Audiovisual detection at different intensities and delays

3 *Chandramouli Chandrasekaran*

4 Neuroscience Institute & Department of Psychology, Princeton University, USA

5 Department of Electrical Engineering, Stanford University, USA

6 *Matthias Gondan*

7 Department of Psychology, University of Copenhagen, Denmark

8 Faculty for Psychology, University of Vienna, Austria

9 **Correspondence address**

10 Matthias Gondan

11 Faculty for Psychology

12 University of Vienna

13 Liebiggasse 5

14 1010 Vienna

15 Austria

16 Email: matthias.gondan-rochon@univie.ac.at

17 **Running Title**

18 Audiovisual detection

Abstract

In divided-attention tasks with two classes of target stimuli (e.g., auditory and visual), redundancy gains are typically observed if both targets are presented simultaneously, as compared with single-target presentation. Different models explain such redundant-signals effects, including race and coactivation models. Here we generalize one such coactivation model, the superposition model, and show that restricting this evidence accumulation process to a limited time interval, that is a temporal deadline, provides the model the ability to describe the detection behavior of two monkeys across a wide range of intensities and stimulus onset asynchronies. We present closed-form solutions for the mean absorption times and probabilities for the two-stage diffusion process with drift towards a single barrier in the presence of a temporal deadline.

Keywords

Multisensory processing; Monkey; Decision making; Diffusion processes; Reaction times.

Background

31

32   Integrating information from different senses improves perception and action, speeds up response

33   time (RT), enhances detection and discrimination accuracy, and facilitates correct arm movements

34   (Cluff, Crevecoeur, & Scott, 2015; Crevecoeur, Munoz, & Scott, 2016; Diederich & Colonius, 2004a;

35   Miller, 1982; Nickerson, 1973; Sakata, Yamamori, & Sakurai, 2004; Schwarz, 1989, 1994; Seilheimer,

36   Rosenberg, & Angelaki, 2014; Stein & Stanford, 2008; van Atteveldt, Murray, Thut, & Schroeder,

37   2014). Despite many studies of multisensory integration, several questions about how the behavioral

38   benefits of multisensory integration emerge remain unanswered (Chandrasekaran, 2016; Cluff et al.,

39   2015; Crevecoeur et al., 2016).

40   In this study, we focused on the question of how the behavioral benefits of multisensory

41   integration vary as a function of both the sensory intensities as well as delays between the stimuli.

42   With few exceptions (Gondan, Götze, & Greenlee, 2010), past studies have largely investigated these

43   two aspects in isolation (Cluff et al., 2015; Crevecoeur et al., 2016; Diederich & Colonius, 2004a; Dixon

44   & Spitz, 1980; Holmes, 2009; Meredith, Nemitz, & Stein, 1987; Stein, Meredith, Huneycutt, &

45   McDade, 1989; van Wassenhove, Grant, & Poeppel, 2007). In this study, we extended our recent

46   paradigm where monkeys detect visual, auditory, and audiovisual vocalizations in a background of

47   noise (Chandrasekaran, Lemus, & Ghazanfar, 2013; Chandrasekaran, Lemus, Trubanova, Gondan, &

48   Ghazanfar, 2011; Miller, 1986) by incorporating delays between the sensory stimuli (Miller, 1986). We

49   measured the response times (RTs) and detection accuracy of the monkeys and then modeled the

50   behavioral patterns by expanding on the coactivation modeling framework (Diederich, 1995; Miller,

51   1982; Schwarz, 1989, 1994, 2006).

52   We demonstrate that the behavioral benefits accrued by monkeys when detecting vocalizations

53   are a lawful function of both the intensities and the delay between sensory stimuli (Cluff et al., 2015;

54   Crevecoeur et al., 2016; Dixon & Spitz, 1980; Holmes, 2009; Meredith et al., 1987; Stein et al., 1989;

55   van Wassenhove et al., 2007). After ensuring that separate activation models are insufficient to

56   explain the behavior of the monkeys (Miller, 1986; Raab, 1962) we show that the superposition

57   framework that assumes additivity of visual and auditory inputs and accumulation to a bound

58   provides an approximate description of the RTs but fails to describe the accuracy of the monkeys

59   performing the task (Diederich, 1995; Schwarz, 1989, 1994). We, therefore, expanded on the

60   diffusion superposition model by incorporating a deadline for accumulation. We derive closed-form

61   solutions for the mean reaction times (RTs) and accuracy rates for the two-stage diffusion process

62   with drift towards a single barrier in the presence of a temporal deadline. We then show that such an

63   expansion of the classical coactivation model provides an excellent description of both accuracies as

64   well as RTs of these monkeys.

65   The manuscript is organized as follows. We provide a brief background on the diffusion

66   superposition model. We then describe the details of and results from our behavioral experiments in

67   monkeys and show that the classical model provides an unsatisfactory fit to the accuracy and

68   response times. We then show that extending the model by incorporating a deadline for

69   accumulation provides an better account of the observed behavior of the monkeys. We end by

70   discussing the ramifications of this model.

71   **Redundant Signals Tasks**

72   The redundant signals task is a simple and powerful experimental paradigm for investigating

73   multisensory perception. In this task, subjects are asked to respond in the same way to stimuli of two

74   sensory modalities, for example, auditory and visual stimuli (A, V, Diederich & Colonius, 1987; Miller,

75   1982, 1986). The typical finding is that if signals from both modalities are present (redundant signals,

76   AV), average responses are faster than average responses to targets from any single modality

77   (unimodal targets). This redundant signals effect, by itself, is not necessarily indicative of any special

78   mechanism that integrates the information of the different channels. Redundancy gains may arise

79   from a race between the channel-specific detection processes ("race model," Miller, 1982; Raab,

80   1962). In audiovisual detection, however, redundancy gains are typically larger than predicted by race

81   models and these are thought to be better explained by "coactivation models" that assume some

82   kind of integration of the information provided by the two sensory systems (Diederich, 1995;

83   Diederich & Colonius, 2004b; Miller, 1986; Schwarz, 1989, 1994). One such model is the diffusion

84   superposition model, which we explain below.

85 **The diffusion superposition model**

86 The single-barrier diffusion superposition model (Schwarz, 1994) is a coactivation model that

87 describes redundancy gains assuming additive superposition of channel-specific diffusion processes.

88 In diffusion models, the assumption is that the presentation of a stimulus leads to a buildup of

89 evidence that is described by a noisy diffusion process $\mathbf{X}(t)$ with drift μ and variance $\sigma^2 > 0$ (Ratcliff,

90 1988; Ratcliff & McKoon, 2008; Ratcliff, Smith, Brown, & McKoon, 2016; Smith & Ratcliff, 2004). The

91 stimulus is 'detected' as soon as an evidence criterion $c > 0$ is met for the first time. The density $g(t)$

92 and distribution $G(t)$ of the first-passage times $\mathbf{D}$ are well known ('inverse Gaussian', Cox & Miller,

93 1965),

94

$$g(t \mid c, \mu, \sigma^2) = \frac{c}{\sqrt{2\pi\sigma^2 t^3}} \cdot \exp\left[-\frac{(c-\mu t)^2}{2\sigma^2 t}\right] \tag{1}$$

$$G(t \mid c, \mu, \sigma^2) = \Phi(\mu t \mid c, \sigma^2 t) + \exp\left(\frac{2c\mu}{\sigma^2}\right) \Phi(-\mu t \mid c, \sigma^2 t) \tag{2}$$

97

98 with $\Phi(x \mid m, s^2)$ denoting the Normal distribution with mean m and variance s^2. The expected

99 detection time $E(\mathbf{D})$ is obtained by integrating $\int_0^\infty t \cdot g(t \mid c, \mu, \sigma^2)\, dt$, which simplifies to

100

$$E(\mathbf{D}) = \frac{c}{\mu} \tag{3}$$

102

103 Predictions for the detection times for unimodal stimuli A and V are, therefore, easily obtained,

104 c/μ_A and c/μ_V, respectively. When two stimuli are presented simultaneously, coactivation occurs.

105 The model assumes that the two modality-specific processes superimpose linearly, $\mathbf{X}_{AV}(t) = \mathbf{X}_A(t) +$

106 $\mathbf{X}_V(t)$. The new process $\mathbf{X}_{AV}$ is, therefore, again a diffusion process with drift $\mu_{AV} = \mu_A + \mu_V$ and

107 variance $\sigma_{AV}^2 = \sigma_A^2 + \sigma_V^2 + 2\rho_{AV}\sigma_A\sigma_V$ (under the assumption that $\mathbf{X}_A$ and $\mathbf{X}_V$ are uncorrelated, the

108 covariance term will be zero). Since the drift parameters add up, $\mathbf{X}_{AV}$ reaches the response criterion

109 earlier than any of its single constituents, resulting in faster responses to redundant stimuli,

110

111 $$E(\mathbf{D}_{AV}) = \frac{c}{\mu_A + \mu_V}$$

112   What happens in stimuli presented with onset asynchrony (SOA, e.g., V160A, i.e., the visual

113 stimulus component is presented first, and then the auditory stimulus component follows the visual

114 stimulus component with a delay τ, here 160 ms)? During the first τ ms, the sensory evidence is

115 accumulated by the visual channel alone. If the criterion is reached within this interval, the stimulus is

116 detected, and a response is initiated. This case occurs with probability

117 $P\big(\mathbf{D}_{V(\tau)A} \leq \tau\big) = G(\tau \mid c, \mu_V, \sigma_V^2)$ given by Eq. (2). If detection occurs before a time of τ has elapsed, it

118 is expected to occur within

119

120     $$E\big(\mathbf{D}_{V(\tau)A} \mid \mathbf{D}_{V(\tau)A} \leq \tau\big) = \frac{1}{P(\mathbf{D}_{V(\tau)A} \leq \tau)} \int_0^\tau t \cdot g(t \mid c, \mu_V, \sigma_V^2)\, dt \qquad (4)$$

121

122   The solution for the integral is given by Schwarz (1994, Eq. 6). In the other case, the process has

123 attained a subthreshold activation level $\mathbf{X}(\tau) = x < c$, with probability density described by

124

125     $$w(x, \tau \mid c, \mu_V, \sigma_V^2) = \phi(x \mid \mu_V\tau, \sigma_V^2\tau) - \exp\left(\frac{2c\mu_V}{\sigma_V^2}\right) \cdot \phi(x \mid 2c + \mu_V\tau, \sigma_V^2\tau), \qquad (5)$$

126

127 (e.g., Schwarz, 1994, Eq. 7) with $\phi(x \mid \mu, \sigma^2)$ denoting the Normal density with mean μ and variance

128 σ^2. Because activation level x has already been attained, less 'work' has to be done to reach the

129 criterion. Starting at time τ, both stimuli contribute to the buildup of activity, resulting in an

130 aggregate process drifting with μ_{AV} towards a residual barrier $c - x$:

131

132     $$E\big(\mathbf{D}_{V(\tau)A} \mid \mathbf{X}(\tau) = x\big) = \tau + \frac{c - x}{\mu_{AV}}. \qquad (6)$$

133

134   This expectation must be integrated for all possible levels of activation $x < c$, weighted by the

135 density (5) that level x has been reached by time $t = \tau$:

136

137 $$E\big(\mathbf{D}_{V(\tau)A} \mid \mathbf{D}_{V(\tau)A} > \tau\big) = \int_{-\infty}^{c} w(x, \tau \mid c, \mu_V, \sigma_V^2) \cdot E\big(\mathbf{D}_{V(\tau)A} \mid \mathbf{X}(\tau) = x\big)\, dx. \qquad (7)$$

138

139 An analytic solution for the overall, unconditional expected first-passage time $E\big(\mathbf{D}_{V(\tau)A}\big)$ has been

140 derived by Schwarz (1994, Eq. 10).

141   The diffusion process only describes the 'detection' latency $\mathbf{D}$, or processing time. To derive a

142 prediction for the observed response time $\mathbf{T}$, an additional variable $\mathbf{M}$ is typically introduced that

143 summarizes everything not described by the diffusion process (e.g., motor execution), such that

144 $\mathbf{T} = \mathbf{D} + \mathbf{M}$. Therefore, the model prediction for the mean response time is

145

146 $$E(\mathbf{T}) = E(\mathbf{D}) + \mu_M \qquad (8)$$

147

148   where the additional parameter denotes the expectation of $\mathbf{M}$. Schwarz (1994) demonstrated

149 that such a model of additive superposition quite accurately describes the mean response times and

150 variances reported for Participant B.D. from (Miller, 1986) in a speeded response task with

151 13 different SOAs (at a single intensity).

152 **Present behavioral experiment**

153 Our stated goal is to understand how the benefits in redundant signals tasks depend on both the

154 intensity of the sensory stimuli as well as the SOA. To that end, we trained monkeys to detect visual,

155 auditory and audiovisual vocal signals in a constant background of auditory noise (Chandrasekaran et

156 al., 2013; Chandrasekaran et al., 2011). We chose a free response paradigm (without explicit trial

157 markers , Egan, Greenberg, & Schulman, 1961; Shub & Richards, 2009) because it mimics natural

158 audiovisual communication—faces are usually continuously visible and move during vocal production.

159 The task was a typical redundant signals task (Miller, 1982, 1986); stimuli were chosen to

160 approximate natural face-to-face vocal communication. This task was heavily inspired by the

161 observation that besides providing benefits for discrimination of speech sounds (Besle, Fort,

162 Delpuech, & Giard, 2004; Grant, Walden, & Seitz, 1998), visual speech enhances the detection of

163 auditory speech (Grant, 2001; Grant & Seitz, 2000; Schwartz, Berthommier, & Savariaux, 2004). In

164   such settings, the vocal components of the communication signals are degraded by environmental

165   noise. The motion of the face, on the other hand, is usually perceived clearly. In the task, monkeys

166   detected the onset of 'coo' calls that are affiliative vocalizations commonly produced by macaque

167   monkeys in a variety of contexts (Hauser & Marler, 1993; Rowell & Hinde, 1962). The coo calls were

168   presented at three different levels of sound intensity and were embedded in a constant background

169   noise. The visual signals were videos of monkey avatars opening their mouth to make a coo

170   vocalization. The size of the mouth opening was in accordance with the intensity of the associated

171   vocalization: greater sound intensity was coupled to larger mouth openings by the dynamic face.

172   Finally, we generated audiovisual stimuli by combining the videos with the coo vocalizations. The task

173   of the monkeys was to detect the visual motion of the mouth or the onset of the coo vocalization.

174   Audiovisual stimuli were presented either in synchrony or at 10 different SOAs.

175   **Methods**

176   **Subjects**

177   Nonhuman primate subjects were two adult male macaques (S and B, *Macaca fascicularis*). The

178   monkeys were born in captivity and housed socially. The monkeys underwent sterile surgery for the

179   implantation of a painless head restraint (see Chandrasekaran, Turesson, Brown, & Ghazanfar, 2010).

180   All experiments and surgical procedures were performed in compliance with the guidelines of the

181   Princeton University Institutional Animal Care and Use Committee.

182   **Procedure**

183   Experiments were conducted in a sound attenuating radio frequency enclosure. The monkey sat in a

184   primate chair fixed 74 cm opposite a 19 inch CRT color monitor with a 1280 × 1024 screen resolution

185   and 75 Hz refresh rate. The screen subtended a visual angle of ~25° horizontally and 20° vertically. All

186   stimuli were centrally located on the screen and occupied a total area (including blank regions) of

187   640 × 653 pixels. For every session, the monkeys were placed in a restraint chair and head-posted. A

188   depressible lever (ENV-610M, Med Associates) was located at the center-front of the chair. Both

189   monkeys spontaneously used their left hand for responses. Stimulus presentation and data collection

190   were performed using Presentation (Neurobehavioral Systems).

Stimuli

192   Coo vocalizations could be one of three different levels of sound intensity (85 dB, 68 dB, 53 dB) and

193   were embedded in a constant background noise of ~63 dB SPL (giving us a range of signal to noise

194   ratios, SNR, Fig. 1A). We used coo calls from two macaques as the auditory components of

195   vocalizations; these were recorded from individuals that were unknown to the monkey subjects. The

196   auditory vocalizations were resized to a constant duration of 400 ms using a MATLAB implementation

197   of a phase vocoder (Flanagan & Golden, 1966) and normalized in amplitude (Fig. 1A). The visual

198   components of the vocalizations were 400 ms long videos of synthetic monkey agents making a coo

199   vocalization. The animated stimuli were generated using 3D Studio Max 8 (Autodesk) and Poser Pro

200   (Smith Micro), and were extensively modified from a stock model made available by DAZ Productions

201   (Silver key 3D monkey, Figs. 1B, C). Further details of the generation of these visual avatars are

202   available in a prior study (Chandrasekaran et al., 2011).

203   The audiovisual stimuli were generated by presenting both the visual and auditory components

204   either in synchrony or at 10 different stimulus onset asynchronies (SOAs). Audition could precede

205   vision by 240, 160, 120, 80 or 40 ms (Fig. 1D) and vice versa (Fig. 1E). Intensities were always paired;

206   that is the weak auditory stimulus was always paired with a small mouth opening. We, therefore, had

207   3 pairs of intensities and 11 SOAs, which resulted in 33 AV conditions in total. Each block also had 3

208   auditory intensities, 3 visual intensities. Catch trials (C) were used to discourage from spontaneous

209   responses and to be able to control for fast guesses in the analysis of the RT distributions.

Task

211   During the task (Fig. 1F), an avatar face (e.g., Avatar 1) was continuously present on the screen; the

212   background noise was also continuous. In the visual-only condition (V), Avatar 1 moved its mouth

213   without any corresponding auditory component. In the auditory-only condition (A), the vocalization

214   paired with Avatar 2 was presented with the static face of Avatar 1. Finally, in the audiovisual

215   condition (AV), Avatar 1 moved its mouth accompanied by the corresponding vocalization of Avatar 1

216   and in accordance with its intensity. We, therefore, had two AV stimuli. (A1V1 and A2V2). In the even

217   blocks, the avatar face was the still frame of V1, A2 was the auditory sound played, and A1V1 was the

218   audiovisual stimulus. The other block had the opposite configuration. This task design avoids the

219   conflict between hearing a vocalization with the corresponding avatar face not moving.
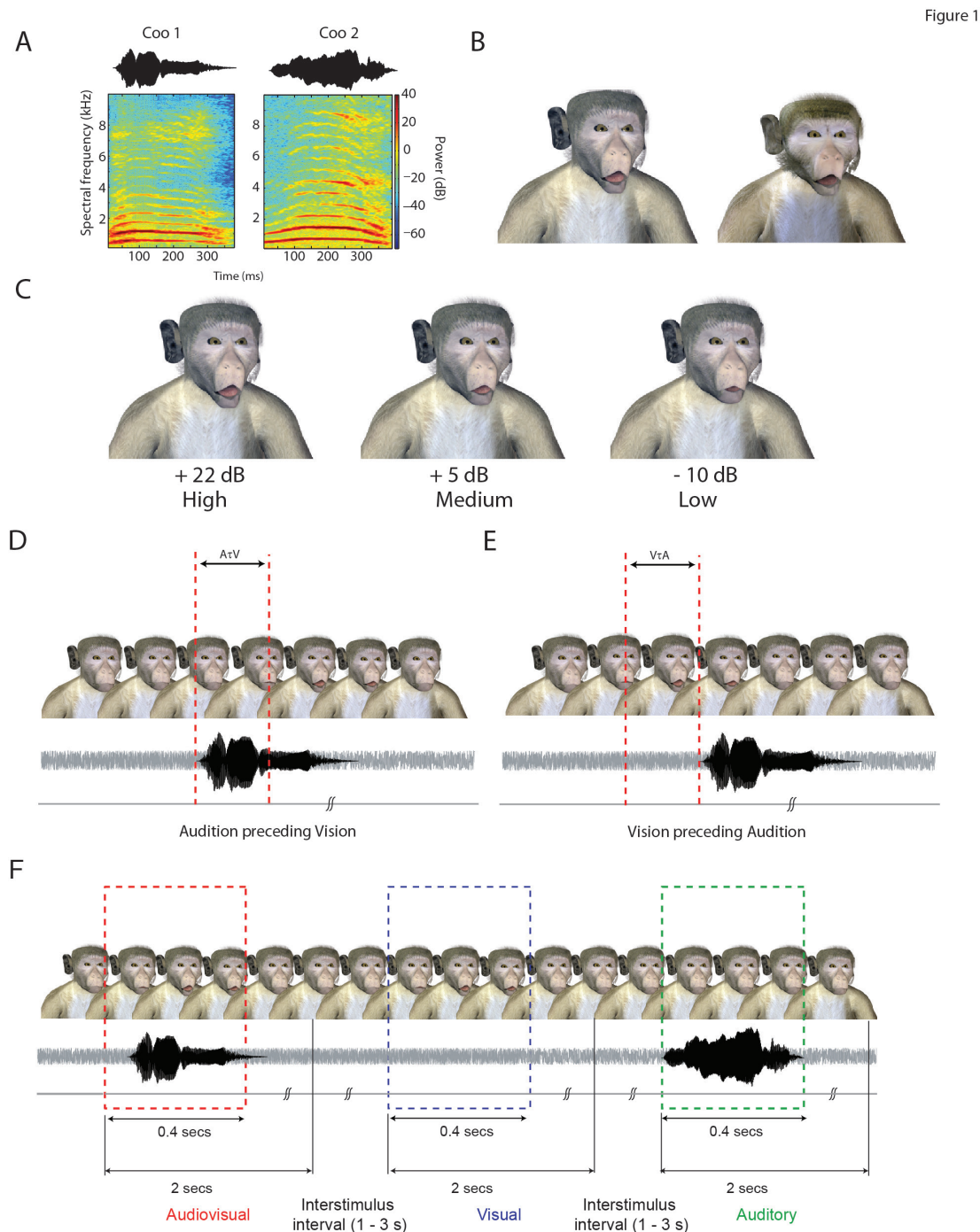

Figure 1

Figure 1 (previous page): Stimuli, task structure, and behavior. A: Waveform and spectrogram of coo vocalizations detected by the monkeys. **B:** Frames of the two monkey avatars at the point of maximal mouth opening for the largest SNR. **C**: Frames with maximal mouth opening from one of the monkey avatars for three different SNRs of +22 dB (High), +5 dB (Medium) and −10 dB (Low). **D:** An illustration of the A(τ)V audiovisual condition. In this stimulus, the onset of the vocalization precedes the onset of the mouth motion. **E:** An illustration of the V(τ)A audiovisual condition. In this stimulus, visual cues such as mouth opening precede the onset of the vocalization. **F:** Task structure for monkeys. An avatar face was always on the screen. Visual, auditory and audiovisual stimuli were randomly presented with an inter stimulus interval of 1–3 seconds drawn from a shifted and truncated exponential distribution. Responses within a 2 sec window after stimulus onset were considered hits. Responses in the inter-stimulus interval are considered to be response errors and led to timeouts.

220   Stimuli of each condition (V, A, AV, C) were presented after a variable inter stimulus interval

221   between 1 and 3 seconds (drawn from a shifted and truncated exponential distribution). Monkeys

222   indicated the detection of a V, A or AV event by pressing the lever within 2 seconds following the

223   onset of the stimulus. In the case of hits, the ISI was started immediately following a juice reward. In

224   the case of misses, the ISI began after the two second response window.

225   After every block of 126 trials (33 AV stimuli × 3 trials + 3 A × 3 trials, 3 V × 3 trials, 9 catch

226   trials), a brief pause (~10 to 12 seconds) was imposed. Then, a new block was started in which, the

227   avatar face, and the identity of the coo sound used for the auditory-only condition were switched.

228   Within a block, all the conditions were randomly interleaved with one another.

Training

230   Monkeys were initially trained over many sessions to respond to the coo vocalization events in visual,

231   auditory or audiovisual conditions while withholding responses when no stimuli were presented. A

232   press of the lever within a window starting 150 ms after onset of the vocalization event and within

233   two seconds led to a juice reward and was defined as a hit. An omitted response in this two-second

234   window was classified as a miss similar to the studies of free response tasks (without explicit trial

235   markers , Egan et al., 1961; Shub & Richards, 2009). Lever presses to catch trials were defined as false

236   alarms. In addition, the random presses during the interstimulus interval (ISI) were discouraged by

237   enforcing a timeout where no stimuli were presented. The timeout was chosen randomly from a

238 distribution between 3.0 and 5.5 s. The monkeys had to wait for the entire duration of this timeout

239 period before a new stimulus was presented. Any lever press during the timeout period led to a

240 renewal of the timeout with the duration again randomly drawn from the same distribution. Monkeys

241 were trained until the erroneous presses in this ISI period were less than or equal to 10% of trials in

242 any given session.

243 **Statistical analysis of behavioral performance**

244 Hit rate was defined as the ratio of hits to the total number of targets. For each SNR and condition,

245 the accuracy was defined as the ratio of hits to hits plus misses expressed as a percentage. The false

246 alarm rate (i.e., presses to the catch stimuli) was defined as the number of false alarms divided by the

247 total number of catch trials. Mean RTs and SDs were computed for the correct responses; confidence

248 intervals were obtained by resampling the observed RT distributions (including omitted responses

249 and false alarms) with replacement 1000 times and estimating the standard deviation of the mean of

250 the resampled data.

251 **Test of the race model inequality**

252 An important model class for redundant signals effects is the so-called race model, or separate

253 activation model (Colonius & Diederich, 2006; Gondan & Minakata, 2016; Miller, 1982, 1986; Raab,

254 1962). According to the race model, redundancy benefits are not due to an actual integration of visual

255 and auditory signals but because of parallel processing of both signals. In the bimodal stimulus, the

256 two channels engage in a race-like manner ("parallel first-terminating model", Townsend & Ashby,

257 1983), so that the probability for fast responses is increased because slow processing times are

258 canceled out by the other channel. The redundancy gains obtained in the race model are limited,

259 however, and a classical test of whether this mechanism can explain the observed reaction times is

260 the well- known race model inequality (Miller, 1982), stating that the RT distribution for redundant

261 stimuli $F_{AV}(t)$ never exceeds the sum of the RT distributions for the unisensory stimuli $F_A(t), F_V(t)$,

262

263 $\qquad F_{AV}(t) \leq F_A(t) + F_V(t)$, for all t. $\hfill$ (9)

264

265   (Eriksen, 1988, see also Gondan and Heckel, 2008) demonstrated that this inequality could be refined

266   by taking into account anticipatory responses to catch trials (C),

267

268        $$F_{AV}(t) + F_C(t) \leq F_A(t) + F_V(t), \text{ for all } t. \tag{10}$$

269

270   For redundant targets presented with SOA τ, the inequality generalizes to (Miller, 1986)

271

272        $$F_{V(\tau)A}(t) + F_C(t - \tau) \leq F_V(t) + F_A(t - \tau), \text{ for all } t. \tag{11}$$

273

274   If this inequality is violated in a given data set, then parallel self-terminating processing cannot

275   account for the benefits observed for multisensory stimuli, suggesting an explanation based on the

276   integration of the signals (e.g., the superposition model described above; see also (Luce, 1986; Miller,

277   2016) for a discussion of the assumption of context independence).

278   For each condition, we determined the empirical cumulative distribution functions (eCDFs) and

279   then computed the maximum violation, that is, the maximum difference between the left-hand side

280   and the right-hand side of Inequality 11. A bootstrap technique (Miller, 1986), was used to assess the

281   statistical significance of the observed violations of the race model inequality.

282   **Test of the diffusion superposition model**

283   We fit the predictions of the diffusion superposition model to the mean RTs from the two monkeys.

284   We used the analytic equation provided in Schwarz (1994, Eq. 10). To perform model fitting, we

285   computed for each monkey, an approximate χ^2_{26} goodness-of-fit statistic given by the sum of the

286   squared standardized deviations of the predicted and the observed average response times (e.g.,

287   Schwarz, 2006). The 26 degrees of freedom are given by the difference between the number of

288   conditions (3 intensities × 13 SOAs) minus the number of model parameters: 12 parameters are due

289   to the drift and variance for each visual and auditory SNR (3 SNRs × 2 modalities × 2 parameters). The

290   thirteenth parameter is the average residual non-decision time (Eq. 8).

Results

Figures 2A, C show the accuracy of the two monkeys in the detection task for different SNRs and SOAs. The accuracy (see Figure 2A, C, see also Supplementary Tables S1–S4) showed lawful decreases with changes in SNR for the auditory components of vocalizations (e.g., only about 55% detection rate with auditory stimuli). In contrast, changes in mouth opening size for the visual component of the vocalizations, which were meant to match the changes in auditory intensity, had only minimal impact on the accuracy of the animals. The mean RTs (Figures 2B,D) showed the wing shaped pattern typically observed in redundant signals tasks with SOA manipulations (Miller, 1986; Ulrich & Miller, 1997).
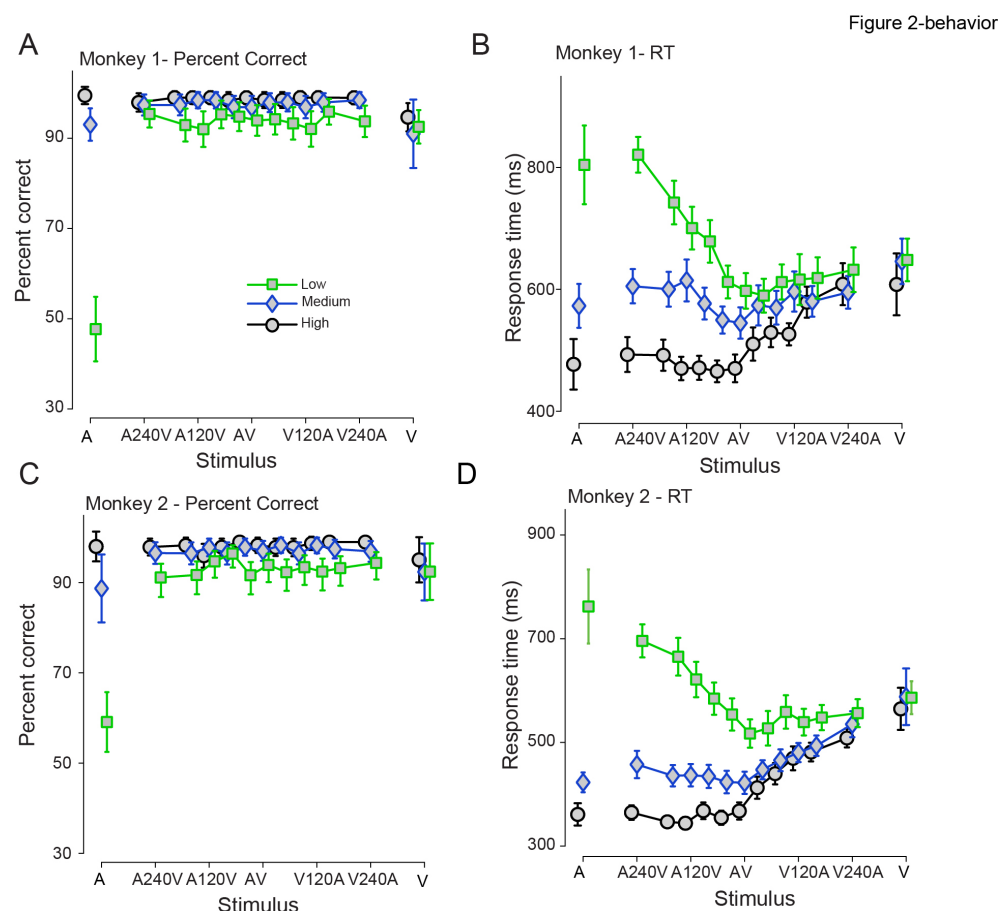

Figure 2: Monkeys integrate visual and auditory cues across a large range of SOAs and SNRs
A, B: Mean accuracy and RTs for audiovisual, visual-only and auditory-only conditions for Monkey 1 (A, B) and Monkey 2 (C, D) as a function of different stimulus conditions and SNR. X-axes depict stimuli. Y-axes in the left panel depict the accuracy in percent. Y-axes in the right panel denote response times in ms. Error bars denote confidence intervals around the mean (2 × SEM).

301   In both monkeys, statistically significant violations of the upper bound of the race model

302   inequality for RTs (Inequality 11) occurred for a large range of SOAs, indicating that the observed

303   redundancy gains were inconsistent with a race model. For Monkey 1, statistically significant

304   violations of the RMI (at the criterion of $P < 0.05$) were observed in 28 out of the 33 multisensory

305   conditions, for Monkey 2, this was observed in all 33 conditions (Tables S5, S6 in online supplement).

306

307   We first examined if the diffusion superposition model could describe the behavior of the

308   monkeys during this task. We found that the superposition model could describe the overall pattern

309   of the mean RTs of the monkeys (Figures 3A, B). However, the overall fit was unsatisfactory and

310   especially poor for the lowest SNRs (Monkey 1: $\chi^2_{26} = 49.2, P = 0.004$; Monkey 2: $\chi^2_{26} = 53.9, P =$
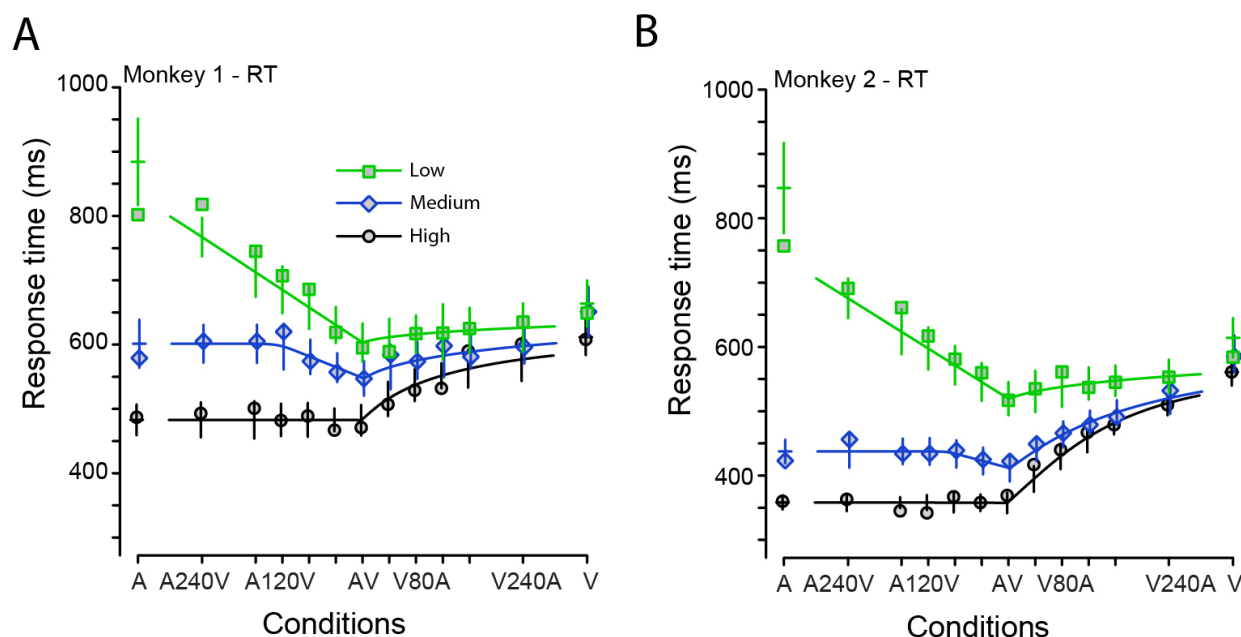

Figure 3: The diffusion superposition model provides an incomplete description of the detection behavior of monkeys
Audiovisual, visual-only and auditory-only RTs for both monkeys along with the predicted RT shown in lines according to the diffusion superposition model as a function of SNR (squares = low, diamonds = medium, circles = high)and SOA. Error bars denote confidence intervals (2 × standard error) around predicted mean RTs. A shows the RT for monkey 1; B shows the RT for monkey 2.

311   0.001). The monkeys also omitted a substantial proportion of responses to weak auditory stimuli

312   (detection accuracy rates for the lowest auditory intensity were ~55% and 60%), which is, by design,

313   not accounted for by the superposition model described above. The integral in Equation 3 ranges

314   from 0 to infinity, such that, absorption at the upper barrier is a certain event given enough time. This

315   means that the superposition model always predicts ceiling level accuracies for all intensities, a

316   prediction clearly inconsistent with the observed behavioral data.

A coactivation model with a deadline

318   An unrealistic assumption of the model described above is that accumulation will always complete,

319   which in the single-barrier diffusion model implies that the monkeys have 100 percent detection

320   accuracy. Given enough time, a diffusion process with drift $\mu > 0$ will almost certainly reach the

321   criterion. From an experimental perspective, this has several implications: the intensity of the

322   stimulus components must be sufficiently high to ensure detection rates of 100% and the temporal

323   window for responding is infinitely long to guarantee that all responses are collected.

324   However, if the temporal window for stimulus detection is limited by a deadline d (we assume that

325    $d > \tau$) the proportion of correct responses is given by the distribution of the detection times at $t =$

326   d. For unimodal stimuli and synchronous audiovisual stimuli, this probability corresponds to the

327   inverse Gaussian distribution at time d, $P(\mathbf{D}_i \leq d) = G(d \mid c, \mu_i, \sigma_i^2)$, with $i \in \{A, V, AV\}$, depending

328   on the modality. The expected detection time, conditional on stimulus detection before d, is again

329   obtained by integration of $t \cdot g(t)$ from $t = 0$ to d (see Eq. 4). The solution has been originally

330   presented in (Schwarz, 1994, Eq. 6)

331

332   In bimodal stimuli with onset asynchrony $0 < \tau < d$ (say, without loss of generality, V(τ)A), it is

333   necessary to distinguish the intervals $[0 \dots \tau]$ and $[\tau \dots d]$ during which the drift (the variance) of the

334   diffusion process amount to μ_V (σ_V^2) and μ_{AV} (σ_{AV}^2), respectively. The probability for correct

335   detections amounts to the sum of the detections within $[0 \dots \tau]$ in which only the first stimulus

336   contributes to the buildup of evidence; and the detections within $[\tau \dots d]$ in which both stimuli are

337   active.

338

$$P\big(\mathbf{D}_{V(\tau)A} \leq d\big) = P\big(\mathbf{D}_{V(\tau)A} \leq \tau\big) + P\big(\tau < \mathbf{D}_{V(\tau)A} \leq d\big)$$

$$= P(\mathbf{D}_V \leq \tau) + \int_{-\infty}^{c} P[\mathbf{X}(\tau) = x] \cdot P(\mathbf{D}_{AV} \leq d - \tau \mid \mathbf{X}(\tau) = x)\, dx$$

$$= G(\tau \mid c, \mu_V, \sigma_V^2) + \int_{-\infty}^{c} w(x, \tau \mid c, \mu_V, \sigma_V^2) \cdot G(d - \tau \mid c - x, \mu_{AV}, \sigma_{AV}^2)\, dx \quad (12)$$

342

with $P[\mathbf{X}(\tau) = x] = w(x, \tau \mid \cdots)$ denoting the density of the activation of the processes not yet

absorbed at $t = \tau$ (Eq. 5). The integrand decomposes into a sum of four terms of the form $\phi(x) \cdot$

$\Phi(x \mid m, s^2)$ that can be integrated using the bivariate Normal distribution (Owen, 1980, Eq.

10,010.1, see Appendix A). For the predicted amount of correct responses, a χ_{26}^2 statistic is obtained

(e.g.,Schwarz, 2006) by the squared difference between the predicted and observed proportion of

responses, divided by the variance for binomial proportions $\pi(1 - \pi)/N$, with $\pi = P(\mathbf{D} \leq d)$

denoting the probabilities for correct detections. To avoid numerical problems close to zero or one, π

was bounded within $[0.01, 0.99]$.

351

The expected detection time, conditional on stimulus detection before the deadline, amounts to

353

$$E\big(\mathbf{D}_{V(\tau)A} \mid \mathbf{D}_{V(\tau)A} \leq d\big) = \frac{1}{P\big(\mathbf{D}_{V(\tau)A} \leq d\big)} \times$$

$$\big[P\big(\mathbf{D}_{V(\tau)A} \leq \tau\big) \cdot E\big(\mathbf{D}_{V(\tau)A} \mid \mathbf{D}_{V(\tau)A} \leq \tau\big) + P\big(\tau < \mathbf{D}_{V(\tau)A} \leq d\big) \cdot$$

$$E\big(\mathbf{D}_{V(\tau)A} \mid \tau < \mathbf{D}_{V(\tau)A} \leq d\big)\big] \quad (13)$$

357

The normalization factor $P\big(\mathbf{D}_{V(\tau)A} \leq d\big)$ is the same as in Equation 12. The first term within the

square brackets is given by Equation 4,

360

$$P\big(\mathbf{D}_{V(\tau)A} \leq \tau\big) \cdot E\big(\mathbf{D}_{V(\tau)A} \mid \mathbf{D}_{V(\tau)A} \leq \tau\big) = P(\mathbf{D}_V \leq \tau) \cdot E(\mathbf{D}_V \mid \mathbf{D}_V \leq \tau) =$$

$$\int_0^{\tau} t \cdot g(t \mid c, \mu_V, \sigma_V^2)\, dt. \quad (14)$$

363

The second term is more complicated,

365

$$P\big(\tau < \mathbf{D}_{V(\tau)A} \le d\big) \cdot E\big(\mathbf{D}_{V(\tau)A} \mid \tau < \mathbf{D}_{V(\tau)A} \le d\big)$$

$$= \int_{-\infty}^{c} P[\mathbf{X}(\tau) = x] \cdot P(\mathbf{D}_{AV} \le d - \tau \mid \mathbf{X}(\tau) = x) \cdot \{\tau + E[\mathbf{D}_{AV} \mid \mathbf{X}(\tau) = x, \mathbf{D}_{AV} \le d - \tau]\} \, dx$$

$$= \int_{-\infty}^{c} w(x, \tau \mid c, \mu_V, \sigma_V^2) \cdot G(d - \tau \mid c - x, \mu_{AV}, \sigma_{AV}^2) \, dx \cdot \tau$$

$$+ \int_{-\infty}^{c} w(x, \tau \mid c, \mu_V, \sigma_V^2) \cdot \int_{0}^{d-\tau} t \cdot g(t \mid c - x, \mu_{AV}, \sigma_{AV}^2) \, dt \, dx. \tag{15}$$

370

371   The $w \cdot G$ term corresponds to Equation 12, multiplied by τ. The double integral decomposes into

372   four $\int_{-\infty}^{y} \phi(x \mid m_1, s_1^2) \cdot \Phi(x \mid m_2, s_2^2) \, dx$ terms (Owen, 1980, Eq. 10,010.1, see Eq. 12 above) and

373   another four terms of the form $\int_{-\infty}^{y} x \cdot \phi(x \mid m_1, s_1^2) \cdot \Phi(x \mid m_2, s_2^2) \, dx$ that match with (Owen,

374   1980, eqs. 10,010.1 and 10,011.1). Details are given in Appendix B, and R code (R Core Team, 2017) is

375   available as online supplementary material.

376

377   For the observable mean response time, we assume again an SOA invariant mean residual μ_M,

378

379    $$E(\mathbf{T}) = E(\mathbf{D}) + \mu_M. \tag{16}$$

380

381   For each monkey, an approximate χ_{25}^2 goodness-of-fit statistic can be determined by the sum of

382   the squared standardized deviations of the predicted and the observed average response times

383   (e.g.,Schwarz, 2006). Compared to the model without deadline, one degree of freedom is lost

384   because the deadline is adjusted to the data. Because the χ^2 statistics for the mean RTs and

385   proportions of correct responses are not independent, we did not combine them but instead

386   transformed them into P-values and maximized the minimum of the P-values as a conservative fitting

387   criterion.

388   **Results for the deadline model**

389   Figure 4 show the results from fitting the diffusion superposition model with the deadline to the

390   behavioral performance of the monkeys (accuracy and mean RTs) as a function of the signal-to-noise

391 ratio and the SOAs. The additional deadline parameter improves the model fits and provides a very

392 good description of both accuracy and RTs of the monkeys. For Monkey 1 the model provided an

393 excellent fit to the data. (Accuracy: $\chi^2_{26} = 34.15$; mean RT: $\chi^2_{25} = 32.99$, $P_{\min} = 0.131$). In Monkey

394 2, the model fit was less convincing (Accuracy: $\chi^2_{26} = 64.63$; mean RT: $\chi^2_{25} = 63.08$, $P_{\min} < 0.001$),

395 but still acceptable given the conservative fitting procedure where we try to jointly fit both the RTs

396 and accuracy of the monkeys. The best fit parameter estimates are shown in Table 1.

397

Table 1. Model parameters

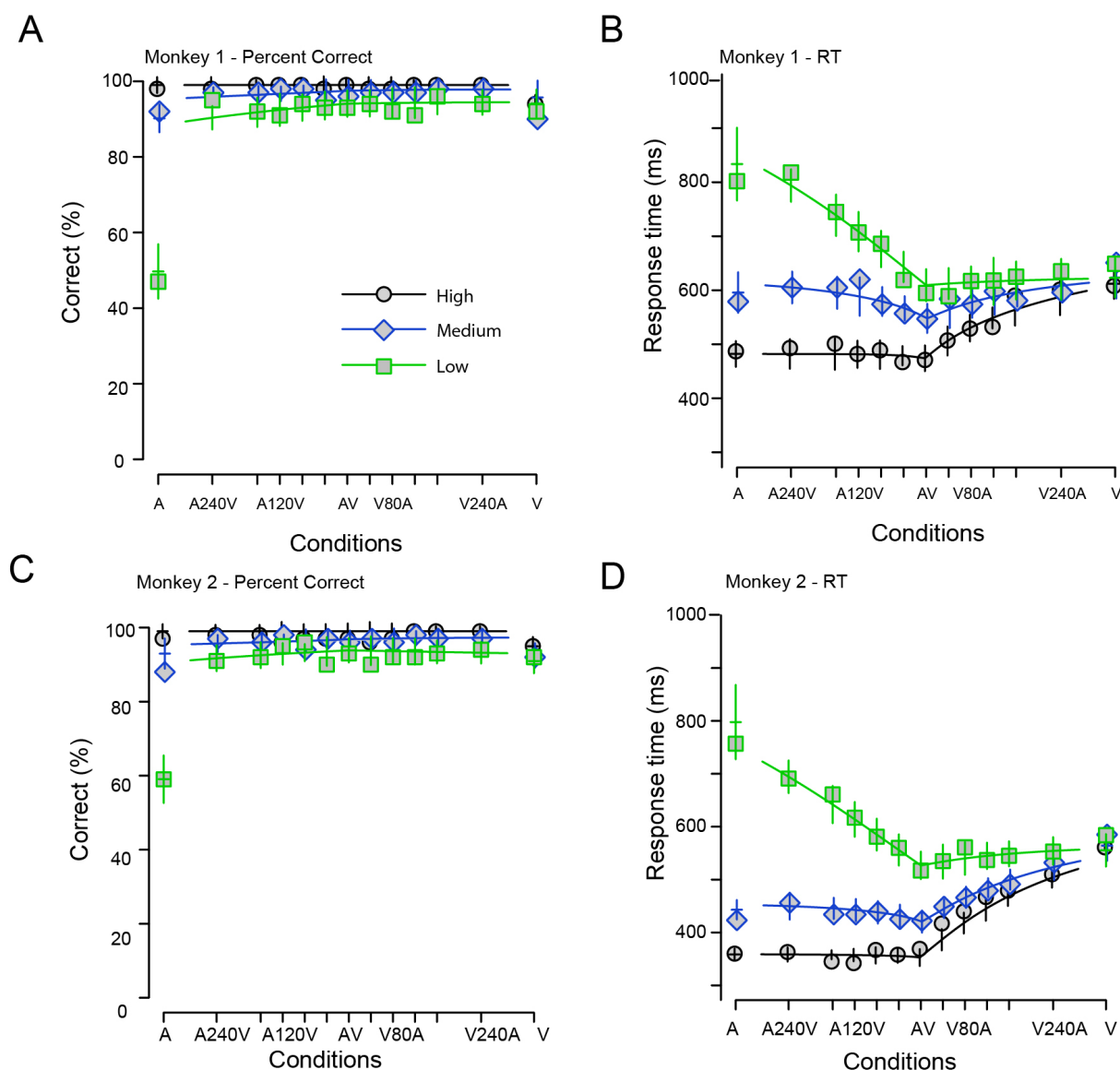| Parameter	Monkey 1	Monkey 2
μ_A (low, medium, high)	0.01, 0.25, 1.53	0.09, 0.41, 9.74
σ_A^2 (low, medium, high)	18.5, 130.3, 74.4	11.7, 379.6, 10.7
μ_V (low, medium, high)	0.35, 0.37, 0.34	0.30, 0.34, 0.37
σ_V^2 (low, medium, high)	73.2, 68.0, 92.8	67.8, 73.1, 41.9
μ_M (msec)	419	343
Deadline d (msec)	951	828

398
399
Figure 4: A diffusion superposition model with a time deadline predicts monkey RTs and response accuracy to audiovisual stimuli
Accuracy and RT of the two monkeys along with the fits from the diffusion superposition model with a deadline. **A, C** Accuracy of the monkeys as a function of SNR and SOA. X-axes show SOA; Y-axes show the percent correct. **B, D** Response time of the monkeys as a function of SNR And SOA. X-axes show SOA; Y-axes show Response Time in ms. In both panels, the high SNRS are shown in black circles, the medium SNRs in blue diamonds and low SNRs in green squares.

Discussion

407

408 The goal of our study was to test if the single-barrier diffusion superposition model (Schwarz, 1989,

409 1994; Diederich, 1995) can describe accuracy and RTs to audiovisual vocalizations of different

410 intensities in a detection task. In the auditory modality the intensity manipulation was effective

411 (Figures 2 and 3). In line with this, the drift estimates μ_A monotonically increased with SNR (Table 1;

412 the variance estimates showed a less regular pattern). In the visual modality, drift estimates μ_V and

413 variance estimates σ_V^2 were more or less equal for the different intensities, which is consistent with

414 the converging pattern of the mean RTs for positive SOAs (Figure 2). The residual μ_M was similar in

415 the two animals, reflecting their overall response speed and the fact that stimulus detection is

416 probably just one of several stages of the overall response process.

417 Consistent with many earlier results in bimodal divided attention (Diederich & Colonius, 2004a;

418 Molholm et al., 2002; Murray et al., 2005), separate activation (a.k.a. race) models were insufficient

419 to explain the behavioral patterns we observed (Miller, 1982, 1986). We used many SOA conditions

420 and thus the majority of the stimuli used in the present study were audiovisual stimuli. Enrichment of

421 audiovisual conditions rules out trial history effects and modality shift effects as the *only* driving force

422 of coactivation effects (Gondan, Lange, Rösler, & Röder, 2004; Miller, 1986; Otto, Dassy, &

423 Mamassian, 2013; Otto & Mamassian, 2012).

424 We have focused on describing and modeling the mean RTs and mean response accuracy for a

425 detection task across different SNRs and SOAs (Cluff et al., 2015; Crevecoeur et al., 2016; Dixon &

426 Spitz, 1980; Holmes, 2009; Meredith et al., 1987; Stein et al., 1989; van Wassenhove et al., 2007).

427 Some studies have addressed the effect of sensory reliability which is related to the sensory intensity

428 manipulation we performed here on benefits of multisensory integration but did not modulate the

429 delay between the sensory stimuli (Drugowitsch, DeAngelis, Klier, Angelaki, & Pouget, 2014). Other

430 studies have examined the dependence on sensory delays but not on sensory intensity (Crevecoeur et

431 al., 2016). Experiments that simultaneously vary stimulus intensity, stimulus reliability and SOA are

432 needed to fully understand the relative roles of these factors in multisensory discrimination(Gondan

433 et al., 2010) .

434   Our study is focused on describing the behavior of monkeys performing an audiovisual

435   detection task. Our model is likely to apply to humans performing similar tasks. We previously

436   showed that the behavior of monkeys and humans in a simpler version of this task that only involved

437   variation in intensity of the sensory stimuli was very similar (Chandrasekaran et al., 2011). Our

438   analysis here expands on a classical coactivation model which has been previously used to

439   successfully model the behavior of a human participant across a range of SOAs (Schwarz, 1994).

440   We have shown that an accumulator, which integrates visual and auditory inputs to a bound,

441   explained the behavioral benefits from multisensory integration. However, in our task design, no

442   explicit trial onset information was provided to the animals. Instead, the stimulus arrived in a

443   continuous ongoing stream. This paradigm has several advantages because it mimics a natural flow of

444   stimuli in the real world and avoids sharp transients in visual stimuli. But, it raises the important

445   question of how an integrator knows when to begin integrating the sensory evidence? One plausible

446   solution is that a neural circuit resets the integrator after either the last behavioral action by the

447   animal (false alarm/correct detection) or after some time has elapsed (Janssen & Shadlen, 2005). The

448   fits may improve by incorporating previous ideas which propose to jointly model the inter stimulus

449   interval as well as the responses to sensory stimuli.

450   The superposition model with a deadline predicts RTs and accuracy of monkeys when they

451   detect *dynamic* visual and auditory stimuli (vocalizations). In other contexts, generalized variants of

452   these coactivation models have been used with dynamic stimuli (Drugowitsch et al., 2014). We

453   believe these types of models may also be applicable to static audiovisual stimuli for two reasons.

454   First, Diffusion models are commonly used with static visual stimuli (Ratcliff, Thapar, & McKoon,

455   2003; Voss, Rothermund, & Brandtstädter, 2008). Second,the superposition model has been used to

456   explain discrimination behavior for *static* audiovisual stimuli (Gondan et al., 2010; Schwarz, 1989,

457   1994). Applying these models to both static and dynamic multisensory stimuli in the same study may

458   help test proposals that there are different mechanisms for the processing of static and dynamic

459   multisensory stimuli (Raposo, Sheppard, Schrater, & Churchland, 2012).

460

461   The key contribution was to show that a model with additive superposition of the channel-specific

462   evidence explains the benefits of integrating faces and voices in animal perception across a wide

463   range of SOAs and SNRs. This class of coactivation models has previously been used to explain

464   response times of human participants in auditory-visual detection tasks (Diederich, 1995; Schwarz,

465   1989, 1994). The emphasis of these additive coactivation models (or more general versions, e.g.,

466   (Drugowitsch et al., 2014)) seems *prima facie* at odds with classical reports promoting superadditive

467   multisensory interaction (Stein & Meredith, 1993). In these studies, superadditivity, and other

468   nonlinear mechanisms were considered fundamental for mediating benefits from multisensory

469   integration. However, as a series of studies have shown, the majority of neurons in classical

470   multisensory brain regions such as the superior colliculus accumulate their synaptic input in a linear

471   manner for a range of stimulus intensities, and nonlinearities are observed only at very low intensities

472   (Dahl, Logothetis, & Kayser, 2010; Populin & Yin, 2002; Skaliora, Doubell, Holmes, Nodal, & King,

473   2004; Stanford, Quessy, & Stein, 2005; Stanford & Stein, 2007; Stein & Stanford, 2008) . Stated

474   differently, additive combination is the norm. For conflicting stimuli (e.g., in temporal order

475   judgment, where participants are asked to report which modality came first), linear summation may

476   occur in the other direction, with the overall evidence corresponding to the difference between the

477   channel-specific activations (Schwarz, 2006).

478

479   Besides linearity of multisensory integration in single neurons, studies increasingly demonstrate that

480   ensembles of neurons (which might encode stimuli nonlinearly at the single neuron level) can

481   perform linear computations (Ma, Beck, Latham, & Pouget, 2006). We believe that our abstract

482   behavioral model might be implemented by adopting frameworks such as probabilistic population

483   codes. For example, computationally, at the population level, linear summation of neural activation is

484   possible and yields optimal solutions for a very general class of computational problems (Beck et al.,

485   2008; Ma et al., 2006). Extensions of this model showed that assuming Poisson-like distributions of

486   spike counts allows biological networks to accumulate evidence while choosing the most likely action

487   (Beck et al., 2008). We believe our description of behavioral data by this linear summation model will

488   assist in relating neurophysiological and modeling studies of multisensory detection and broadly

489   integration (Chandrasekaran, 2016; Fetsch, DeAngelis, & Angelaki, 2013; Ma et al., 2006; Seilheimer

490   et al., 2014).

Acknowledgments

Supplemental material

500   The online supplement to this article includes mean RTs and accuracy rates for the two animals

501   (Tables S1–S4), bootstrap test of the race model (Tables S5–S6), as well as a commented R script with

502   an implementation of the deadline model.

503

504

505

506

507

508 # Tables S1-S6

509

510 *Table S1*. Mean accuracy and bootstrap standard error for different conditions for Monkey 1

	A	A240V	A160V	A120V	A80V	A40V	AV	V40A	V80A	V120A	V160A	V240A	V
High	98(1)	98(1)	99(1)	99(1)	99(1)	98(1)	99(1)	98(1)	98(1)	99(1)	99(1)	99(1)	94(2)
Medium	92(2)	97(1)	97(1)	98(1)	98(1)	95(2)	96(1)	97(1)	97(1)	97(1)	98(1)	98(1)	90(2)
Low	47(4)	95(2)	92(2)	91(2)	94(2)	93(2)	93(2)	94(2)	92(2)	91(2)	96(2)	94(2)	92(2)

511
512 *Table S2*. Mean RT and Standard Error of RTs for different conditions for Monkey 1.

	A	A240V	A160V	A120V	A80V	A40V	AV	V40A	V80A	V120A	V160A	V240A	V
High	487(12)	493(14)	501(15)	482(13)	489(14)	467(9)	471(12)	507(14)	529(13)	532(10)	590(15)	601(17)	608(14)
Medium	579(19)	605(15)	605(15)	620(18)	574(14)	557(12)	547(14)	584(18)	574(15)	598(17)	581(14)	596(14)	651(20)
Low	802(34)	818(15)	745(20)	707(19)	686(17)	619(15)	595(15)	589(15)	617(15)	618(22)	625(18)	635(20)	649(19)

513
514 *Table S3*. Mean accuracy and bootstrap standard error for different conditions for Monkey 2

	A	A240V	A160V	A120V	A80V	A40V	AV	V40A	V80A	V120A	V160A	V240A	V
High	97(1)	98(1)	98(1)	96(1)	97(1)	97(1)	97(1)	96(1)	97(1)	99(1)	99(1)	100(0)	95(1)
Medium	88(2)	97(1)	96(1)	98(1)	94(2)	97(1)	96(1)	97(1)	96(1)	98(1)	97(1)	97(1)	92(2)
Low	59(3)	91(2)	92(2)	95(1)	96(1)	90(2)	93(2)	90(2)	92(2)	92(2)	93(2)	94(2)	92(2)

515
516 *Table S4*. Mean RT and Standard Error of RTs for different conditions for Monkey 2.

	A	A240V	A160V	A120V	A80V	A40V	AV	V40A	V80A	V120A	V160A	V240A	V
High	360(6)	363(7)	345(4)	342(6)	367(8)	358(7)	369(8)	417(10)	440(10)	467(12)	479(9)	510(10)	561(11)
Medium	423(10)	456(13)	434(10)	434(11)	439(11)	425(11)	422(11)	449(9)	466(10)	479(9)	491(10)	532(13)	585(15)
Low	757(36)	691(16)	661(18)	617(17)	581(16)	560(15)	517(13)	535(17)	561(16)	537(13)	545(12)	553(13)	584(16)

517
518 *Table S5*. *P* values for race model test (Monkey 1)

	A240V	A160V	A120V	A80V	A40V	AV	V40A	V80A	V120A	V160A	V240A
High	0.057	0.062	0.014	0.039	0.038	0.031	0.018	0.000	0.000	0.000	0.000
Medium	0.125	0.013	0.029	0.062	0.293	0.002	0.000	0.000	0.000	0.000	0.000
Low	0.043	0.000	0.000	0.000	0.000	0.001	0.000	0.040	0.000	0.007	0.000

519
520 *Table S6*. *P* values for race model test for Monkey 2

	A240V	A160V	A120V	A80V	A40V	AV	V40A	V80A	V120A	V160A	V240A
High	0.001	0.001	0.000	0.004	0.005	0.033	0.000	0.000	0.000	0.000	0.000
Medium	0.008	0.003	0.011	0.041	0.000	0.000	0.004	0.000	0.000	0.000	0.000
Low	0.019	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.000

Appendix A: Probability of absorption

Here we derive the explicit expressions for the probability of absorption in bimodal stimuli with onset asynchrony $0 < \tau < d$ (Eq. 12). Without loss of generality, we consider the case $V(\tau)A$ in which the visual stimulus is presented first. Between $t = 0$ and $t = \tau$, only the visual channel contributes to the build-up of evidence, so the probability of absorption within the interval $(0, \tau)$ is given by

$$P\big(\mathbf{D}_{V(\tau)A} \leq \tau\big) = G(\tau \mid c, \mu_V, \sigma_V^2) \tag{A.1}$$

with G denoting the inverse Gaussian distribution (Eq. 2). Later, within the time interval (τ, d), the probability of absorption is the mixture of absorption probabilities of those processes still active at time τ, with the barrier depending on the activation $\mathbf{X}(\tau) < c$, weighted by the density of processes at $\mathbf{X}(\tau)$. Let $d' = d - \tau$. Then,

$$P\big(\tau < \mathbf{D}_{V(\tau)A} \leq d\big) = \int_{-\infty}^{c} w(x, \tau \mid c, \mu_V, \sigma_V^2) \cdot G(d' \mid c - x, \mu_{AV}, \sigma_{AV}^2) \, dx, \tag{A.2}$$

with w given by Equation 5. The integrand in (A.2) can be transformed into four integrals of the form $q \int_{-\infty}^{c} \exp(rx) \, \phi(x \mid m_1, s_1^2) \cdot \Phi(x \mid m_2, s_2^2) \, dx$:

$$\int_{-\infty}^{c} w(x, \tau \mid c, \mu_V, \sigma_V^2) \cdot G(d' \mid c - x, \mu_{AV}, \sigma_{AV}^2) \, dx$$

$$= \int_{-\infty}^{c} \phi(x \mid \mu_V \tau, \sigma_V^2 \tau) \, \Phi(x \mid c - \mu_{AV} d', \sigma_{AV}^2 d') \, dx$$

$$+ \exp\left(\frac{2c\mu_{AV}}{\sigma_{AV}^2}\right) \int_{-\infty}^{c} \exp\left(-\frac{2\mu_{AV}}{\sigma_{AV}^2} x\right) \phi(x \mid \mu_V \tau, \sigma_V^2 \tau) \Phi(x \mid c + \mu_{AV} d', \sigma_{AV}^2 d') \, dx$$

$$- \exp\left(\frac{2c\mu_V}{\sigma_V^2}\right) \int_{-\infty}^{c} \phi(x \mid 2c + \mu_V \tau, \sigma_V^2 \tau) \, \Phi(x \mid c - \mu_{AV} d', \sigma_{AV}^2 d') \, dx$$

$$- \exp\left(\frac{2c\mu_V}{\sigma_V^2} + \frac{2c\mu_{AV}}{\sigma_{AV}^2}\right) \int_{-\infty}^{c} \exp\left(-\frac{2\mu_{AV}}{\sigma_{AV}^2} x\right) \phi(x \mid 2c + \mu_V \tau, \sigma_V^2 \tau) \, \Phi(x \mid c + \mu_{AV} d', \sigma_{AV}^2 d') \tag{A.3}$$

By completing the square, we have $\exp(rx) \, \phi(x \mid m_1, s_1^2) = \exp\left(rm_1 + \frac{r^2 s_1^2}{2}\right) \phi(x \mid m_1', s_1^2)$, with $m_1' = m_1 + r s_1^2$. Let $u = \frac{x - m_1'}{s_1}$, or $x = u s_1 + m_1'$. Then, the integral can be rewritten as

547 $\quad \int_{-\infty}^{c} \phi(x \mid m_1', s_1^2) \cdot \Phi(x \mid m_2, s_2^2) \, dx = \int_{-\infty}^{(c-m_1')/s_1} \phi(u) \cdot \Phi\left(u \mid \frac{m_2 - m_1'}{s_1}, \frac{s_2^2}{s_1^2}\right) du$

548 $\quad\quad = \int_{-\infty}^{(c-m_1')/s_1} \phi(u) \cdot \Phi\left(\frac{s_1}{s_2} u + \frac{m_1' - m_2}{s_2}\right) du.$ (A.4)

549

550 The form of (A.4) now matches Owen (1980, Eq. 10,010.1) which can be determined by the bivariate

551 Normal distribution, $\int_{-\infty}^{y} \phi(u) \cdot \Phi(a + bu) \, du = \Phi\Phi\left(\frac{a}{\sqrt{1+b^2}}, y \mid \varrho = -\frac{b}{\sqrt{1+b^2}}\right)$. An implementation in R (R

552 Core Team, 2017) is available as online supplementary material.

553 **Appendix B: Conditional mean response time**

554 Here we derive the explicit expressions for the conditional mean response time for bimodal stimuli with

555 onset asynchrony τ, conditional on absorption before the deadline (Eqs. 13–15). We consider again the

556 case $V(\tau)A$ in which the visual stimulus is presented first. Between $t = 0$ and $t = \tau$, only the visual channel

557 contributes to the build-up of evidence, so the conditional mean RT is given by Schwarz (1994, Equation

558 A.2).

559

560 $\quad P(\mathbf{D}_{V(\tau)A} \leq \tau) \cdot E(\mathbf{D}_{V(\tau)A} \mid \mathbf{D}_{V(\tau)A} \leq \tau)$

561 $\quad\quad = \int_0^{\tau} t \cdot g(t \mid c, \mu_V, \sigma_V^2) \, dt = \frac{c}{\mu_V}\left[\Phi(\mu_V \tau \mid c, \sigma_V^2 \tau) - \exp\left(\frac{2c\mu_V}{\sigma_V^2}\right) \Phi(-\mu_V \tau \mid c, \sigma_V^2 \tau)\right]$ (B.1)

562

563 Later, within the time interval (τ, d), the expected detection time is again a mixture of expected detection

564 times for the processes still active at time τ, weighted by the density of processes at $\mathbf{X}(\tau)$. These processes

565 now have increased drift μ_{AV} and variance σ_{AV}^2; the work to be done (i.e., the barrier) depends on the

566 activation $\mathbf{X}(\tau) < c$. Note that τ milliseconds have already passed since stimulus onset, hence the

567 remaining time is $d' = d - \tau$.

568

569 $\quad P(\tau < \mathbf{D}_{V(\tau)A} \leq d) \cdot E(\mathbf{D}_{V(\tau)A} \mid \tau < \mathbf{D}_{V(\tau)A} \leq d)$

570 $\quad\quad = \int_{-\infty}^{c} P[\mathbf{X}(\tau) = x] \cdot P(\mathbf{D}_{AV} \leq d' \mid \mathbf{X}(\tau) = x) \cdot \{\tau + E[\mathbf{D}_{AV} \mid \mathbf{X}(\tau) = x, \mathbf{D}_{AV} \leq d']\} \, dx$

571 $\quad\quad = \int_{-\infty}^{c} w(x, \tau \mid c, \mu_V, \sigma_V^2) \cdot G(d' \mid c - x, \mu_{AV}, \sigma_{AV}^2) \, dx \cdot \tau$

572 $\quad\quad\quad + \int_{-\infty}^{c} w(x, \tau \mid c, \mu_V, \sigma_V^2) \cdot \int_0^{d'} t \cdot g(t \mid c - x, \mu_{AV}, \sigma_{AV}^2) \, dt \, dx$ (B.2)

573

574   The first term corresponds exactly to (A.2), multiplied by the onset asynchrony τ. See again Schwarz (1994,

575   Equation A.2) for $\int_0^{d'} t \cdot g(t \mid c - x, \mu_{AV}, \sigma_{AV}^2)\, dt$. The double integral in (B.2) can then be rewritten as

576

577      $\displaystyle\int_{-\infty}^{c} \left[\phi(x \mid \mu_V \tau, \sigma_V^2 \tau) - \exp\left(\frac{2c\mu_V}{\sigma_V^2}\right) \cdot \phi(x \mid 2c + \mu_V \tau, \sigma_V^2 \tau) \right]$

578          $\times \dfrac{c-x}{\mu_{AV}} \cdot \left[\Phi(x \mid c - \mu_{AV} d', \sigma_{AV}^2 d') - \exp\left(\frac{2(c-x)\mu_{AV}}{\sigma_{AV}^2}\right) \Phi(x \mid c + \mu_{AV} d', \sigma_{AV}^2 d') \right]\, dx$

579          $= \dfrac{c}{\mu_{AV}} \displaystyle\int_{-\infty}^{c} \phi(x \mid \mu_V \tau, \sigma_V^2 \tau)\, \Phi(x \mid c - \mu_{AV} d', \sigma_{AV}^2 d')\, dx$

580          $- \dfrac{c}{\mu_{AV}} \exp\left(\frac{2c\mu_V}{\sigma_V^2}\right) \displaystyle\int_{-\infty}^{c} \phi(x \mid 2c + \mu_V \tau, \sigma_V^2 \tau)\, \Phi(x \mid c - \mu_{AV} d', \sigma_{AV}^2 d')\, dx$

581          $- \dfrac{c}{\mu_{AV}} \exp\left(\frac{2c\mu_{AV}}{\sigma_{AV}^2}\right) \displaystyle\int_{-\infty}^{c} \exp\left(-\frac{2\mu_{AV}x}{\sigma_{AV}^2}\right) \phi(x \mid \mu_V \tau, \sigma_V^2 \tau)\, \Phi(x \mid c + \mu_{AV} d', \sigma_{AV}^2 d')\, dx$

582          $+ \dfrac{c}{\mu_{AV}} \exp\left(\frac{2c\mu_V}{\sigma_V^2} + \frac{2c\mu_{AV}}{\sigma_{AV}^2}\right) \displaystyle\int_{-\infty}^{c} \exp\left(-\frac{2\mu_{AV}x}{\sigma_{AV}^2}\right) \phi(x \mid 2c + \mu_V \tau, \sigma_V^2 \tau)\, \Phi(x \mid c + \mu_{AV} d', \sigma_{AV}^2 d')\, dx$

583          $- \dfrac{1}{\mu_{AV}} \displaystyle\int_{-\infty}^{c} x\, \phi(x \mid \mu_V \tau, \sigma_V^2 \tau)\, \Phi(x \mid c - \mu_{AV} d', \sigma_{AV}^2 d')\, dx$

584          $+ \dfrac{1}{\mu_{AV}} \exp\left(\frac{2c\mu_V}{\sigma_V^2}\right) \displaystyle\int_{-\infty}^{c} x\, \phi(x \mid 2c + \mu_V \tau, \sigma_V^2 \tau)\, \Phi(x \mid c - \mu_{AV} d', \sigma_{AV}^2 d')\, dx$

585          $+ \dfrac{1}{\mu_{AV}} \exp\left(\frac{2c\mu_{AV}}{\sigma_{AV}^2}\right) \displaystyle\int_{-\infty}^{c} x \exp\left(-\frac{2\mu_{AV}x}{\sigma_{AV}^2}\right) \phi(x \mid \mu_V \tau, \sigma_V^2 \tau)\, \Phi(x \mid c + \mu_{AV} d', \sigma_{AV}^2 d')\, dx$

586          $- \dfrac{1}{\mu_{AV}} \exp\left(\frac{2c\mu_V}{\sigma_V^2} + \frac{2c\mu_{AV}}{\sigma_{AV}^2}\right) \displaystyle\int_{-\infty}^{c} x \exp\left(-\frac{2\mu_{AV}x}{\sigma_{AV}^2}\right) \phi(x \mid 2c + \mu_V \tau, \sigma_V^2 \tau)\, \Phi(x \mid c + \mu_{AV} d', \sigma_{AV}^2 d')\, dx$ (B.3)

587

588   The first four terms correspond to (A.3), multiplied by a constant $(\pm c/\mu_{AV})$. By completing the square, we

589   have again $\exp(rx)\, \phi(x \mid m_1, s_1^2) = \exp\left(r m_1 + \frac{r^2 s_1^2}{2}\right) \phi(x \mid m_1', s_1^2)$, with $m_1' = m_1 + r s_1^2$.

590   Let $u = \dfrac{x - m_1'}{s_1}$, or $x = m_1' + u s_1$. Then, the integral can be rewritten as

591

592      $\displaystyle\int_{-\infty}^{c} x \cdot \phi(x \mid m_1', s_1^2) \cdot \Phi(x \mid m_2, s_2^2)\, dx$

593          $= \displaystyle\int_{-\infty}^{(c - m_1')/s_1} (m_1' + u s_1) \cdot \phi(u \mid 0, 1) \cdot \Phi\left(u \,\Big|\, \frac{m_2 - m_1'}{s_1}, \frac{s_2^2}{s_1^2}\right) du$

594 $$= m_1' \cdot \int_{-\infty}^{(c-m_1')/s_1} \phi(u \mid 0,1) \cdot \Phi\left(\frac{s_1}{s_2}u + \frac{m_1'-m_2}{s_2} \,\middle|\, 0,1\right) du$$

595 $$+ s_1 \cdot \int_{-\infty}^{(c-m_1')/s_1} u \cdot \phi(u \mid 0,1) \cdot \Phi\left(\frac{s_1}{s_2}u + \frac{m_1'-m_2}{s_2} \,\middle|\, 0,1\right) du \tag{B.4}$$

596

597 The first term of (B.4) matches again (Owen, 1980, Eq. 10,010.1). The second term matches Eq. 10,010.1

598 from (Owen, 1980, Eq. 10,010.1) and is calculated by

599 $\int_{-\infty}^{y} u \cdot \phi(u) \cdot \Phi(a + bu)\, du = \frac{b}{\sqrt{1+b^2}} \phi\left(\frac{b}{\sqrt{1+b^2}}\right) \Phi\left(y\sqrt{1+b^2} + \frac{ab}{\sqrt{1+b^2}}\right) - \phi(y)\Phi(a + bu)$. A

600 implementation in R (R Core Team, 2017) is available as an online supplement.

Author Contributions

CC and MG designed the experiment. CC performed experiments. CC and MG analyzed the data. MG developed the formulation of the deadline model. CC and MG wrote the paper. Both authors evaluated data and modeling results.

Accuracy and mean response time in a diffusion superposition model with deadline

Online supplement for "Audiovisual detection at different intensities and delays"

This online supplement provides implementation details on the derivation of the predictions for mean response time and accuracy for the diffusion superposition model (Diederich, 1995; Schwarz, 1994) with a deadline. For simplicity, we reiterate the relevant parts of the methods section here and then add code in R statistical language for the different equations. The R code (R Core Team, 2017) includes the necessary defaults that allow testing and deployment in other analyses.

Libraries

The code requires the inverse Gaussian distribution package SuppDists (Wheeler, 2013, available from CRAN). In addition, package mvtnorm (Genz et al., 2014) is used for the bivariate Normal distribution.

```
#
# Implementation of the inverse Gaussian distribution (Wheeler, 2016, install from CRAN first)
#
library(SuppDists)

#
# Multivariate Normal distribution (Genz et al., 2016)
#
library(mvtnorm)
```

Accuracy

For unimodal/synchronous stimuli, accuracy is given by the inverse Gaussian distribution at the deadline d. For example, Monkey 1's accuracy in Condition v (low intensity) is given by

acc_sync(d=1000, c=100, mu=0.13, sigma2=4.3^2).

```
#
# Accuracy in unimodal and synchronous stimuli
#
acc_sync = function(d, c, mu, sigma2)
{
        pinvGauss(d, nu=c/mu, lambda=c*c/sigma2)
```

```
}
```

For stimuli with onset asynchrony τ, accuracy is given by (12) which is the sum of the inverse Gaussian distribution at time τ and four integrals of the form $q \cdot \int_{-\infty}^{c} \exp(rx) \cdot \phi(x \mid m_1, s_1^2) \cdot \Phi(x \mid m_2, s_2^2) \, dx$:

1. $q = 1, r = 0, m_1 = \mu_V \tau, s_1^2 = \sigma_V^2 \tau, m_2 = c - \mu_{AV} d', s_2^2 = \sigma_{AV}^2 d'.$

2. $q = \exp\left(\frac{2c\mu_{AV}}{\sigma_{AV}^2}\right), r = -\frac{2\mu_{AV}}{\sigma_{AV}^2}, m_1 = \mu_V \tau, s_1^2 = \sigma_V^2 \tau, m_2 = c + \mu_{AV} d', s_2^2 = \sigma_{AV}^2 d'.$

3. $q = -\exp\left(\frac{2c\mu_V}{\sigma_V^2}\right), r = 0, m_1 = 2c + \mu_V \tau, s_1^2 = \sigma_V^2 \tau, m_2 = c - \mu_{AV} d', s_2^2 = \sigma_{AV}^2 d'$

4. $q = -\exp\left(\frac{2c\mu_V}{\sigma_V^2} + \frac{2c\mu_{AV}}{\sigma_{AV}^2}\right), r = -\frac{2\mu_{AV}}{\sigma_{AV}^2}, m_1 = 2c + \mu_V \tau, s_1^2 = \sigma_V^2 \tau, m_2 = c + \mu_{AV} d', s_2^2 = \sigma_{AV}^2 d'$

with $d' = d - \tau$. Three convenience functions evaluate these integrals using Eq. 10,010.1 in (Owen, 1980):

$$\int_{-\infty}^{c} \exp(rx) \, \phi(x \mid m_1, s_1^2) \Phi(x \mid m_2, s_2^2) \, dx = \exp\left(rm_1 + \frac{r^2 s_1^2}{2}\right) \int_{-\infty}^{c} \phi(x \mid m_1', s_1^2) \cdot \Phi(x \mid m_2, s_2^2) \, dx,$$

$$\int_{-\infty}^{c} \phi(x \mid m_1', s_1^2) \cdot \Phi(x \mid m_2, s_2^2) \, dx = \int_{-\infty}^{(c-m_1')/s_1} \phi(u) \cdot \Phi\left(\frac{s_1}{s_2} u + \frac{m_1' - m_2}{s_2}\right) du,$$

$$\int_{-\infty}^{y} \phi(u) \cdot \Phi(a + bu) \, du = \Phi\Phi\left(\frac{a}{\sqrt{1+b^2}}, y \,\middle|\, \varrho = -\frac{b}{\sqrt{1+b^2}}\right), \text{ with } m_1' = m_1 + rs_1^2.$$

```
#
# Integrate dnorm(x) * pnorm(a + b*x) from -Inf to y (Owen, 1980, Eq. 10,010.1)
#
owen10_010.1 = function(y, a, b)
{
        rho = -b/sqrt(1 + b*b)
        pmvnorm(lower=c(-Inf, -Inf), upper=c(a/sqrt(1+b*b), y), corr=matrix(c(1, rho, rho, 1), nrow=2))
}


#
# Integrate dnorm(x | mu1, sigma1) * pnorm(x | mu2, sigma2) from -Inf to y
#
owen10_010.1b = function(y, mu1, sigma1, mu2, sigma2)
{
        owen10_010.1(y=(y - mu1)/sigma1, a=(mu1 - mu2)/sigma2, b=sigma1/sigma2)
}

#
```

```
# Integrate exp(rx) * dnorm(x | mu1, sigma1) * pnorm(x | mu2, sigma2) from -Inf to y
#
owen10_010.1c = function(y, r, mu1, sigma1, mu2, sigma2)
{
        exp(r * mu1 + r^2 * sigma1^2 / 2) * owen10_010.1b(y, mu1 + r * sigma1^2, sigma1, mu2, sigma2)
}


acc_async = function(d, c, mua, sigma2a, mub, sigma2b, tau)
{
        muab = mua + mub
        sigma2ab = sigma2a + sigma2b

        # Probability of absorption within 0...tau
        p0 = acc_sync(d=tau, c=c, mu=mua, sigma2=sigma2a)

        # 1st term of Equation A.3
        q = 1
        r = 0
        mu1 = mua * tau
        sigma1=sqrt(sigma2a * tau)
        mu2 = c - muab * (d - tau)
        sigma2 = sqrt(sigma2ab * (d - tau))
        p1 = q * owen10_010.1c(c, r, mu1, sigma1, mu2, sigma2)

        # 2nd integral
        q = exp(2 * c * muab / sigma2ab)
        r = -2 * muab / sigma2ab
        mu1 = mua * tau
        sigma1 = sqrt(sigma2a * tau)
        mu2 = c + muab * (d - tau)
        sigma2 = sqrt(sigma2ab * (d - tau))
        p2 = q * owen10_010.1c(c, r, mu1, sigma1, mu2, sigma2)

        # 3rd integral
        q = -exp(2 * c * mua / sigma2a)
        r = 0
        mu1 = 2 * c + mua * tau
        sigma1 = sqrt(sigma2a * tau)
        mu2 = c - muab * (d - tau)
        sigma2 = sqrt(sigma2ab * (d - tau))
        p3 = q * owen10_010.1c(c, r, mu1, sigma1, mu2, sigma2)

        # 4th integral
        q = -exp(2 * c * mua / sigma2a + 2 * c * muab / sigma2ab)
        r = -2 * muab / sigma2ab
```

```
        mu1 = 2 * c + mua * tau
        sigma1 = sqrt(sigma2a * tau)
        mu2 = c + muab * (d - tau)
        sigma2 = sqrt(sigma2ab * (d - tau))
        p4 = q * owen10_010.1c(c, r, mu1, sigma1, mu2, sigma2)


        p0 + p1 + p2 + p3 + p4
}
```

For example, Monkey 1's accuracy in Condition v100a (low intensity) is given by

acc_async(d=1000, c=100, mua=0.13, sigma2a=4.3^2, mub=0.34, sigma2b=11.7^2, tau=67).

Mean response time

For unimodal/synchronous stimuli, the mean RT is given by Equation 4 which integrates the product of the

time and the inverse Gaussian density from zero until the deadline d. For example, Monkey 1's mean RT in

Condition v (low intensity) is given by mrt_sync(d=1000, c=100, mu=0.13, sigma2=4.3^2).

```
#
# Mean RT in unimodal and synchronous stimuli
#
mrt_sync = function(d, c, mu, sigma2)
{
        # Integral t * density from 0 to d (Schwarz, 1994, Equation A.2)
        m = c / mu * {pnorm(mu*d, c, sqrt(sigma2*d)) -
                exp(2*c*mu / sigma2) * pnorm(pnorm(-mu*d, c, sqrt(sigma2*d))}

        # Normalize with detection accuracy
        m / acc_sync(d, c, mu, sigma2)
}
```

For stimuli with onset asynchrony τ, mean RT is given by Equations 13–15 which boil down to Equations 4 and

12, and four integrals of the form $q \cdot \int_{-\infty}^{c} \exp(rx) \cdot \phi(x \mid m_1, s_1^2) \cdot \Phi(x \mid m_2, s_2^2) \, dx$, as well as four integrals

of the form $q \cdot \int_{-\infty}^{c} x \cdot \exp(rx) \cdot \phi(x \mid m_1, s_1^2) \cdot \Phi(x \mid m_2, s_2^2) \, dx$

1. $q = \frac{c}{\mu_{AV}}, r = 0, m_1 = \mu_V \tau, s_1^2 = \sigma_V^2 \tau, m_2 = c - \mu_{AV} d', s_2^2 = \sigma_{AV}^2 d'.$

2. $q = -\frac{c}{\mu_{AV}} \cdot \exp\left(-\frac{2c\mu_{AV}}{\sigma_{AV}^2}\right), r = -\frac{2\mu_{AV}}{\sigma_{AV}^2}, m_1 = \mu_V \tau, s_1^2 = \sigma_V^2 \tau, m_2 = c + \mu_{AV} d', s_2^2 = \sigma_{AV}^2 d'.$

3. $q = -\frac{c}{\mu_{AV}} \cdot \exp\left(\frac{2c\mu_V}{\sigma_V^2}\right), r = 0, m_1 = 2c + \mu_V \tau, s_1^2 = \sigma_V^2 \tau, m_2 = c - \mu_{AV} d', s_2^2 = \sigma_{AV}^2 d'$

4. $q = \frac{c}{\mu_{AV}} \cdot \exp\left(\frac{2c\mu_V}{\sigma_V^2} + \frac{2c\mu_{AV}}{\sigma_{AV}^2}\right), r = -\frac{2\mu_{AV}}{\sigma_{AV}^2}, m_1 = 2c + \mu_V\tau, s_1^2 = \sigma_V^2\tau, m_2 = c + \mu_{AV}d', s_2^2 = \sigma_{AV}^2 d'$

5. $q = \frac{-1}{\mu_{AV}}, r = 0, m_1 = \mu_V\tau, s_1^2 = \sigma_V^2\tau, m_2 = c - \mu_{AV}d', s_2^2 = \sigma_{AV}^2 d'.$

6. $q = \frac{1}{\mu_{AV}} \cdot \exp\left(\frac{2c\mu_V}{\sigma_V^2}\right), r = 0, m_1 = 2c + \mu_V\tau, s_1^2 = \sigma_V^2\tau, m_2 = c - \mu_{AV}d', s_2^2 = \sigma_{AV}^2 d'$

7. $q = \frac{1}{\mu_{AV}} \cdot \exp\left(-\frac{2c\mu_{AV}}{\sigma_{AV}^2}\right), r = -\frac{2\mu_{AV}}{\sigma_{AV}^2}, m_1 = \mu_V\tau, s_1^2 = \sigma_V^2\tau, m_2 = c + \mu_{AV}d', s_2^2 = \sigma_{AV}^2 d'.$

8. $q = \frac{-1}{\mu_{AV}} \cdot \exp\left(\frac{2c\mu_V}{\sigma_V^2} + \frac{2c\mu_{AV}}{\sigma_{AV}^2}\right), r = -\frac{2\mu_{AV}}{\sigma_{AV}^2}, m_1 = 2c + \mu_V\tau, s_1^2 = \sigma_V^2\tau, m_2 = c + \mu_{AV}d', s_2^2 = \sigma_{AV}^2 d'$

We defined again convenience functions that transform these integrals to two expressions that match Equations 10,010.1 and 10,011.1 in (Owen, 1980):

$$\int_{-\infty}^{c} x \exp(rx)\,\phi(x \mid m_1, s_1^2)\Phi(x \mid m_2, s_2^2)\,dx$$

$$= \exp\left(rm_1 + \frac{r^2 s_1^2}{2}\right) \int_{-\infty}^{c} x \cdot \phi(x \mid m_1', s_1^2) \cdot \Phi(x \mid m_2, s_2^2)\,dx,$$

$$\int_{-\infty}^{c} x \cdot \phi(x \mid m_1', s_1^2) \cdot \Phi(x \mid m_2, s_2^2)\,dx$$

$$= \int_{-\infty}^{\frac{(c-m_1')}{s_1}} (s_1 u + m_1') \cdot \phi(u) \cdot \Phi\left(\frac{s_1}{s_2}u + \frac{m_1' - m_2}{s_2}\right) du$$

$$= m_1' \int_{-\infty}^{\frac{c-m_1'}{s_1}} \phi(u) \cdot \Phi\left(\frac{s_1}{s_2}u + \frac{m_1' - m_2}{s_2}\right) du$$

$$+ s_1 \int_{-\infty}^{\frac{c-m_1'}{s_1}} u \cdot \phi(u) \cdot \Phi\left(\frac{s_1}{s_2}u + \frac{m_1' - m_2}{s_2}\right) du$$

$$\int_{-\infty}^{y} \phi(u) \cdot \Phi(a + bu)\,du = \text{(see above)}$$

$$\int_{-\infty}^{y} u \cdot \phi(u) \cdot \Phi(a + bu)\,du = \frac{b}{\sqrt{1+b^2}} \cdot \phi\left(\frac{a}{\sqrt{1+b^2}}\right) \cdot \Phi\left(u\sqrt{1+b^2} + \frac{ab}{\sqrt{1+b^2}}\right) - \phi(u) \cdot \Phi(a + bu)$$

```
#
# Integrate dnorm(x) * pnorm(a + b*x) from -Inf to y (Owen, 1980, Eq. 10,011.1)
#
owen10_011.1 = function(y, a, b)
{
        bb = sqrt(1 + b*b)
        b/bb * dnorm(a/bb) * pnorm(y*bb + a*b/bb) - pnorm(a + b*y)*dnorm(y)
}


#
# Integrate x * dnorm(x | mu1, sigma1) * pnorm(x | mu2, sigma2) from -Inf to y
```

```
#
owen10_011.1b = function(y, mu1, sigma1, mu2, sigma2)
{
        owen10_011.1(y=(y - mu1)/sigma1, a=(mu1 - mu2)/sigma2, b=sigma1/sigma2) * sigma1 +
           mu1 * owen10_010.1(y=(y - mu1)/sigma1, a=(mu1 - mu2)/sigma2, b=sigma1/sigma2)
}


#
# Integrate x * exp(rx) * dnorm(x | mu1, sigma1) * pnorm(x | mu2, sigma2) from -Inf to y
#
owen10_010.1c = function(y, r, mu1, sigma1, mu2, sigma2)
{
        exp(mu1 * v + sigma1^2 * v^2 / 2) * owen10_011.1b(y, mu1 + v * sigma1^2, sigma1, mu2, sigma2)
}
```

For example, Monkey 1's mean RT in Condition v100a (low intensity) is predicted to

mrt_async(d=1000, c=100, mua=0.13, sigma2a=4.3^2, mub=0.34, sigma2b=11.7^2, tau=67).

```
mrt_async = function(d, c, mua, sigma2a, mub=0.34, sigma2b, tau)
{
        muab = mua + mub
        sigma2ab = sigma2a + sigma2b
        d_ = d - tau

        # Integral t * density from 0 to tau (Schwarz, 1994, Equation A.2)
        m0 = c / mua * {pnorm(mua*tau, c, sqrt(sigma2a*tau)) -
               exp(2*c*mua / sigma2a) * pnorm(pnorm(-mua*tau, c, sqrt(sigma2a*tau))}

        # First term of Eq. 15 (tau * second term of Eq. 12)
        mtau = tau*{acc_async(d, c, mua, sigma2a, mub, sigma2b, tau) - acc_sync(d, c, mua, sigma2a)}

        # 1st integral in B.3
        q = c / mua
        r = 0
        mu1 = mua * tau
        sigma1=sqrt(sigma2a * tau)
        mu2 = c - muab * d_
        sigma2 = sqrt(sigma2ab * d_)
        m1 = q * owen10_010.1c(c, r, mu1, sigma1, mu2, sigma2)

        # 2nd integral
        q = -c / muab * exp(2 * c * muab / sigma2ab)
        r = -2 * muab / sigma2ab
```

```
mu1 = mua * tau
sigma1 = sqrt(sigma2a * tau)
mu2 = c + muab * d_
sigma2 = sqrt(sigma2ab * d_)
m2 = q * owen10_010.1c(c, r, mu1, sigma1, mu2, sigma2)

# 3rd integral
q = -c / muab * exp(2 * c * mua / sigma2a)
r = 0
mu1 = 2 * c + mua * tau
sigma1 = sqrt(sigma2a * tau)
mu2 = c - muab * d_
sigma2 = sqrt(sigma2ab * d_)
m3 = q * owen10_010.1c(c, r, mu1, sigma1, mu2, sigma2)

# 4th integral
q = c / muab * exp(2 * c * mua / sigma2a + 2 * c * muab / sigma2ab)
r = -2 * muab / sigma2ab
mu1 = 2 * c + mua * tau
sigma1 = sqrt(sigma2a * tau)
mu2 = c + muab * d_
sigma2 = sqrt(sigma2ab * d_)
m4 = q * owen10_010.1c(c, r, mu1, sigma1, mu2, sigma2)

# 5th integral in B.3
q = -1 / mua
r = 0
mu1 = mua * tau
sigma1 = sqrt(sigma2a * tau)
mu2 = c - muab * d_
sigma2 = sqrt(sigma2ab * d_)
m5 = q * owen10_011.1c(c, r, mu1, sigma1, mu2, sigma2)

# 6th integral
q = 1 / muab * exp(2 * c * muab / sigma2ab)
r = -2 * muab / sigma2ab
mu1 = mua * tau
sigma1 = sqrt(sigma2a * tau)
mu2 = c + muab * d_
sigma2 = sqrt(sigma2ab * d_)
m6 = q * owen10_011.1c(c, r, mu1, sigma1, mu2, sigma2)

# 7th integral
q = 1 / muab * exp(2 * c * mua / sigma2a)
r = 0
```

```
    mu1 = 2 * c + mua * tau
    sigma1 = sqrt(sigma2a * tau)
    mu2 = c - muab * d_
    sigma2 = sqrt(sigma2ab * d_)
    m7 = q * owen10_011.1c(c, r, mu1, sigma1, mu2, sigma2)

    # 8th integral
    q = -1 / muab * exp(2 * c * mua / sigma2a + 2 * c * muab / sigma2ab)
    r = -2 * muab / sigma2ab
    mu1 = 2 * c + mua * tau
    sigma1 = sqrt(sigma2a * tau)
    mu2 = c + muab * d_
    sigma2 = sqrt(sigma2ab * d_)
    m8 = q * owen10_011.1c(c, r, mu1, sigma1, mu2, sigma2)

    # Return value: normalized integral t * density from 0 to d (Eq. 13)
    p = acc_async(d, c, mua, sigma2a, mub, sigma2b, tau)
    (m0 + mtau + m1 + m2 + m3 + m4 + m5 + m6 + m7 + m8) / p
}
```

References

Beck, J. M., Ma, W. J., Kiani, R., Hanks, T., Churchland, A. K., Roitman, J., et al. (2008). Probabilistic population codes for Bayesian decision making. *Neuron,* 60(6), 1142-1152.

Besle, J., Fort, A., Delpuech, C., & Giard, M.-H. (2004). Bimodal speech: early suppressive visual effects in human auditory cortex. *Eur J Neurosci,* 20(8), 2225-2234.

Chandrasekaran, C. (2016). Computational Models and Principles of Multisensory Integration. *Current Opinion in Neurobiology,* 43(Neurobiology of Learning and Plasticity).

Chandrasekaran, C., Lemus, L., & Ghazanfar, A. A. (2013). Dynamic faces speed up the onset of auditory cortical spiking responses during vocal detection. *Proc Natl Acad Sci U S A,* 110(48), E4668-4677.

Chandrasekaran, C., Lemus, L., Trubanova, A., Gondan, M., & Ghazanfar, A. (2011). Monkeys and Humans Share a Common Computation for Face/Voice Integration. *PLoS Comput Biol,* 7(9), e1002165.

Chandrasekaran, C., Turesson, H. K., Brown, C. H., & Ghazanfar, A. A. (2010). The Influence of Natural Scene Dynamics on Auditory Cortical Activity. *J. Neurosci.,* 30(42), 13919-13931.

Cluff, T., Crevecoeur, F., & Scott, S. H. (2015). A perspective on multisensory integration and rapid perturbation responses. *Vision Res,* 110(Pt B), 215-222.

Colonius, H., & Diederich, A. (2006). The race model inequality: interpreting a geometric measure of the amount of violation. *Psychol Rev,* 113(1), 148-154.

Cox, D. R., & Miller, H. D. (1965). *The Theory of Stochastic Processes*. New York: Wiley.

Crevecoeur, F., Munoz, D. P., & Scott, S. H. (2016). Dynamic Multisensory Integration: Somatosensory Speed Trumps Visual Accuracy during Feedback Control. *J Neurosci,* 36(33), 8598-8611.

Dahl, C. D., Logothetis, N. K., & Kayser, C. (2010). Modulation of visual responses in the superior temporal sulcus by audio-visual congruency. *Front Integr Neurosci,* 4, 10.

Diederich, A. (1995). Intersensory Facilitation of Reaction Time: Evaluation of Counter and Diffusion Coactivation Models. *Journal of Mathematical Psychology,* 39(2), 197-215.

Diederich, A., & Colonius, H. (1987). Intersensory facilitation in the motor component? *Psychological Research,* 49(1), 23-29.

Diederich, A., & Colonius, H. (2004a). Bimodal and trimodal multisensory enhancement: effects of stimulus onset and intensity on reaction time. *Percept Psychophys,* 66(8), 1388-1404.

Diederich, A., & Colonius, H. (2004b). Modeling the time course of multimodal interaction in manual and saccadic responses. In G. Calvert, C. Spence & B. Stein (Eds.), *Handbook of Multisensory Processes* (pp. 395 - 408). Cambridge, MA: MIT Press,.

Dixon, N. F., & Spitz, L. (1980). The Detection of Auditory Visual Desynchrony. *Perception*, 719-721.

Drugowitsch, J., DeAngelis, G. C., Klier, E. M., Angelaki, D. E., & Pouget, A. (2014). Optimal multisensory decision-making in a reaction-time task. *Elife (Cambridge)*, e03005.

Egan, J. P., Greenberg, G. Z., & Schulman, A. I. (1961). Operating Characteristics, Signal Detectability, and the Method of Free Response. *The Journal of the Acoustical Society of America,* 33(8), 993-1007.

Eriksen, C. W. (1988). A source of error in attempts to distinguish coactivation from separate activation in the perception of redundant targets. *Percept Psychophys,* 44(2), 191-193.

Fetsch, C. R., DeAngelis, G. C., & Angelaki, D. E. (2013). Bridging the gap between theories of sensory cue integration and the physiology of multisensory neurons. *Nat Rev Neurosci,* 14(6), 429-442.

Flanagan, J. L., & Golden, R. M. (1966). Phase Vocoder. *Bell System Technical Journal*, 1493-1509.

Gondan, M., Götze, C., & Greenlee, M. W. (2010). Redundancy gains in simple responses and go/no-go tasks. *Atten Percept Psychophys,* 72(6), 1692-1709.

Gondan, M., & Heckel, A. (2008). Testing the race inequality: A simple correction procedure for fast guesses. *Journal of Mathematical Psychology,* 52(5), 322-325.

Gondan, M., Lange, K., Rösler, F., & Röder, B. (2004). The redundant target effect is affected by modality switch costs. *Psychon Bull Rev,* 11(2), 307-313.

Gondan, M., & Minakata, K. (2016). A tutorial on testing the race model inequality. *Atten Percept Psychophys,* 78(3), 723-735.

Grant, K. W. (2001). The effect of speechreading on masked detection thresholds for filtered speech. *The Journal of the Acoustical Society of America,* 109(5), 2272-2275.

Grant, K. W., & Seitz, P.-F. (2000). The use of visible speech cues for improving auditory detection of spoken sentences. *The Journal of the Acoustical Society of America,* 108(3), 1197-1208.

Grant, K. W., Walden, B. E., & Seitz, P. F. (1998). Auditory-visual speech recognition by hearing-impaired subjects: Consonant recognition, sentence recognition, and auditory-visual integration. *The Journal of the Acoustical Society of America,* 103(5), 2677-2690.

Hauser, M. D., & Marler, P. (1993). Food-associated calls in rhesus macaques (Macaca mulatta): I. Socioecological factors. *Behav. Ecol.,* 4(3), 194-205.

Holmes, N. P. (2009). The principle of inverse effectiveness in multisensory integration: some statistical considerations. *Brain Topogr,* 21(3-4), 168-176.

Janssen, P., & Shadlen, M. N. (2005). A representation of the hazard rate of elapsed time in macaque area LIP. *Nat Neurosci,* 8(2), 234-241.

Luce, R. D. (1986). *Response times: Their role in inferring elementary mental organization*: Oxford University Press on Demand.

Ma, W. J., Beck, J. M., Latham, P. E., & Pouget, A. (2006). Bayesian inference with probabilistic population codes. *Nat Neurosci,* 9(11), 1432-1438.

Meredith, M. A., Nemitz, J. W., & Stein, B. E. (1987). Determinants of multisensory integration in superior colliculus neurons. I. Temporal factors. *J Neurosci,* 7(10), 3215-3229.

Miller, J. (1982). Divided attention: Evidence for coactivation with redundant signals. *Cognitive Psychology,* 14(2), 247-279.

Miller, J. (1986). Timecourse of coactivation in bimodal divided attention. *Percept Psychophys,* 40(5), 331-343.

Miller, J. (2016). Statistical facilitation and the redundant signals effect: What are race and coactivation models? *Atten Percept Psychophys,* 78(2), 516-519.

Molholm, S., Ritter, W., Murray, M. M., Javitt, D. C., Schroeder, C. E., & Foxe, J. J. (2002). Multisensory auditory-visual interactions during early sensory processing in humans: a high-density electrical mapping study. *Cognitive Brain Research,* 14(1), 115-128.

Murray, M. M., Molholm, S., Michel, C. M., Heslenfeld, D. J., Ritter, W., Javitt, D. C., et al. (2005). Grabbing your ear: rapid auditory-somatosensory multisensory interactions in low-level sensory cortices are not constrained by stimulus alignment. *Cereb Cortex,* 15(7), 963-974.

Nickerson, R. S. (1973). Intersensory facilitation of reaction time: Energy summation or preparation enhancement? *Psychological Review,* 80(6), 489-509.

Otto, T. U., Dassy, B., & Mamassian, P. (2013). Principles of multisensory behavior. *J Neurosci,* 33(17), 7463-7474.

Otto, T. U., & Mamassian, P. (2012). Noise and correlations in parallel perceptual decision making. *Curr Biol,* 22(15), 1391-1396.

Owen, D. B. (1980). A table of Normal integrals. *Communications in Statistics – Simulation and Computation,* 9, 389–419.

Populin, L. C., & Yin, T. C. (2002). Bimodal interactions in the superior colliculus of the behaving cat. *J Neurosci,* 22(7), 2826-2834.

R Core Team. (2017). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.

Raab, D. H. (1962). Statistical facilitation of simple reaction times. *Trans N Y Acad Sci,* 24, 574-590.

Raposo, D., Sheppard, J. P., Schrater, P. R., & Churchland, A. K. (2012). Multisensory decision-making in rats and humans. *J Neurosci,* 32(11), 3726-3735.

Ratcliff, R. (1988). Continuous versus discrete information processing modeling accumulation of partial information. *Psychol Rev,* 95(2), 238-255.

Ratcliff, R., & McKoon, G. (2008). The diffusion decision model: theory and data for two-choice decision tasks. *Neural Comput,* 20(4), 873-922.

Ratcliff, R., Smith, P. L., Brown, S. D., & McKoon, G. (2016). Diffusion Decision Model: Current Issues and History. *Trends Cogn Sci,* 20(4), 260-281.

Ratcliff, R., Thapar, A., & McKoon, G. (2003). A diffusion model analysis of the effects of aging on brightness discrimination. *Percept Psychophys,* 65(4), 523-535.

Rowell, T. E., & Hinde, R. A. (1962). Vocal communication by the rhesus monkey (Macaca mulatta). *Proceedings of the Zoological Society London,* 138, 279-294.

Sakata, S., Yamamori, T., & Sakurai, Y. (2004). Behavioral studies of auditory-visual spatial recognition and integration in rats. *Experimental Brain Research,* 159(4), 409-417.

Schwartz, J.-L., Berthommier, F., & Savariaux, C. (2004). Seeing to hear better: evidence for early audio-visual interactions in speech identification. *Cognition,* 93(2), B69-B78.

Schwarz, W. (1989). A new model to explain the redundant-signals effect. *Percept Psychophys,* 46(5), 498-500.

Schwarz, W. (1994). Diffusion, Superposition and the Redundant-Targets Effect. *Journal of Mathematical Psychology,* 38(4), 504-520.

Schwarz, W. (2006). On the relationship between the redundant signals effect and temporal order judgments: parametric data and a new model. *J Exp Psychol Hum Percept Perform,* 32(3), 558-573.

Seilheimer, R. L., Rosenberg, A., & Angelaki, D. E. (2014). Models and processes of multisensory cue combination. *Curr Opin Neurobiol,* 25, 38-46.

Shub, D. E., & Richards, V. M. (2009). Psychophysical spectro-temporal receptive fields in an auditory task. *Hear Res,* 251(1-2), 1-9.

Skaliora, I., Doubell, T. P., Holmes, N. P., Nodal, F. R., & King, A. J. (2004). Functional topography of converging visual and auditory inputs to neurons in the rat superior colliculus. *J Neurophysiol,* 92(5), 2933-2946.

Smith, P. L., & Ratcliff, R. (2004). Psychology and neurobiology of simple decisions. *Trends Neurosci,* 27(3), 161-168.

Stanford, T. R., Quessy, S., & Stein, B. E. (2005). Evaluating the operations underlying multisensory integration in the cat superior colliculus. *Journal of Neuroscience,* 25(28), 6499-6508.

Stanford, T. R., & Stein, B. E. (2007). Superadditivity in multisensory integration: putting the computation in context. *Neuroreport,* 18(8), 787-792.

Stein, B. E., & Meredith, M. A. (1993). *Merging of the Senses*. Cambridge, MA: MIT Press.

Stein, B. E., Meredith, M. A., Huneycutt, W. S., & McDade, L. (1989). Behavioral Indices of Multisensory Integration: Orientation to Visual Cues is Affected by Auditory Stimuli. *Journal of Cognitive Neuroscience,* 1(1), 12-24.

Stein, B. E., & Stanford, T. R. (2008). Multisensory integration: current issues from the perspective of the single neuron. *Nat Rev Neurosci,* 9(4), 255-266.

Townsend, J. T., & Ashby, F. G. (1983). *Stochastic modeling of elementary psychological processes*: CUP Archive.

Ulrich, R., & Miller, J. (1997). Tests of Race Models for Reaction Time in Experiments with Asynchronous Redundant Signals. *J Math Psychol,* 41(4), 367-381.

van Wassenhove, V., Grant, K. W., & Poeppel, D. (2007). Temporal window of integration in auditory-visual speech perception. *Neuropsychologia,* 45(3), 598-607.

van Atteveldt, N., Murray, Micah M., Thut, G., & Schroeder, Charles E. (2014). Multisensory Integration: Flexible Use of General Operations. *Neuron,* 81(6), 1240-1253.

Voss, A., Rothermund, K., & Brandtstädter, J. (2008). Interpreting ambiguous stimuli: Separating perceptual and judgmental biases. *Journal of Experimental Social Psychology,* 44(4), 1048-1056.