



**HAL**  
open science

# Humans (*Homo sapiens*) but not baboons (*Papio papio*) demonstrate crossmodal pitch-luminance correspondence

Konstantina Margiotoudi, Joel Fagot, Adrien Meguerditchian, Isabelle Dautriche

## ► To cite this version:

Konstantina Margiotoudi, Joel Fagot, Adrien Meguerditchian, Isabelle Dautriche. Humans (*Homo sapiens*) but not baboons (*Papio papio*) demonstrate crossmodal pitch-luminance correspondence. *American Journal of Primatology*, 2024, 86 (5), 10.1002/ajp.23613 . hal-04694282

**HAL Id: hal-04694282**

**<https://hal.science/hal-04694282v1>**

Submitted on 23 Sep 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34

Humans (*Homo sapiens*) but not baboons (*Papio papio*) demonstrate crossmodal  
pitch-luminance correspondence

Konstantina Margiotoudi<sup>1,2</sup>, Joel Fagot<sup>1,2</sup>, Adrien Meguerditchian<sup>1,2</sup> and Isabelle  
Dautriche<sup>1\*</sup>

<sup>1</sup> Laboratoire de Psychologie Cognitive, UMR7290, LPC, CNRS, Aix-Marseille  
Université, Marseille, France

<sup>2</sup> Station de Primatologie-Celphedia UAR846, CNRS, Rousset, France

\* corresponding author: [isabelle.dautriche@cnrs.fr](mailto:isabelle.dautriche@cnrs.fr)

35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65  
66  
67  
68

Abstract

Humans spontaneously and consistently map information coming from different sensory modalities. Surprisingly, the phylogenetic origin of such cross-modal correspondences has been under-investigated. A notable exception is the study of Ludwig et al (2011) which reports that both humans and chimpanzees spontaneously map high-pitched sounds with bright objects and low-pitched sounds with dark objects. Our pre-registered study aimed to directly replicate this research on both humans and baboons (*Papio papio*), an old world monkey which is more phylogenetically distant from humans than chimpanzees. Following Ludwig et al., participants were presented with a visual classification task where they had to sort black and white square (low and high luminance), while background sounds (low or high-pitched tones) were playing. Whereas we replicated the finding that humans' performance on the visual task was affected by congruency between sound and luminance of the target, we did not find any of those effects on baboons' performance. These results question the presence of a shared cross-modal pitch-luminance mapping in other non-human primates.

Keywords: crossmodal correspondence; comparative cognition; iconicity

Abbreviations: RFID: Radio Frequency Identification Device

69 Humans (*Homo sapiens*) but not baboons (*Papio papio*) demonstrate crossmodal  
70 pitch-luminance correspondence

71

72 Humans tend to spontaneously form cross-modal correspondences, i.e., associations  
73 between modality-specific features (Spence, 2011). For instance, they tend to  
74 associate certain sound frequency to shape properties: a high-pitched tone is typically  
75 associated with small, spiky and bright objects, while a low-pitched tone is generally  
76 associated with big, round, and dark objects (Spence, 2019). Cross-modal biases  
77 have been experimentally demonstrated across a wide range of modalities, including  
78 between odors and visual stimuli (Hanson-Vaux et al., 2013), sound and taste  
79 (Knöferle & Spence, 2012), and taste and touch (Christensen, 1980), among others  
80 (Calvert et al., 2004).

81

82 Cross-modal biases are considered a crucial mechanism for multisensory integration  
83 and language evolution. In perception, cross-modal correspondences may help to  
84 constrain the cross-modal binding problem, i.e., the challenge of determining which  
85 sensory information to integrate together (Ernst, 2007). In language, cross-modal  
86 biases to associate sounds with other sensory modalities may have facilitated the  
87 evolution of spoken language by allowing for the expression of sensory experiences  
88 through a single auditory channel (Cuskley & Kirby, 2013; Ramachandran & Hubbard,  
89 2001). Surprisingly though, despite the potential evolutionary importance of cross-  
90 modal correspondences, the investigation of their phylogenetic origins remains limited.

91

92 It is plausible that cross-modal correspondences may be constrained in non-human  
93 animals due to their lack of linguistic experience. According to the semantic/linguistic  
94 account (Martino & Marks, 1999), the perceived connection between different sensory  
95 stimuli may be explained by the use of common linguistic terms to describe them. For  
96 example, in English, tones are characterized as 'high' and 'low' in pitch, employing  
97 terms associated with spatial elevation. This linguistic connection may be at the origin  
98 of the well-documented cross-modal association between pitch and height (Ben-Artzi  
99 & Marks, 1995; Dolscheid et al., 2013; Getz & Kubovy, 2018; Occelli et al., 2009;  
100 Trimble, 1934). However, cross-cultural (Parkinson et al., 2012) and developmental  
101 studies (Walker et al., 2010) but see (Dolscheid et al., 2023) have provided evidence  
102 challenging the purely linguistic nature of the pitch-height association. Infants and

103 speakers from cultures where pitch is not described using spatial elevation terms still  
104 exhibit a bias to associate higher-pitched sounds with higher spatial elevation and  
105 lower-pitched sounds with lower spatial elevation.

106

107 Comparative studies provide an immediate way to test whether cross-modal  
108 associations are a purely linguistic phenomenon. To our knowledge, only few studies  
109 have examined the existence of these cross-modal mappings in other animals. Two  
110 studies provide evidence that domestic dogs prefer to associate higher frequencies  
111 with high spatial positions (a pitch-elevation mapping (Korzeniowska et al., 2019) and  
112 with larger stimuli (a pitch-size mapping, Korzeniowska et al., 2022), Another study  
113 shows that chicks prefer to associate stimuli situated at their right with brighter stimuli  
114 (a space-luminance mapping, Loconsole et al., 2021). Lastly, the pitch-luminance  
115 mapping—where higher pitch is associated with bright stimuli and lower pitch to dark  
116 stimuli— has been found in chimpanzees (Ludwig et al., 2011), but not in chicks  
117 (Loconsole et al., 2022). Collectively, these results suggest that certain cross-modal  
118 correspondences have non-linguistic origins.

119

120 What is then the origin of these cross-modal correspondences? Some may be  
121 acquired through exposure to consistent patterns in the environment. For instance, the  
122 pitch-elevation correspondence might be explained by the environmental statistics of  
123 auditory scenes, where higher frequency sounds typically originate from elevated  
124 sources (Parise et al., 2014). Likewise, the pitch-size correspondence may be due to  
125 the observation that large objects tend to produce low frequency sounds (e.g., when  
126 dropped), whereas small objects generate high-pitched sounds (Coward & Stevens,  
127 2004; Gaver, 1993). In contrast, the pitch-luminance and the space-luminance  
128 mappings are difficult to explain based on regularities found in the environment alone  
129 (but see Spence & Deroy, 2012). Given that chimpanzees have no language abilities,  
130 it has been proposed that the pitch-luminance mapping is a structural correspondence  
131 that is hard-wired in the brain (Walker et al., 2010). Both visual high luminance and  
132 loudness have a shared increase in stimulus intensity which is suggested to be  
133 encoded by the same brain structure dedicated to magnitude (Lewkowicz & Turkewitz,  
134 1980; Walsh, 2003), This shared feature in neural encoding could serve as the  
135 foundation for correspondences based on intensity, such as the pitch-luminance  
136 mapping. If this account is correct, intensity-based correspondences may be a

137 characteristic common with other species that share the same neural mechanism  
138 (Gazes et al., 2023).

139

140 The results of Ludwig et al., (2011) are thus particularly significant as they represent  
141 the only demonstration that non-human animals can match features across modalities  
142 in a way that is not directly attributable to associative learning, suggesting that some  
143 cross-modal mappings may have a structural, innate, origin. In that study, 6  
144 chimpanzees and 33 human participants were trained in a visual speeded  
145 classification task on a touch screen, where they had to match a sample square (either  
146 black or white) to the matching stimulus (i.e., an identical black or white square).  
147 During testing, both species performed the same task while passively listening to  
148 background sounds (high and low-pitched tones), which were crossmodally either  
149 congruent or incongruent to the luminance of the sample square. Both species  
150 performed better when the frequency of the sound matched the luminance of the  
151 sample square, i.e. when the white square was congruent with the high-pitched tone  
152 and the black square was congruent with the low-pitched tone.

153

154 Previous attempts to test other sound-shape mappings in great apes (i.e., gorillas,  
155 chimpanzees and bonobos) have failed to show any cross-modal effects in these  
156 species (Margiotoudi et al., 2019, 2022). In a forced choice matching task, humans  
157 but not great apes, were more likely to select a rounded shaped upon being presented  
158 with the wordform 'maluma', and an angular shape when presented with the wordform  
159 'takete' (the 'maluma-takete' or 'bouba-kiki' effect (Köhler, 1967). Even, Kanzi, a  
160 language-trained bonobo, showed no evidence of such a sound-shape sensitivity  
161 suggesting language exposure/learning alone cannot account for the observed  
162 differences between humans and great apes, leading to the proposition that  
163 neuroanatomical differences between human and nonhuman primates could underlie  
164 this effect in humans (Margiotoudi et al., 2022).

165

166 Considering the weak experimental evidence for the existence of cross-modal  
167 mappings in non-human primates, the purpose of the current pre-registered study was  
168 to directly replicate Ludwig et al. (2011)'s study on the pitch-luminance  
169 correspondence, in humans and baboons (*Papio papio*) using the same methods and  
170 analyses. Humans and baboons were trained to classify squares based on their

171 luminance in a visual speeded classification task. During the testing phase, both  
172 species completed the same speeded classification task, while being presented with  
173 background sounds (low vs high pitched tones). If baboons and humans have an  
174 inherent tendency to match high pitch with luminance and low pitch with dark  
175 luminance, we expect participants to be faster to classify dark and white squares when  
176 a congruent sound is presented, compared to when an incongruent sound is  
177 presented. The outcomes of this experiment would be significant in two aspects.  
178 Firstly, it will provide more support to the claim that possibly innate hardwired cross-  
179 modal mappings exist in non-human species, a claim currently grounded in a single  
180 study (Ludwig et al., 2011). Second, a positive result in baboons would indicate that  
181 the sensitivity to pitch-luminance correspondence is a shared feature of the primate  
182 lineage, that could be part of an innate predisposition to associate intensity across  
183 dimensions.

184

185

## 186 **Methods**

187

### 188 **Data availability statement**

189 The methods, design and analysis were pre-registered at: <https://osf.io/vsnp4>.

190 All the data and stimulus material are openly available at:  
191 [https://osf.io/ng6y3/?view\\_only=b3797de197744bb79a1bcc544f0cf69a](https://osf.io/ng6y3/?view_only=b3797de197744bb79a1bcc544f0cf69a).

192

### 193 **Ethical standards**

194 All methods for the human study were approved by the Ethics Committee of Aix-  
195 Marseille University and for the baboon study received approval from the French  
196 Ministère de l'Éducation Nationale (approval no. APAFIS-2717-2015111708173794-  
197 V3).

198

### 199 **Participants and apparatus**

200 Humans: Twenty-seven, right-handed human adults (21 females, age  $M=22.70$  years  
201  $\pm 5.55$ ) participated in the study. The sample size was determined from a power  
202 analysis based on an effect size of  $d=0.57$ , as reported in the human study of Ludwig  
203 et al. (2011), a power of 0.8, and an alpha level of  $\alpha=0.05$ . All the participants were  
204 native speakers of French and had corrected to normal vision and normal hearing.

205 Participants were recruited from announcements at the Aix-Marseille University  
206 campus and internal email lists. Testing took place at an experimental booth at the  
207 Laboratoire de Psychologie in the same campus. Participants received a monetary  
208 compensation of 10 euros.

209

210 Baboons: Fifteen Guinea baboons (*Papio papio*), (12 females, age  $M=11.4$  years  $\pm$   
211 5.64) from the CNRS primate center of Rousset-sur-Arc completed the study. A power  
212 analysis based on an effect size of  $d=0.87$ , as reported in the chimpanzee study of  
213 Ludwig et al. (2011), a power of 0.8, and an alpha level of  $\alpha=0.05$ , indicated that we  
214 should test at least 13 individuals to replicate an effect of similar magnitude. Because  
215 of our testing apparatus, the baboons cannot be tested away from their social group,  
216 so the study was presented to all the individuals ( $n = 23$ ). Baboons live in two social  
217 groups and had access to 13 ALDMs (Automated Computer Learning devices) (Fagot  
218 & Bonté, 2010), that are freely accessible from their enclosures. Each device has a  
219 food dispenser, a touchscreen, two speakers and a radio frequency identification  
220 (RFID tag) of the participants. The RFID tag allows us to test the individuals without  
221 capturing them, which improves animal welfare in experimental research (Fagot et al.,  
222 2014).

223

#### 224 **Apparatus.**

225 Humans were tested on a on a 23-inch LCD monitor Iiyama ProLite (screen refresh  
226 rate: 60Hz, screen resolution: 1920 x 1080). The sounds were presented via two  
227 Logitech S-20 speakers, placed at each side of the screen. Baboons were tested on  
228 LCD capacitive touchscreens, for which a built-in secure glass of 3 mm wide allowed  
229 screen protection. The speakers were positioned in front of the baboons, above them,  
230 and directed toward them. The human experiment was designed on E-prime 3.0  
231 software (Psychology Software Tools, Inc., Pittsburg, PA, USA), while the baboon  
232 experiment on E-prime 1.2.

233

#### 234 **Stimuli**

235 Both auditory and visual stimuli were the same as in the original study (Ludwig et al.,  
236 2011). The auditory stimuli consisted of computer-generated tones, with frequencies  
237 of either 1,047 Hz (high-pitched) or 175 Hz (low-pitched). As for the visual stimuli, the  
238 sample stimulus was 99 x 99 pixel square ( $2.5 \times 2.5$  cm), displayed at the screen's



239 center against a gray background (RGB: 128, 128, 128]). This square was black (RGB:  
240 0, 0, 0) or white (RGB: 255, 255, 255) depending on the trial. Each of the two choice  
241 buttons measured 250 × 250 pixels (6.5 × 6.5 cm). For half of the participants (humans  
242 and baboons), the right-side button was white and for the other half, it was black.

243

### 244 **Procedure and design**

245 Humans and baboons were required to perform an identity match to sample task (MTS  
246 task), where they had to match a sample square (black or white) to one of the two  
247 alternatives that was identical to the sample. The experiment was divided into two  
248 phases, a training and a testing phase. During the training phase, participants  
249 performed the MTS task with no sound. During the testing phase, participants had to  
250 perform the same task while a tone, either high or low-pitched, was played.  
251 Participants received feedback in both the training and testing phases. The design  
252 was identical to the one described in Ludwig et al. (2011), with the sole difference on  
253 the feedback screens (see supplementary Figure S1). In the present study, we used  
254 visual feedback, instead of audio feedback (Ludwig et al., 2011), as this group of  
255 baboons had previously been accustomed to training with visual feedback while  
256 engaging in touch-screen tasks (Fagot & Bonté, 2010). After a correct response, a  
257 black screen appeared for 3s combined with food reward (only for the baboons),  
258 whereas after an incorrect response, a green screen appeared for 3s followed by a  
259 black screen of 3s.

260



283 they had to achieve a performance of at least 80% in both categories (i.e., black/white)  
284 (White:  $M=0.80\%$ ;  $SE=0.02\%$ , Black:  $M=0.88\%$ ;  $SE=0.02$ ). Blocks were repeated until  
285 the criterion was reached (average number of training blocks =2.71; min=1; max=11;  
286  $SE=0.31$ ). The details of the baboon training are available in the supplementary  
287 material. After that, the testing phase started. First baboons performed 4 practice test  
288 trials with background sounds, then, a block of 160 testing trials followed. In total  
289 baboons completed 480 testing trials across the 3 experimental sessions.

290

291 Testing blocks consisted of the same number of possible sample-sound combinations  
292 following the description by Ludwig et al. (2011). That is for a testing block of 160 trials,  
293 there were 80 congruent trials (40 trials presenting a white sample and a high pitch  
294 and 40 trials presenting a black sample and a low pitch) and 80 incongruent trials (40  
295 trials presenting a white sample and a low pitch and 40 trials presenting a black sample  
296 and a high pitch). The same sound category, luminance sample or congruency  
297 category could not be presented in more than three consecutive trials (as in Ludwig et  
298 al., 2011).

299

### 300 **Analysis**

301 All analyses were performed in R [29]. We used the lme4 package for the mixed model  
302 analysis (Bates et al., 2014) and the package emmeans ([https://cran.r-](https://cran.r-project.org/web/packages/emmeans/index.html)  
303 [project.org/web/packages/emmeans/index.html](https://cran.r-project.org/web/packages/emmeans/index.html)) was used for post-hoc tests. T-test  
304 are all paired t-test.

305

### 306 **Results**

307

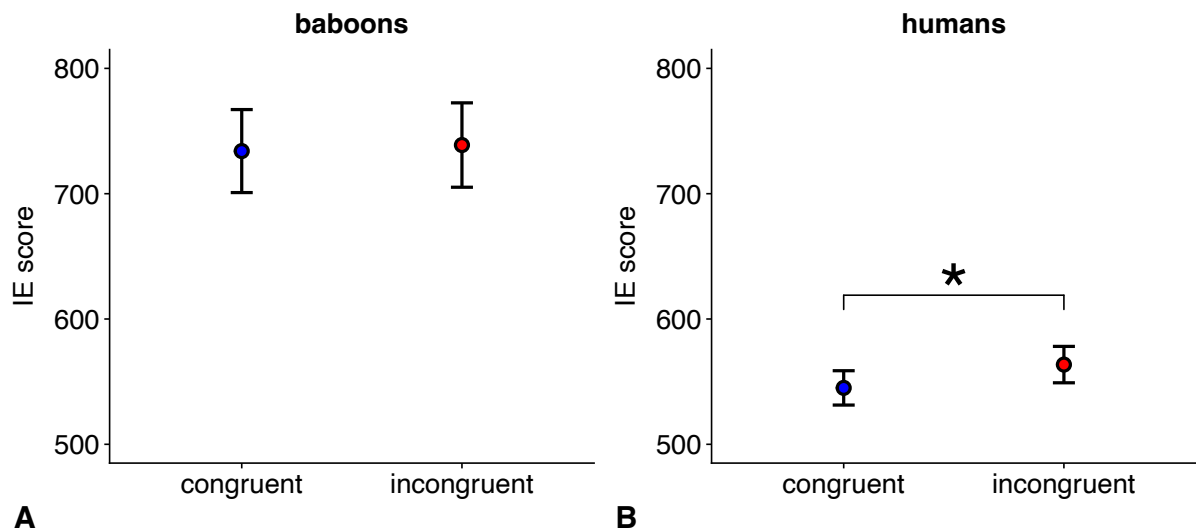
308 We conducted the analysis described by Ludwig et al., (2011). We first looked at the  
309 median RT and the error rates for congruent and incongruent trials in both species.  
310 Our results confirm the findings of Ludwig et al. (2011) in humans but not in non-  
311 human primates. In Ludwig et al. (2011), chimpanzees made more errors in  
312 incongruent than in congruent trials, while, in baboons, no differences were observed  
313 on the mean error rates between congruent ( $M= 10.58\%$ ,  $SD= 3.16\%$ ) and incongruent  
314 trials ( $M = 10.58\%$ ,  $SD= 4.16\%$ ). Similarly, there was no difference on the median RTs  
315 between congruent ( $M = 653\text{ms}$ ,  $SD = 97.6\text{ms}$ ) and incongruent trials ( $M = 657\text{ms}$ ,  $SD$

316 = 102.5ms;  $t(14)=-1.31$ ,  $p=0.20$ ,  $r = 0.33$ ) for baboons, mirroring the findings of Ludwig  
317 et al. (2011) in chimpanzees.

318 Humans exhibited no significant difference in mean error rates between congruent ( $M$   
319 = 0.18%,  $SD = 0.04\%$ ) and incongruent trials ( $M = 0.46\%$ ,  $SD = 0.06\%$ ;  $t(26)=-0.94$ ,  
320  $p=0.35$ ,  $r = 0.18$ ). While the humans were overall faster than in Ludwig et al. (2011)'s  
321 study (see supplemental material), they were similarly affected by trial congruency: as  
322 in Ludwig et al. (2011), they were faster for congruent trials ( $M = 544$ ms;  $SD = 71.8$   
323 ms) than for incongruent trials ( $M = 560$ ms,  $SD = 69.8$ ms;  $t(26)=-4.88$ ,  $p<0.001$ ,  $r =$   
324 0.69).

325

326 Following Ludwig et al. (2011), we then calculated the speed-accuracy trade-off for  
327 congruent and incongruent trials in both species, by calculating the Inverse Efficiency  
328 score (IE) (Townsend & Ashby, 1983). To compute the IE scores, we divided the  
329 median RT by the proportion of correct responses, for each participant, for the two  
330 conditions (congruent vs. incongruent). Humans had a significantly lower IE score (i.e.,  
331 indicating better performance) ( $t(26)=-4.15$ ,  $p<0.001$ ;  $r=0.69$ ) for congruent (IE = 545,  
332  $SD = 71.2$ ) than for incongruent trials (IE = 564,  $SD = 75.5$ ). In contrast to the findings  
333 of Ludwig et al. (2011) on chimpanzees, the results on the baboons revealed no  
334 significant difference ( $t(14)=-0.54$ ,  $p=0.60$ ) on the IE score between congruent (mean  
335 IE: 734,  $SD = 128$ ) and incongruent trials (mean IE: 739,  $SD = 130$ ) (see Fig.2).  
336 Overall, the IE scores were lower for humans (who were faster and performed better  
337 at the task) than baboons. This was similar in Ludwig et al. (2011): humans displayed  
338 lower IE scores than chimpanzees. Based on these findings, we replicated the findings  
339 of Ludwig et al. (2011) in humans, but did not extend it in baboons (for details see  
340 results in supplementary material).



341

342 **Figure 2.** Mean IE scores for congruent (blue) and incongruent (red) trials in A.  
 343 baboons and B. humans. The error bars represent standard errors of the mean. The  
 344 asterisk indicates significance between conditions in the human group ( $p < 0.001$ ).

345

346 Discussion

347

348 Following the protocol of Ludwig et al. (2011), we tested the pitch-luminance  
 349 crossmodal correspondence in humans and baboons. In the original study, humans  
 350 exhibited longer response times when the sound was incongruent to the luminance of  
 351 the sample stimulus. In chimpanzees, no interference effect was observed on  
 352 response time, but they did make more errors on cross-modally incongruent trials  
 353 compared to congruent trials. In the present study, although we replicated the human  
 354 results, our results did not show any evidence of sound interference effects on either  
 355 response time or error rate in baboons.

356

357 How can we account for the difference between chimpanzees (Ludwig et al., 2011)  
 358 and the current baboon results? Certainly, primate species differences might suggest  
 359 that the mapping of pitch-to-luminance is unique to chimpanzees and humans, but  
 360 does not extend to more phylogenetically distant relatives such as the Old World  
 361 monkeys, including the baboons. According to this hypothesis, such a cross-modal  
 362 correspondence might thus have emerged and be inherited from the last common  
 363 ancestor of the Hominini group.

364

365 An alternative possibility is that the absence of any systemic pitch-to-luminance  
366 correspondence in baboons, questions the existence of this mapping not only in this  
367 species but also in any other nonhuman species. There are several reasons to  
368 question the robustness of the chimpanzee's result. First, while the study of Ludwig et  
369 al. (2011) showed a pitch interference effect in a visual speeded classification task in  
370 chimpanzees, this effect was based on a 2.7% difference between mean error rates  
371 for congruent and incongruent trials and was not visible on median response times,  
372 contrary to human participants. Second, a common problem in animal studies is the  
373 low statistical power due to small sample size (6 individuals in Ludwig et al., 2011), an  
374 aspect known to increase type I errors (Christley, 2010). Third, whether the  
375 chimpanzees tested in the Ludwig et al., (2011) had previous experiences on auditory  
376 discrimination between high- and low-pitched tones, is not specified in the original  
377 study. Any such previous experience could have increased sensitivity to the auditory  
378 stimuli of the present paradigm. This is of particular importance as testing non-human  
379 primates on auditory paradigms is not a trivial task (Hashiya & Kojima, 2001; Martinez  
380 & Matsuzawa, 2009).

381

382 Our results are consistent with previous results showing no evidence of structural  
383 crossmodal mappings in non-human primates (Margiotoudi et al., 2019, 2022) or in  
384 other animals such as chicks (Loconsole et al., 2022). This could suggest that some  
385 crossmodal correspondences might not be shared with other nonhuman species, thus  
386 reopening the possibility that those mappings, and in particular the pitch-luminance  
387 mapping, could be mediated by language, in English the same adjectives to refer to  
388 luminance and pitch, namely 'high' and 'low'. While the pitch-luminance effect has  
389 been evidenced robustly in humans (Hubbard, 1996; Ludwig et al., 2011; Marks et al.,  
390 1987), it has not been systematically examined in speakers of languages that do not  
391 describe luminance and pitch in the same way, nor in 'pre-linguistic' infants. Such  
392 investigations will be critical to determine whether the pitch-luminance mapping truly  
393 reflect a universally predisposed structural correspondence.

394

395 As we noted, the absence of crossmodal effects in baboons and in great apes  
396 (Margiotoudi et al., 2019, 2022) could have plausibly resulted from attention or  
397 discriminatory troubles on auditory stimuli. In addition, factors such as previous  
398 habituation to visual tasks, training history, protocol design, or age of the individuals

399 appear to play an important role in testing the performance of nonhuman primates on  
400 tasks including auditory stimuli (Ennaji, 2022). More generally, this highlights the need  
401 for a broad comparative approach, investigating the ability of various species to  
402 spontaneously associate different stimuli (much like has been done in human adults,  
403 see (Getz & Kubovy, 2018; Parise & Spence, 2009) in order to derive a cartography  
404 of cross-modal correspondences and of where they are found in phylogeny.

405

406 To conclude, evidence of the presence of a pitch-luminance mapping in other  
407 nonhuman species appears to be weak, suggesting that language, culture or human-  
408 specific neuro-anatomical features may be fundamental for the emergence of the  
409 pitch-luminance mapping. Future studies will be necessary to investigate the reliability  
410 of this hypothesis.

411

#### 412 Fundings

413 This work was supported by Fyssen Foundation, by the European Research Council  
414 under the European Union's Horizon 2020 research and innovation program grant  
415 agreement No 716931 - GESTIMAGE - ERC-2016-STG, by the “Agence Nationale de  
416 le Recherche” ANR-16-CONV-0002 (ILCB) and the Excellence Initiative of Aix-  
417 Marseille University (A\*MIDEX).

418

#### 419 References

420

421 Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). *Fitting Linear Mixed-Effects*

422 *Models using lme4* (arXiv:1406.5823). arXiv. <http://arxiv.org/abs/1406.5823>

423 Ben-Artzi, E., & Marks, L. E. (1995). Visual-auditory interaction in speeded

424 classification: Role of stimulus difference. *Perception & Psychophysics*, *57*(8),

425 1151–1162. <https://doi.org/10.3758/BF03208371>

426 Calvert, G. A., Spence, C., & Stein, B. E. (Eds) (2004). *The handbook of*

427 *multisensory processes*. Cambridge, MA: MIT Press.

428 Christensen, C. M. (1980). Effects of taste quality and intensity on oral perception of  
429 viscosity. *Perception & Psychophysics*, 28(4), 315–320.  
430 <https://doi.org/10.3758/BF03204390>

431 Christley, R. M. (2010). Power and error: Increased risk of false positive results in  
432 underpowered studies. *The Open Epidemiology Journal*, 3(1).

433 Coward, S. W., & Stevens, C. J. (2004). Extracting meaning from sound: Nomic  
434 mappings, everyday listening, and perceiving object size from frequency. *The*  
435 *Psychological Record*, 54, 349–364.

436 Cuskley, C., & Kirby, S. (2013). Synesthesia, cross-modality, and language  
437 evolution. In *Oxford handbook of synesthesia* (eds J Simner, EM Hubbard),  
438 pp. 869 - 907. Oxford, UK: Oxford University Press

439 Dolscheid, S., Çelik, S., Erkan, H., Küntay, A., & Majid, A. (2023). Children’s  
440 associations between space and pitch are differentially shaped by language.  
441 *Developmental Science*, 26(5), e13341. <https://doi.org/10.1111/desc.13341>

442 Dolscheid, S., Shayan, S., Majid, A., & Casasanto, D. (2013). The thickness of  
443 musical pitch: Psychophysical evidence for linguistic relativity. *Psychological*  
444 *Science*, 24(5), 613–621. <https://doi.org/10.1177/0956797612457374>

445 Ennaji, F.-E. (2022). *Voice perception in non-human primates: A behavioral study of*  
446 *categorization of vocal and non-vocal sounds in Guinea baboon (Papio papio)*  
447 [PhD Thesis, Aix-Marseille].

448 Ernst, M. O. (2007). Learning to integrate arbitrary signals from vision and touch.  
449 *Journal of Vision*, 7(5), 7:1-14.

450 Fagot, J., & Bonté, E. (2010). Automated testing of cognitive performance in  
451 monkeys: Use of a battery of computerized test systems by a troop of semi-



452 free-ranging baboons (*Papio papio*). *Behavior Research Methods*, 42(2), 507–  
453 516. <https://doi.org/10.3758/BRM.42.2.507>

454 Fagot, J., Gullstrand, J., Kemp, C., Defilles, C., & Mekaouche, M. (2014). Effects of  
455 freely accessible computerized test systems on the spontaneous behaviors  
456 and stress level of Guinea baboons ( *Papio papio* ). *American Journal of*  
457 *Primatology*, 76(1), 56–64. <https://doi.org/10.1002/ajp.22193>

458 Gaver, W. W. (1993). What in the World Do We Hear?: An Ecological Approach to  
459 Auditory Event Perception. *Ecological Psychology*, 5(1), 1–29.  
460 [https://doi.org/10.1207/s15326969eco0501\\_1](https://doi.org/10.1207/s15326969eco0501_1)

461 Gazes, R. P., Templer, V. L., & Lazareva, O. F. (2023). Thinking about order: A  
462 review of common processing of magnitude and learned orders in animals.  
463 *Animal Cognition*, 26(1), 299–317.

464 Getz, L. M., & Kubovy, M. (2018). Questioning the automaticity of audiovisual  
465 correspondences. *Cognition*, 175, 101–108.

466 Hanson-Vaux, G., Crisinel, A.-S., & Spence, C. (2013). Smelling shapes:  
467 Crossmodal correspondences between odors and shapes. *Chemical Senses*,  
468 38(2), 161–166.

469 Hashiya, K., & Kojima, S. (2001). Acquisition of auditory–visual intermodal matching-  
470 to-sample by a chimpanzee (*Pan troglodytes*): Comparison with visual—visual  
471 intramodal matching. *Animal Cognition*, 4, 231–239.

472 Hubbard, T. L. (1996). Synesthesia-like mappings of lightness, pitch, and melodic  
473 interval. *The American Journal of Psychology*, 219–238.

474 Knöferle, K., & Spence, C. (2012). Crossmodal correspondences between sounds  
475 and tastes. *Psychonomic Bulletin & Review*, 19(6), 992–1006.  
476 <https://doi.org/10.3758/s13423-012-0321-z>

477 Köhler, W. (1967). Gestalt psychology. *Psychologische Forschung*, 31(1), XVIII–  
478 XXX.

479 Korzeniowska, A. T., Root-Gutteridge, H., Simner, J., & Reby, D. (2019). Audio–  
480 visual crossmodal correspondences in domestic dogs (*Canis familiaris*).  
481 *Biology Letters*, 15(11), 20190564. <https://doi.org/10.1098/rsbl.2019.0564>

482 Korzeniowska, A. T., Simner, J., Root-Gutteridge, H., & Reby, D. (2022). High-pitch  
483 sounds small for domestic dogs: Abstract crossmodal correspondences  
484 between auditory pitch and visual size. *Royal Society Open Science*, 9(2),  
485 211647. <https://doi.org/10.1098/rsos.211647>

486 Lewkowicz, D. J., & Turkewitz, G. (1980). Cross-modal equivalence in early infancy:  
487 Auditory–visual intensity matching. *Developmental Psychology*, 16(6), 597–  
488 607.

489 Loconsole, M., Gasparini, A., & Regolin, L. (2022). Pitch–luminance crossmodal  
490 correspondence in the baby chick: An investigation on predisposed and  
491 learned processes. *Vision*, 6(2), 24.

492 Loconsole, M., Pasculli, M. S., & Regolin, L. (2021). Space-luminance crossmodal  
493 correspondences in domestic chicks. *Vision Research*, 188, 26–31.

494 Ludwig, V. U., Adachi, I., & Matsuzawa, T. (2011). Visuoauditory mappings between  
495 high luminance and high pitch are shared by chimpanzees (*Pan troglodytes*)  
496 and humans. *Proceedings of the National Academy of Sciences*, 108(51),  
497 20661–20665. <https://doi.org/10.1073/pnas.1112605108>

498 Margiotoudi, K., Allritz, M., Bohn, M., & Pulvermüller, F. (2019). Sound symbolic  
499 congruency detection in humans but not in great apes. *Scientific Reports*,  
500 9(1), 12705. <https://doi.org/10.1038/s41598-019-49101-4>

501 Margiotoudi, K., Bohn, M., Schwob, N., Tagliatela, J., Pulvermüller, F., Epping, A.,  
502 Schweller, K., & Allritz, M. (2022). Bo-NO-bouba-kiki: Picture-word mapping  
503 but no spontaneous sound symbolic speech-shape mapping in a language  
504 trained bonobo. *Proceedings of the Royal Society B: Biological Sciences*,  
505 289(1968), 20211717. <https://doi.org/10.1098/rspb.2021.1717>

506 Marks, L. E., Hammeal, R. J., Bornstein, M. H., & Smith, L. B. (1987). Perceiving  
507 similarity and comprehending metaphor. *Monographs of the Society for*  
508 *Research in Child Development*, i–100.

509 Martinez, L., & Matsuzawa, T. (2009). Visual and auditory conditional position  
510 discrimination in chimpanzees (*Pan troglodytes*). *Behavioural Processes*,  
511 82(1), 90–94.

512 Martino, G., & Marks, L. E. (1999). Perceptual and Linguistic Interactions in Speeded  
513 Classification: Tests of the Semantic Coding Hypothesis. *Perception*, 28(7),  
514 903–923. <https://doi.org/10.1068/p2866>

515 Occelli, V., Spence, C., & Zampini, M. (2009). Compatibility effects between sound  
516 frequency and tactile elevation. *Neuroreport*, 20(8), 793–797.

517 Parise, C. V., Knorre, K., & Ernst, M. O. (2014). Natural auditory scene statistics  
518 shapes human spatial hearing. *Proceedings of the National Academy of*  
519 *Sciences*, 111(16), 6104–6108. <https://doi.org/10.1073/pnas.1322705111>

520 Parise, C. V., & Spence, C. (2009). ‘When birds of a feather flock together’:  
521 Synesthetic correspondences modulate audiovisual integration in non-  
522 synesthetes. *PloS One*, 4(5), e5664.

523 Parkinson, C., Kohler, P. J., Sievers, B., & Wheatley, T. (2012). Associations  
524 between auditory pitch and visual elevation do not depend on language:

525 Evidence from a remote population. *Perception*, 41(7), 854–861.  
526 <https://doi.org/10.1068/p7225>

527 Ramachandran, V. S., & Hubbard, E. M. (2001). Synaesthesia—a window into  
528 perception, thought and language. *Journal of Consciousness Studies*, 8(12),  
529 3–34.

530 Spence, C. (2011). Crossmodal correspondences: A tutorial review. *Attention*,  
531 *Perception*, & *Psychophysics*, 73(4), 971–995.  
532 <https://doi.org/10.3758/s13414-010-0073-7>

533 Spence, C. (2019). On the relative nature of (pitch-based) crossmodal  
534 correspondences. *Multisensory Research*, 32(3), 235–265.

535 Spence, C., & Deroy, O. (2012). Crossmodal Correspondences: Innate or Learned?  
536 *I-Perception*, 3(5), 316–318. <https://doi.org/10.1068/i0526ic>

537 Townsend, J. T., & Ashby, F. G. (1983). *Stochastic modeling of elementary*  
538 *psychological processes*. CUP Archive.  
539 [https://books.google.com/books?hl=en&lr=&id=pYw6AAAAIAAJ&oi=fnd&pg=](https://books.google.com/books?hl=en&lr=&id=pYw6AAAAIAAJ&oi=fnd&pg=PR13&dq=Stochastic+Modelling+of+Elementary+Psychological+Processes+&ots=bajR9cCOSL&sig=-9aVxhd-Nnwy9RE44U6bVbY6qJM)  
540 [PR13&dq=Stochastic+Modelling+of+Elementary+Psychological+Processes+&](https://books.google.com/books?hl=en&lr=&id=pYw6AAAAIAAJ&oi=fnd&pg=PR13&dq=Stochastic+Modelling+of+Elementary+Psychological+Processes+&ots=bajR9cCOSL&sig=-9aVxhd-Nnwy9RE44U6bVbY6qJM)  
541 [ots=bajR9cCOSL&sig=-9aVxhd-Nnwy9RE44U6bVbY6qJM](https://books.google.com/books?hl=en&lr=&id=pYw6AAAAIAAJ&oi=fnd&pg=PR13&dq=Stochastic+Modelling+of+Elementary+Psychological+Processes+&ots=bajR9cCOSL&sig=-9aVxhd-Nnwy9RE44U6bVbY6qJM)

542 Trimble, O. C. (1934). Localization of sound in the anterior-posterior and vertical  
543 dimensions of auditory space. *British Journal of Psychology*, 24(3), 320.

544 Walker, P., Bremner, J. G., Mason, U., Spring, J., Mattock, K., Slater, A., & Johnson,  
545 S. P. (2010). Preverbal Infants' Sensitivity to Synaesthetic Cross-Modality  
546 Correspondences. *Psychological Science*, 21(1), 21–25.  
547 <https://doi.org/10.1177/0956797609354734>

548 Walsh, V. (2003). A theory of magnitude: Common cortical metrics of time, space  
549 and quantity. *Trends in Cognitive Sciences*, 7(11), 483–488.

550  
551  
552  
553