classification component.

338 Here we computed a single trial classification of visual motion direction across all cue conditions. The
339 time course of classification performance exhibited three peaks (P1 at 0.093s; P2 at 0.32s and P3 at
340 0.64s; **Fig. 2A**). A statistical test based on a permutation procedure and controlling for multiple
341 comparisons along time (at $p < 0.05$) revealed a significant cluster (P2) between 0.28s and 0.4s
342 ($p = 0.014$; $t_{\text{cluster}} = 0.15$). We then focused on the EEG component defined by the LDA projection
343 obtained at P2 to ask whether the encoding of visual motion direction reflected by the underlying
344 neural activity was significantly affected by each cue (using a cluster-based randomization procedure
345 contrasting the LDA evidence between congruent and incongruent trials, correcting for multiple
346 comparisons along time at $p < 0.05$; **Fig. 2B,C**). The visual post-cue condition here served as a control,
347 as one should not expect significant congruency effects given that the cue was presented following
348 the stimulus. We found a significant congruency effect on visual motion encoding for the acoustic

349 ($p=0.032$, $t_{\text{cluster}}=16.7$, from 0.38 s to 0.44s, Cohen's d at peak = 0.8) and the visual pre-cue conditions
350 ($p=0.005$, $t_{\text{cluster}}=34.3$, from 0.26s to 0.37s, $d=1.0$). No effect was found for the visual-post cue, as
351 expected ($p>0.05$, $d=0.4$). These results replicate previous findings of a sound-driven enhancement of
352 visual motion encoding in EEG activity after 300ms post-stimulus onset (Kayser et al., 2017; Kayser
353 and Kayser, 2018), and demonstrate that a within-modality visual symbolic cue influences brain
354 activity captured by the same EEG component.

355 The above data suggest that the congruency benefit emerges earlier for the visual pre-cue than the
356 acoustic cue by about 130ms. To corroborate this result using a direct statistical analysis we
357 implemented a group-level percentile bootstrap test on the difference in the latencies of the first
358 significant congruency effects (**Fig. 2D**). The likelihood of observing a positive (i.e. visual pre-cue
359 being earlier) latency difference was 99.6% (i.e. $p=0.004$ for the observed difference, with the mean
360 of the bootstrapped distribution serving as measure of effect size of 96ms).

361 To obtain a better understanding of the cerebral sources giving rise to the motion sensitive EEG
362 component we obtained a source localization of this LDA component (at time=0.32s). This revealed
363 strongest contribution from left occipital regions (**Fig. 2E**; the peak source was at MNI: -43, -78 -5
364 around the inferior occipital gyrus, as determined using the AAL atlas), in line with the previous
365 studies.

366 As a control analysis, and to determine how specific this result was for this specific EEG component
367 (i.e. the LDA projection obtained at time=0.32s), we repeated the congruency analysis for the two
368 EEG components defined by the LDAs at the two local peaks that were not significant with respect to
369 motion direction classification (P1 and P3; c.f. Fig. 2A). For these, we found no significant congruency
370 effects for any of the cue conditions (at $p<0.05$ corrected for multiple comparisons along time for
371 each condition, with the peak effect size across cue conditions and the two LDA components being
372 $d=0.41$).

373

374 *Pre-stimulus activity shapes stimulus-cue interactions*

375 Previous work has shown that patterns of rhythmic pre-stimulus activity can influence how
376 multisensory information is combined (Keil et al., 2012; Rohe et al., 2019). To understand the
377 influence of pre-stimulus activity in the present paradigm we asked whether the power of different
378 frequency bands shapes the influence of visual motion and of the respective cue on participants'
379 choice. For this analysis, we derived the single trial spectral power from the visual motion-sensitive
380 EEG component (at P2) to avoid the statistical burden arising from a full electrode by time by
381 frequency analysis. We then modelled participants single trial responses (choice) based on a logistic
382 GLM including the visual motion direction, the direction indicated by the respective cue, their
383 interaction and their interactions with pre-stimulus power as predictors (**Fig. 3A**). This revealed a
384 significant contribution to participants behaviour of visual motion as expected per experimental
385 design ($t(19) = 13.5, 15.9, 16.9$ for the acoustic, visual-pre and visual post-cues; all at least $p<10^{-5}$,
386 Cohen's $d> 3$), a significant contribution of each cue confirming the above results ($t(19)=6.6, 2.4$ and
387 4.5 for each cue condition respectively, all $p<0.05$, $d= 1.4, 0.5$ and 1.0 respectively), but no
388 interaction between the visual stimulus and any of the cues ($t(19) = -0.25, -0.09$ and -1.05
389 respectively, all $p>0.05$, $d\leq 0.24$).

390 Concerning the pre-stimulus activity we found that alpha power influenced how visual motion and
391 the acoustic cue were combined. In the auditory condition alpha power significantly positively
392 modulated the impact of the visual stimulus (7-8Hz, -0.25s to -0.07s, $p=0.05$, $t_{\text{cluster}}=41$; Cohen's d at
393 peak = 0.68, **Fig. 3A**) and negatively modulated the influence of the acoustic cue (7-9Hz, -0.2s to -

394 0.03s, $p=0.023$, $t_{\text{cluster}} = -65$, $d=0.88$). Across individuals, the positive and negative alpha effects (i.e.
395 the regression beta's) were anti-correlated ($r(19)=-0.45$, $p=0.038$, Spearman rank correlation),
396 suggesting that these arise from a common neural generator. This result was confirmed in a separate
397 model, where we predicted performance (response accuracy, rather than choice) based on pre-
398 stimulus power and its interaction with the congruency of visual motion and the acoustic cue (**Fig.**
399 **3B**). This revealed no effect of power (no significant cluster; maximal $d=0.53$) but a significant
400 interaction of power and congruency in the alpha band, as expected ($p=0.045$, $t_{\text{cluster}} = -50$, $d=0.8$).

401 For the visual pre-cue alpha activity modulated how visual motion shaped behavior (5-7Hz -0.18s to -
402 0.01s, $p=0.029$, $t_{\text{cluster}} = -53$, $d=0.82$) but there was no significant interaction of pre-stimulus activity
403 with the cue (max $d=0.23$). The influence of alpha band activity on behavior in this condition was
404 hence restricted to an influence of alpha power on how the visual information shaped responses, but
405 did not depend on the congruency between visual motion and the cue. The latter conclusion was also
406 confirmed by an absence of a modulatory influence of alpha power on congruency in a direct
407 regression model of performance on power and congruency. Here we confirmed the expected
408 influence of alpha power (**Fig. 3B**; $p=0.022$, $t_{\text{cluster}} = -64$, $d=0.79$) and found no effect of the interaction
409 (no significant cluster, maximal $d=0.61$). Because the cue preceded the time window in which pre-
410 stimulus activity was quantified, we also asked whether the symbolic cue had a direct influence on
411 oscillatory power: we contrasted alpha power between the two directions of the visual-pre cue but
412 found no significant effect ($t(19)=0.64$, $p>0.05$, $d=0.004$), suggesting that the visual pre-cue did not
413 influence the subsequent alpha band activity. Finally, and as expected, for the visual post-cue
414 condition there were no significant effects (all $p>0.05$, maximal $d=0.55$).

415 Given that the analysed LDA component reflects a mixture of different EEG sensors, we also
416 computed these regression models at the sensor level, focusing only on the alpha frequencies and
417 time bins revealed by the above analysis (**Fig. 3A; insets**). This confirmed a contribution of left
418 occipital and fronto-central electrodes to the alpha effect.

419 To better understand how alpha activity influences behavioural responses in the audio-visual
420 condition, we sorted trials into groups composed of particularly low and high alpha power (median
421 split; using the intersection of the two alpha clusters for the acoustic cue condition from Fig. 3A) and
422 compared how alpha related to response accuracy depending on stimulus-cue congruency. That is,
423 we tested how behavioural performance co-varies with alpha in congruent and incongruent trials.
424 For the acoustic cue this revealed a significant dependence of performance on alpha in incongruent
425 ($t(19)=2.8$, $d=0.62$, $p_{\text{adj}} = 0.04$; paired two-sided t-test, FDR adjusted for multiple comparisons
426 using the Benjamini & Hochberg procedure), but not congruent trials ($t(19)=0.69$, $d=0.15$, $p_{\text{adj}} =$
427 0.39). This suggests that the influence of the acoustic cue is shaped by alpha power particularly when
428 the cue mismatches the visual information. For the visual pre-cue there was no significant
429 dependence of performance on alpha in either trial type (all $\text{abs}(t(19))<1.8$, $p_{\text{adj}}>0.05$), in line with
430 the above null finding for an interaction of alpha and the symbolic cue.

431

432 *Discussion*

433 Both abstract sensory cues inducing specific expectations as well as multisensory information can
434 improve perceptual performance. Using a visual motion discrimination task we here compared the
435 effect of within-modality symbolic cues and multisensory combination. The behavioural data showed
436 that both a prior visual symbolic cue and a visual symbolic cue following the task-relevant stimulus
437 can induce similar behavioural benefits, in line with previous work (Bang and Rahnev, 2017).
438 However, the congruency benefit from a congruent acoustic cue was significantly stronger compared
439 to these symbolic cues, although all cue types had the same level of accuracy in predicting the

440 correct visual motion direction, and the visual pre-cue was already presented before the task
441 stimulus. It remains to be investigated whether there is a genuine benefit for multisensory
442 information, or whether potential differences in attention for within and between-modality
443 associations contributed to these differences in cue efficiency (Talsma et al., 2010).

444 *Multisensory and expectation-based facilitation of behavior*

445 We asked whether the perceptual benefit arising from a prior symbolic cue and from multisensory
446 information arise from similar neurophysiological correlates. Concerning the acoustic cue, the
447 present data reproduce previous findings, in that congruent acoustic information facilitates the
448 encoding of visual motion in occipital brain regions at latencies of around 300ms post-stimulus onset.
449 These results support the notion of multisensory influences in low-level sensory regions (Rohe and
450 Noppeney, 2015; Cao et al., 2019), fit with previous neuroimaging studies reporting multisensory
451 influences in visual motion regions (Poirier et al., 2005; Lewis and Noppeney, 2010; Alink et al.,
452 2012), and directly reproduce previous results obtained using a very similar task (Kayser et al., 2017;
453 Kayser and Kayser, 2018). The minor differences in the stimulus classification performance obtained
454 here and that previous study may have two potential explanations: first, we here used a stimulus
455 with dynamic motion coherence, as opposed to fixed motion coherence used previously (Kayser et
456 al., 2017; Kayser and Kayser, 2018). Second, the present task emphasized response accuracy while
457 the previous study used speeded reactions.

458 Concerning the symbolic cue, our results suggest that the benefits induced by priming and
459 multisensory information at least partly arise from a shared neural substrate. This conclusion is
460 supported by the observation that the same visual motion sensitive EEG component was affected by
461 a prior visual cue and by multisensory congruency. Still, the priming benefit emerged significantly
462 earlier in time. Possibly, the information provided by the symbolic cue was already fully processed at
463 the time of visual motion onset, while the acoustic cue was presented simultaneously with the visual
464 motion. Previous studies have shown that expectation can influence the activity of neurons in
465 regions processing visual motion (Schlack and Albright, 2007; Kok et al., 2013), in line with the
466 localization of the cue-stimulus congruency benefit in the present study to an occipital EEG
467 component that is sensitive to visual motion.

468 We did not include a neutral condition, that is, a sound not inducing any motion percept or a neutral
469 symbolic cue. As a result, the present study cannot dissociate whether congruent or incongruent
470 cues enhance or reduce performance compared to a baseline condition (de Lange et al., 2013; Bang
471 and Rahnev, 2017). While a neutral condition is required to judge the specific influence of each
472 individual cue, and to understand how predictive vs. un-predictive cues influence cerebral
473 processing, the comparison of between and within-modality cues, which is of relevance here, is not
474 affected by the lack of a neutral condition.

475 Expectation has been suggested to act via two types of mechanisms, one reflecting a change in post-
476 sensory decision criteria and one reflecting an improvement in the encoding of sensory information
477 in low-level brain regions (Summerfield and Egner, 2009; Summerfield and de Lange, 2014; Bang and
478 Rahnev, 2017; Rungratsameetaweemana et al., 2018). The behavioural results obtained from the
479 visual post-cue condition confirm the relevance of post-sensory contributions: in this condition the
480 cue was presented only after the offset of the task-relevant stimulus, but the perceptual benefit was
481 comparable to the prior visual symbolic cue. At the same time our data provide a direct cerebral
482 correlate of the benefit induced by the prior symbolic cue in a generator of physiological activity that
483 can be interpreted as arising from low-level (visual motion-encoding) brain regions. The latency of
484 the congruency effect was longer than 200ms, making it difficult to speak of early (short-latency)
485 neural effects, and leaving the possibility that this congruency benefit arises in a top-down manner.

486 Importantly, the same question of whether a benefit induced by auxiliary information arises from
487 sensory- or decision-level mechanisms is discussed in the context of multisensory perception (Bizley
488 et al., 2016). Recent work has shown that multisensory integration arises in a hierarchical and
489 distributed manner, which makes it difficult to pinpoint a specific single underlying mechanisms or
490 brain region underlying this (Rohe and Noppeney, 2015; Aller and Noppeney, 2019; Cao et al., 2019;
491 Rohe et al., 2019). A similar conclusion may hold for general cueing effects. Hence, a parsimonious
492 interpretation is that both sensory-level and post-sensory mechanisms contribute to the symbolic
493 cueing benefit in the present paradigm, with the sensory-level contribution arising from neural
494 processes that are also susceptible to multisensory influences.

495

496 *Auxiliary multisensory effects and cueing*

497 In the literature on multisensory perception two types of multisensory effects have been described.
498 On the one hand there is sensory integration, also known as fusion, where the partly redundant task-
499 informative evidence from two modalities is combined, possibly in a bottom-up manner (Ernst and
500 Bulthoff, 2004; Angelaki et al., 2009; Cao et al., 2019). On the other hand, so called auxiliary
501 multisensory interactions have been described. Here, the perception of a task-relevant stimulus is
502 affected by the presentation of a stimulus in another modality, which by itself does not offer primary
503 task-relevant information (Odgaard et al., 2003; Jaekl and Harris, 2009; Caclin et al., 2011; Chen and
504 Spence, 2011; Gleiss and Kayser, 2014a; b). One example is the enhanced contrast sensitivity when a
505 visual stimulus is accompanied by an irrelevant sound (Lippert et al., 2007). In these paradigms the
506 auxiliary stimulus could be interpreted as an additional ‘cue’ that does not carry information about
507 the specific task-relevant variable (e.g. motion direction, or stimulus lateralization) but carries other
508 information, such as about stimulus timing, that nevertheless helps to facilitate performance.

509 Interestingly, one study reported such an auxiliary multisensory interaction in the context of visual
510 motion detection: this study showed that the detection of which of two stimulus intervals carries
511 coherent (as opposed to random) motion can be enhanced when both intervals are accompanied by
512 the same acoustic motion (Kim et al., 2012). Although the sound was identical across stimulus
513 intervals, performance was significantly enhanced when the sound was congruent to the visual
514 motion. One could interpret this benefit as resulting from the enhancement of visual motion
515 encoding by acoustic information in a bottom-up manner (Kim et al., 2012). However, an EEG study
516 capitalizing on the same paradigm linked the behavioural improvement to changes in post-stimulus
517 parietal alpha-band activity, which is most parsimoniously understood as reflecting an attentional
518 effect (Gleiss and Kayser, 2014b). In line with this, several auxiliary multisensory influences have
519 been considered to arise from changes in sensory saliency or attentional gain (Lippert et al., 2007;
520 Chen and Spence, 2011; Gleiss and Kayser, 2014a).

521 The present paradigm differed from this two-interval task (Kim et al., 2012) in that the additional
522 multisensory or symbolic cues were directly predictive of the likely motion direction, and hence these
523 could be used in principle (by guessing in proportion to the cue validity) to solve the task. In that
524 sense, the present paradigm more directly corresponds to a sensory integration task, where one
525 task-relevant stimulus (here visual motion) is accompanied by a second task-informative stimulus
526 (here the cue). Further work is required to better understand which perceptual and neural
527 mechanisms are common to abstract cueing, multisensory fusion and other types of multisensory
528 effects, and which perceptual and neural mechanisms differentiate these paradigms. The analysis
529 used here could provide one experimental approach to address this question in the future.

530

531 *Pre-stimulus brain state and cue combination*

532 To understand how pre-stimulus activity influences the manner in which two pieces of sensory
533 evidence are combined, we modelled the influence of oscillatory power on how visual motion and
534 the cue are combined. This revealed that alpha power significantly interacted with the impact of
535 visual motion on behavior regardless of the nature of the cue, and was predictive of how an acoustic
536 but not a visual symbolic cue influenced behavior. In the multisensory condition, stronger alpha
537 power apparently enhances the influence of the visual stimulus but reduces the influence of the
538 sound, seemingly reflecting a relative weighting of the two stimuli. This apparent influence of pre-
539 stimulus activity was stronger for incongruent trials, where participant's accuracy was worse when
540 alpha power was low. This is in line with the recent suggestion that reduced pre-stimulus parietal
541 alpha reflects the *a priori* tendency with which participants combine multisensory information (Rohe
542 et al., 2019). With this interpretation in mind, the analysed alpha band activity is reflective of a high-
543 level amodal process arbitrating between sensory integration and segregation.

544 In the visual pre-cue condition reduced alpha power interacted with the influence of the visual
545 stimulus, in line with reduced alpha reflecting enhanced visual attention (Thut et al., 2006; Romei et
546 al., 2007). However, we did not find a significant interaction of alpha with the visual symbolic cue,
547 nor was alpha power itself directly influenced by the preceding cue. This is in contrast to previous
548 studies proposing a role of pre-stimulus alpha in mediating the influence of prior expectations (Bauer
549 et al., 2014; Mayer et al., 2016; Samaha et al., 2018). One possibility is that alpha mediates
550 expectations most strongly for spatial or temporal stimuli (Worden et al., 2000; Samaha et al., 2015;
551 Samaha et al., 2016), although effects for symbolic cues have been reported as well (Mayer et al.,
552 2016). Another possibility is that the analysed alpha activity in the two conditions originated from
553 distinct neural structures. This may seem unlikely given that the analysis of spectral activity was
554 based on the same electrode combination ('virtual signal') for both conditions, and given that the
555 respective sensor-level topographies were similar between the two cue types. However, it remains
556 debated whether parietal alpha band activity arises from a single supramodal mechanism or distinct
557 and modality specific generators (Banerjee et al., 2011). Previous work has linked parietal alpha to
558 multiple sensory or cognitive processes, including spatial attention or sensory arbitration (Thut et al.,
559 2006; Romei et al., 2007; Rohe et al., 2019), without clear evidence for how to experimentally
560 separate these potentially different mechanisms reflected in the same physiological activity pattern.
561 As a result, future work specifically needs to address the relevance of attention and alpha band
562 activity when investigating the possibly shared or distinct mechanisms underlying abstract cueing,
563 multisensory fusion and other types of multisensory effects (Summerfield and Egner, 2009; Talsma et
564 al., 2010; Zuanazzi and Noppeney, 2019).

565

566

567

568

569

570 Figure legends

571

572 **Figure 1. Paradigm and behavioural data.** **A)** Participants discriminated the direction of visual random-dot
573 motion (800ms stimulus duration). Across three conditions, the visual motion was accompanied by different
574 auxiliary cues: simultaneous acoustic motion, a visual symbolic cue preceding the visual motion, or a visual
575 symbolic cue following the motion stimulus. Each cue was congruent with the random-dot motion on 66% of
576 trials. **B)** For each cue type response accuracy was higher for congruent trials, while reaction times (rel. to a
577 response cue) were not affected by congruency (mean and s.e.m. across participants, dots represent single
578 participant data). **C)** Psychophysical response templates (weights) were calculated to quantify the influence of
579 the moment-by-moment motion energy on behavior. These templates were significant for most time points
580 (upper panel; thick black line: grand average, thin black line 5% bootstrap confidence interval across conditions;
581 thick coloured lines: condition wise group means), but were not affected by cue-stimulus congruence (lower
582 panel; thin lines condition wise group means, dashed lines 5% bootstrap confidence intervals).

583

584 **Figure 2. Electrophysiological correlates of cue-stimulus congruency.** **A)** Single trial linear discriminant analysis
585 (LDA) was used to determine EEG components sensitive to the direction of random-dot motion, that is, sensitive
586 to the task-relevant visual information. The curve shows the classifier performance as area under the receiver
587 operator characteristic (mean and s.e.m. across participants), which was significant (cluster-based permutation
588 test, $p < 0.05$) around 300ms (peak at P2). The topographies show the forward models of the LDA classifiers at
589 three local peaks. **B)** Effect of cue-stimulus congruency on the single trial LDA evidence in the component
590 derived at time P2, separately for each cue condition and shown as group-level t-value. Epochs with a
591 congruency effect are indicated (cluster-based permutation test, $p < 0.05$). **C)** Congruency effects on LDA
592 evidence at the time points of the respective global peaks (mean and s.e.m. across participants). **D)** Test for a
593 latency difference of the congruency effect in the acoustic and visual-pre conditions. The graph shows the
594 group-level bootstrap distribution (2000 samples) of the latency difference between the congruency effects. **E)**
595 Source localization of the LDA component (at P2) determined as source-level correlation (z-scored) between
596 grid-wise activity and the LDA activity (group-averaged, shown at MNI Z=-5).

597

598 **Figure 3. Time-frequency analysis of pre-stimulus influences on behavior.** **A)** For each condition we modelled
599 single trial choice based on the visual stimulus, the cue, their interaction with each other, and the interaction of
600 each with pre-stimulus power. Graphs display the group-level t-values obtained from the respective single-
601 participant regression betas for the interaction of power with visual motion and the cue. Pre-stimulus activity
602 was analysed within the LDA component derived at P2 (c.f. Figure 2). The topographies (insets) show the
603 electrode-wise t-values for the alpha band cluster in the respective panel obtained from single-electrode
604 regression models. Cluster-averaged regression coefficients (betas) for individual participants are shown on the
605 right. Black outlines indicate significant clusters (cluster-based permutation test, $p < 0.05$). **B)** Same as in A) but
606 for a regression model of performance (response accuracy) based on stimulus-cue congruency, pre-stimulus
607 power and the interaction of power with congruency.

608

609

610 *References*

611

612

- 613 Alink A, Euler F, Kriegeskorte N, Singer W, Kohler A (2012) Auditory motion direction encoding in
614 auditory cortex and high-level visual cortex. *Hum Brain Mapp* 33: 969-978.
- 615 Aller M, Noppeney U (2019) To integrate or not to integrate: Temporal dynamics of hierarchical
616 Bayesian causal inference. *PLoS Biol* 17: e3000210.
- 617 Angelaki DE, Gu Y, DeAngelis GC (2009) Multisensory integration: psychophysics, neurophysiology,
618 and computation. *Curr Opin Neurobiol* 19: 452-458.
- 619 Banerjee S, Snyder AC, Molholm S, Foxe JJ (2011) Oscillatory alpha-band mechanisms and the
620 deployment of spatial attention to anticipated auditory and visual target locations:
621 supramodal or sensory-specific control mechanisms? *J Neurosci* 31: 9923-9932.
- 622 Bang JW, Rahnev D (2017) Stimulus expectation alters decision criterion but not sensory signal in
623 perceptual decision making. *Sci Rep* 7: 17072.
- 624 Bauer M, Stenner M-P, Friston KJ, Dolan RJ (2014) Attentional Modulation of Alpha/Beta and Gamma
625 Oscillations Reflect Functionally Distinct Processes. *The Journal of Neuroscience* 34: 16117-
626 16125.
- 627 Bizley JK, Jones GP, Town SM (2016) Where are multisensory signals combined for perceptual
628 decision-making? *Curr Opin Neurobiol* 40: 31-37.
- 629 Brainard DH (1997) The Psychophysics Toolbox. *Spat Vis* 10: 433-436.
- 630 Caclin A, Bouchet P, Djoulah F, Pirat E, Pernier J, Giard MH (2011) Auditory enhancement of visual
631 perception at threshold depends on visual abilities. *Brain Res* 1396: 35-44.
- 632 Cao Y, Summerfield C, Park H, Giordano BL, Kayser C (2019) Causal Inference in the Multisensory
633 Brain. *Neuron* 102: 1076-1087.
- 634 Chauvin A, Worsley KJ, Schyns PG, Arguin M, Gosselin F (2005) Accurate statistical tests for smooth
635 classification images. *J Vis* 5: 659-667.
- 636 Cheadle S, Egnér T, Wyart V, Wu C, Summerfield C (2015) Feature expectation heightens visual
637 sensitivity during fine orientation discrimination. *J Vis* 15: 14.
- 638 Chen YC, Spence C (2011) The crossmodal facilitation of visual object representations by sound:
639 evidence from the backward masking paradigm. *J Exp Psychol Hum Percept Perform* 37:
640 1784-1802.
- 641 Cichy RM, Kriegeskorte N, Jozwik KM, van den Bosch JF, Charest I (2019) The spatiotemporal neural
642 dynamics underlying perceived similarity for real-world objects. *Neuroimage* 194: 12-24.
- 643 de Lange FP, Heilbron M, Kok P (2018) How Do Expectations Shape Perception? *Trends Cogn Sci* 22:
644 764-779.
- 645 de Lange FP, Rahnev DA, Donner TH, Lau H (2013) Prestimulus oscillatory activity over motor cortex
646 reflects perceptual expectations. *J Neurosci* 33: 1400-1410.
- 647 Eckstein MP, Ahumada AJ, Jr. (2002) Classification images: a tool to analyze visual strategies. *J Vis* 2:
648 1.
- 649 Ernst MO, Bulthoff HH (2004) Merging the senses into a robust percept. *Trends Cogn Sci* 8: 162-169.
- 650 Gleiss S, Kayser C (2014a) Acoustic Noise Improves Visual Perception and Modulates Occipital
651 Oscillatory States. *J Cogn Neurosci* 26: 699-711.
- 652 Gleiss S, Kayser C (2014b) Oscillatory mechanisms underlying the enhancement of visual motion
653 perception by multisensory congruency. *Neuropsychologia* 53: 84-93.
- 654 Green DM, Sweets JA (1966) *Signal Detection Theory and Psychophysics*. Place, Wiley.
- 655 Grootswagers T, Cichy RM, Carlson TA (2018) Finding decodable information that can be read out in
656 behaviour. *Neuroimage* 179: 252-262.
- 657 Haufe S, Meinecke F, Gorgen K, Dahne S, Haynes JD, Blankertz B, et al. (2014) On the interpretation
658 of weight vectors of linear models in multivariate neuroimaging. *Neuroimage* 87: 96-110.

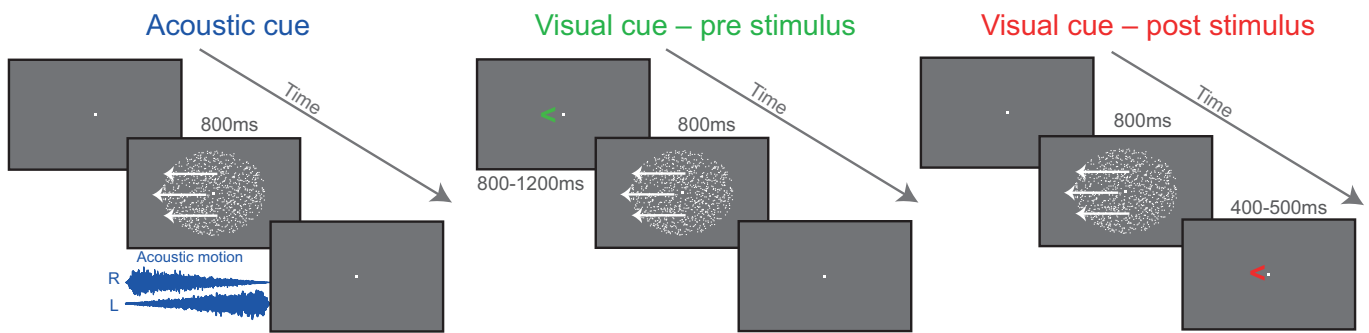
15

- 659 Hipp JF, Siegel M (2013) Dissociating neuronal gamma-band activity from cranial and ocular muscle
660 activity in EEG. *Front Hum Neurosci* 7: 338.
- 661 Jaekl PM, Harris LR (2009) Sounds can affect visual perception mediated primarily by the
662 parvocellular pathway. *Vis Neurosci* 26: 477-486.
- 663 Jiang J, Summerfield C, Egner T (2013) Attention sharpens the distinction between expected and
664 unexpected percepts in the visual brain. *J Neurosci* 33: 18438-18447.
- 665 Kaiser M, Senkowski D, Busch NA, Balz J, Keil J, Keil J, et al. (2019) Single trial prestimulus oscillations
666 predict perception of the sound-induced flash illusion. *Sci Rep* 9: 5983.
- 667 Kayser C, Logothetis NK (2007) Do early sensory cortices integrate cross-modal information? *Brain*
668 *Struct Funct* 212: 121-132.
- 669 Kayser SJ, Kayser C (2018) Trial by trial dependencies in multisensory perception and their correlates
670 in dynamic brain activity. *Sci Rep* 8: 3742.
- 671 Kayser SJ, McNair SW, Kayser C (2016) Prestimulus influences on auditory perception from sensory
672 representations and decision processes. *Proc Natl Acad Sci U S A* 113: 4842-4847.
- 673 Kayser SJ, Philiastides MG, Kayser C (2017) Sounds facilitate visual motion discrimination via the
674 enhancement of late occipital visual representations. *Neuroimage* 148: 31-41.
- 675 Keil J, Muller N, Ihssen N, Weisz N (2012) On the variability of the McGurk effect: audiovisual
676 integration depends on prestimulus brain states. *Cereb Cortex* 22: 221-231.
- 677 Keren AS, Yuval-Greenberg S, Deouell LY (2010) Saccadic spike potentials in gamma-band EEG:
678 characterization, detection and suppression. *Neuroimage* 49: 2248-2263.
- 679 Kiani R, Hanks TD, Shadlen MN (2008) Bounded integration in parietal cortex underlies decisions
680 even when viewing duration is dictated by the environment. *J Neurosci* 28: 3017-3029.
- 681 Kim R, Peters MA, Shams L (2012) 0 + 1 > 1: How adding noninformative sound improves performance
682 on a visual task. *Psychol Sci* 23: 6-12.
- 683 Kline RB (2004) *Beyond Significance Testing: Reforming Data Analysis Methods in Behavioral*
684 *Research*. Place, American Psychological Association.
- 685 Kok P, Brouwer GJ, van Gerven MA, de Lange FP (2013) Prior expectations bias sensory
686 representations in visual cortex. *J Neurosci* 33: 16275-16284.
- 687 Kok P, Jehee JF, de Lange FP (2012) Less is more: expectation sharpens representations in the
688 primary visual cortex. *Neuron* 75: 265-270.
- 689 Lange J, Oostenveld R, Fries P (2013) Reduced occipital alpha power indexes enhanced excitability
690 rather than improved visual perception. *J Neurosci* 33: 3212-3220.
- 691 Lewis R, Noppeney U (2010) Audiovisual synchrony improves motion discrimination via enhanced
692 connectivity between early visual and auditory areas. *J Neurosci* 30: 12329-12339.
- 693 Lippert M, Logothetis N, Kayser C (2007) Improvement of visual contrast detection by a simultaneous
694 sound. *Brain Res* 102-109.
- 695 Maris E, Oostenveld R (2007) Nonparametric statistical testing of EEG- and MEG-data. *J Neurosci*
696 *Methods* 164: 177-190.
- 697 Marmarelis V (1978) *Analysis of Physiological Systems: The White-Noise Approach*. Place, Springer
698 US.
- 699 Mayer A, Schwiedrzik CM, Wibrall M, Singer W, Melloni L (2016) Expecting to See a Letter: Alpha
700 Oscillations as Carriers of Top-Down Sensory Predictions. *Cereb Cortex* 26: 3146-3160.
- 701 Meyer GF, Wuerger SM (2001) Cross-modal integration of auditory and visual motion signals.
702 *Neuroreport* 12: 2557-2560.
- 703 Moore BC (2003) *An Introduction to the Psychology of Hearing*. Place, Emerald Group Publishing Ltd.
- 704 Neri P, Heeger DJ (2002) Spatiotemporal mechanisms for detecting and identifying image features in
705 human vision. *Nat Neurosci* 5: 812-816.
- 706 Nichols TE, Holmes AP (2002) Nonparametric permutation tests for functional neuroimaging: a
707 primer with examples. *Hum Brain Mapp* 15: 1-25.
- 708 O'Beirne GA, Patuzzi RB (1999) Basic properties of the sound-evoked post-auricular muscle response
709 (PAMR). *Hear Res* 138: 115-132.

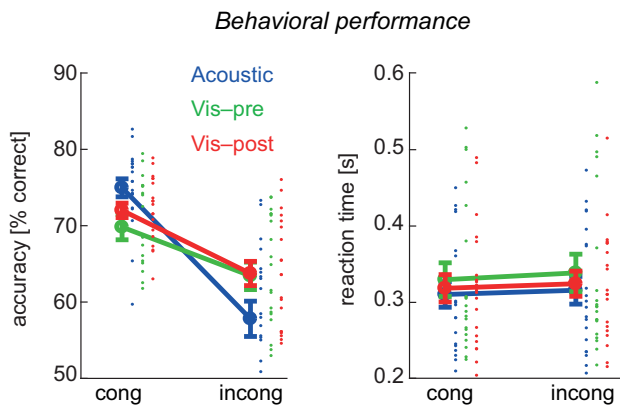
- 710 Odgaard EC, Arieh Y, Marks LE (2003) Cross-modal enhancement of perceived brightness: sensory
711 interaction versus response bias. *Percept Psychophys* 65: 123-132.
- 712 Oldfield RC (1971) The assessment and analysis of handedness: the Edinburgh inventory.
713 *Neuropsychologia* 9: 97-113.
- 714 Oostenveld R, Fries P, Maris E, Schoffelen JM (2011) FieldTrip: Open source software for advanced
715 analysis of MEG, EEG, and invasive electrophysiological data. *Comput Intell Neurosci* 2011:
716 156869.
- 717 Parra LC, Spence CD, Gerson AD, Sajda P (2005) Recipes for the linear analysis of EEG. *Neuroimage*
718 28: 326-341.
- 719 Philiastides MG, Heekeren HR, Sajda P (2014) Human scalp potentials reflect a mixture of decision-
720 related signals during perceptual choices. *J Neurosci* 34: 16877-16889.
- 721 Poirier C, Collignon O, Devolder AG, Renier L, Vanlierde A, Tranduy D, et al. (2005) Specific activation
722 of the V5 brain area by auditory motion processing: an fMRI study. *Brain Res Cogn Brain Res*
723 25: 650-658.
- 724 Quax SC, Dijkstra N, van Staveren MJ, Bosch SE, van Gerven MAJ (2019) Eye movements explain
725 decodability during perception and cued attention in MEG. *Neuroimage* 195: 444-453.
- 726 Rohe T, Ehli AC, Noppeney U (2019) The neural dynamics of hierarchical Bayesian causal inference in
727 multisensory perception. *Nat Comm* 10: 1907.
- 728 Rohe T, Noppeney U (2015) Cortical Hierarchies Perform Bayesian Causal Inference in Multisensory
729 Perception. *PLoS Biol* 13: e1002073
- 730 Rolfs M (2009) Microsaccades: small steps on a long way. *Vision Res* 49: 2415-2441.
- 731 Romei V, Murray MM, Merabet LB, Thut G (2007) Occipital transcranial magnetic stimulation has
732 opposing effects on visual and auditory stimulus detection: implications for multisensory
733 interactions. *J Neurosci* 27: 11465-11472.
- 734 Rungratsameetaweemana N, Itthipuripat S, Salazar A, Serences JT (2018) Expectations Do Not Alter
735 Early Sensory Processing during Perceptual Decision-Making. *J Neurosci* 38: 5632-5648.
- 736 Samaha J, Bauer P, Cimaroli S, Postle BR (2015) Top-down control of the phase of alpha-band
737 oscillations as a mechanism for temporal prediction. *Proc Natl Acad Sci U S A* 112: 8439-
738 8444.
- 739 Samaha J, Boutonnet B, Postle BR, Lupyan G, Samaha J, Postle BR (2018) Effects of meaningfulness on
740 perception: Alpha-band oscillations carry perceptual expectations and influence early visual
741 responses. *Sci Rep* 8: 6606.
- 742 Samaha J, Sprague TC, Postle BR (2016) Decoding and Reconstructing the Focus of Spatial Attention
743 from the Topography of Alpha-band Oscillations. *J Cogn Neurosci* 28: 1090-1097.
- 744 Schlack A, Albright TD (2007) Remembering visual motion: neural correlates of associative plasticity
745 and motion recall in cortical area MT. *Neuron* 53: 881-890.
- 746 Simmons JP, Nelson LD, Simonsohn U (2011) False-positive psychology: undisclosed flexibility in data
747 collection and analysis allows presenting anything as significant. *Psychol Sci* 22: 1359-1366.
- 748 Stein BE, Meredith MA (1993) Merging of the Senses. *MIT Press, Cambridge*
- 749 Stein BE, Stanford TR (2008) Multisensory integration: current issues from the perspective of the
750 single neuron. *Nat Rev Neurosci* 9: 255-266.
- 751 Summerfield C, de Lange FP (2014) Expectation in perceptual decision making: neural and
752 computational mechanisms. *Nat Rev Neurosci* 15: 745-756.
- 753 Summerfield C, Egnér T (2009) Expectation (and attention) in visual cognition. *Trends Cogn Sci* 13:
754 403-409.
- 755 Talsma D, Senkowski D, Soto-Faraco S, Woldorff MG (2010) The multifaceted interplay between
756 attention and multisensory integration. *Trends Cogn Sci* 14: 400-410.
- 757 Thut G, Nietzel A, Brandt SA, Pascual-Leone A (2006) Alpha-band electroencephalographic activity
758 over occipital cortex indexes visuospatial attention bias and predicts visual target detection. *J*
759 *Neurosci* 26: 9494-9502.
- 760 Urai AE, Braun A, Donner TH (2017) Pupil-linked arousal is driven by decision uncertainty and alters
761 serial choice bias. *Nat Commun* 8: 14637.

- 762 Worden MS, Foxe JJ, Wang N, Simpson GV (2000) Anticipatory biasing of visuospatial attention
763 indexed by retinotopically specific alpha-band electroencephalography increases over
764 occipital cortex. *J Neurosci* 20: Rc63.
- 765 Zuanazzi A, Noppeney U (2019) Distinct Neural Mechanisms of Spatial Attention and Expectation
766 Guide Perceptual Inference in a Multisensory World. *J Neurosci* 39: 2301-2312.
- 767

A



B



C

