1   **Sound-driven enhancement of vision: Disentangling detection from decisional level**

2   **contributions**

3

4   Alexis Pérez-Bellido[1], Salvador Soto-Faraco[2,3] & Joan López-Moliner[1,4,*]

5

6   [1]Departament de Psicologia Bàsica, Universitat de Barcelona

7   [2]Departament de Tecnologies de la Informació i les Comunicacions, Universitat Pompeu Fabra

8   [3]Institució Catalana de Recerca i Estudis Avançats (ICREA)

9   [4]Institute for Brain Cognition and Behavior (IR3C)

10   *Corresponding author: j.lopezmoliner@ub.edu

11

12

13

14

15

16

17

18

19

20

21   **Abstract**

22   Cross-modal enhancement can be mediated both by higher-order effects due to attention and

23   decision making, as well as by detection level stimulus-driven interactions. However, the

24   contribution of each of these sources to behavioural improvements has not been conclusively

25   determined and quantified separately. Here, we apply psychophysical analysis based on Piéron

26   functions in order to separate stimulus-dependent changes from those accounted by decisional

27   level contributions. Participants performed a simple visual speeded detection task on Gabor

28   patches of different spatial frequencies and contrast values, presented with and without

29   accompanying sounds. On the one hand, we identified an additive cross-modal improvement in

30   mean reaction times across all types of visual stimuli that would be well explained by

31   interactions not strictly based on stimulus-driven modulations (e.g. due to reduction of temporal

32   uncertainty and motor times). On the other hand, we singled out an audio-visual benefit that

33   strongly depended on stimulus features such as frequency and contrast. This particular

34   enhancement was selective to low visual spatial frequency stimuli, optimized for magnocellular

35   sensitivity. We therefore conclude that interactions at detection stages and at decisional

36   processes in response selection contribute in audio-visual enhancement can be separated online,

37   and express on partly different aspects of visual processing.

38

39   Key words: sensory integration, reaction time, psychophysics, audio-visual, magnocellular, race

40   models

41

42

43

44

45

46 **Introduction**

47 Watching a movie is not only a visual experience, attending to a concert is not only acoustic and

48 stroking our pet is not only tactile. Instead, perception in everyday life typically involves input

49 from various sensory modalities. In consequence, the nervous system of complex animals,

50 including humans, has evolved to maximize the information by exploiting correlations between

51 multimodal inputs. In recent years, a growing body of literature indicates the existence of

52 multisensory interactions in a wide variety of contexts. For example, neurophysiological

53 evidence from animal studies has described multisensory responses of neurons in the deep

54 layers of the superior colliculus (SC), a sub-cortical structure that receives ascending inputs

55 from auditory, visual, and somatosensory pathways and, subserves reflexive stimulus-driven

56 behaviours such as orienting the head and eye movements (e.g., Meredith & Stein, 1983;

57 Meredith & Stein, 1986). Some of these neurons are characterized by their enhanced responses

58 to multisensory compared to unisensory stimulation, typically when sensory inputs coming from

59 different modalities are in approximate temporal synchrony and spatially overlapping (Meredith

60 & Stein, 1996; Wallace, Meredith, & Stein, 1993, 1998). Complementary behavioural studies

61 have demonstrated that multisensory stimulation leads to benefits such as lower detection

62 thresholds (e.g. Noesselt et al., 2010; Wuerger, Hofbauer, & Meyer, 2003), increased precision

63 (e.g. Ernst & Bülthoff, 2004), and quicker motor responses (e.g. Morrell, 1972; Nozawa,

64 Reuter-Lorenz, & Hughes, 1994). One of the most often cited, yet still controversial cases of

65 multisensory facilitation is the enhancement of visual perception as a function of concurrent

66 acoustic input. Here, we examine the basis of this acoustic facilitation of visual processing in

67 behaviour.

68 Behavioural evidence for auditory enhancement of vision is illustrated by improvements in

69 orienting responses to audio-visual, as compared to visual-only stimuli (in cats, Stein, Meredith,

70 Huneycutt, & McDade1989); faster and more accurate saccadic reactions(Colonius & Arndt,

71 2001; Corneil, Van Wanrooij, Munoz, & Van Opstal, 2002; Diederich & Colonius, 2004);

72 increases in brightness judgments of a visual stimulus accompanied by a sound (Stein, London,

73   Wilkinson, & Price, 1996); and, improvements in sensitivity to different aspects of visual

74   stimuli (Bolognini, Frassinetti, Serino, & Làdavas, 2005; Frassinetti, Bolognini, & Làdavas,

75   2002; Jaekl & Soto-Faraco, 2010; Manjarrez, Mendez, Martinez, Flores, & Mirasso, 2007).

76   However, some controversy still surrounds the underlying sources of these audio-visual

77   enhancements. Despite these improvements are generally attributed to changes in early sensory

78   processes at detection stage as a result of hard-wired integrative mechanisms (Lakatos, Chen,

79   O'Connell, Mills, & Schroeder, 2007; Stein, 1998), many of these cross-modal enhancements

80   cannot be dissociated from effects at decisional stages of processing that are more general in

81   scope (Doyle & Snowden, 2001; Lippert, Logothetis, & Kayser, 2007; Odgaard, Arieh, &

82   Marks, 2003). For example, several authors have argued that sound-induced enhancement of

83   vision can be attributed to a reduction in uncertainty within the spatial and temporal domains,

84   without necessarily implying early interactions at sensory level. Indeed, generic mechanisms

85   such as the involuntary orienting of spatial attention to the location of a sound have been shown

86   to enhance early perceptual processing within the visual modality (e.g. McDonald, Teder-

87   Sälejärvi, & Hillyard, 2000; McDonald, Teder-Sälejärvi, Di Russo, & Hillyard, 2005; Spence &

88   Driver, 1994), without necessarily involving any early sensory interaction between sound and

89   vision. Likewise, in the temporal domain, an abrupt sound can considerably reduce the temporal

90   uncertainty about the time of visual target presentation (e.g. Lippert, Logothetis, & Kayser,

91   2007).Thus, according to this alternative view, cross-modal enhancement would be explained

92   by cross-modal cueing effects, which have been widely demonstrated in the attention literature

93   (e.g. Driver & Spence, 1998).Discerning the contribution of these cross-modal cueing effects

94   from early cross-modal convergence leading to enhanced detection, as a classical sensory

95   integration account would propose, is critical when characterizing the nature of multisensory

96   interactions (Driver & Noesselt, 2008; McDonald, Teder-Sälejärvi, & Ward, 2001).

97   The nature of auditory-induced enhancement of visual processing has previously been addressed

98   by measuring detection thresholds or changes in visual sensitivity, but as discussed above, with

99   mixed conclusions (Frassinetti et al., 2002; Jaekl & Soto-Faraco, 2010; Lippert et al., 2007;

100 McDonald et al., 2000; Noesselt, Bergmann, Hake, Heinze, & Fendrich, 2008). Studying the

101 temporal dynamics of audio-visual perception has also provided interesting insights about the

102 nature of such interactions. For example, at a neurophysiological level, multisensory stimulation

103 leads to shorter neuronal latencies (Rowland, Quessy, Stanford, & Stein, 2007). Rowland et al.

104 found that initial neural responses in individual superior colliculus neurons of anesthetized cats

105 speed up in response to audio-visual stimuli. Additionally, Giard & Peronnet (1999) provided

106 physiological evidence in humans that audio-visual stimuli produced neural activity in the

107 visual cortex as early as 40 ms post-stimulus, above and beyond that evoked by visual-only

108 stimuli. It is remarkable that, considering that possible sensory-based effects may express at

109 (and lead to) shorter processing latencies, previous behavioural studies addressing the nature of

110 auditory-induced visual enhancement (see above) have typically used off-line detection tasks

111 that are not so sensitive to latency shifts. Note that, as we discuss below, the many studies using

112 the redundant signals approach on audio-visual RTs cannot tell apart whether the speed up in

113 latencies is due to enhancement of vision specifically. Here, we decided to apply a reaction time

114 paradigm specifically focused on measuring visual processing, with the aim to capture any

115 latency aspect involved in auditory-induced visual enhancement.

116 When two or more signals are presented together in different channels, simple detection

117 responses are faster on average than when only one signal is present – a well-known

118 phenomenon termed the redundant signals effect (RSE; Kinchla, 1974). Previous studies have

119 modelled such multisensory reaction time advantages as a function of unisensory RTs applying

120 the race-model inequality test (Colonius & Arndt, 2001; Miller, 1982; but see Otto &

121 Mamassian, 2012, for a criticism). According to many authors, this paradigm allows one to infer

122 how these signals are used in detection. If the auditory and the visual signal are processed

123 independently at sensory level and combined during a decision stage, the distribution of RTs to

124 audio-visual events cannot be significantly faster than the probability summation model

125 predictions based on the RTs to unisensory targets (Raab, 1962). If, on the other hand, audio-

126 visual RTs are faster than those predicted by the probabilistic summation (race-) model, then it

127   is inferred that these signals are effectively integrated within a common channel prior to

128   decision. This latter kind of interaction has been accurately described by linear or co-activation

129   models (Colonius & Arndt, 2001; Miller, 1982; Nickerson, 1973). This race-model inequality

130   test has been extensively applied in the multisensory domain (see Cappe, Thut, Romei, &

131   Murray, 2009; Colonius & Arndt, 2001; Laurienti, Kraft, Maldjian, Burdette, & Wallace, 2004;

132   Leo, Bertini, di Pellegrino, & Làdavas, 2008). However, a key aspect of this approach is that the

133   task is not modality selective (that is, participants respond to any signal regardless of whether it

134   is auditory or visual). Therefore, regardless of recent criticisms (Otto & Mamassian, 2012), the

135   RSE model is simply not adequate when the goal is to ascertain how, specifically one sensory

136   modality, is enhanced by the presence of the other such as is the case of the present study. For

137   this reason, we started by using a different RT paradigm well-known in visual literature that

138   exploits the fact that RTs to visual events decrease monotonically with stimulus intensity

139   following a power function, typically referred to as the Piéron function (Piéron, 1914):

140    $RT = kC^{-\alpha} + t_0$ (1)

141   The function in Equation (1) captures different separate contributions to response latency across

142   variations in stimulus contrast (C). In this model, two parameters (k and α) modulate the decay

143   of RTs caused by stimulus-dependent variables, while t_0 modulations reflects the variation in

144   RTs due to stimulus-independent factors (Pins & Bonnet, 1996; Plainis & Murray, 2000). In

145   particular, k accounts for the time it takes for sensory information to reach threshold, that is, the

146   gain rate in the accumulation of sensory evidence. In turn, α is an exponent that characterizes a

147   given sensory modality. Finally, in this framework t_0 represents the asymptotic RT, which

148   reflects a time constant that includes processing latencies intrinsic of the sensory pathway and

149   motor time of the effector system (Pins & Bonnet, 2000). One important aspect for the present

150   study is that variations in response latencies caused by differences in stimulus uncertainties, task

151   complexity or decision making, cumulate additively and equally across all levels of contrast or

152   any other stimulus variation, thus adding to the parameter t_0 (Pins & Bonnet, 1996). This

153   assumption has been explicitly tested by Carpenter (2004) using visual stimuli. In particular,

154   using this framework Carpenter showed that the contributions of contrast variations to overall

155   RT are associated to detection processes, whereas variations concerning higher-order factors

156   (such as the probability of the stimulus requiring a response or not) are linked to decision

157   processes leading to shifts in t_0. Applying this approach to the cross-modal case, we will address

158   how the different parameters of the Piéron function account for changes in RT following the

159   addition of sound to an imperative visual stimulus. This paradigm will therefore allow us to

160   discern higher-level effects from interactions occurring due to stimulus driven processing at the

161   detection level between vision and sound (see the Data analysis section).

162   Given that previous studies have suggested that differences in multisensory integration may

163   depend on the particular visual pathways involved (Jaekl & Harris, 2009; Jaekl & Soto-Faraco,

164   2010; Leo et al., 2008), we included a range of spatial frequencies putatively engaging the

165   Magno (M) and Parvo-cellular (P) pathways to different extents. We chose spatial frequency

166   values of 0.3 c/deg, 0.74 c/deg and 5.93 c/deg given that frequencies above 4 c/deg are thought

167   to stimulate sustained channels more specifically than frequencies below (Legge, 1978; Levi,

168   Harwerth, & Manny, 1979). Therefore the low spatial frequencies 0.3 c/deg and 0.74 c/deg in

169   our stimuli are assumed to primarily stimulate the magnocellular pathway whilst the

170   intermediate frequency 5.93 c/deg is predicted to elicit relatively more activation in the

171   parvocellular channel.

172   We hypothesize that sound-induced visual enhancements that are dependent of contrast (i.e.

173   detection level) should be modulated by spatial frequency and will therefore lead to differences

174   in the gain parameter k across spatial frequencies. In particular, stimulus-driven benefits should

175   concentrate in the low spatial frequencies. On the other hand, interactions produced by the

176   sound that are not contrast dependent (e.g., decision stage, Carpenter 2004) should in principle

177   accumulate equally across all contrast levels, being mainly reflected in changes of the

178   asymptotic parameter t_0. Moreover, this sensory unspecific improvement driven by high-level

179   processes is not only expected to occur invariantly across contrast levels, but also across spatial

180   frequencies too.

181 Note that, in strict sense, it is not logically possible to discard the influence of sensory-level

182 interactions on t_0, so that the non-sensory interpretation of changes in t_0 is just a conservative

183 interpretation. Of course, this does not invalidate the interpretation of the k parameter, which

184 would remain representative of the rate at which evidence is accumulated at the detection stage

185 across levels of contrast (i.e., gain).

186

187 **Experiment 1**

188

189 **Methods**

190 *Participants*

191 Two of the authors and three naive volunteers (24-42 years old, 3 female) from the University

192 of Barcelona participated in this experiment. All subjects had normal hearing and normal or

193 corrected-to-normal vision by self-report, and gave informed consent prior to participating. The

194 study received the approval of the University of Barcelona ethical committee.

195

196 *Materials*

197 Visual stimuli consisted of vertical Gabor patches (Gaussian window standard deviation=0.25;

198 mean luminance = 23.4 cd/m^2) subtending a visual angle of 2 deg., with spatial frequency (SF)

199 of 0.3, 0.74 or 5.93 c/deg., depending on the experimental condition. They were presented at the

200 centre of a Philips 22'' CRT monitor (Brilliance 202P4) placed at 1 m viewing distance. These

201 Gabor patches were presented at varying levels of contrast in six steps ranging from 0.05 to

202 0.822 (Michelson contrast) on a logarithmic scale. Auditory stimuli were 340 ms broadband

203 white noise bursts (58 dB (A) SPL; 10 ms on/off ramps) delivered from two loudspeaker cones

204   located at either side of the monitor. Auditory and visual stimulus timing accuracy and precision

205   was calibrated using a Black Box Toolkit oscilloscope (Black Box Toolkit Ltd).

206

207   *Procedure*

208   Prior to each session, participants adapted to the room lighting conditions and to the screen

209   luminance for 5 minutes, and then were presented with a training block of 20 trials. Participants

210   were asked to press a joystick button as soon as they detected a visual stimulus (Gabor patch) at

211   the centre of the monitor. Each block contained 180 trials including 20% catch-trials, where no

212   visual stimulus was present. Catch-trials encouraged participants to selectively attend visually,

213   and discourage responses based on the auditory component. Half of the total trials (including

214   half of the catch-trials) contained an auditory stimulus. All types of trials (sound vs. no sound,

215   with vs. without visual stimulus, and the contrast level of the visual stimulus) were randomly

216   intermixed throughout each block while the SF was blocked. The beginning of each trial was

217   signalled by a central fixation cross (which remained throughout the trial) and, after a variable

218   delay (randomly chosen from 500 ms to 1500 ms), the target stimulus was presented for 340 ms

219   (Fig 1). Following a response deadline of 1500 ms, the screen blanked and a 2000 ms delay led

220   to the beginning of the next trial. Subjects typically completed four blocks of 180 trials of one

221   spatial frequency per session, with short resting breaks between blocks. Each session was run on

222   a different day, and the experiment lasted 6 sessions (2 sessions for each spatial frequency). The

223   order of the sessions was randomized to prevent learning effects. Overall, each participant

224   performed an average of 80 (a minimum of 60) valid trials per condition (with and without

225   sound) and level of contrast.

226

227   == **FIG 1** ==

228

229 **Data analysis**

230 RT to visual stimuli is determined by contrast level and spatial frequency (Breitmeyer, 1975;

231 Harwerth & Levi, 1978). The decrease in RTs to increments in contrast is well accounted for by

232 the Piéron function in Equation (1). As discussed in the introduction, the slope of this function

233 (determined by the parameters k and α in Equation (1), captures variations of RT caused by

234 variations in stimulus properties. In many cases it is possible to express the slope using a single

235 parameter by assuming an α exponent of -1, which is a particular case of the general function

236 often applied in visual psychophysics (see Equation (2); Murray & Plainis, 2003; Plainis &

237 Murray, 2000; see the fitting of the α parameter section below)

238 $$RT = k \cdot \frac{1}{C} + t_0 \qquad (2)$$

239 In this case, the slope k can be directly interpreted as the gain (rate) at which the system

240 accumulates sensory evidence about the presence of a stimulus (k is inversely proportional to

241 sensitivity, so that sensitivity $=1/k$). Interestingly, within this framework the parameter k has

242 been shown to be modulated by the different sensitivity of the P- and the M-pathways to

243 different levels of contrast and spatial frequencies (Murray & Plainis, 2003; Plainis & Murray,

244 2000). Specifically, low values of k (indicating higher sensitivity) have been reported for low

245 spatial frequencies as compared to higher spatial frequencies (for a more detailed discussion of

246 the paradigm see Murray & Plainis, 2003; Plainis & Murray, 2000).

247 In the present study, we extrapolate this framework to the multisensory domain in order to

248 investigate how sound modulates the processing of a visual (Gabor) stimulus. We aim to reveal

249 potential frequency-specific modulations by looking at RTs as a function of stimulus contrast.

250 That is, we contend that efficient visual processing will lead to shallower slopes (lower k,

251 indicating higher sensitivity). If present, this reduction in RTs is expected to be stronger at low

252 contrast than at high contrast visual stimuli, because there is more room for improvement in

253 sensory processing in that case.

254 In terms of cross-modal enhancement, it is predicted that if audio-visual interactions produce

255 changes in the rate of accumulation of sensory evidence, this will be reflected as reductions in k

256 (i.e., the slope) [Footnote 1] with respect to visual-only trials. Moreover, we hypothesize that

257 this slope modulation should reveal some specificity for low spatial frequencies in agreement

258 with the results of previous investigations (Jaekl & Soto-Faraco, 2010; Leo et al., 2008). Other

259 benefits due to sound that are not strictly modulating sensitivity are also expected, but in this

260 case the modulation should concern the constant t_0. Abrupt sounds are indeed known to speed

261 up motor responses independently of the stimulus features (Nickerson, 1973; Valls-Solé et al.,

262 1995),via reductions in temporal uncertainty (Lippert et al., 2007) or altering the decision

263 making process (Doyle & Snowden, 2001; Odgaard et al., 2003). These mechanisms are

264 assumed to be equally effective across different levels of contrast. Hence, we contend that

265 decreases in RTs caused by these higher order decisional processes not originated in the

266 detection phase will thus be captured by the parameter t_0, in agreement with Carpenter (2004)

267 and Pins & Bonnet (1996). As mentioned above, the t_0 term of Equation (2) reflects the intercept

268 of the function. Pins & Bonnet (1996; see also Carpenter 2004) already tested how the slope and

269 the asymptote of the Piéron function vary under task manipulations which involved, for

270 example, stimulus uncertainty, to demonstrate that high order effects produce changes in the

271 asymptotic parameter t_0 (whilst the slope remained unaltered).

272

273 **Fitting α parameter**

274 == **FIG 2** ==

275 As it has been pointed out before, the oft-used simplification of the original Piéron function

276 shown in Equation (2) allows one to relate RTs to the reciprocal of contrast (C^{-1}) with slope k

277 and intercept t_0. This simplification involves the assumption that the parameter α in the original

278 Piéron functions is -1. We thought it important to test this initial assumption, and did so in two

279 ways. First, we fitted the Piéron function in its original form to our data, leaving α as a free

280 parameter and calculated its variability with a nonlinear method (Marquardt–Levenberg

281 algorithm). The result was in general compatible with an α = -1 (with only four exceptions from

282 the thirty possible combinations, see Fig. 2). To secure further the viability of this assumption

283 and considering that the non-linear fitting of the exponent is not fully conclusive could depend

284 on the algorithm used, we also proceeded with an approach based on linear fitting. We tested for

285 the linearity of the data points when plotted against the reciprocal of the contrast (therefore, this

286 time keeping α fixed to -1). The linearity of the distributions was assessed using a likelihood

287 ratio test (per each subject and condition; $p<0.05$ in all cases). Since both approaches led to

288 converging evidence, it was safe to accept an α = -1.

289

290 **Results and discussion**

291 == **FIG 3** ==

292 Mean RTs were calculated after filtering out individual data points falling two standard

293 deviations away from the mean for each distribution of RTs per condition and participant

294 (<4.5% of the original data were rejected). Simple visual inspection of the averaged data in Fig

295 3 shows how responses in audio-visual trials were systematically faster and less variable than in

296 visual-only ones (see also, Table 1).

297 We adjusted the RT data as a linear function of the reciprocal of contrast (Equation (2)) for each

298 different visual frequency and modality combination (Fig.3) in order to estimate the

299 corresponding slopes (i.e., the gain parameter k) and t_0. Below, we report statistically significant

300 results only, and unless otherwise noted, all other effects and interactions were non-significant.

301 First, we ran simple paired t-tests on the k parameter across spatial frequencies within the visual

302 modality (threshold for significance set to $p=.033$ for multiple comparisons adjustment,

303 Benjamini & Hochberg, 1995). This was done for comparison with previous visual studies. As

304 expected, maximal sensitivity (low k parameter) was seen for SF=0.74 c/deg compared to

305 SF=0.3 c/deg ($t(8)= -3.344$, $p=.01$) and SF=5.93 c/deg ($t(8)=9.694$, $p<.001$). These data are thus

306 consistent with Plainis & Murray (2000) and in agreement with the typical contrast sensitivity

307 curves for spatial frequencies in visual literature.

308   In order to test the effect of sound on visual sensitivity, we then ran an ANOVA on the

309 gain parameter k with modality of presentation (visual vs. audio-visual) and visual spatial

310 frequency (0.3, 0.74 and 5.93 c/deg) as independent variables. This analysis revealed a main

311 effect of modality, confirming that responses to audio-visual trials were faster than to visual

312 ones ($F_{1,4}=14.79$, p= .018), and a significant interaction between modality and spatial frequency

313 ($F_{2,8}= 4.5925$, p=.046). Following up on this interaction, we ran paired t tests between the

314 audio-visual and visual conditions at each level of SF. These revealed a significant increases in

315 sensitivity (i.e., lower k parameter) in the audio-visual condition with respect to the visual-only

316 condition for SF=0.3 c/deg ($t(8)= -9.352$,p<.001) and 0.74 c/deg ($t(8)=-5.548$, p<.001), but not

317 for 5.93 c/deg stimuli (Fig 4.a). This reduction in slope is compatible with audio-visual

318 interaction at detection stages; namely, the system accumulates visual evidence at a faster rate

319 when the sound is present, thereby increasing sensitivity. Moreover, our data suggest that this

320 enhancement results stronger at low contrast values.

321 == **FIG 4** ==

322

323 We ran another ANOVA on the intercepts (t_0) extracted from the fits to Equation (2), using the

324 same independent variables as before (SF and modality). This analysis revealed a significant

325 main effect of presentation modality caused by a shift in t_0 indicating an overall RT decrement

326 (a speed up of 35 ms on average) in the audio-visual condition with respect to the visual-only

327 condition ($F_{1,4}=97.24$, p<.001). This enhancement was equivalent for all spatial frequencies

328 tested (see Fig 4.b), as revealed by the lack of significant interaction between spatial frequency

329 and modality ($F_{1,4}=.418$, p= .672). The main effect of spatial frequency was also insignificant

330 ($F_{1,4}=.770$, p= .494).

331

332   == **FIG 5** ==

333   One possible concern with the modulation in slope arising from these analyses is the potential

334   impact of erroneous responses triggered by the sound, which are only possible in the audio-

335   visual condition. To address this, we analysed the proportion of false alarms (FAs,

336   corresponding to erroneous responses to catch-trials). The percentage of FAs in the no-sound

337   catch-trials was very close to 0% for all the participants and conditions (no significant

338   differences). However, participants committed an average of 17% FAs in the sound-present

339   catch-trials (no differences between SFs). In any case, the linear regression between parameter k

340   and FA rate (Fig 5) did not reveal any significant or near-significant positive or negative

341   correlation, suggesting that these FAs to sound were not causing variations in the slope (k

342   parameter) of the functions. On the other hand, the percentage of misses (failing to respond to a

343   target when present) was very low overall (0.016% on average) and statistically equivalent

344   across SFs and presentation modality (as expected from the supra-threshold nature of the

345   stimuli). Please note that using a signal detection approach (i.e., calculating parameter d') is not

346   applicable here, given that the paradigm attempts to address changes in visual sensitivity from

347   differences in processing time (RT) and not from accuracy measurements (i.e., FAs are mostly

348   due motor anticipations, not to incorrect detections).

349

350   The results described above are clear in that RTs to visual events are speeded up by sounds, and

351   that this enhancement comprises both a gain in stimulus-driven information towards detection,

352   as well as a more generic, across the board facilitation that can be attributed to higher-order

353   mechanisms. The former source of gain is sensitive to stimulus parameters such as spatial

354   frequency, and aligns well with previous results suggesting a stronger engagement of the

355   mango-cellular visual pathway in multisensory integration. The latter source of enhancement,

356   on the other hand, captures well the interpretations of some previous studies showing the role of

357   accessory sounds in reducing uncertainty (Lippert et al., 2007)and speeding up motor pathways

358   (Valls-Solé et al., 1995).

359   Previous literature has often characterized the mechanism underlying multisensory

360   enhancement by testing if the data from the cross-modal condition surpasses the prediction of a

361   probability summation model based on the unisensory data (e.g. Arnold, Tear, Schindel, &

362   Roseboom, 2010; Meyer, Wuerger, Röhrbein, & Zetzsche, 2005). Despite some recent

363   fundamental criticism depending on the involved paradigm (Otto & Mamassian, 2012; Pannunzi

364   et al., 2012), this approach has been used frequently and in a variety of contexts to evaluate the

365   nature of multisensory enhancement. Probability summation assumes that the speed up in audio-

366   visual trials arises from a statistical advantage when responses can be based on two independent

367   inputs that race toward reaching a threshold (Green, 1958; Quick, 1974; Treisman, 1998;

368   Wuerger, Hofbauer, & Meyer, 2003). Given that, in the present experiment, participants were

369   asked to respond exclusively to the visual stimulus and not to the sounds, probability

370   summation would not appear adequate to model the current task. We nevertheless consider it

371   relevant to bridge our present results with this common test for integration, as it has been quite

372   influential in the literature so far. We decided to do so in two ways: First, we used the FA data

373   from Experiment 1 to simulate auditory RTs in order to be able to calculate the outcome of a

374   probability summation model (see analysis below). Second, we reproduced the stimulus

375   conditions of Experiment 1 in a new test where participants were asked to respond to both

376   modalities, thus effectively adopting the standard paradigm of probability summation

377   (Experiment 2, reported in the next section).

378

379   Assessing probability summation

380   Judging from the FAs to sound-only trials (17% on average; Fig 5), a percentage of the correct

381   responses in the audio-visual RT distributions could have been triggered by uncontrolled

382   responses toward the auditory stimulus. We decided to evaluate the possible impact of

383 probability summation in our results assuming that a percentage of the RTs to the audio-visual

384 trials originated from anticipations to sound. First, we obtained the averages and standard

385 deviations (SD) of the individual RTs to the visual-only stimuli and those of FAs in catch-trials

386 with sound (Table 1; note that false alarm RT distributions were filtered in the same manner as

387 the correct RTs in Experiment 1).

388 == **TABLE 1** ==

389 In order to compute the prediction of the probabilistic summation model, we applied a modified

390 version of the LATER model. LATER assumes that sensory signals start a race to threshold in

391 an ergodic rate, with the winning signal triggering the response (Carpenter & Williams, 1995;

392 Logan, Cowan, & Davis, 1984). We adapted the model to the current experimental design by

393 assuming that only in a percentage of the audio-visual trials, intrusions from FAs would win the

394 race to threshold against the visual stimulus (this proportion was adjusted for each spatial

395 frequency and observer, according to the empirical data from individual auditory catch-trial

396 conditions). Data of the probabilistic summation model were simulated by randomly selecting

397 RTs from the visual and the FA distributions (9000 simulations per point). Sensitivity (1/k) was

398 computed from the newly modelled chronometric distributions (represented, in Fig.3, by the

399 continuous thin black line). The empirical audio-visual k parameter was significantly lower

400 (higher sensitivity) than the theoretical k calculated from the probability summation model for

401 the 0.3 c/deg and 0.74 c/deg spatial frequencies ($t(8)$=-14.91 p<.0005 and $t(8)$= -3.484 p=.025

402 respectively), but there were no significant differences for the 5.93 c/deg condition. Thus,

403 according to this analysis, while a probability summation model based on intrusive false alarms

404 to sound may eventually account for the RT shift in the 5.93 c/deg spatial frequency, it cannot

405 account for the gains produced in the remaining, lower SFs.

406 The analysis above helps reassure, in a new fashion, that audio-visual responses in this

407 experiment were genuinely benefited from interactions at the stage of detection, not (only)

408 mediated by intrusive FAs to sound. However, there is an established literature suggesting that

409   error and consequent physiological error negativity (which reflects the manifestation of an error

410   detection system that checks actual behavior against an internal goal) have a marked effect on

411   physiological processes, even when participants are not aware of their mistakes (e.g.,

412   Nieuwenhuis, Ridderinkhof, Blom, Band, & Kok, 2001). For this reason, an RT to FAs might

413   not be completely suitable for a conclusive test/reject of the probability summation

414   hypothesis. Furthermore, the size of the FA reaction times distributions was based on 25 trials

415   on average per participant and SF, thus making for a low reliability of the corresponding mean

416   and SDs. We therefore decided, in Experiment 2, to apply the race-model inequality test in a

417   more traditional form (Miller, 1982; Raab, 1962), in order to validate our paradigm and the

418   results of Experiment 1 against a host of previous literature.

419

420   **Experiment 2**

421   In this experiment we applied the race-model inequality test (RMI; Cappe et al., 2009; Colonius

422   & Arndt, 2001; Laurienti et al., 2004; Leo et al., 2008; Raab, 1962) by directly collecting

423   responses to unisensory visual and auditory trials and estimating Miller's bound (Ulrich, Miller,

424   & Schröter, 2007) for our stimuli. As noted above, recent studies (e.g., Otto & Mamassian,

425   2012) have argued that the Miller's bound might not be adequate to reveal sensory interactions

426   beyond purely statistical advantage. It is not within the scope of this study to claim otherwise,

427   but this paradigm should suffice to back the results from Experiment 1 by comparison with a

428   known and widely used approach to multisensory interactions.

429   According to past literature, Miller's bound is the maximal reaction time gain that can be

430   explained by purely statistical facilitation when decisions are based on two independent sources

431   of information. This is represented mathematically by Equation (3), where the probability of

432   detecting an audio-visual signal (AV) is the sum of the individual probabilities of detecting the

433   visual (V) and the auditory (A) signals.

434    $p(T<t|AV) \leq p(T<t|V) + p(T<t|A)$ (3),

435   If the results from Experiment 1 were to be confirmed in this paradigm, we would expect to find

436   the largest specific RT enhancement for the audio-visual RTs (relatively to the unimodal ones)

437   at the lowest spatial frequency conditions.

438

439   **Methods**

440   *Participants*

441   Seven new volunteers naïve to the purpose of the experiment (23-29 years old, 4 female) were

442   recruited amongst students from the University of Barcelona. All had normal hearing and

443   normal or corrected-to-normal vision by self-report, and gave informed consent prior to

444   participation.

445

446   *Materials and procedure*

447   The experimental apparatus, time line and stimuli were identical to the one used in the

448   Experiment 1, except that we simplified the design by using only two SFs (0.3 and 5.93 c/deg,

449   lowest and highest in Experiment 1, respectively), and two levels of contrast (0.05 and 0.822

450   Michelson contrast units, the lowest and the highest in Experiment 1).

451   Participants were asked to respond as fast as possible to any stimulus (visual, auditory or audio-

452   visual, indistinctly). Each participant ran 600 trials divided in 3 blocks of 200 trials (45 minutes

453   in total) where all SFs and contrasts appeared randomly intermixed. Prior to the experimental

454   session, participants dark-adapted for 5 min and performed 40 warm-up trials. In total every

455   subject performed 45 valid trials for each visual, audio-visual, SF and contrast condition, and

456   180 auditory trials (in order to have the same number of visual, auditory and audio-visual trials

457   per block). Each block also included catch-trials (amounting to 10% of the total) where no

458   stimulus was presented, thus participants had to refrain from responding.

459

Results and discussion

461 Just like in Experiment 1, we filtered RTs below and above 2SD from the mean in each

462 condition and participant (leading to the rejection of <6 % of the data), and obtained the

463 empirical cumulative distribution functions. Participants responded to less than 4 % of the

464 catch-trials, therefore erroneous anticipations can be ruled out as a having a major contribution

465 to the pattern of results. A 2 x 2 x 3 (Contrast x SF x Modality) repeated measures ANOVA was

466 conducted on the RTs. This analysis revealed a significant main effect of SF ($F_{1,6} = 9.15$, $p =$

467 .0023), contrast level ($F_{1,6} = 208.87$, $p < .0001$), and modality ($F_{2,12} = 168.62$, $p < .0001$). All

468 the interactions were also significant (all, $p < .001$, except SF by Contrast which was marginally

469 significant, $F_{1,6} = 0.439$, $p = .0532$).

470 == **FIG 6** ==

471

472 As can be seen in Fig 6, RTs were faster for high contrast and for low SF visual stimuli, in

473 consonance with previous literature (Breitmeyer, 1975; Legge, 1978) and with the results of

474 Experiment 1. Moreover, RTs were also generally faster for the auditory-only stimuli than for

475 the visual-only ones. According to our expectations, responses to audio-visual stimuli were

476 significantly faster than the visual-only and the auditory-only ones in all the conditions with the

477 exception of the low contrast 5.93 c/deg SF condition, where audio-alone RTs were as fast as

478 audio-visual ones. Interestingly the audio-visual absolute enhancement was larger at the high

479 contrast conditions, contradicting what the inverse effectiveness law would predict. However, it

480 is important to point out that considering direct RTs to make the comparison might be

481 misleading. For instance, if we consider the cumulative RT distribution functions presented at

482 the bottom part of Fig 6, it is also evident that the probabilistic overlap between visual and

483 auditory unisensory conditions is also larger at high contrast conditions. This means that just by

484 means of probability summation, overall audio-visual response speed up should be larger, but

485   mostly because of statistical reasons. Thus, we calculated the relative index of enhancement for

486   each type of visual stimulus by obtaining the Miller's bound (for a specific description of the

487   method, see Ulrich et al., 2007). This is normally considered the maximal possible speed up in

488   reaction time obtained by simple probabilistic advantage when two signals are independent.

489   == **FIG 7** ==

490   Data were binned in 10 quantiles (10 % increments), and at each quantile (see Fig 7) we tested

491   the alternative hypothesis that audio-visual RTs were faster than the prediction of the Miller's

492   bound. The shaded areas in Fig 7 indicate the quantiles where the race-model was violated (i.e.

493   the audio-visual RT was significantly faster). The 0.3 c/deg SF manifested the highest

494   proportion of such violations, with significant differences in 6 quantiles in the low contrast

495   condition and 4 in the high contrast condition. The analysis of the 5.93 c/deg SF did not reflect

496   any enhancement for the low contrast condition and only 3 significant violations in the high

497   contrast condition. In all cases, all significant violations of the race model encompassed

498   adjacent time bins.

499

500   The results of Experiment 2 shed new light on the interpretation of the results of Experiment 1.

501   In particular, the reduction in the slope at low SFs in Experiment 1, according to the analysis of

502   the Piéron functions, seems to correspond to the specific sensitivity enhancement found for the

503   low SF in Experiment 2 accomplishing a convergence criterion.

504   From previous literature, changes in the slope of the Piéron function are highly determined by

505   the low contrast values (Murray & Plainis, 2003). The inverse effectiveness principle, often

506   proposed in multisensory integration, might be a logical candidate to explain our results, given

507   that the Miller's bound violation in Experiment 2 was greater (for the SF = 0.3 c/deg) at the low

508   contrast condition.[Footnote 2]

509   This conclusion is in principle independent of the potential problems for the claim that the race-

510   model may not be a good test of probability summation as recently pointed out by Otto &

511   Mamassian (2012). In short, we use here the race model as a relative index to quantify the

512   benefits produced across SFs in audio-visual trials. The main conclusion to be extracted from

513   this experiment is that our results support the paradigm applied in Experiment 1, showing that

514   multisensory interactions are larger when visual stimuli correspond to low SFs and contrast

515   levels.

516

517   **General discussion**

518   The brain combines incoming sensory information from various modalities in order to optimize

519   the organism's interaction with the environment. Various animal and human studies have

520   highlighted the putative neural mechanisms supporting these enhancements (Calvert, Hansen,

521   Iversen, & Brammer, 2001; Meredith & Stein, 1986; Schroeder & Foxe, 2005; Smiley &

522   Falchier, 2009; Watkins, Shams, Tanaka, Haynes, & Rees, 2006). However, one important, and

523   yet unresolved question refers to the specific case of the putative visual enhancement produced

524   by concurrent sounds. As highlighted in the introduction, previous reaction time studies

525   addressing audio-visual integration (Cappe et al., 2009; Colonius & Arndt, 2001; Laurienti et

526   al., 2004; Leo et al., 2008; Miller, 1982) have reported improvements in detection which cannot

527   be accounted solely as a result of a statistical advantage. Nevertheless, while these studies have

528   consistently revealed the existence of cross-modal enhancements, they cannot dissociate and

529   quantify the different sources contributing to the cross-modal benefit. One important constrain

530   related to the classical approach based on the redundant signal effect (RSE) is that responses in

531   this task are given to any stimulus indiscriminately (i.e., visual or auditory), and consequently

532   the RSE paradigm fails to describe with precision detection changes over one specific sensory

533   modality. Hence, we decided to apply a different paradigm based on the analysis of Piéron

534   functions on visual response latencies, in order to study specific changes in visual processing

535   resulting from stimulus-driven interactions with a concurrent sound. Our results provide

536   evidence for the coexistence of two functionally distinct mechanisms that support the

537   enhancement of vision by sound: First, presenting a sound synchronously with a visual stimulus

538   can decrease RTs as a result of processes not necessarily supported by sensory interactions. This

539   source of enhancement may be a consequence of the superior temporal resolution of the

540   auditory system (Lippert et al., 2007), in addition to (or in combination with) the potential

541   alerting effects of sounds leading to increases in overt-motor activation, shortening response

542   latency (startle effect; Valls-Solé et al., 1995). In line with the non-sensory specific nature of

543   these mechanisms, their contribution to RT speed up was equivalent across all intensities of

544   visual contrast and spatial frequencies tested. Such non-sensory specific enhancement was

545   clearly reflected in the reduction in the asymptotic term of the Pièron functions.

546   Most importantly however, our results also provide strong evidence another source of

547   enhancement, based on sensory interactions between visual and auditory inputs at the detection

548   stage. This interaction was clearly dependent on stimulus variations and, as we content, it would

549   be produced by a modulation of the rate at which visual sensory evidence is accrued (though,

550   alternatively, could also be based on a reduction of the sensory threshold needed to elicit a

551   response, see below). In any case, this effect was reflected in the change of the slope of the

552   Pièron function (gain parameter k) which is interpreted as an increase in the sensitivity of the

553   visual system. From a computational point of view, this advantage in detection can be explained

554   in at least two alternative ways. One the one hand, visual and auditory signals may be integrated

555   following a linear summation model which assumes convergence of the sensory inputs from the

556   two modalities into a single physiological or psychophysical channel before a decision

557   regarding the presence of the stimulus(Meyer et al., 2005; López-Moliner & Soto-Faraco 2007;

558   Morrone, Burr, & Vaina, 1995). A second possibility is that the sound reduces the sensory

559   threshold (for example, by increasing visual cortex excitability, Romei, Murray, Cappe, & Thut,

560   2009) to trigger the response. (Please note that although the main conclusions of this paper are

561   not affected by which of these two particular alternatives is true, we attempted to tell apart these

562   two models of enhancement at the detection level without conclusive results, see, Appendix).

563   What remains clear so far is that, regardless of the particular mechanism, an important part of

564   the auditory-induced enhancement of vision reported here owes to changes at the detection

565   stages of stimulus processing.

566

567   In order to validate the original paradigm applied in Experiment 1, and bridge our results with

568   extant multisensory literature, in Experiment 2 we followed the classical RSE approach. We

569   showed that sensory interactions conformed well to the pattern of findings in Experiment 1.

570   Interestingly, and in stark contrast with the non-specific component of the audio-visual

571   enhancement, both experiments converged in showing that improvements at the detection level

572   of processing were selective for low spatial frequencies. Further analyses of the slopes in

573   Experiment 1 allowed us to address this audio-visual enhancement across the different levels of

574   contrast. As demonstrated by Murray & Plainis (2003), the low intensity contrast region

575   determines the slope, hence the marked reduction of the slope in the sound-present condition

576   could be construed as a behavioural analogue of the inverse effectiveness rule, often cited in

577   multisensory integration literature (Meredith & Stein, 1983; Wallace, Meredith, & Stein, 1998).

578   Namely, the lower the contrast intensity, the higher the benefit produced by the sound (Fig 3;

579   Fig 7), and in consequence, the audio-visual gain. Congruently, Experiment 2 also yielded

580   larger sensory interactions for the critical low spatial frequency condition at the low contrast

581   stimuli.

582   One might perhaps think of the reductions in the RT in the audio-visual conditions could be

583   accounted in terms of response preparation (Nickerson, 1973). According to this perspective,

584   when one target and one accessory stimulus are presented in temporal synchrony, the longer is

585   the processing time for the target, the more time has the accessory stimulus to work as a

586   warning signal and prepare the system for responding to that target, leading to advantages in

587   RTs. However, our present results in Experiments 1 (Table 1) and 2 (Fig 6) showed that while

588   visual latencies are slower for the high SF, the multisensory benefits in reaction time are more

589   likely to occur at the low SFs, thus effectively rejecting response preparation as a viable account

590   for the present pattern of results.

591   Most previous studies conducted in multisensory research have not directly addressed

592   possible differences in audio-visual integration as a function of the well-known functional

593   division in visual processing between Magno- and Parvo-cellular pathways. In the present

594   study, the component of enhancement linked to the detection-stage revealed here, was

595   constrained to low spatial frequencies and thus the present findings bear on the potentially

596   important role of the low frequency-tuned magno-cellular visual pathway in multisensory

597   detection (see Jaekl & Soto-Faraco, 2010, for related findings). This fits well with prior

598   knowledge concerning the physiology underlying low-level audio-visual interaction in the M-

599   pathway and its functionality in terms of mediating detection of intensity changes. Moreover,

600   from a neurophysiological perspective this dominance of the M-pathway in multisensory

601   interaction for detection is indirectly supported by two additional lines of evidence. First, the

602   superior colliculus (SC) which has been frequently considered as determinant for multisensory

603   enhancement (Leo et al., 2008; Meredith & Stein, 1986, 1996), receives connections from the

604   magnocellular layers of the lateral geniculate nucleus (LGN) of the Thalamus via cortical area

605   V1 (Berson, 1988) whereas evidence for connections from the P-cellular pathway is scant

606   (Livingstone & Hubel, 1988; Merigan & Maunsell, 1993). Second, anatomical inputs into V1

607   from the auditory cortex occur primarily in the representation of the peripheral visual field

608   (Falchier, Clavagnier, Barone, & Kennedy, 2002), a region of space which is mainly

609   represented by the M-pathway (Silveira & Perry, 1991). Third, another relevant feature of the

610   M-pathway is its specific sensitivity to transient stimulation. Some recent studies have

611   suggested that pure transient information is integrated (Andersen & Mamassian, 2008) and it

612   could be determinant for multisensory associations (Van der Burg, Cass, Olivers, Theeuwes, &

613   Alais, 2010; Werner & Noppeney, 2011).

614 Although our results support the important contribution of the M-pathway to multisensory

615 interactions in detection, it is important to take into account that probably the features of the

616 experiment prioritize fast detection responses and, in consequence, the processing under this

617 particular visual pathway, which has shorter latencies. Therefore, our results do not necessarily

618 negate a possible role of the P-pathway in driving multisensory interactions in other kinds of

619 tasks more specifically tapping at parvocellular processing (i.e., cross-modal object recognition;

620 Giard & Peronnet, 1999; Molholm et al., 2002, Jaekl and Harris, 2009). Nevertheless, more

621 research specifically addressing this dissociation should be considered in future research.

622

623 Conclusion

624   The present psychophysical analysis of RTs allowed us to shed light on the contribution

625 of different sources of interaction leading to sound-induced enhancement of vision, one of the

626 most oft-cited but still controversial cases of cross-modal gain. First, the results of this study

627 provide clear evidence showing that concurrent auditory signals result in an increase in

628 sensitivity to visual stimuli, at a detection-specific level. This increase in sensitivity is largest at

629 low spatial frequency and low contrast values of the visual stimulus. Second, the present

630 paradigm allowed us to isolate and quantify the reduction in reaction times to audio-visual

631 stimuli as a consequence of more general processes that are separable from the reduction

632 resulting from interactions at the detection-stage. Although previous literature had hinted at the

633 potential mixed contribution of detection-level and the decision-level processes in multisensory

634 enhancement, clear-cut evidence for the contribution of each type of source within the same

635 paradigm was scant.

636

637

638

639   *Appendix*

640   *Linear summation hypothesis*

641   Considering an explanation based on the linear summation hypothesis, the reduction in slope

642   (gain parameter k) observed for the lower SFs is described as the stimulus energy reaching the

643   detection threshold θ (see appendix) at a faster rate, due to the concomitant presence of a sound

644   that is linearly added to the visual signal. The slope ($k = \theta/c$) then consists of a threshold θ and

645   the rate c that determines the accumulation to this threshold when the channel is stimulated by a

646   given amount of energy E. According to this description an accurate estimate of k will depend

647   on how well the energy values (E) of the stimuli used to fit the model are chosen. In our main

648   analyses of Experiment 1 we modelled the RTs as a function of vision (visual contrast) alone, as

649   this was the modality the participants were asked to respond to. Hence, here we evaluate the

650   possibility that the race to the threshold in the audio-visual condition was also determined by

651   the auditory signals, at least for the lower spatial frequencies.

652   To determine if the present data are consistent with a linear summation model, we estimated the

653   specific contribution of the auditory stimulus to the different audio-visual conditions.

654   Specifically, we fitted the Equation (A.8; see the mathematical derivation of the linear

655   summation model) and included an additional term (Ea) as the only free parameter to account

656   for the auditory stimulus contribution.

657   We set ($k = \theta/c$) for each SF to their respective values estimated from the visual alone condition

658   under the assumption that the threshold and the gain of the channel do not depend on the nature

659   of the stimulation. That is, we consider that the threshold θ used to trigger the response is the

660   same for both stimulation conditions (visual only and audio-visual). We set t_0 to the values

661   obtained with the Piéron equation in the audio-visual condition, as this reflects the non-specific

662   enhancement on RT due to sound. The linear model with a single parameter fitted the data

663   perfectly. As dictated by the physical stimuli shown to the participants, Ea denotes a constant

664   stimulation (i.e., the sound was always the same). However, the estimated contribution of this

665   parameter to the final RT is inversely proportional to the contrast for each SF and importantly,

666   inversely proportional to SF, considering that the estimated values for *Ea* were 0.053 and 0.035

667   for the 0.3 and 0.74 c/deg SFs, while for 5.93 was nearly 0. In other words, the contribution of

668   the sound to visual detection is greatest at low spatial frequencies. Importantly this selective

669   contribution accurately describes the data while keeping the slope (channel gain) at the same

670   value of that obtained in the visual condition.

671

672   *Mathematical derivation of the linear summation model*

673   The means of visual reaction times measured as a function of stimulus contrast are usually well
674   described by the Piéron function:

675

676    $RT = kC^{-\alpha} + t_0$ (A.1)

677

678   Let us assume that the response mechanism is based on an accumulator model. The stimulus-
679   dependent latency component ($t = kC^{-\alpha}$) is then a function of the time it takes for the mechanism
680   to accumulate evidence (E, contrast in our case) until it reaches a threshold (θ):

681

682    $t = \dfrac{\theta}{cE}$ (A.2)

683

684   Where *cE* denotes the rate at which the system accumulates the evidence. A more general form
685   of (2) underlies decisional response mechanisms like the LATER model (Carpenter & Williams,
686   1995). Reaction time can then be expressed as:

687

688    $RT = \dfrac{\theta}{cE} + t_0$ (A.3)

689

690   Combining (2) and (3) and taking the logarithms we obtain:

691

692 $$\log t = \log(RT - t_0) = \log\frac{\theta}{c} - \log E \qquad \text{(A.4)}$$

693

694 A more general form of (A.4) is easily obtained with adding the extra parameter $\alpha = -1$ and
695 letting $k = \theta/c$:

696

697 $$\log(RT - t_0) = \log(t) = \log(k) - \alpha \log E \qquad \text{(A.5)}$$

698

699 By taking the exponential at both sides of equation (A.5) we obtain the Piéron function (A.1):

700

701 $$RT = k \cdot E^{-\alpha} + t_0 \qquad \text{(A.6)}$$

702

703 Therefore the larger the rate at which the sensory evidence accumulates the smaller the
704 parameter k.

705

706 Any specific sensory effect must then affect the stimulus-dependent part of equation (A.1):

707 $$t = \frac{\theta}{c} E^{-1} = \frac{\theta}{cE} \text{(A.7)}$$

708

709 If a visual (E) and an auditory signal (Ea, constant in our case) were integrated within a single
710 channel during the race to the threshold θ without changing the rate c in a linear summation
711 fashion:

712 $$t = \frac{\theta}{c(E + Ea)} = \frac{\theta}{c}(E + Ea)^{-1} \qquad \text{(A.8)}$$

713

714 The true sensory benefit could be captured in the audio-visual condition by a reaction time
715 model that takes into account an auditory signal in the input with a slope $k = \theta/c$ not different
716 from the visual slope and an additive parameter t_0 that is not different from the estimate in
717 equation (1,6) in which the auditory system is not considered.

718

719   *Visual threshold reduction by sound hypothesis*

720   Alternatively, the reduction of the slope value in the lower spatial frequencies could be caused

721   by an increase in the excitably of the system by the sound, leading to a reduction (*Ha*) of the

722   threshold θ in the audio-visual conditions.

723   $RT = ((\theta - Ha)/c) + t_0$ (A.9)

724   Therefore, we fitted Equation (A.9) with fixed values of θ and c whose ratio (θ/c) matched the

725   slope observed in the visual condition (*Ha*=0.0). In this way we can obtain an approximation of

726   *Ha* for the different spatial frequencies. The estimated values were 2.07 and 1.977 for 0.3 and

727   0.74 c/deg respectively and nearly 0.0 for the high frequency condition 5.93 c/deg. This pattern

728   was very similar to the one shown by assuming changes in the rate of sensory evidences

729   accumulation, however, after statistical comparison of the goodness of fit for the linear

730   summation against the varying threshold model, the linear summation model resulted slightly

731   better in all the SFs but only in a significant way for the 0.3 c/deg spatial frequency ($F_{2,3}$ =

732   15.64, p = .023). However, we do not interpret these fittings as being conclusive about what is

733   the particular mechanism best explaining the results, nevertheless we can claim that the changes

734   in the slope are linked to sensory processing independently of the model applied, and are

735   specifically larger for the low SF visual channels than for the high SF channels, highlighting

736   their stimulus dependency.

737

738

739   [footnote 1: Note that this pattern can be loosely related to the rule of inverse effectiveness,

740   often present in multisensory literature (Rowland et al., 2007; Stein, Stanford, Ramachandran,

741   Perrault, & Rowland, 2009). However, the conditions to put this idea to test are not optimal,

742   since in this case the strength of the acoustic event is not manipulated particularly weak.]

743

744 [Footnote 2: Note that there was also a Millers bound violation at the SF = 5.93 c/deg high

745 contrast condition. Although it might seem contradictory with what inverse effectiveness

746 postulate, this result can be accommodated because high contrast stimuli, regardless of

747 frequency, tend to engage the magnocellular visual pathway as well (as pointed out by different

748 authors; Mitov& Totev, 2005; Thomas, Fagerholm, & Bonnet, 1999).]

749

756

757

758 **References**

759 Andersen, T. S., & Mamassian, P. (2008). Audiovisual integration of stimulus transients. *Vision*
760 *research*, *48*(25), 2537–44. doi:10.1016/j.visres.2008.08.018

761 Arnold, D. H., Tear, M., Schindel, R., & Roseboom, W. (2010). Audio-visual speech cue
762 combination. *PloS one*, *5*(4), e10217. doi:10.1371/journal.pone.0010217

763 Benjamini, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and
764 Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B*
765 *(Methodological)*, *57*(1), 289–300. Retrieved from http://www.jstor.org/stable/2346101

766 Berson, D. M. (1988). Retinal and cortical inputs to cat superior colliculus: composition,
767 convergence and laminar specificity. *Progress in brain research*, *75*, 17–26. Retrieved
768 from http://www.ncbi.nlm.nih.gov/pubmed/3055056

Bolognini, N., Frassinetti, F., Serino, A., & Làdavas, E. (2005). "Acoustical vision" of below threshold stimuli: interaction among spatially converging audiovisual inputs. *Experimental brain research. Experimentelle Hirnforschung. Expérimentation cérébrale*, *160*(3), 273–82. doi:10.1007/s00221-004-2005-z

Breitmeyer, B. G. (1975). Simple reaction time as a measure of the temporal response properties of transient and sustained channels. *Vision research*, *15*(12), 1411–2. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/1210028

Calvert, G. a, Hansen, P. C., Iversen, S. D., & Brammer, M. J. (2001). Detection of audio-visual integration sites in humans by application of electrophysiological criteria to the BOLD effect. *NeuroImage*, *14*(2), 427–38. doi:10.1006/nimg.2001.0812

Cappe, C., Thut, G., Romei, V., & Murray, M. M. (2009). Selective integration of auditory-visual looming cues by humans. *Neuropsychologia*, *47*(4), 1045–52. doi:10.1016/j.neuropsychologia.2008.11.003

Carpenter, R. H. S. (2004). Contrast , Probability , and Saccadic Latency : Evidence for Independence of Detection and Decision. *Current*, *14*, 1576–1580. doi:10.1016/j

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

Colonius, H., & Arndt, P. (2001). A two-stage model for visual-auditory interaction in saccadic latencies. *Perception & psychophysics*, *63*(1), 126–47. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11304009

Corneil, B. D., Van Wanrooij, M., Munoz, D. P., & Van Opstal, a J. (2002). Auditory-visual interactions subserving goal-directed saccades in a complex scene. *Journal of neurophysiology*, *88*(1), 438–54. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12091566

Diederich, A., & Colonius, H. (2004). Bimodal and trimodal multisensory enhancement: effects of stimulus onset and intensity on reaction time. *Perception & psychophysics*, *66*(8), 1388–404. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/15813202

Doyle, M. C., & Snowden, R. J. (2001). Identification of visual stimuli is improved by accompanying auditory stimuli: The role of eye movements and sound location. *Perception*, *30*(7), 795–810. doi:10.1068/p3126

Driver, J, & Spence, C. (1998). Cross-modal links in spatial attention. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, *353*(1373), 1319–31. doi:10.1098/rstb.1998.0286

Driver, Jon, & Noesselt, T. (2008). Multisensory interplay reveals crossmodal influences on "sensory-specific" brain regions, neural responses, and judgments. *Neuron*, *57*(1), 11–23. doi:10.1016/j.neuron.2007.12.013

Ernst, M. O., & Bülthoff, H. H. (2004). Merging the senses into a robust percept. *Trends in cognitive sciences*, *8*(4), 162–9. doi:10.1016/j.tics.2004.02.002

807 Falchier, A., Clavagnier, S., Barone, P., & Kennedy, H. (2002). Anatomical Evidence of
808   Multimodal Integration in Primate Striate Cortex. *Area*, *22*(13), 5749–5759.

809 Frassinetti, F., Bolognini, N., & Làdavas, E. (2002). Enhancement of visual perception by
810   crossmodal visuo-auditory interaction. *Experimental brain research. Experimentelle*
811   *Hirnforschung. Expérimentation cérébrale*, *147*(3), 332–43. doi:10.1007/s00221-002-
812   1262-y

813 Giard, M. H., & Peronnet, F. (1999). Auditory-visual integration during multimodal object
814   recognition in humans: a behavioral and electrophysiological study. *Journal of cognitive*
815   *neuroscience*, *11*(5), 473–90. Retrieved from
816   http://www.ncbi.nlm.nih.gov/pubmed/10511637

817 Green, D. M. (1958). Detection of multiple component signals in noise.

818 Harwerth, R. S., & Levi, D. M. (1978). Reaction time as a measure of suprathreshold grating
819   detection. *Vision research*, *18*(11), 1579–86. Retrieved from
820   http://www.ncbi.nlm.nih.gov/pubmed/726307

821 Jaekl, P. M., & Harris, L. R. (2009). Sounds can affect visual perception mediated primarily by
822   the parvocellular pathway. *Visual neuroscience*, *26*(5-6), 477–86.
823   doi:10.1017/S0952523809990289

824 Jaekl, P. M., & Soto-Faraco, S. (2010). Audiovisual contrast enhancement is articulated
825   primarily via the M-pathway. *Brain research*, *1366*, 85–92.
826   doi:10.1016/j.brainres.2010.10.012

827 Kinchla, R. A. (1974). Detecting target elements in multielement arrays : A confusability model.
828   *Perception*, *15*(1).

829 Lakatos, P., Chen, C.-M., O'Connell, M. N., Mills, A., & Schroeder, C. E. (2007). Neuronal
830   oscillations and multisensory interaction in primary auditory cortex. *Neuron*, *53*(2), 279–
831   92. doi:10.1016/j.neuron.2006.12.011

832 Laurienti, P. J., Kraft, R. a, Maldjian, J. a, Burdette, J. H., & Wallace, M. T. (2004). Semantic
833   congruence is a critical factor in multisensory behavioral performance. *Experimental*
834   *brain research. Experimentelle Hirnforschung. Expérimentation cérébrale*, *158*(4), 405–14.
835   doi:10.1007/s00221-004-1913-2

836 Legge, G. E. (1978). Sustained and transient mechanisms in human vision: temporal and spatial
837   properties. *Vision research*, *18*(1), 69–81. Retrieved from
838   http://www.ncbi.nlm.nih.gov/pubmed/664278

839 Leo, F., Bertini, C., di Pellegrino, G., & Làdavas, E. (2008). Multisensory integration for orienting
840   responses in humans requires the activation of the superior colliculus. *Experimental brain*
841   *research. Experimentelle Hirnforschung. Expérimentation cérébrale*, *186*(1), 67–77.
842   doi:10.1007/s00221-007-1204-9

843 Levi, D. M., Harwerth, R. S., & Manny, R. E. (1979). Suprathreshold spatial frequency detection
844   and binocular interaction in strabismic and anisometropic amblyopia. *Investigative*

845 *ophthalmology & visual science*, *18*(7), 714–25. Retrieved from
846 http://www.ncbi.nlm.nih.gov/pubmed/447470

847 Lippert, M., Logothetis, N. K., & Kayser, C. (2007). Improvement of visual contrast detection by
848 a simultaneous sound. *Brain research*, *1173*, 102–9. doi:10.1016/j.brainres.2007.07.050

849 Livingstone, M. S., & Hubel, D. H. (1988). Do the relative mapping densities of the magno- and
850 parvocellular systems vary with eccentricity? *The Journal of neuroscience : the official*
851 *journal of the Society for Neuroscience*, *8*(11), 4334–9. Retrieved from
852 http://www.ncbi.nlm.nih.gov/pubmed/3183725

853 Logan, G. D., Cowan, W. B., & Davis, K. a. (1984). On the ability to inhibit simple and choice
854 reaction time responses: a model and a method. *Journal of experimental psychology.*
855 *Human perception and performance*, *10*(2), 276–91. Retrieved from
856 http://www.ncbi.nlm.nih.gov/pubmed/6232345

857 Manjarrez, E., Mendez, I., Martinez, L., Flores, A., & Mirasso, C. R. (2007). Effects of auditory
858 noise on the psychophysical detection of visual signals: cross-modal stochastic resonance.
859 *Neuroscience letters*, *415*(3), 231–6. doi:10.1016/j.neulet.2007.01.030

860 McDonald, J J, Teder-Sälejärvi, W. A., & Hillyard, S. a. (2000). Involuntary orienting to sound
861 improves visual perception. *Nature*, *407*(6806), 906–8. doi:10.1038/35038085

862 McDonald, J J, Teder-Sälejärvi, W. a, & Ward, L. M. (2001). Multisensory integration and
863 crossmodal attention effects in the human brain. *Science (New York, N.Y.)*, *292*(5523),
864 1791. doi:10.1126/science.292.5523.1791a

865 McDonald, John J, Teder-Sälejärvi, W. a, Di Russo, F., & Hillyard, S. a. (2005). Neural basis of
866 auditory-induced shifts in visual time-order perception. *Nature neuroscience*, *8*(9), 1197–
867 202. doi:10.1038/nn1512

868 Meredith, M. A., & Stein, B. E. (1986). Visual, auditory, and somatosensory convergence on
869 cells in superior colliculus results in multisensory integration. *J Neurophysiol*, *56*(3), 640–
870 662. Retrieved from http://jn.physiology.org/cgi/content/abstract/56/3/640

871 Meredith, M. A., & Stein, B. E. (1996). Spatial determinants of multisensory integration in cat
872 superior colliculus neurons. *J Neurophysiol*, *75*(5), 1843–1857. Retrieved from
873 http://jn.physiology.org/cgi/content/abstract/75/5/1843

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

876 Merigan, W. H., & Maunsell, J. H. (1993). How parallel are the primate visual pathways?
877 *Annual review of neuroscience*, *16*, 369–402. doi:10.1146/annurev.ne.16.030193.002101

878 Meyer, G. F., Wuerger, S. M., Röhrbein, F., & Zetzsche, C. (2005). Low-level integration of
879 auditory and visual motion signals requires spatial co-localisation. *Experimental brain*
880 *research. Experimentelle Hirnforschung. Expérimentation cérébrale*, *166*(3-4), 538–47.
881 doi:10.1007/s00221-005-2394-7

882   Miller, J. (1982). Divided attention: evidence for coactivation with redundant signals. *Cognitive*
883   *psychology*, *14*(2), 247–79. Retrieved from
884   http://www.ncbi.nlm.nih.gov/pubmed/7083803

885   Mitov, D., & Totev, T. (2005). How many pathways determine the speed of grating detection?
886   *Vision research*, *45*(7), 821–5. doi:10.1016/j.visres.2004.09.033

887   Molholm, S., Ritter, W., Murray, M. M., Javitt, D. C., Schroeder, C. E., & Foxe, J. J. (2002).
888   Multisensory auditory – visual interactions during early sensory processing in humans : a
889   high-density electrical mapping study. *Cognitive Brain Research*, *14*, 115–128.

890   Morrell, F. (1972). Visual System's View of Acoustic Space. *Nature*, *238*(5358), 44–46.
891   doi:10.1038/238044a0

892   Morrone, M. C., Burr, D. C., & Vaina, L. M. (1995). Two stages of visual processing for radial
893   and circular motion. *Nature*, *376*(6540), 507–9. doi:10.1038/376507a0

894   Murray, I. J., & Plainis, S. (2003). Contrast coding and magno / parvo segregation revealed in
895   reaction time studies. *Vision Research*, *43*, 2707–2719. doi:10.1016/S0042-
896   6989(03)00408-5

897   Nickerson, R S. (1973). Intersensory facilitation of reaction time: energy summation or
898   preparation enhancement? *Psychological review*, *80*(6), 489–509. Retrieved from
899   http://www.ncbi.nlm.nih.gov/pubmed/4757060

900   Nieuwenhuis, S., Ridderinkhof, K. R., Blom, J., Band, G. P., & Kok, a. (2001). Error-related brain
901   potentials are differentially related to awareness of response errors: evidence from an
902   antisaccade task. *Psychophysiology*, *38*(5), 752–60. Retrieved from
903   http://www.ncbi.nlm.nih.gov/pubmed/11577898

904   Noesselt, T., Tyll, S., Boehler, C. N., Budinger, E., Heinze, H.-J., & Driver, J. (2010). Sound-
905   Induced Enhancement of Low-Intensity Vision: Multisensory Influences on Human
906   Sensory-Specific Cortices and Thalamic Bodies Relate to Perceptual Enhancement of
907   Visual Detection Sensitivity. *Journal of Neuroscience*, *30*(41), 13609–13623.
908   doi:10.1523/JNEUROSCI.4524-09.2010

909   Noesselt, Toemme, Bergmann, D., Hake, M., Heinze, H.-J., & Fendrich, R. (2008). Sound
910   increases the saliency of visual events. *Brain research*, *1220*, 157–63.
911   doi:10.1016/j.brainres.2007.12.060

912   Nozawa, G., Reuter-Lorenz, P. A., & Hughes, H. C. (1994). Parallel and serial processes in the
913   human oculomotor system: bimodal integration and express saccades. *Biological*
914   *Cybernetics*, *72*(1), 19–34. doi:10.1007/BF00206235

915   Odgaard, E. C., Arieh, Y., & Marks, L. E. (2003). Cross-modal enhancement of perceived
916   brightness: sensory interaction versus response bias. *Perception & psychophysics*, *65*(1),
917   123–32. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12699315

918   Otto, T. U., & Mamassian, P. (2012). Noise and Correlations in Parallel Perceptual Decision
919   Making. *Current biology : CB*, 1–6. doi:10.1016/j.cub.2012.05.031

920 Pannunzi, M., Pérez-Bellido, A., Baños, A. P., López-Moliner, J., Deco, G., & Soto-Faraco, S.
921   (2012, January 1). Scrutinizing integrative effects in a multi-stimuli detection task. *Seeing*
922   *and Perceiving*. Ice Virtual Library. doi:10.1163/187847612X647324

923 Pins, D, & Bonnet, C. (2000). The Piéron function in the threshold region. *Perception &*
924   *psychophysics*, *62*(1), 127–36. Retrieved from
925   http://www.ncbi.nlm.nih.gov/pubmed/10703261

926 Pins, Delphine, & Bonnet, C. (1996). On the relation between stimulus intensity and processing
927   time : Pieron ' s law and choice reaction time. *Perception*, *58*(3), 390–400.

928 Piéron, H. (1914). Recherches sur les lois de variation des temps de latence sensorielle en
929   fonction des intensités excitatrices. *L'Année Psychologique*, *20*(5), 17–96.

930 Plainis, S., & Murray, I. J. (2000). Neurophysiological interpretation of human visual reaction
931   times: effect of contrast, spatial frequency and luminance. *Neuropsychologia*, *38*(12),
932   1555–64. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11074078

933 Quick, R. F. (1974). A vector-magnitude model of contrast detection. *Kybernetik*, *16*(2), 65–7.
934   Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/4453110

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

937 Romei, V., Murray, M. M., Cappe, C., & Thut, G. (2009). Preperceptual and stimulus-selective
938   enhancement of low-level human visual cortex excitability by sounds. *Current biology :*
939   *CB*, *19*(21), 1799–805. doi:10.1016/j.cub.2009.09.027

940 Rowland, B. A., Quessy, S., Stanford, T. R., & Stein, B. E. (2007). Multisensory integration
941   shortens physiological response latencies. *The Journal of neuroscience : the official*
942   *journal of the Society for Neuroscience*, *27*(22), 5879–84. doi:10.1523/JNEUROSCI.4986-
943   06.2007

944 Schroeder, C. E., & Foxe, J. (2005). Multisensory contributions to low-level, "unisensory"
945   processing. *Current opinion in neurobiology*, *15*(4), 454–8.
946   doi:10.1016/j.conb.2005.06.008

947 Silveira, L. C., & Perry, V. H. (1991). The topography of magnocellular projecting ganglion cells
948   (M-ganglion cells) in the primate retina. *Neuroscience*, *40*(1), 217–37. Retrieved from
949   http://www.ncbi.nlm.nih.gov/pubmed/1711175

950 Smiley, J. F., & Falchier, A. (2009). Multisensory connections of monkey auditory cerebral
951   cortex. *Hearing research*, *258*(1-2), 37–46. Retrieved from
952   http://cat.inist.fr/?aModele=afficheN&cpsidt=22352861

953 Spence, C., & Driver, J. (1994). Covert orienting in audition: Endogenous and exogenous
954   mechanisms. *Journal of Experimental Psychology: Human Perception and Performance*.
955   Retrieved from http://discovery.ucl.ac.uk/100237/

956 Stein, B E. (1998). Neural mechanisms for synthesizing sensory information and producing
957   adaptive behaviors. *Experimental brain research. Experimentelle Hirnforschung.*

958 *Expérimentation cérébrale*, *123*(1-2), 124–35. Retrieved from
959 http://www.ncbi.nlm.nih.gov/pubmed/9835401

960 Stein, Barry E, Stanford, T. R., Ramachandran, R., Perrault, T. J., & Rowland, B. a. (2009).
961 Challenges in quantifying multisensory integration: alternative criteria, models, and
962 inverse effectiveness. *Experimental brain research. Experimentelle Hirnforschung.*
963 *Expérimentation cérébrale*, *198*(2-3), 113–26. doi:10.1007/s00221-009-1880-8

964 Stein, Barry E., London, N., Wilkinson, L. K., & Price, D. D. (1996). Enhancement of Perceived
965 Visual Intensity by Auditory Stimuli: A Psychophysical Analysis. *Journal of Cognitive*
966 *Neuroscience*, *8*(6), 497–506. doi:10.1162/jocn.1996.8.6.497

967 Stein, Barry E., Meredith, M. A., Huneycutt, W. S., & McDade, L. (1989). Behavioral Indices of
968 Multisensory Integration: Orientation to Visual Cues is Affected by Auditory Stimuli.
969 *Journal of Cognitive Neuroscience*, *1*(1), 12–24. doi:10.1162/jocn.1989.1.1.12

970 Thomas, J. P., Fagerholm, P., & Bonnet, C. (1999). One spatial filter limits speed of detecting
971 low and middle frequency gratings. *Vision research*, *39*(9), 1683–93. Retrieved from
972 http://www.ncbi.nlm.nih.gov/pubmed/10343860

973 Treisman, M. (1998). Combining information: Probability summation and probability averaging
974 in detection and discrimination. *Psychological Methods*, *3*(2), 252–265.
975 doi:10.1037/1082-989X.3.2.252

976 Ulrich, R., Miller, J., & Schröter, H. (2007). Testing the race model inequality: an algorithm and
977 computer programs. *Behavior research methods*, *39*(2), 291–302. Retrieved from
978 http://www.ncbi.nlm.nih.gov/pubmed/17695357

979 Valls-Solé, J., Solé, A., Valldeoriola, F., Muñoz, E., Gonzalez, L. E., & Tolosa, E. S. (1995).
980 Reaction time and acoustic startle in normal human subjects. *Neuroscience letters*,
981 *195*(2), 97–100. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/7478277

982 Van der Burg, E., Cass, J., Olivers, C. N. L., Theeuwes, J., & Alais, D. (2010). Efficient visual
983 search from synchronized auditory signals requires transient audiovisual events. *PloS*
984 *one*, *5*(5), e10664. doi:10.1371/journal.pone.0010664

985 Wallace, M T, Meredith, M. a, & Stein, B. E. (1993). Converging influences from visual,
986 auditory, and somatosensory cortices onto output neurons of the superior colliculus.
987 *Journal of neurophysiology*, *69*(6), 1797–809. Retrieved from
988 http://www.ncbi.nlm.nih.gov/pubmed/8350124

989 Wallace, M T, Meredith, M. a, & Stein, B. E. (1998). Multisensory integration in the superior
990 colliculus of the alert cat. *Journal of neurophysiology*, *80*(2), 1006–10. Retrieved from
991 http://www.ncbi.nlm.nih.gov/pubmed/9705489

992 Wallace, Mark T., Meredith, M. A., & Stein, B. E. (1998). Multisensory Integration in the
993 Superior Colliculus of the Alert Cat. *J Neurophysiol*, *80*(2), 1006–1010. Retrieved from
994 http://jn.physiology.org/cgi/content/abstract/80/2/1006

995   Watkins, S., Shams, L., Tanaka, S., Haynes, J.-D., & Rees, G. (2006). Sound alters activity in
996   human V1 in association with illusory visual perception. *NeuroImage*, *31*(3), 1247–56.
997   doi:10.1016/j.neuroimage.2006.01.016

998   Werner, S., & Noppeney, U. (2011). The contributions of transient and sustained response
999   codes to audiovisual integration. *Cerebral cortex (New York, N.Y. : 1991)*, *21*(4), 920–31.
1000   doi:10.1093/cercor/bhq161

1001 Wuerger, S. M., Hofbauer, M., & Meyer, G. F. (2003). The integration of auditory and visual
1002   motion signals at threshold. *Perception & psychophysics*, *65*(8), 1188–96. Retrieved from
1003   http://www.ncbi.nlm.nih.gov/pubmed/14710954

1004

1005

1006

1007 **Figure captions/Legends:**

1008

1009 Figure 1. Schematic representation of a trial sequence. After a randomly chosen interval

1010 following the onset of the fixation cross (500-1500 ms), a Gabor stimulus (visual target) was

1011 presented on 80% of the trials (therefore, this is the proportion of trials requiring response). The

1012 remaining 20% were catch-trials (no visual target). In 50% of the trials (across target-bearing

1013 and catch-trials) contained a sound temporally aligned with the visual event, thus making an

1014 equal proportion of visual and audio-visual trials. (The relative size of the Gabor has been

1015 scaled up for clarity of presentation, see details in text.)

1016

1017 Figure 2. This plot represents individual alpha parameter values and their confidence intervals

1018 across visual spatial frequencies (SF) for the visual (top row) and audio-visual (bottom row)

1019 conditions for the data in Experiment 1. These values correspond to fits of the Pieron function

1020 when leaving α as a free parameter (see text for details). Each column represents one

1021   participant. Except for four of the thirty fitted values, the α parameter was not significantly

1022   different from -1.

1023

1024   Figure 3. RTs as a function of the reciprocal of contrast (1/C) for each spatial frequency (plotted

1025   in different panels) and modality combination (plotted in different shades of grey, see legend) in

1026   Experiment 1. Small symbols represent individual means of each participant (each based on an

1027   average of 80 reaction times, min=60), whereas large symbols represent the inter-participant

1028   average. The lines represent the linear least square regression fits (thin dotted and thick solid

1029   lines for the individual and average data, respectively). It can be appreciated that that the slope

1030   of these fits (k parameter) is shallower in the AV condition compared to the V condition, at the

1031   low spatial frequencies. The thin solid black line in each plot represents the prediction of the

1032   probability summation model according to the LATER model.

1033

1034   Figure 4. **(a)** Plot of the inter-participant averages for audio-visual (grey) and visual (black)

1035   sensitivities ($1/k$) in Experiment 1. **(b)** Average audio-visual and visual values for the intercept

1036   (t_0). The results are presented separately for each SF (see legend). Each dot represents averaged

1037   inter-participant intercepts and error bars represent the associated SD.

1038

1039   Figure 5. Correlation plot between visual (black) and audio-visual (grey) individual sensitivity

1040   ($1/k$) and the percentage of false alarms in Experiment 1.

1041

1042   Figure 6. Inter-participant average RTs in Experiment 2 (error bars represent confidence

1043   intervals) plotted as a function of contrast and modality (see legend). The two tested SFs are

1044   represented in different columns. Insets contain plots of cumulative Gaussians fitted to the

1045 empirical cumulative distribution functions for the RTs in the auditory (light grey), visual

1046 (medium grey) and audio-visual (dark grey) conditions.

1047

1048 Figure 7. Race model inequality tests in Experiment 2. The black squares represent the

1049 empirical average RTs in the audio-visual condition binned in 10 quantiles, whilst the grey

1050 squares represent the prediction of the probability summation model obtained from the auditory-

1051 only and visual-only RT distributions (Miller's bound). Error bars represents the standard

1052 errors. Shaded areas indicate quantiles where the race model was significantly violated

1053 (p<0.05).

1054

1055

1056 Table 1. Average and standard deviation of visual and audio-visual RTs, and latency of FA

1057 responses in auditory-only catch-trials. Although this is an averaged representation of the data,

1058 in the simulations we applied the LATER model considering each participant individually.

1059

1060

1061

1062