



HAL
open science

Multisensory processes in birds: From single neurons to the influence of social interactions and sensory loss

Manon Rolland, Catherine del Negro, Nicolas Giret

► To cite this version:

Manon Rolland, Catherine del Negro, Nicolas Giret. Multisensory processes in birds: From single neurons to the influence of social interactions and sensory loss. *Neuroscience and Biobehavioral Reviews*, 2022, 143, pp.104942. 10.1016/j.neubiorev.2022.104942 . hal-04171727

HAL Id: hal-04171727

<https://cnrs.hal.science/hal-04171727>

Submitted on 19 Oct 2023

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 4.0 International License

1 **Multisensory processes in birds: from single neurons to the influence of social interactions and**
2 **sensory loss**

3 Manon Rolland, Catherine Del Negro, Nicolas Giret*

4 Institut des Neurosciences Paris Saclay, UMR9197 - Université Paris Saclay

5 151 route de la Rotonde

6 91400 SACLAY

7 France

8 *Corresponding author : nicolas.giret@universite-paris-saclay.fr

9

10 Abstract: World experiences involve multisensory stimulation that arises simultaneously from multiple
11 sources. Yet, we experience a coherent and unified world. Many studies have focused on how sensory
12 information from distinct modalities are integrated and showed that numerous behavioural and
13 cognitive benefits are provided by multisensory processes. Much work has been done with mammalian
14 models but research on avian species also expands our knowledge on multisensory processes. Avian
15 species exhibit a great diversity of behaviours and these species have provided evidence that
16 multisensory processes benefit by the learning that occurs in natural situations. Cross-modal
17 influences on the control of sensorimotor processes occur in circumstances of sensory loss. Also,
18 studies suggest pervasive multisensory influences throughout the avian brain. This review summarizes
19 research done on the imprinting behaviour of precocial bird species, on the ability of barn owls to
20 detect prey and on the vocal communication of songbirds.

21

22 Keywords: Multisensory integration, cross-modality, imprinting, vocal communication, prey detection

23

24 **1. Introduction**

25 We experience a unified external world from a vast amount of information processed through multiple
26 sensory systems. This leads us to assume convergence and integration of information from these
27 different modalities to create this coherent world. Within this framework, studies on humans and non-
28 human animals have emphasized numerous perceptual and behavioural benefits from experiencing
29 multiple sensory inputs (Stein et al., 2020). This includes a reduction in ambiguity, especially in noisy
30 environments, lower sensory thresholds, enhancement of the speed and/or accuracy of stimulus
31 detection, as well as facilitating the acquisition of higher cognitive functions and sensorimotor skills
32 such as those involved in language and speech (Frassinetti et al., 2002; Frens et al., 1995; Lovelace et
33 al., 2003; McDonald et al., 2000; Ross et al., 2007; Stein et al., 1989; Sumby and Pollack, 1954). The
34 mechanism by which multisensory information is combined by neural circuits has been studied for
35 many years. The number of identified multisensory anatomical circuits is continuously increasing and
36 suggests that the perceptual benefit of multisensory integration is supported by functional interaction
37 between sensory areas of different modalities from the early stages of sensory processing to more
38 integrative levels (Cappe et al., 2009; Choi et al., 2018; Ghazanfar and Schroeder, 2006; Lohse et al.,
39 2021a; Stein and Stanford, 2008a). Our understanding of neurophysiological mechanisms underlying
40 multisensory processes has also increased (see Box 1 and Figure 1) and has revealed a set of empirical
41 principles by which two sensory inputs interact to modify their neural responses (Stein and Stanford,
42 2008a; van Atteveldt et al., 2014).

43 Much of our knowledge of how multisensory processes are mediated, and at what level of hierarchy
44 processing they occur, has been achieved by mammal studies. Surveys on bird species have also been
45 useful in multisensory integration-related research. One of the main advantages of birds has been the
46 opportunity to exploit their naturally occurring behaviour and the species-specific abilities they use in
47 natural contexts. Bird groups exhibit a great diversity of behaviour, especially in their learning
48 behaviour and/or specific abilities. As some of the abilities that bird groups can exhibit are not shared
49 by mammalian species used as experimental animal models, birds enlarge the perspective on
50 multisensory processing function. They also allow highlighting shared computations and neural

51 mechanisms with mammals, including humans. This report aims to summarize the current knowledge
52 of multisensory processes based on behavioural, neuroanatomical and electrophysiological studies
53 made on different avian species. Multisensory processes embrace various forms of complexity, from
54 the “simple” cross-modal cue integration to top-down modulations by high-level cognitive functions
55 (see box 1) (Meijer et al., 2019). We thus emphasized the synergistic effects of multisensory and
56 learning processes in birds. We summarized the detection performance in barn owls because of their
57 remarkable abilities to perform cue integration of auditory and visual information to detect a prey. We
58 then described the neuronal mechanisms underlying cross-modal cue integration in the learning
59 context of conspecific (parent) recognition by focusing on the imprinting behaviour exhibited by
60 precocial avian species. Finally, we focus on the vocal communication in songbirds that has striking
61 parallels with human speech. In songbirds, considered as auditory specialists, recent studies have
62 opened up a new line of research. They have revealed the potential control of sensory inputs, other
63 than auditory, on song behaviour, even in deaf songbirds. They have also extended our current
64 knowledge on multisensory processes. They have provided evidence that cross-modal integration can
65 serve not only sensory perception but also motor control of vocal production, making possible to
66 finely adjust vocal acoustic features. Finally, we discuss how avian research can be informative on
67 corresponding processes in the mammal brain.

68 **2. Interconnected sensory pathways. Multisensory anatomical pathways in birds**

69 The main limitation of using birds as a model organism system was once thought to be the large
70 differences in brain organization between avian and mammalian species, including a lack of a
71 neuroarchitecture comparable to that of the cerebral cortex. However, studies have now revealed many
72 analogies or homologies in both brain structures and networks between birds and mammalian species.
73 A cortex-like organization based on connectivity patterns has demonstrated that the avian brain
74 organization is more similar to that of mammals than previously thought (Medina and Reiner, 2000a;
75 Reiner et al., 2004; Stacho et al., 2020).

76 Multisensory processes in mammalian brains have been thought to result from the convergence of the
77 different sensory streams in interconnected high-order cortical areas. The idea that multisensory

78 interaction is limited to high stages of brain processing has been challenged by studies demonstrating
79 such interactions in the midbrain, particularly in superior colliculus, to cortical brain regions including
80 primary sensory areas, classically considered as unisensory brain regions (Cappe et al., 2009; Stein
81 and Stanford, 2008a).

82 In the avian brain, heteromodal connections at the different stages of sensory processing have also
83 been described (Figure 2). Connections between the auditory and the visual pathways already occur in
84 the midbrain: the auditory nucleus mesencephalic lateralis, pars dorsalis (MLd), also known as inferior
85 colliculus and the optic tectum (OT), a midbrain visual area considered to be the avian homolog of the
86 superior colliculus (Gutfreund et al., 2002; Hyde and Knudsen, 2000; Knudsen and Knudsen, 1983;
87 Luksch et al., 2000). In many mammalian species, the generation and control of eye and head
88 movement involves the superior colliculus, characterized by aligned maps of space for auditory, visual
89 and tactile modalities (King and Hutchings, 1987; Sterbing et al., 2002; Wallace et al., 1996; Wallace
90 and Stein, 1996). As in the superior colliculus of mammals, multisensory neurons that respond both to
91 visual and auditory stimuli are common in the avian OT. In barn owls, auditory maps are located in the
92 OT. This shows a remarkable topographic resolution that is aligned to the visual one (Wagner et al.,
93 2013). Auditory or visual stimuli that originate from a single location in space activate the same set of
94 neurons in the avian superior colliculus (Knudsen, 1982) suggesting that neurons in the TeO nucleus
95 exhibit multisensory properties.

96 Importantly, multisensory convergence is not limited to the midbrain, similarly to mammals.
97 Connections between the auditory and visual pathways have been identified in the avian pallium,
98 where a cortex-like canonical circuitry exists. The analogue of the primary auditory cortex, so called
99 Field L, connects with the homologue of the primary visual cortex, the visual part of the Wulst,
100 (Medina and Reiner, 2000b; Stacho et al., 2020; Wild and Williams, 1999). At a higher stage, the non-
101 sensory associative area, the nidopallium caudolaterale (NCL), considered as the avian analogue of the
102 mammalian prefrontal cortex, receives inputs from both auditory and visual areas, as does its
103 mammalian counterpart (Güntürkün, 2005; Kröner and Güntürkün, 1999). Also, tangential projections
104 connect distant somatosensory, visual and auditory territories (Stacho et al., 2020). Thus, the

105 connectivity pattern of the avian brain can mediate multisensory processing at different stages of the
106 brain organization.

107 **3. Multisensory cue integration in an auditory specialist bird species: the barn owl**

108 Cross-modal cue integration, which benefits stimulus detection or discrimination, is considered as a
109 form of multisensory processing which mostly relies on bottom-up neural mechanisms. In this section,
110 we focus on one bird species, the barn owl (*Tyto alba*), who uses cross-modal cue integration to detect
111 preys. Neuronal mechanisms underlying this behaviour are in described in the light of mechanisms
112 observed in mammals, including humans.

113 **3.1 The prey detection behaviour**

114 The barn owl (*Tyto alba*) provides an appropriate model system for investigating interactions between
115 visual and auditory information in the context of natural behaviour. This nocturnal bird species is
116 characterized by its high and precise efficiency to detect small prey by hearing (Konishi, 1973;
117 Hausmann *et al.*, 2008; Singheiser *et al.*, 2010). The barn owl is clearly a species specialized in
118 auditory processing. Yet, barn owls use cross-modal cue integration of auditory and visual signals
119 when light conditions are suitable (Knudsen *et al.*, 1979; Konishi, 1973; Wagner *et al.*, 2013). When
120 barn owls hear or see a potentially salient signal, their head turns in a quick saccadic movement to face
121 that signal. Barn owls can use auditory or visual cues alone to detect their prey (Hausmann *et al.*,
122 2008; Konishi, 1973; Singheiser *et al.*, 2010). They are also able to exploit this signal combination
123 when it occurs simultaneously. In such a case, they use the fastest sensory signal, the auditory one, to
124 initiate their behaviour, *i.e.*, the saccadic movement and the most precise signal, the visual one, to
125 refine the location of the emitter. Consequently, this results in similar short response latency and high
126 spatial precision than when auditory and visual cues are presented alone, respectively (see more details
127 in a review Wagner *et al.*, 2013; Whitchurch and Takahashi, 2006). Beyond this potential benefit from
128 an audio-visual stimulus, the audio-visual combination can impact the target selection. Barn owls can
129 adjust their visual search behaviour by shifting their gaze, according to the location of the sound
130 source. They select the visual target closest to the sound source, suggesting that auditory and visual

131 information, when combined, are used for selecting the most relevant target rather than for simply
132 localizing stimuli (Hazan et al., 2015).

133 **3.2 Multisensory neuronal processes at play for detecting prey**

134 In barn owls, cross-modal interactions influence spiking activity of TeO neurons (Knudsen and
135 Brainard, 1991). Visual information occurring at a given location in the visual field enhances neuronal
136 responses to sound originating from the same location while no such modulation is observed when
137 visual stimuli originate elsewhere (Zahar et al., 2009). Also, bimodal responses of multisensory
138 neurons exceed responses to unimodal stimuli (Zahar et al., 2009). However, this bimodal
139 enhancement of neuronal responsiveness cannot be predicted by responses to unimodal stimuli since
140 the number of spikes driven by the audio-visual compound is not higher than the sum of spikes driven
141 by each unimodal stimulus (Zahar et al., 2009). In terms of the temporal pattern of spikes, bimodal
142 compounds evoke responses that are better phase locked to stimuli presentation than responses based
143 on the summation of unimodal responses. This suggests that multisensory tectal neurons exhibit a non-
144 linear integration of auditory and visual information which is a property shared with neurons of the
145 mammalian superior colliculus (Meredith and Stein, 1986). This might allow neurons to better detect
146 events (Zahar et al., 2009). Cross-modal interactions also influence auditory maps in barn owls
147 (Knudsen and Brainard, 1991). During development, manipulation of visual information, based on a
148 prismatic effect, induces an experience-dependent shift in the auditory space map (Knudsen and
149 Brainard, 1995, 1991). A similar effect has been reported in the mammalian superior colliculus:
150 visually depriving ferrets during development leads to an abnormal auditory space map in this brain
151 region (King, 1993; King and Carlile, 1993). This refinement of the auditory space map according to
152 visual inputs in these two evolutionary distant animal species illustrates a crucial role of interactions
153 between auditory and visual spatial information in the shaping of neurons. It also suggests a required
154 matching of the representation of both auditory and visual space that depends on the experience. The
155 adaptive alignment of the auditory space map sheds light on the importance of multisensory
156 processes in coherent perception of the space.

157 The TeO, where many neurons are multisensory, is part of the tectofugal pathway, one of the two main
158 avian visual pathways in addition to the thalamofugal pathway (Fig 1; Knudsen, 1982; Knudsen and
159 Brainard, 1991; Reches and Gutfreund, 2009; Zahar et al., 2009). It sends its efferents to the thalamic
160 nucleus rotundus which, in turn, projects to the telencephalic entopallium. This region is considered to
161 be an analogue of the extrastriate cortex (Jarvis et al., 2005; Karten and Shimizu, 1989). The TeO
162 receives inputs from the retinae but also from the external nucleus of the inferior colliculus, where an
163 auditory space map has been described (Knudsen et al., 1987; Wagner et al., 1987). The external
164 nucleus of the inferior colliculus projects to the main auditory nucleus of the thalamus of the
165 ascending auditory system (the nucleus ovoïdalis). Given the hypothetical convergence of auditory
166 and visual inputs, studies have examined multisensory properties of neurons along the tectofugal
167 pathway. Neurons in both entopallium and nucleus rotundus respond to both auditory and visual
168 stimuli (Reches and Gutfreund, 2009). In the entopallium, neurons show enhanced and earlier peaked
169 responses to spatiotemporally congruent multisensory stimuli in comparison to responses to unimodal
170 or incongruent stimuli. Responses of entopallium neurons to bimodal stimuli exceed the simple linear
171 summation of unimodal responses, reflecting a non-linear integration of visual and auditory stimuli
172 (Reches and Gutfreund, 2009). Therefore, the entopallium appears to be a brain region where
173 multisensory properties exhibit some of the principles that characterize multisensory integration and
174 that were reported in mammals (Stein and Stanford, 2008b; Wallace et al., 1996). Auditory responses
175 in the entopallium were examined, after pharmacological inactivation of the TeO, to determine their
176 origin (Reches and Gutfreund, 2009). A complete loss of auditory responses in the entopallium
177 supports the hypothesis that auditory information ascends through the tectofugal pathway, which is a
178 visual pathway in other species. In the context of this information *i.e.* bimodal responses in areas
179 whose homolog regions in mammals are known to be visual, these areas might also be multisensory in
180 other species but with a contribution of the auditory modality varying among species (Reches and
181 Gutfreund, 2009). This view is consistent with the possibility that cortical brain areas are multisensory
182 rather than sensory specific (Ghazanfar and Schroeder, 2006). In the chicken, a species not considered
183 to be an auditory specialist, spatial auditory information is conveyed to the TeO. Therefore, the barn
184 owl, as a species considered as specialized in auditory processing, led researchers to reconsider how

185 sensory information is processed throughout the brain (Reches and Gutfreund, 2009). Interestingly, the
186 stimulation of a premotor area involved in the control of gaze direction induces changes in the
187 auditory space map of the TeO (Winkowski and Knudsen, 2006), thus illustrating a top-down
188 modulation that is reminiscent of the modulation of visual responses by stimulation of the frontal
189 cortex in non-human primates (Moore and Armstrong, 2003).

190 In a framework for examining the importance of multimodalities on perception and detection of
191 stimuli, another neural phenomenon is worth consideration. The stimulus-specific adaptation (SSA)
192 phenomenon occurs in neurons that show a decrease in their responsiveness to frequent stimuli while
193 rare stimuli reinstate strong responses. This neuronal adaptation has been proposed to allow deviance
194 detection or surprise saliency by emphasizing unexpected stimuli relative to expected ones. The
195 finding that neurons in the entopallium exhibit a SSA effect for visual stimuli suggests that SSA may
196 be enhanced by pairing visual with auditory stimuli. One study addressed this issue and reported that
197 rare bimodal stimuli evoke stronger responses than rare visual stimuli. However, no such difference
198 was observed in responses to frequent stimuli (Reches et al., 2010). In addition, responses were
199 enhanced when the visual stimulus was presented with an auditory stimulus that matched spatially but
200 not when the bimodal stimuli were spatially incongruent. In a study based on a simple adaptation
201 paradigm where the same stimulus was repeated several times, the first unexpected bimodal stimulus
202 triggered a multisensory enhancement in the TeO while responses to subsequent stimuli did not show
203 clear multisensory enhancement (Zahar et al., 2009). These findings provide evidence that the
204 integration of visual and auditory information can benefit detection of unexpected or salient events and
205 is achieved, as in mammals, by a network including midbrain and cortical areas, most probably in a
206 bottom-up manner (Choi et al., 2018).

207 **4. An example of cue integration learning: the imprinting behaviour**

208 The bottom-up transfer of information in the sensory cortex can be modulated by top-down signals
209 from higher-order areas involved in various cognitive functions. Studies in mammals, including
210 humans (Robinson and Sloutsky, 2004), have shown that learning induces changes in the perception of
211 sensory stimuli and that top-down signals contribute to modulate the neuronal responses of sensory

212 cortex (Polley et al., 2006; Poort et al., 2015). In birds, the imprinting behaviour is a remarkable
213 paradigm used for the study of neural mechanisms underlying learning processes. It has provided
214 evidence that multisensory experience promotes learning, which might rely on the modulation of low-
215 level areas by high-level areas through changes in neuronal responses to sensory stimuli.

216 Imprinting has been described in precocial bird species such as chickens, ducks, and quail where
217 neonate individuals are relatively mature and capable of locomotion in the immediate postnatal period
218 (Lorenz, 1937). It involves the young bird approaching and following a stimulus, *i.e.* an animated
219 object such as a moving red dot, or an individual such as a sibling (or mother, father), learning the
220 characteristics of this stimulus, and consequently expressing a preference for that stimulus compared
221 to a novel stimulus. Filial imprinting has been interpreted as being adaptive as it presumably increases
222 survival of the young birds.

223 Initial surveys mainly investigated the imprinting of young birds to a unimodal stimulus such as a
224 visual stimulus (Bateson, 1978; Kovach and Hickox, 1971). The observation that mothers emit calls
225 after egg hatching led researchers to expose neonate birds to either an audio-visual compound
226 stimulus, *i.e.* a visual stimulus presented simultaneously with the maternal call or a visual stimulus
227 alone (Bolhuis and Honey, 1994; Honey and Bolhuis, 1997). In domestic chicks, the presentation of
228 the maternal call during training selectively enhanced subsequent chick preference for the visual
229 training stimulus over a new one in comparison to conditions where the visual stimuli were presented
230 alone (Kampen and Bolhuis, 1991; Van Kampen and Bolhuis, 1993). Reciprocally, simultaneous
231 exposure to bimodal stimuli of the training compound affected the approach behaviour toward the
232 training auditory stimulus (Kampen and Bolhuis, 1991). An enhanced preference for the visual
233 stimulus was not observed when stimuli were presented sequentially or when the visual stimulus
234 preceded the audio-visual compound (Honey and Bolhuis, 1997). This indicated the importance of the
235 temporal congruency in the exposure to multisensory stimuli (Bolhuis, 1999). Altogether, these studies
236 have provided evidence that learning in early life can benefit from multisensory information.

237 A recent study that focused on recognition and generalization abilities also provided evidence that
238 multisensory information benefits learning and perceptual abilities (Versace et al., 2017).

239 Generalization through imprinting behaviour enables chicks to recognize familiar individuals in spite
240 of changes in their appearance (e.g., mother hen and siblings seen from different angles). In this case,
241 the natural imprinting stimulus was a single object such as a particular individual. Yet, an object can
242 also consist in a temporal series of multiple elements characterized by their sequential organization,
243 e.g., a series of multiple distinct sounds as, for example, words or sentences in human speech. Thus,
244 the question was whether chicks are able to learn to recognize stimuli organized according to either an
245 XX or XY pattern and to generalize to novel stimuli where X and Y are novel items. Multisensory
246 information helps to identify and recognize objects or scenes (Shams and Seitz, 2008), enhancing the
247 perceptual representation of sensory features of stimuli used. For example, a visual stimulus can
248 enhance the representation of a temporally paired sound stimulus in the auditory cortex (Atilgan et al.,
249 2018). This study demonstrated that multisensory information also helps to extract sequential patterns
250 available for further generalization and categorization processes. The presentations of XX and XY
251 patterns, where both X and Y consisted of audio-visual stimuli, was more effective than patterns in the
252 visual or auditory modality for discriminating between novel stimuli according to their sequential
253 pattern. This study provided evidence that multisensory information promotes the capacity to extract
254 the sequential structure of stimuli (Versace et al., 2017).

255 This study raises the question of how multisensory association benefits learning. It is unclear whether,
256 and to what extent, multimodal learning differs from unisensory learning. In imprinting studies, a
257 significant behavioural approach to the training stimuli was observed in tests where presentations of
258 only one sensory stimulus of the audiovisual compound, either the auditory stimulus or the visual
259 stimulus, was presented (Kampen and Bolhuis, 1991; Van Kampen and Bolhuis, 1993). This
260 suggested imprinting-induced changes in the representation of each unisensory information. Thus, it is
261 unclear if these changes in the representation of each unisensory information stimulus are
262 accompanied by the emergence of a representation of the multisensory compound that could form a
263 unitary object. Some studies have addressed this issue by examining the neural mechanisms
264 underlying imprinting. Recordings of neuronal responses were performed in birds trained with a visual
265 stimulus associated with a taped maternal call (Brown and Horn, 1994; Horn et al., 2001; Nicol et al.,

266 1995). In chicks, the intermediate and medial mesopallium (IMM), a multisensory region that receives
267 afferents from all sensory systems and projects to motor areas (Brown and Horn, 1994; Horn, 1985;
268 Town and McCabe, 2011), is a critical area for the imprinting behaviour (McCabe et al., 1992, 1981).
269 Strong evidence implicates the IMM in recognition memory that underpins imprinting (Horn, 2004).
270 Neuronal recordings showed that imprinting using a bimodal stimulus affected both the number of
271 IMM neurons responsive to the visual or auditory component when presented alone, in comparison to
272 a novel stimulus of the same modality. This provided evidence for alterations in unisensory
273 representations of stimuli areas (Brown and Horn, 1994; Town and McCabe, 2011). Town and
274 McCabe (2011) also compared neuronal responses using different types of audiovisual stimuli,
275 including the training ones, novel or incongruent stimuli, the latter combining the training stimulus of
276 one sensory with a novel stimulus in the other modality. In this case, the training visual stimulus was
277 coupled with an auditory stimulus. Neuronal activity did not show any effect of training on responses
278 driven by the audiovisual training stimulus. Responses to the training compound stimulus were similar
279 to that of a novel one, although chicks were able to behaviourally discriminate between the imprinted
280 and novel audiovisual stimuli (Town and McCabe, 2011). Yet, results also indicated that the
281 incongruent audiovisual stimulus drove more responses from IMM neurons in imprinted than in naïve
282 chicks (Town and McCabe, 2011) suggesting that IMM could detect mismatching between the
283 auditory and the visual components of the imprinted stimuli. This suggests an encoding of the
284 multisensory training stimuli, but elsewhere in the avian brain than in IMM. It does not rule out the
285 possibility of a multisensory representation of the compound. These differences between imprinted
286 and naïve birds also illustrate that learning yields to changes in response properties of sensory neurons,
287 suggesting a local effect of learning and/or a top-down modulation exerted by IMM neurons on
288 sensory neurons.

289 In conclusion, precocial avian species through their imprinting behaviour have provided support for
290 the possibility that multimodal information promotes learning and also offered the opportunity to
291 study how multisensory inputs may lead to a unified representation of the multimodal object.

292 **5. Multisensory influences in a model system for studying vocal communication: the**
293 **songbirds.**

294 Songbird research supports the idea that being an auditory specialist does not exclude the influence of
295 other modalities, especially visual information. As an animal model system, songbirds have also
296 provided unique evidence that cues from multisensory modalities can act in concert to promote the
297 perceptual and sensorimotor skill learning required in vocal communication.

298 Songbirds are a group of more than 4000 species with complex vocalization behaviour and high
299 listening sensitivity. Birdsong is an acoustically complex communication signal that provides rich
300 information about the singer and plays a crucial role in social interactions, especially in the contexts of
301 mate attraction and territorial defence. Songbirds are one of the few non-human animal groups that
302 learn their singing behaviour. During a critical period of at least several weeks, immatures need to hear
303 an adult vocal model, called the tutor song, to learn to copy that song. They use auditory feedback to
304 drive and maintain what they vocally produce. Eliminating or disrupting auditory feedback during
305 song development results in abnormal songs (Iyengar and Bottjer, 2002; Konishi, 1965, 1964; Marler
306 and Sherman, 1983). Deprivation of normal auditory feedback in adult songbirds also results in
307 gradual changes in song structure (Brainard and Doupe, 2001, 2000; Horita et al., 2008; Lombardino
308 and Nottebohm, 2000; Nordeen and Nordeen, 1992). Song behaviour has remarkable parallels to
309 human speech given the strong dependence of hearing one's own vocal signals in both infants and
310 adults (Doupe and Kuhl, 1999). However, human speech perception can be affected by visual
311 information, as in the famous McGurk effect (McGurk and Macdonald, 1976) and ventriloquism effect
312 (Slutsky and Recanzone, 2001): visual information provided by movements of the mouth and face
313 strongly influences what is perceived (Kuhl, 2010, 2007). Also, human speech learning critically relies
314 on auditory information perception although multisensory integration through information given by
315 caregivers participates to learning progress (Mason et al., 2019). While literature from studies
316 performed in mammalian models is rich on data providing evidence for multisensory processing in
317 audiovisual perception and learning, knowledge remains sparse on how multimodal integration can
318 shape motor production, *i.e.* lead to changes in acoustic features of vocalizations. Songbird vocal

319 behaviour offers the great opportunity to shed light on this role that remains to be explored. Songbird
320 research has mostly focused on the role of auditory signals on vocal learning, perception and
321 production. However, the influence of other sensory signals, including visual ones, has long been
322 assumed (Hultsch et al., 1999; Todt et al., 1979).

323 **5.1 Multisensory perception in songbirds**

324 In songbirds, the auditory processing of vocal communicative signals involves a substantial number of
325 brain regions including auditory areas as well as the interconnected neural networks specialized in
326 song production and song learning, known as the song system (Fig. 3) in which neurons respond to
327 song playbacks (e.g. Dave and Margoliash, 2000; Giret et al., 2014; Prather et al., 2008). Little is
328 known about the neural pathways and mechanisms underlying multisensory processes in song
329 behaviour. Some studies have reported anatomical evidence that nuclei of the song system receive
330 inputs from the visual system (Paterson and Bottjer, 2017; Wild, 1994). Physiological studies showed
331 that neurons in the song nucleus, at the departure of both the motor and the learning pathways, known
332 as the HVC (proper name), respond to light flashes (Bischof and Engelage, 1985; Seki and Okanoya,
333 2006). Indirect evidence comes from investigations focused on the impact of social conditions at the
334 neuronal level (Cousillas et al., 2006; George et al., 2011, 2012; Hessler and Doupe, 1999a; Menardy
335 et al., 2014). Some of these investigations attempted to examine the respective contribution of visual
336 and auditory components in neuronal responses (Avey et al., 2005; Cousillas et al., 2006). Social
337 conditions may affect neuronal responsiveness in auditory areas (Menardy et al., 2014; Woolley and
338 Doupe, 2008) including the auditory area considered as a homologue to the primary auditory cortex of
339 mammals and consequently as a unisensory area, known as the Field L complex (Cousillas et al.,
340 2006; George et al., 2012). In European starlings (*Sturnus vulgaris*), exposure to an audio-visual
341 combination consisting of playback of conspecific songs and a picture of an individual evoked
342 neuronal responses that differed from those evoked by the song playback alone (Cousillas et al., 2006;
343 George et al., 2012). Therefore, even in a species known as a powerful model system for exploring the
344 auditory processing of complex signals, visual stimuli may tune the primary auditory cortex.
345 Furthermore, the modulation of auditory responses of single neurons in the Field L by visual stimuli

346 depends on the degree of familiarity of the stimuli: seeing a familiar or an unfamiliar conspecific
347 yields to a suppression or an enhancement of auditory responses, respectively (George et al., 2011).
348 Factors of social cognition may thus have an impact on the integration of multisensory signals and this
349 impact may occur in primary sensory areas.

350 The avian primary auditory cortex, along with other sensory areas, projects to the NCL (Fig. 1),
351 considered to be analogous to the mammalian prefrontal cortex (Güntürkün, 2005). As such, the NCL
352 is a good candidate to investigate the neurophysiological mechanisms of multisensory integration at
353 the single cell scale. Electrophysiological recordings of NCL neurons were performed in carrion crows
354 (*Corvus corone*) previously trained to form arbitrary associations between a sound and a colour.
355 During the learning, a set of NCL neurons was tuned to one or the other association between the
356 auditory and visual stimuli (Moll and Nieder, 2015), independently of the ordering of the stimuli (first
357 visual or auditory) (Moll and Nieder, 2017), such that the bird response could be predicted based on
358 the neuronal activity (Moll and Nieder, 2015). Corvids are songbirds who exhibit impressive cognitive
359 abilities related to interindividual competition and cooperation that depend on the processing of both
360 vocal and visual signals (Bugnyar, 2021). They live in complex fission-fusion societies in which
361 interindividual recognition is critical for group maintenance. This recognition involves discrimination
362 among conspecifics which can rely on multisensory signals, as in large-billed crows (*Corvus*
363 *macrorhynchos*) who use both the calls and the appearance of conspecifics to discriminate group
364 members from non-group members (Kondo et al., 2012). The integration of multisensory signals is
365 thus essential for social cognition, and the NCL may be a core area for this process (Brecht and
366 Nieder, 2020). The question remains open on how individual neurons, either in unisensory (Field L) or
367 high-order (NCL) areas acquire their multisensory properties. One possibility is that multisensory
368 properties actually arise in the NCL and contribute to modulate neuronal activity within sensory areas
369 in a top-down manner. Another hypothesis is that unisensory areas (e.g. Field L) acquire their
370 multisensory properties that eventually converge through bottom-up processes to higher-order areas,
371 such as in the NCL. Further experiments are required to assess more precisely these processes.

372 Multisensory processes also play a role in female behaviour, particularly in their courtship behaviour
373 that they display as receiver of songs (Amy et al., 2015). Receiver behaviour has been described as
374 influenced by signals from different modalities in various avian species (Candolin, 2003; Mitoyen et
375 al., 2019; Partan and Marler, 2005). In the non-songbird pigeon (*Columba livia*), an audio-videotape
376 was more effective in triggering courtship displays from females than auditory or visual signals alone
377 (Partan et al., 2005). In the brown-headed cowbird, the preference of females was assessed through
378 their courtship displays to compound stimuli made from videos of a male paired with song playbacks.
379 Manipulation of visual or acoustic parameters of the compound stimulus indicated that changing the
380 intensity of the video affected the attractiveness of the song playback and provided additional evidence
381 for how visual signals influence song processing (Ronald et al., 2018, 2017). Female zebra finches
382 elicit strong preference for the courtship songs compared to non-courtship songs, although the acoustic
383 differences between these two song types are very subtle (Chen et al., 2017; Clayton, 1988; Miller,
384 1979; Woolley and Doupe, 2008). While singing for a female, male zebra finches also perform a
385 stereotyped dance including head and body movements (Ullrich et al., 2016). Zebra finch songs may
386 thus convey information on the social context in which they were produced (the male singing while
387 alone or while socially interacting with a female). Interestingly, an fMRI study revealed that courtship
388 and non-courtship songs induce different levels of activation of the caudocentral nidopallium (NCC), a
389 sensory area outside of the classic avian auditory pathway, suggesting that this area plays a key role in
390 the assessment of social, putatively multisensory, signals (Van Ruijssevelt et al., 2018). In canaries,
391 females produce multimodal signals, a copulation solicitation display and calls, in response to
392 conspecific songs (Amy et al., 2015). Lesions of the HVC nucleus in females change their behavioural
393 responses to conspecific and heterospecific songs, suggesting that the HVC belongs to a neural
394 network involved in the control of courtship cues (Del Negro et al., 1998). Anatomical surveys suggest
395 that this neural network extend to downstream areas including premotor nuclei (RA and the respiratory
396 premotor nucleus retroambigualis (RAm), Wild and Botelho, 2015). This overview suggests that
397 feedforward mechanisms from sensory areas to premotor nuclei drive the multimodal courtship cues
398 provided by females in response to male courtship songs.

399 **5.2 A multisensory approach of vocal learning**

400 Humans are not the only species in which vocal learning and perception are affected by social
401 interactions. Song learning is also shaped by social interactions that rely on sensory inputs from
402 several modalities. This suggests that information from multiple sources may be integrated or may
403 interact to influence the perceiver. Surveys focused on birdsong learning often compare the learning
404 performances of juveniles that experienced social interaction to those of juveniles that did not. For
405 example, auditory information through tape tutoring is as efficient as live tutoring in song sparrows
406 (*Melospiza melodia*) and swamp sparrows (*Melospiza georgiana*), but this is not true for other species
407 (Catchpole and Slater, 2003). For highly social birds like zebra finches (*Taeniopygia guttata*), song is
408 better learned from live tutors than from song playback (Eales, 1989; Slater, 1988). This raises the
409 question of whether visual and acoustic information provided by the adult tutor during interactions can
410 promote the song learning process. Such a question must then consider that other aspects can influence
411 vocal learning processes. The motivation to learn or the attention to the stimuli source (e.g. a talking
412 parent or a singing bird) are involved in vocal learning (Chen et al., 2016). In an attempt to control for
413 these factors, a focus on basic perceptual mechanisms indicated that, in nightingales (*Luscinia*
414 *megarhynchos* L.), juveniles can be tutored with a tape but their vocal learning performance is
415 improved when strobe flashes are tuned to be synchronous with song notes (Hultsch et al., 1999).
416 Other studies have addressed the issue by examining song acquisition in young birds that are allowed
417 either to visually and acoustically interact with a live tutor, or to other conditions that include
418 interacting only acoustically with a tutor or passively hearing the adult model (Chen et al., 2016;
419 Derégnaucourt et al., 2013; Varkevisser et al., 2022). Tutoring for as little as one day in zebra finches
420 is sufficient to highlight the benefit of multimodal information on song learning: interacting both
421 acoustically and visually with tutors made songs more similar to the tutor's songs than interacting only
422 acoustically or passively hearing the adult models (Chen et al., 2016). Also, young zebra finches that
423 are either housed with, could see and hear, or only hear the same tutor showed that having access only
424 to auditory information delays the timing of, and reduces, the vocal learning (Varkevisser et al., 2022).
425 Song learning is consistently enhanced when the juvenile's own songs are contingently paired with a
426 video playback of a female exhibiting a special behaviour, *i.e.* a wing stroke, that is considered as a

427 non-vocal feedback since it is commonly observed in response to attractive songs (Miller et al., 2008;
428 Smith et al., 2000; West and King, 1988). Interestingly, young birds who received non contingent
429 visual displays of a female feedback sang less accurate songs (Carouso-Peck and Goldstein, 2019).
430 Another interesting approach is to use a stuffed model, or even better, robotic models that can be
431 controlled by the experimenter (Araguas et al., 2022; Bolhuis et al., 1999). A robot can for example
432 provide vocal responses (songs or calls) to the vocalization produced by the juvenile either in a
433 contingent or non-contingent way, allowing to control for social interactions (Araguas et al., 2022). In
434 humans, visual and acoustic information provided by adults during social interactions with infants
435 enhanced acquisition of speech sounds (Goldstein et al., 2003; Kuhl, 2010). Babbling infants who
436 received contingent feedback rapidly restructured their vocalizations (Goldstein and Schwade, 2008).
437 Therefore, evidence that visual cues act in concert with auditory cues or influence the auditory
438 processing to promote vocal learning in songbirds as well as in humans suggest that multimodal
439 information plays a key role in vocal communication (Goldstein et al., 2003).

440 Birdsong learning has received attention from neuroscientists (e.g. Gobes et al., 2010; Okubo et al.,
441 2015; Ölveczky et al., 2011; Shank and Margoliash, 2009). However, it still remains to investigate
442 mechanisms underlying the impact of multisensory integration during development at the neuronal or
443 network level. A network of brain areas has been proposed to mediate multimodal integration
444 processes during song ontogeny, based on anatomical data (Paterson and Bottjer, 2017). Now,
445 functional characterization remains to be carried out to better understand how multimodal integration
446 maps onto the network.

447 **5.3 A multisensory approach on the neuronal bases of birdsong production**

448 Beside vocal perception and learning, vocal production in adults is another cognitive function in which
449 multisensory processes likely play a role (Tremblay et al., 2003). In human speech, the monitoring of
450 what is said does not exclusively rely on auditory information (Ito and Ostry, 2010; Lametti et al.,
451 2012; Tremblay et al., 2003). Songbirds as an animal model system provide insights into this issue. In
452 adult songbirds, social interactions based on multiple sensory modalities may shape singing behaviour.
453 For example, male brown-headed cowbirds (*Molothrus ater*) rapidly adjust their song according to a

454 visual display, such as a wing stroke, provided by females (West and King, 1988). However, no study
455 has investigated the extent to which socially-induced variations in song production depend on the
456 visual component. In songbirds, the social environment is not the sole condition in which changes in
457 vocal outputs may occur. Plastic changes in song features may be induced in a reinforcement learning
458 paradigm. This paradigm commonly consists in providing auditory feedback, e.g. a loud white noise
459 contingent to the fundamental frequency and the pitch of one song element (called a syllable)
460 (Andalman and Fee, 2009; Canopoli et al., 2014; Tumer and Brainard, 2007; Warren et al., 2011).
461 Singing birds alter the pitch of the target syllable allowing them to avoid the white noise delivery.
462 Investigations based on this pitch-shifting learning paradigm have primarily investigated the role of
463 auditory information in the monitoring of vocal signals. Recent studies have examined whether non-
464 auditory feedback can substitute for the white noise. They provided the first evidence that a visual (a
465 light-off or a light-on) as well as a tactile stimulus (applied to the scalp), respectively, can guide
466 changes in the pitch value (McGregor et al., 2021; Zai et al., 2020). These results support the idea that
467 the ability to carefully adjust one's own vocalizations, shared by humans and songbirds, can efficiently
468 rely on information from several different modalities. These experiments have opened a new line of
469 perspective supporting the idea that vocal production relies on multimodal processes.

470 Support for this hypothesis is provided by other studies. Multisensory processes are commonly
471 considered to be susceptible to circumstances producing sensory loss. Zai and coll. (2020) showed that
472 exposure of adult birds to visual stimuli in the pitch-shifting learning paradigm could drive vocal
473 changes in hearing birds, as previously mentioned, but also in deafened birds. Male deafened zebra
474 finches after bilateral cochlea removal are able to shift the pitch of one target syllable in the paradigm
475 in which a light was switched off every time the pitch produced was lower (or higher) than a threshold
476 value (Zai et al., 2020). These results demonstrate that visual signals in songbirds may exert an
477 efficient control on vocal production when the auditory feedback is lacking. Importantly, birds with
478 normal hearing and deafened birds exhibit differences in their behaviour adaptation. Deafened birds
479 shifted the syllable pitch to increase the rate of visual stimuli delivery while hearing birds did the
480 opposite. In both cases, somatosensory signals were available to control how to sing (Suthers et al.,

481 2002). Bird behaviour varied depending on the number of sensory modalities that conveyed
482 information. Visual and somatosensory information was perceived in deafened birds while auditory
483 information was also available to hearing birds. The magnitude of pitch changes induced by visual
484 information also differed between hearing and deaf birds, being lower in hearing than in deafened
485 birds (Zai et al., 2020). This lower magnitude of song changes in hearing bird could reflect a form of
486 auditory dominance which minimizes the influence of visual signals. Similar sensory dominance was
487 reported in an audio-visual fear conditioning paradigm in which the conditioned fear-response of
488 rodents was weakened when a loud noise was simultaneously presented (Iurilli et al., 2012). In this
489 experiment, the very high amplitude of the auditory stimulus provides more salient or reliable
490 information than the visual stimulus. The auditory stimulus is thus prioritized over the visual stimulus,
491 resulting in a segregation of sensory modalities in a bottom-up manner (Meijer et al., 2019). In the
492 experiment performed in birds, hearing birds had access to both auditory and visual feedback while
493 deafened birds only had access to visual feedback (Zai et al., 2020). The loss of auditory feedback may
494 thus reveal the salience or the relevance of the visual feedback, that is constantly overshadowed by
495 auditory feedback in hearing birds. Since hearing birds are able to adjust their song less efficiently
496 than deafened birds, it suggests a dominance of one sensory (here auditory) modality in hearing birds
497 and so a partial segregation of sensory modalities in a bottom-up manner. Given the fine motor control
498 that the pitch shifting learning task requires, it might also be hypothesized that top-down processes
499 from high order brain areas do occur to adjust the behavioural strategy (changing the song to get more
500 or less visual stimuli). Further experiments will be necessary to understand how the respective
501 influences of the different sensory modalities interact and control vocal production.

502 A specialized neural pathway is required for auditory-guided song learning during development and
503 song maintenance in adults and is analogous to the basal ganglia-thalamo-cortical circuit of mammals.
504 This pathway is involved in the drive of vocal plasticity induced either by visual or tactile stimuli in
505 the pitch-shifting learning paradigm (McGregor et al., 2022; Zai et al., 2020). Lesioning Area X or
506 LMAN, two nuclei of the avian basal ganglia-thalamo-cortical circuit, prevent the non-auditory-
507 induced vocal plasticity (McGregor et al., 2022; Zai et al., 2020) demonstrating that this circuit is not

508 solely involved in the processing of auditory feedback in songbirds. Other evidence comes from
509 assessment of the impact of social context focusing on this pathway (Kao and Brainard, 2006). In
510 zebra finches, adult males produce a highly stereotyped song with slight variations between renditions.
511 In the presence of a female, males produce courtship songs that becomes even more stereotyped
512 (Sossinka and Bohner, 1980; Dunn & Zann, 1996; Hessler & Doupe, 1999a; Kao et al, 2005). The
513 basal ganglia-thalamo-cortical pathway is strongly affected by social cues (Hessler and Doupe, 1999a;
514 Jarvis et al., 1998; Kao et al., 2008). Neural recordings showed that the pattern of singing-related
515 neural activity in two regions, the Area X and the LMAN, is dependent on whether the bird is singing
516 in isolation or to a female (Hessler and Doupe, 1999a, 1999b; Kao et al., 2008). Therefore, as
517 information from different modalities may guide learning-induced changes in song structure, the avian
518 basal ganglia-thalamo-cortical pathway could be viewed as mediating vocal changes from different
519 sensory information and not exclusively from auditory information. Multisensory neurons in the basal
520 ganglia of mammals exhibit the ability to integrate multisensory information (Nagy et al., 2006; Reig
521 and Silberberg, 2014). Multisensory integration may help in the effective processing of sensory events
522 and the changes in the environment during motor actions controlled by the basal ganglia. Interestingly,
523 an early survey reported a positive correlation between the volume of the area X and of the visual
524 nucleus rotundus and the rate of courtship songs, but not with the rate of non-courtship songs in male
525 brown-headed cowbirds. This result thus highlights that the female-directed singing is a multimodal
526 courtship behaviour not only recruiting a song-related nucleus (Area X), but also a visual area (nucleus
527 rotundus).

528 In summary, research findings highlight the powerful influence of multisensory conditions on
529 behaviour, in juveniles as well as in adult males and females. Birdsong, as human speech, plays a
530 critical role in the social life of individuals. The findings demonstrate that social interactions, through
531 multisensory signals, also strongly influence vocal behaviour. Neurophysiological studies have
532 identified many brain areas where signals from multiple sensory sources may potentially converge.
533 Yet, it is unclear how the neural mechanisms by which information from different sensory modalities
534 interact or are integrated to influence songbird behaviour. There is currently great interest in the basal-

535 ganglia forebrain circuits for deciphering neural mechanisms underlying the learning and generation of
536 song behaviour. Recent evidence indicates that these circuits are required for non-auditory information
537 to drive changes in the song features of adult songbirds (McGregor et al., 2022; Zai et al., 2020).
538 Future research will help resolve the neural mechanisms by which non-auditory information exerts its
539 influence.

540 **Conclusion and Perspectives**

541 This review focused on various avian taxa, behaviours and brain networks and highlighted fields of
542 research that deserve more investigations to better understand the processing and control of
543 multisensory signals (Figure 4). Our main goal was to highlight the current state of the art on
544 multisensory processing in birds. Recent reviews have been focused on the neuronal substrates
545 underlying multisensory processes but they were essentially restricted to research in mammalian
546 models (Meijer et al., 2019; Stein et al., 2020). Yet, a clear strength of the research on multisensory
547 processing in birds is, to our view, related to the type of behaviours being investigated. Prey detection
548 as well as imprinting or vocal communication are behaviours observed in natural conditions. Although
549 controlling all factors is a tremendous task when studying natural behaviours, they shed light on
550 specific processes that could not be revealed by simpler behavioural paradigms (Khastkhodaei et al.,
551 2016; Meijer et al., 2019). The research findings for example support the idea that facing a given
552 object through more than one sensory modality, as is often the case in the natural world, makes
553 cognitive abilities, including the evaluation of the object salience as well as acquisition of
554 sensorimotor skills, more effective.

555 This review aimed at highlighting the relevance of studying multisensory processing in another group
556 of animals than mammals. To this end, we have attempted to present multisensory processes in the
557 light of what is known from mammalian models, when data were available. This comparative
558 approach has allowed us to establish that some of the neuronal mechanisms identified in mammalian
559 models are also implemented in birds, thus arguing for evolutionary conserved strategies within brain
560 networks to achieve multisensory integration. Several examples indeed argue for bottom-up or top-
561 down modulation various forms of multisensory processing. In birds as well as in mammalian species,

562 multisensory processes affect numerous brain areas and pathways, ranging from the midbrain to
563 primary sensory or high-order sensory cortical areas including brain regions of the basal ganglia-
564 thalamo-cortical pathways. The availability of the range of techniques (e.g. large scale
565 electrophysiological recordings, optogenetic methods) used in birds now allows to investigate not only
566 neuronal dynamics but also to assess causal relationships between neuronal activity and behaviours
567 that may depend on multisensory integration (Guitchounts et al., 2013; Roberts et al., 2012).

568 Beyond similarities between birds and mammals highlighted in this review, birds have allowed to raise
569 original questions that mammalian species used as animal models left unresolved. The importance of
570 multisensory signals in the fine motor control required by birdsong production to adjust acoustic
571 features of vocalizations (McGregor et al., 2022; Zai et al., 2020), or in the case of a sensory loss, has
572 opened a new line of research. This new field of research leads to assume that multisensory
573 information may guide changes in vocal production in species that learn to vocalize such as humans.
574 Studies of multisensory processing in birds should provide new insights on mechanisms underlying
575 speech acquisition and production in the human brain.

576

- 578 Amy, M., Salvin, P., Naguib, M., Leboucher, G., 2015. Female signalling to male song in the domestic
579 canary, *Serinus canaria*. Royal Society Open Science 2, 140196.
580 <https://doi.org/10.1098/rsos.140196>
- 581 Andalman, A.S., Fee, M.S., 2009. A basal ganglia-forebrain circuit in the songbird biases motor
582 output to avoid vocal errors. Proc Natl Acad Sci USA 106, 12518–12523.
583 <https://doi.org/10.1073/pnas.0903214106>
- 584 Angelaki, D.E., Gu, Y., DeAngelis, G.C., 2009. Multisensory integration: psychophysics,
585 neurophysiology, and computation. Current Opinion in Neurobiology 19, 452–458.
586 <https://doi.org/10.1016/j.conb.2009.06.008>
- 587 Araguas, A., Guellai, B., Gauthier, P., Richer, F., Montone, G., Chopin, A., Derégnaucourt, S., 2022.
588 Design of a robotic zebra finch for experimental studies on developmental song learning.
589 Journal of Experimental Biology 225, jeb242949. <https://doi.org/10.1242/jeb.242949>
- 590 Atilgan, H., Town, S.M., Wood, K.C., Jones, G.P., Maddox, R.K., Lee, A.K.C., Bizley, J.K., 2018.
591 Integration of Visual Information in Auditory Cortex Promotes Auditory Scene Analysis
592 through Multisensory Binding. Neuron 97, 640–655.e4.
593 <https://doi.org/10.1016/j.neuron.2017.12.034>
- 594 Avey, M.T., Phillmore, L.S., MacDougall-Shackleton, S.A., 2005. Immediate early gene expression
595 following exposure to acoustic and visual components of courtship in zebra finches. Behav
596 Brain Res 165, 247–253. <https://doi.org/10.1016/j.bbr.2005.07.002>
- 597 Bateson, P., 1978. Sexual imprinting and optimal outbreeding. Nature 273, 659–660.
598 <https://doi.org/10.1038/273659a0>
- 599 Bischof, H.-J., Engelage, J., 1985. Flash evoked responses in a song control nucleus of the zebra finch
600 (*Taeniopygia guttata castanotis*). Brain Res 326, 370–374. [https://doi.org/10.1016/0006-8993\(85\)90048-4](https://doi.org/10.1016/0006-8993(85)90048-4)
- 601
- 602 Bizley, J.K., Nodal, F.R., Bajo, V.M., Nelken, I., King, A.J., 2007. Physiological and Anatomical
603 Evidence for Multisensory Interactions in Auditory Cortex. Cerebral Cortex 17, 2172–2189.
604 <https://doi.org/10.1093/cercor/bhl128>
- 605 Bolhuis, J.J., 1999. Early learning and the development of filial preferences in the chick. Behav Brain
606 Res 98, 245–252. [https://doi.org/10.1016/S0166-4328\(98\)00090-4](https://doi.org/10.1016/S0166-4328(98)00090-4)
- 607 Bolhuis, J.J., Honey, R.C., 1994. Within-event learning during filial imprinting. J Exp Psychol Anim
608 Behav Process 20, 240–248. <https://doi.org/10.1037//0097-7403.20.3.240>
- 609 Bolhuis, J.J., Van Mil, D.P., Houx, B.B., 1999. Song learning with audiovisual compound stimuli in
610 zebra finches. Animal Behaviour 58, 1285–1292. <https://doi.org/10.1006/anbe.1999.1266>
- 611 Brainard, M.S., Doupe, A.J., 2001. Postlearning consolidation of birdsong: stabilizing effects of age
612 and anterior forebrain lesions. J. Neurosci. 21, 2501–2517.
- 613 Brainard, M.S., Doupe, A.J., 2000. Interruption of a basal ganglia–forebrain circuit prevents plasticity
614 of learned vocalizations. Nature 404, 762–766. <https://doi.org/10.1038/35008083>
- 615 Brecht, K.F., Nieder, A., 2020. Parting self from others: Individual and self-recognition in birds.
616 Neurosci Biobehav Rev 116, 99–108. <https://doi.org/10.1016/j.neubiorev.2020.06.012>
- 617 Brown, M.W., Horn, G., 1994. Learning-related alterations in the visual responsiveness of neurons in
618 a memory system of the chick brain. Eur J Neurosci 6, 1479–1490.
619 <https://doi.org/10.1111/j.1460-9568.1994.tb01009.x>
- 620 Bugnyar, T., 2021. Raven Social Cognition and Behavior, in: Kaufman, A.B., Kaufman, J.C., Call, J.
621 (Eds.), The Cambridge Handbook of Animal Cognition, Cambridge Handbooks in
622 Psychology. Cambridge University Press, Cambridge, pp. 322–342.
623 <https://doi.org/10.1017/9781108564113.018>
- 624 Candolin, U., 2003. The use of multiple cues in mate choice. Biol Rev Camb Philos Soc 78, 575–595.
625 <https://doi.org/10.1017/s1464793103006158>
- 626 Canopoli, A., Herbst, J.A., Hahnloser, R.H.R., 2014. A higher sensory brain region is involved in
627 reversing reinforcement-induced vocal changes in a songbird. J. Neurosci. 34, 7018–7026.
628 <https://doi.org/10.1523/JNEUROSCI.0266-14.2014>

629 Cappe, C., Rouiller, E.M., Barone, P., 2009. Multisensory anatomical pathways. *Hearing Research*
630 258, 28–36. <https://doi.org/10.1016/j.heares.2009.04.017>

631 Carouso-Peck, S., Goldstein, M.H., 2019. Female social feedback reveals non-imitative mechanisms
632 of vocal learning in zebra finches. *Curr Biol* 29, 631–636.
633 <https://doi.org/10.1016/j.cub.2018.12.026>

634 Chen, Y., Clark, O., Woolley, S.C., 2017. Courtship song preferences in female zebra finches are
635 shaped by developmental auditory experience. *Proc. R. Soc. B* 284, 20170054.
636 <https://doi.org/10.1098/rspb.2017.0054>

637 Chen, Y., Matheson, L.E., Sakata, J.T., 2016. Mechanisms underlying the social enhancement of vocal
638 learning in songbirds. *PNAS* 201522306. <https://doi.org/10.1073/pnas.1522306113>

639 Choi, I., Lee, J.-Y., Lee, S.-H., 2018. Bottom-up and top-down modulation of multisensory
640 integration. *Curr Opin Neurobiol* 52, 115–122. <https://doi.org/10.1016/j.conb.2018.05.002>

641 Clayton, N.S., 1988. Song discrimination learning in zebra finches. *Animal Behaviour* 36, 1016–1024.
642 [https://doi.org/10.1016/S0003-3472\(88\)80061-7](https://doi.org/10.1016/S0003-3472(88)80061-7)

643 Cousillas, H., George, I., Mathelier, M., Richard, J.-P., Henry, L., Hausberger, M., 2006. Social
644 experience influences the development of a central auditory area. *Naturwissenschaften* 93,
645 588–596. <https://doi.org/10.1007/s00114-006-0148-4>

646 Dave, A.S., Margoliash, D., 2000. Song replay during sleep and computational rules for sensorimotor
647 vocal learning. *Science* 290, 812–816. <https://doi.org/10.1126/science.290.5492.812>

648 Dehner, L.R., Keniston, L.P., Clemo, H.R., Meredith, M.A., 2004. Cross-modal Circuitry Between
649 Auditory and Somatosensory Areas of the Cat Anterior Ectosylvian Sulcal Cortex: A ‘New’
650 Inhibitory Form of Multisensory Convergence. *Cerebral Cortex* 14, 387–403.
651 <https://doi.org/10.1093/cercor/bhg135>

652 Del Negro, C., Gahr, M., Leboucher, G., Kreutzer, M., 1998. The selectivity of sexual responses to
653 song displays: effects of partial chemical lesion of the HVC in female canaries. *Behav. Brain*
654 *Res.* 96, 151–159.

655 Derégnaucourt, S., Poirier, C., Kant, A.V. der, Linden, A.V. der, Gahr, M., 2013. Comparisons of
656 different methods to train a young zebra finch (*Taeniopygia guttata*) to learn a song. *J.*
657 *Physiol. Paris*, Special issue: Physiological mechanisms of song learning and production 107,
658 210–218. <https://doi.org/10.1016/j.jphysparis.2012.08.003>

659 Doupe, A.J., Kuhl, P.K., 1999. Birdsong and human speech: common theme and mechanisms. *Ann*
660 *Rev Neurosci* 22, 567–631. <https://doi.org/10.1146/annurev.neuro.22.1.567>

661 Dunn, A.M., Zann, R.A., 1996. Undirected song in wild zebra finch flocks: contexts and effects of
662 mate removal. *Ethology* 102, 529–539. <https://doi.org/10.1111/j.1439-0310.1996.tb01145.x>

663 Eales, L.A., 1989. The influences of visual and vocal interaction on song learning in Zebra finches.
664 *Animal Behaviour* 37, 507–508. [https://doi.org/10.1016/0003-3472\(89\)90097-3](https://doi.org/10.1016/0003-3472(89)90097-3)

665 Frassinetti, F., Bolognini, N., Làdavas, E., 2002. Enhancement of visual perception by crossmodal
666 visuo-auditory interaction. *EXP. BRAIN RES.* 147, 332–343. <https://doi.org/10.1007/s00221-002-1262-y>

667

668 Frens, M.A., Van Opstal, A.J., Van der Willigen, R.F., 1995. Spatial and temporal factors determine
669 auditory-visual interactions in human saccadic eye movements. *Percept Psychophys* 57, 802–
670 816. <https://doi.org/10.3758/bf03206796>

671 Garner, A.R., Keller, G.B., 2022. A cortical circuit for audio-visual predictions. *Nat Neurosci* 25, 98–
672 105. <https://doi.org/10.1038/s41593-021-00974-7>

673 George, I., Cousillas, H., Richard, J.-P., Hausberger, M., 2012. Experience with adults shapes
674 multisensory representation of social familiarity in the brain of a songbird. *PLOS ONE* 7,
675 e38764. <https://doi.org/10.1371/journal.pone.0038764>

676 George, I., Richard, J.-P., Cousillas, H., Hausberger, M., 2011. No need to Talk, I Know You:
677 Familiarity Influences Early Multisensory Integration in a Songbird’s Brain. *Front. Behav.*
678 *Neurosci.* 4. <https://doi.org/10.3389/fnbeh.2010.00193>

679 Ghazanfar, A.A., Schroeder, C.E., 2006. Is neocortex essentially multisensory? *Trends Cogn Sci* 10,
680 278–285. <https://doi.org/10.1016/j.tics.2006.04.008>

681 Giret, N., Kornfeld, J., Ganguli, S., Hahnloser, R.H.R., 2014. Evidence for a causal inverse model in
682 an avian cortico-basal ganglia circuit. *Proc Natl Acad Sci USA* 111, 6063–6068.
683 <https://doi.org/10.1073/pnas.1317087111>

684 Gobes, S.M.H., Zandbergen, M.A., Bolhuis, J.J., 2010. Memory in the making: localized brain
685 activation related to song learning in young songbirds. *Proc R Soc Lond B Biol Sci* 277,
686 3343–3351. <https://doi.org/10.1098/rspb.2010.0870>

687 Goldstein, M.H., King, A.P., West, M.J., 2003. Social interaction shapes babbling: testing parallels
688 between birdsong and speech. *Proc. Natl. Acad. Sci. U.S.A.* 100, 8030–8035.
689 <https://doi.org/10.1073/pnas.1332441100>

690 Goldstein, M.H., Schwade, J.A., 2008. Social Feedback to Infants' Babbling Facilitates Rapid
691 Phonological Learning: *Psychological Science*.

692 Guitchounts, G., Markowitz, J.E., Liberti, W.A., Gardner, T.J., 2013. A carbon-fiber electrode array
693 for long-term neural recording. *J. Neural Eng.* 10, 046016. <https://doi.org/10.1088/1741-2560/10/4/046016>

695 Güntürkün, O., 2005. The avian 'prefrontal cortex' and cognition. *Current Opinion in Neurobiology,*
696 *Motor systems / Neurobiology of behaviour* 15, 686–693.
697 <https://doi.org/10.1016/j.conb.2005.10.003>

698 Gutfreund, Y., Zheng, W., Knudsen, E.I., 2002. Gated visual input to the central auditory system.
699 *Science*. <https://doi.org/10.1126/science.1073712>

700 Hausmann, L., Plachta, D.T.T., Singheiser, M., Brill, S., Wagner, H., 2008. In-flight corrections in
701 free-flying barn owls (*Tyto alba*) during sound localization tasks. *J Exp Biol* 211, 2976–2988.
702 <https://doi.org/10.1242/jeb.020057>

703 Hazan, Y., Kra, Y., Yarin, I., Wagner, H., Gutfreund, Y., 2015. Visual-auditory integration for visual
704 search: a behavioral study in barn owls. *Frontiers in Integrative Neuroscience* 9, 11.
705 <https://doi.org/10.3389/fnint.2015.00011>

706 Hessler, N.A., Doupe, A.J., 1999a. Social context modulates singing-related neural activity in the
707 songbird forebrain. *Nat Neurosci* 2, 209–211. <https://doi.org/10.1038/6306>

708 Hessler, N.A., Doupe, A.J., 1999b. Singing-related neural activity in a dorsal forebrain–basal ganglia
709 circuit of adult zebra finches. *J. Neurosci.* 19, 10461–10481.

710 Honey, R.C., Bolhuis, J.J., 1997. Imprinting, conditioning, and within-event learning. *The Quarterly*
711 *Journal of Experimental Psychology Section B* 50, 97–110. <https://doi.org/10.1080/713932648>

712 Horita, H., Wada, K., Jarvis, E.D., 2008. Early onset of deafening-induced song deterioration and
713 differential requirements of the pallial-basal ganglia vocal pathway. *Eur J Neurosci* 28, 2519–
714 2532. <https://doi.org/10.1111/j.1460-9568.2008.06535.x>

715 Horn, G., 2004. Pathways of the past: the imprint of memory. *Nat Rev Neurosci* 5, 108–120.
716 <https://doi.org/10.1038/nrn1324>

717 Horn, G., 1985. *Memory, Imprinting, and the Brain: An Inquiry into Mechanisms*. Clarendon Press.

718 Horn, G., Nicol, A.U., Brown, M.W., 2001. Tracking memory's trace. *Proc Natl Acad Sci U S A* 98,
719 5282–5287. <https://doi.org/10.1073/pnas.091094798>

720 Hultsch, H., Schleuss, F., Todt, D., 1999. Auditory–visual stimulus pairing enhances perceptual
721 learning in a songbird. *Anim Behav* 58, 143–149. <https://doi.org/10.1006/anbe.1999.1120>

722 Hultsch, H., Todt, D., 2004. Learning to sing, in: Marler, P., Slabbekoorn, H. (Eds.), *Nature's Music*
723 *The Science of Birdsong*. Elsevier Academic Press, San Diego, California, pp. 80–107.

724 Hyde, P.S., Knudsen, E.I., 2000. Topographic projection from the optic tectum to the auditory space
725 map in the inferior colliculus of the barn owl. *J Comp Neurol* 421, 146–160.
726 [https://doi.org/10.1002/\(sici\)1096-9861\(20000529\)421:2<146::aid-cne2>3.0.co;2-5](https://doi.org/10.1002/(sici)1096-9861(20000529)421:2<146::aid-cne2>3.0.co;2-5)

727 Ibrahim, L.A., Mesik, L., Ji, X., Fang, Q., Li, H., Li, Y., Zingg, B., Zhang, L.I., Tao, H.W., 2016.
728 Cross-Modality Sharpening of Visual Cortical Processing through Layer-1-Mediated
729 Inhibition and Disinhibition. *Neuron* 89, 1031–1045.
730 <https://doi.org/10.1016/j.neuron.2016.01.027>

731 Ito, T., Ostry, D.J., 2010. Somatosensory Contribution to Motor Learning Due to Facial Skin
732 Deformation. *J Neurophysiol* 104, 1230–1238. <https://doi.org/10.1152/jn.00199.2010>

733 Iurilli, G., Ghezzi, D., Olcese, U., Lassi, G., Nazzaro, C., Tonini, R., Tucci, V., Benfenati, F., Medini,
734 P., 2012. Sound-Driven Synaptic Inhibition in Primary Visual Cortex. *Neuron* 73, 814–828.
735 <https://doi.org/10.1016/j.neuron.2011.12.026>

736 Iyengar, S., Bottjer, S.W., 2002. The Role of Auditory Experience in the Formation of Neural Circuits
737 Underlying Vocal Learning in Zebra Finches. *J. Neurosci.* 22, 946–958.
738 <https://doi.org/10.1523/JNEUROSCI.22-03-00946.2002>

739 Jarvis, E.D., Güntürkün, O., Bruce, L., Csillag, A., Karten, H., Kuenzel, W., Medina, L., Paxinos, G.,
740 Perkel, D.J., Shimizu, T., Strieder, G., Wild, J.M., Ball, G.F., Dugas-Ford, J., Durand, S.E.,
741 Hough, G.E., Husband, S., Kubikova, L., Lee, D.W., Mello, C.V., Powers, A., Siang, C.,
742 Smulders, T.V., Wada, K., White, S.A., Yamamoto, K., Yu, J., Reiner, A., Butler, A.B.,
743 Consortium, A.B.N., 2005. Avian brains and a new understanding of vertebrate brain
744 evolution. *Nat Rev Neurosci* 6, 151–159.

745 Jarvis, E.D., Scharff, C., Grossman, M.R., Ramos, J.A., Nottebohm, F., 1998. For whom the bird
746 sings: Context-dependent gene expression. *Neuron* 21, 775–788.
747 [https://doi.org/10.1016/S0896-6273\(00\)80594-2](https://doi.org/10.1016/S0896-6273(00)80594-2)

748 Kampen, H.S. van, Bolhuis, J.J., 1991. Auditory learning and filial imprinting in the chick. *Behaviour*
749 117, 303–319.

750 Kao, M.H., Brainard, M.S., 2006. Lesions of an avian basal ganglia circuit prevent context-dependent
751 changes to song variability. *J. Neurophysiol.* 96, 1441–1455.
752 <https://doi.org/10.1152/jn.01138.2005>

753 Kao, M.H., Wright, B.D., Doupe, A.J., 2008. Neurons in a forebrain nucleus required for vocal
754 plasticity rapidly switch between precise firing and variable bursting depending on social
755 context. *J. Neurosci.* 28, 13232–13247. <https://doi.org/10.1523/JNEUROSCI.2250-08.2008>

756 Karten, H.J., Shimizu, T., 1989. The origins of neocortex: connections and lamination as distinct
757 events in evolution. *J Cogn Neurosci* 1, 291–301. <https://doi.org/10.1162/jocn.1989.1.4.291>

758 Kayser, C., Petkov, C.I., Logothetis, N.K., 2008. Visual Modulation of Neurons in Auditory Cortex.
759 *Cerebral Cortex* 18, 1560–1574. <https://doi.org/10.1093/cercor/bhm187>

760 Khastkhodaei, Z., Jurjut, O., Katzner, S., Busse, L., 2016. Mice Can Use Second-Order, Contrast-
761 Modulated Stimuli to Guide Visual Perception. *J. Neurosci.* 36, 4457–4469.
762 <https://doi.org/10.1523/JNEUROSCI.4595-15.2016>

763 King, A., 1993. The Wellcome Prize Lecture. A map of auditory space in the mammalian brain: neural
764 computation and development. *Experimental Physiology* 78, 559–590.
765 <https://doi.org/10.1113/expphysiol.1993.sp003708>

766 King, A.J., Carlile, S., 1993. Changes induced in the representation of auditory space in the superior
767 colliculus by rearing ferrets with binocular eyelid suture. *Exp Brain Res* 94, 444–455.
768 <https://doi.org/10.1007/BF00230202>

769 King, A.J., Hutchings, M.E., 1987. Spatial response properties of acoustically responsive neurons in
770 the superior colliculus of the ferret: a map of auditory space. *J Neurophysiol* 57, 596–624.
771 <https://doi.org/10.1152/jn.1987.57.2.596>

772 Knöpfel, T., Sweeney, Y., Radulescu, C.I., Zabouri, N., Doostdar, N., Clopath, C., Barnes, S.J., 2019.
773 Audio-visual experience strengthens multisensory assemblies in adult mouse visual cortex.
774 *Nat Commun* 10, 5684. <https://doi.org/10.1038/s41467-019-13607-2>

775 Knudsen, E.I., 1982. Auditory and visual maps of space in the optic tectum of the owl. *J. Neurosci.* 2,
776 1177–1194. <https://doi.org/10.1523/JNEUROSCI.02-09-01177.1982>

777 Knudsen, E.I., Blasdel, G.G., Konishi, M., 1979. Sound localization by the barn owl (*Tyto alba*)
778 measured with the search coil technique. *J. Comp. Physiol.* 133, 1–11.
779 <https://doi.org/10.1007/BF00663105>

780 Knudsen, E.I., Brainard, M.S., 1995. Creating a Unified Representation of Visual and Auditory Space
781 in the Brain. *Annu Rev Neurosci* 18, 19–43.
782 <https://doi.org/10.1146/annurev.ne.18.030195.000315>

783 Knudsen, E.I., Brainard, M.S., 1991. Visual instruction of the neural map of auditory space in the
784 developing optic tectum. *Science.* <https://doi.org/10.1126/science.2063209>

785 Knudsen, E.I., du Lac, S., Esterly, S.D., 1987. Computational maps in the brain. *Annu Rev Neurosci*
786 10, 41–65. <https://doi.org/10.1146/annurev.ne.10.030187.000353>

787 Knudsen, E.I., Knudsen, P.F., 1983. Space-Mapped auditory projections from the inferior colliculus to
788 the optic tectum in the barn owl (*Tyto alba*). *Journal of Comparative Neurology* 218, 187–196.
789 <https://doi.org/10.1002/cne.902180206>

790 Kondo, N., Izawa, E.-I., Watanabe, S., 2012. Crows cross-modally recognize group members but not
791 non-group members. *Proceedings of the Royal Society B: Biological Sciences.*
792 <https://doi.org/10.1098/rspb.2011.2419>

793 Konishi, M., 1973. How the Owl Tracks Its Prey: Experiments with trained barn owls reveal how their
794 acute sense of hearing enables them to catch prey in the dark. *American Scientist* 61, 414–424.
795 Konishi, M., 1965. Effects of deafening on song development in American robins and black-headed
796 grosbeaks. *Z Tierpsychol* 22, 584–599.
797 Konishi, M., 1964. Effects of deafening on song development in two species of juncos. *Condor* 66,
798 85–102.
799 Kovach, J.K., Hickox, J.E., 1971. Color preferences and early perceptual discrimination learning in
800 domestic chicks. *Dev Psychobiol* 4, 255–267. <https://doi.org/10.1002/dev.420040304>
801 Kröner, S., Güntürkün, O., 1999. Afferent and efferent connections of the caudolateral neostriatum in
802 the pigeon (*Columba livia*): A retro- and anterograde pathway tracing study. *J Comp Neurol*
803 407, 228–260. [https://doi.org/10.1002/\(SICI\)1096-9861\(19990503\)407:2<228::AID-](https://doi.org/10.1002/(SICI)1096-9861(19990503)407:2<228::AID-CNE6>3.0.CO;2-2)
804 [CNE6>3.0.CO;2-2](https://doi.org/10.1002/(SICI)1096-9861(19990503)407:2<228::AID-CNE6>3.0.CO;2-2)
805 Kuhl, P.K., 2010. Brain mechanisms in early language acquisition. *Neuron* 67, 713–727.
806 <https://doi.org/10.1016/j.neuron.2010.08.038>
807 Kuhl, P.K., 2007. Is speech learning “gated” by the social brain? *Dev Sci* 10, 110–120.
808 <https://doi.org/10.1111/j.1467-7687.2007.00572.x>
809 Lakatos, P., Chen, C.-M., O’Connell, M.N., Mills, A., Schroeder, C.E., 2007. Neuronal oscillations
810 and multisensory interaction in primary auditory cortex. *Neuron* 53, 279–292.
811 <https://doi.org/10.1016/j.neuron.2006.12.011>
812 Lametti, D.R., Nasir, S.M., Ostry, D.J., 2012. Sensory preference in speech production revealed by
813 simultaneous alteration of auditory and somatosensory feedback. *J Neurosci* 32, 9351–9358.
814 <https://doi.org/10.1523/JNEUROSCI.0404-12.2012>
815 Lohse, M., Dahmen, J.C., Bajo, V.M., King, A.J., 2021a. Subcortical circuits mediate communication
816 between primary sensory cortical areas in mice. *Nat Commun* 12, 3916.
817 <https://doi.org/10.1038/s41467-021-24200-x>
818 Lohse, M., Dahmen, J.C., Bajo, V.M., King, A.J., 2021b. Subcortical circuits mediate communication
819 between primary sensory cortical areas in mice. *Nat Commun* 12, 3916.
820 <https://doi.org/10.1038/s41467-021-24200-x>
821 Lombardino, A.J., Nottebohm, F., 2000. Age at deafening affects the stability of learned song in adult
822 male zebra finches. *J. Neurosci.* 20, 5054–5064.
823 Lorenz, K.Z., 1937. The companion in the bird’s world. *The Auk* 54, 245–273.
824 <https://doi.org/10.2307/4078077>
825 Lovelace, C.T., Stein, B.E., Wallace, M.T., 2003. An irrelevant light enhances auditory detection in
826 humans: a psychophysical analysis of multisensory integration in stimulus detection. *Brain*
827 *Res Cogn Brain Res* 17, 447–453. [https://doi.org/10.1016/s0926-6410\(03\)00160-5](https://doi.org/10.1016/s0926-6410(03)00160-5)
828 Luksch, H., Gauger, B., Wagner, H., 2000. A Candidate Pathway for a Visual Instructional Signal to
829 the Barn Owl’s Auditory System. *J. Neurosci.* 20, RC70–RC70.
830 <https://doi.org/10.1523/JNEUROSCI.20-08-j0002.2000>
831 Marler, P., Sherman, V., 1983. Song structure without auditory feedback: emendations of the auditory
832 template hypothesis. *J. Neurosci.* 3, 517–531.
833 Mason, G.M., Goldstein, M.H., Schwade, J.A., 2019. The role of multisensory development in early
834 language learning. *Journal of Experimental Child Psychology* 183, 48–64.
835 <https://doi.org/10.1016/j.jecp.2018.12.011>
836 McCabe, B.J., Davey, J.E., Horn, G., 1992. Impairment of learning by localized injection of an N-
837 methyl-D-aspartate receptor antagonist into the hyperstriatum ventrale of the domestic chick.
838 *Behav Neurosci* 106, 947–947.
839 McCabe, B.J., Horn, G., Bateson, P.P.G., 1981. Effects of restricted lesions of the chick forebrain on
840 the acquisition of filial preferences during imprinting. *Brain Res* 205, 29–37.
841 [https://doi.org/10.1016/0006-8993\(81\)90717-4](https://doi.org/10.1016/0006-8993(81)90717-4)
842 McDonald, J.J., Teder-Sälejärvi, W.A., Hillyard, S.A., 2000. Involuntary orienting to sound improves
843 visual perception. *Nature* 407, 906–908. <https://doi.org/10.1038/35038085>
844 McGregor, J.N., Grassler, A., Jaffe, P., Jacob, A., Brainard, M.S., Sober, S.J., 2021. Shared
845 mechanisms of auditory and non-auditory vocal learning in the songbird brain. *bioRxiv*
846 2021.12.09.471883. <https://doi.org/10.1101/2021.12.09.471883>

847 McGregor, J.N., Grassler, A.L., Jaffe, P.I., Jacob, A.L., Brainard, M.S., Sober, S.J., 2022. Shared
848 mechanisms of auditory and non-auditory vocal learning in the songbird brain. *eLife* 11,
849 e75691. <https://doi.org/10.7554/eLife.75691>

850 Mcgurk, H., Macdonald, J., 1976. Hearing lips and seeing voices. *Nature* 264, 746.
851 <https://doi.org/10.1038/264746a0>

852 Medina, L., Reiner, A., 2000a. Do birds possess homologues of mammalian primary visual,
853 somatosensory and motor cortices? *Trends Neurosci* 23, 1–12. [https://doi.org/10.1016/S0166-2236\(99\)01486-1](https://doi.org/10.1016/S0166-2236(99)01486-1)

854
855 Medina, L., Reiner, A., 2000b. Do birds possess homologues of mammalian primary visual,
856 somatosensory and motor cortices? *Trends in Neurosciences* 23, 1–12.
857 [https://doi.org/10.1016/S0166-2236\(99\)01486-1](https://doi.org/10.1016/S0166-2236(99)01486-1)

858 Meijer, G.T., Mertens, P.E.C., Pennartz, C.M.A., Olcese, U., Lansink, C.S., 2019. The circuit
859 architecture of cortical multisensory processing: Distinct functions jointly operating within a
860 common anatomical network. *Progress in Neurobiology* 174, 1–15.
861 <https://doi.org/10.1016/j.pneurobio.2019.01.004>

862 Menardy, F., Giret, N., Del Negro, C., 2014. The presence of an audience modulates responses to
863 familiar call stimuli in the male zebra finch forebrain. *Eur J Neurosci* 40, 3338–3350.
864 <https://doi.org/10.1111/ejn.12696>

865 Meredith, M.A., Stein, B.E., 1986. Visual, auditory, and somatosensory convergence on cells in
866 superior colliculus results in multisensory integration. *J Neurophysiol* 56, 640–662.
867 <https://doi.org/10.1152/jn.1986.56.3.640>

868 Miller, D.B., 1979. The acoustic basis of mate recognition by female Zebra finches (*Taeniopygia*
869 *guttata*). *Animal Behaviour* 27, Part 2, 376–380. [https://doi.org/10.1016/0003-3472\(79\)90172-6](https://doi.org/10.1016/0003-3472(79)90172-6)

870
871 Miller, J.L., King, A.P., West, M.J., 2008. Female social networks influence male vocal development
872 in brown-headed cowbirds, *Molothrus ater*. *Anim Behav* 76, 931–941.
873 <https://doi.org/10.1016/j.anbehav.2008.05.011>

874 Mitoyen, C., Quigley, C., Fusani, L., 2019. Evolution and function of multimodal courtship displays.
875 *Ethology* 125, 503–515. <https://doi.org/10.1111/eth.12882>

876 Moll, F.W., Nieder, A., 2017. Modality-invariant audio-visual association coding in crow endbrain
877 neurons. *Neurobiology of Learning and Memory* 137, 65–76.
878 <https://doi.org/10.1016/j.nlm.2016.11.011>

879 Moll, F.W., Nieder, A., 2015. Cross-Modal Associative Mnemonic Signals in Crow Endbrain
880 Neurons. *Current Biology* 25, 2196–2201. <https://doi.org/10.1016/j.cub.2015.07.013>

881 Moore, T., Armstrong, K.M., 2003. Selective gating of visual signals by microstimulation of frontal
882 cortex. *Nature* 421, 370–373. <https://doi.org/10.1038/nature01341>

883 Morgan, M.L., DeAngelis, G.C., Angelaki, D.E., 2008. Multisensory Integration in Macaque Visual
884 Cortex Depends on Cue Reliability. *Neuron* 59, 662–673.
885 <https://doi.org/10.1016/j.neuron.2008.06.024>

886 Nagy, A., Eördegh, G., Paróczy, Z., Márkus, Z., Benedek, G., 2006. Multisensory integration in the
887 basal ganglia. *European Journal of Neuroscience* 24, 917–924. <https://doi.org/10.1111/j.1460-9568.2006.04942.x>

888
889 Nicol, A.U., Brown, M.W., Horn, G., 1995. Neurophysiological Investigations of a Recognition
890 Memory System for Imprinting in the Domestic Chick. *Eur J Neurosci* 7, 766–776.
891 <https://doi.org/10.1111/j.1460-9568.1995.tb00680.x>

892 Nordeen, K.W., Nordeen, E.J., 1992. Auditory feedback is necessary for the maintenance of
893 stereotyped song in adult zebra finches. *Behav. Neural Biol.* 57, 58–66.

894 Ohshiro, T., Angelaki, D.E., DeAngelis, G.C., 2011. A normalization model of multisensory
895 integration. *Nat Neurosci* 14, 775–782. <https://doi.org/10.1038/nn.2815>

896 Okubo, T.S., Mackevicius, E.L., Payne, H.L., Lynch, G.F., Fee, M.S., 2015. Growth and splitting of
897 neural sequences in songbird vocal development. *Nature* 528, 352–357.
898 <https://doi.org/10.1038/nature15741>

899 Ölveczky, B.P., Otchy, T.M., Goldberg, J.H., Aronov, D., Fee, M.S., 2011. Changes in the neural
900 control of a complex motor sequence during learning. *J Neurophysiol* 106, 386–397.
901 <https://doi.org/10.1152/jn.00018.2011>

902 Partan, S., Yelda, S., Price, V., Shimizu, T., 2005. Female pigeons, *Columba livia*, respond to
903 multisensory audio/video playbacks of male courtship behaviour. *Anim Behav* 70, 957–966.
904 <https://doi.org/10.1016/j.anbehav.2005.03.002>

905 Partan, S.R., Marler, P., 2005. Issues in the classification of multimodal communication signals. *Am*
906 *Nat* 166, 231–245. <https://doi.org/10.1086/431246>

907 Paterson, A.K., Bottjer, S.W., 2017. Cortical inter-hemispheric circuits for multimodal vocal learning
908 in songbirds. *J Comp Neurol* 525, 3312–3340. <https://doi.org/10.1002/cne.24280>

909 Poirier, C., Henry, L., Mathelier, M., Lumineau, S., Cousillas, H., Hausberger, M., 2004. Direct social
910 contacts override auditory information in the song-learning process in starlings (*Sturnus*
911 *vulgaris*). *J Comp Psychol* 118, 179–193. <https://doi.org/10.1037/0735-7036.118.2.179>

912 Polley, D.B., Steinberg, E.E., Merzenich, M.M., 2006. Perceptual learning directs auditory cortical
913 map reorganization through top-down influences. *Journal of Neuroscience* 26, 4970–4982.
914 <https://doi.org/10.1523/JNEUROSCI.3771-05.2006>

915 Poort, J., Khan, A.G., Pachitariu, M., Nemri, A., Orsolich, I., Krupic, J., Bauza, M., Sahani, M., Keller,
916 G.B., Mrsic-Flogel, T.D., Hofer, S.B., 2015. Learning Enhances Sensory and Multiple Non-
917 sensory Representations in Primary Visual Cortex. *Neuron* 86, 1478–1490.
918 <https://doi.org/10.1016/j.neuron.2015.05.037>

919 Prather, J.F., Peters, S., Nowicki, S., Mooney, R., 2008. Precise auditory-vocal mirroring in neurons
920 for learned vocal communication. *Nature* 451, 305–310. <https://doi.org/10.1038/nature06492>

921 Reches, A., Gutfreund, Y., 2009. Auditory and multisensory responses in the tectofugal pathway of
922 the barn owl. *J. Neurosci.* 29, 9602–9613. <https://doi.org/10.1523/JNEUROSCI.6117-08.2009>

923 Reches, A., Netser, S., Gutfreund, Y., 2010. Interactions between stimulus-specific adaptation and
924 visual auditory integration in the forebrain of the barn owl. *J Neurosci* 30, 6991–6998.
925 <https://doi.org/10.1523/JNEUROSCI.5723-09.2010>

926 Reig, R., Silberberg, G., 2014. Multisensory Integration in the Mouse Striatum. *Neuron* 83, 1200–
927 1212. <https://doi.org/10.1016/j.neuron.2014.07.033>

928 Reiner, A., Perkel, D.J., Mello, C.V., Jarvis, E.D., 2004. Songbirds and the Revised Avian Brain
929 Nomenclature. *Ann N Y Acad Sci* 1016, 77–108. <https://doi.org/10.1196/annals.1298.013>

930 Roberts, T.F., Gobes, S.M.H., Murugan, M., Ólveczky, B.P., Mooney, R., 2012. Motor circuits are
931 required to encode a sensory model for imitative learning. *Nat Neurosci* 15, 1454–1459.
932 <https://doi.org/10.1038/nn.3206>

933 Robinson, C.W., Sloutsky, V.M., 2004. Auditory Dominance and Its Change in the Course of
934 Development. *Child Development* 75, 1387–1401. <https://doi.org/10.1111/j.1467-8624.2004.00747.x>

935

936 Ronald, K.L., Fernández-Juricic, E., Lucas, J.R., 2018. Mate choice in the eye and ear of the beholder?
937 Female multimodal sensory configuration influences her preferences. *Proc R Soc Lond B Biol*
938 *Sci* 285, 20180713. <https://doi.org/10.1098/rspb.2018.0713>

939 Ronald, K.L., Zeng, R., White, D.J., Fernández-Juricic, E., Lucas, J.R., 2017. What makes a
940 multimodal signal attractive? A preference function approach. *Behav Ecol* 28, 677–687.
941 <https://doi.org/10.1093/beheco/axx015>

942 Ross, L.A., Saint-Amour, D., Leavitt, V.M., Javitt, D.C., Foxe, J.J., 2007. Do you see what I am
943 saying? Exploring visual enhancement of speech comprehension in noisy environments. *Cereb*
944 *Cortex* 17, 1147–1153. <https://doi.org/10.1093/cercor/bhl024>

945 Seki, Y., Okanoya, K., 2006. Effects of visual stimulation on the auditory responses of the HVC song
946 control nucleus in anesthetized Bengalese Finches. *Ornithol. Sci.* 5, 39–46.
947 <https://doi.org/10.2326/osj.5.39>

948 Shams, L., Seitz, A.R., 2008. Benefits of multisensory learning. *Trends in Cognitive Sciences* 12,
949 411–417. <https://doi.org/10.1016/j.tics.2008.07.006>

950 Shank, S.S., Margoliash, D., 2009. Sleep and sensorimotor integration during early vocal learning in a
951 songbird. *Nature* 458, 73–77. <https://doi.org/10.1038/nature07615>

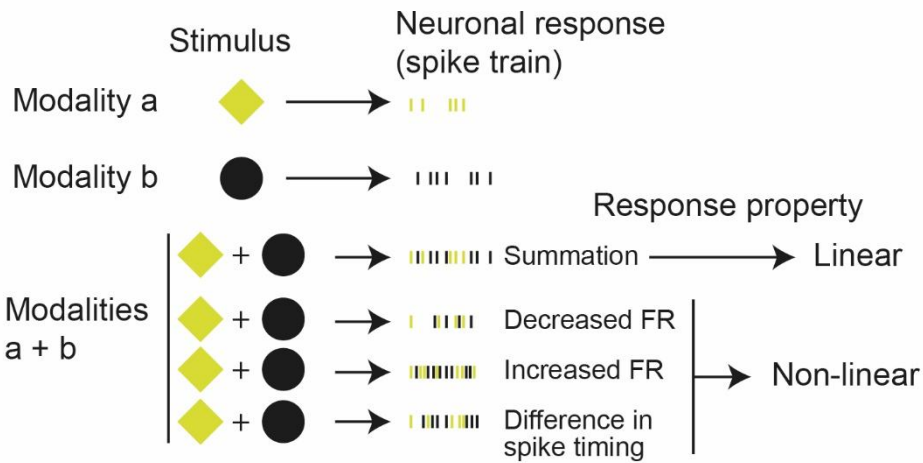
952 Singheiser, M., Plachta, D.T.T., Brill, S., Bremen, P., van der Willigen, R.F., Wagner, H., 2010.
953 Target-approaching behavior of barn owls (*Tyto alba*): influence of sound frequency. *J Comp*
954 *Physiol A* 196, 227–240. <https://doi.org/10.1007/s00359-010-0508-6>

955 Slutsky, D.A., Recanzone, G.H., 2001. Temporal and spatial dependency of the ventriloquism effect.
956 *Neurorep* 12, 7–10. <https://doi.org/10.1097/00001756-200101220-00009>

- 957 Smith, V.A., King, A.P., West, M.J., 2000. A role of her own: female cowbirds, *Molothrus ater*,
958 influence the development and outcome of song learning. *Anim Behav* 60, 599–609.
959 <https://doi.org/10.1006/anbe.2000.1531>
- 960 Stacho, M., Herold, C., Rook, N., Wagner, H., Axer, M., Amunts, K., Güntürkün, O., 2020. A cortex-
961 like canonical circuit in the avian forebrain. *Science* 369.
962 <https://doi.org/10.1126/science.abc5534>
- 963 Stein, B.E., Meredith, M.A., Huneycutt, W.S., McDade, L., 1989. Behavioral Indices of Multisensory
964 Integration: Orientation to Visual Cues is Affected by Auditory Stimuli. *J Cogn Neurosci* 1,
965 12–24. <https://doi.org/10.1162/jocn.1989.1.1.12>
- 966 Stein, B.E., Stanford, T.R., 2008a. Multisensory integration: current issues from the perspective of the
967 single neuron. *Nat Rev Neurosci* 9, 255–266. <https://doi.org/10.1038/nrn2331>
- 968 Stein, B.E., Stanford, T.R., 2008b. Multisensory integration: current issues from the perspective of the
969 single neuron. *Nat Rev Neurosci* 9, 255–266. <https://doi.org/10.1038/nrn2331>
- 970 Stein, B.E., Stanford, T.R., Rowland, B.A., 2020. Multisensory Integration and the Society for
971 Neuroscience: Then and Now. *J. Neurosci.* 40, 3–11.
972 <https://doi.org/10.1523/JNEUROSCI.0737-19.2019>
- 973 Sterbing, S.J., Hartung, K., Hoffmann, K.-P., 2002. Representation of sound source direction in the
974 superior colliculus of the guinea pig in a virtual auditory environment. *Exp Brain Res* 142,
975 570–577. <https://doi.org/10.1007/s00221-001-0963-y>
- 976 Stevenson, R.A., Ghose, D., Fister, J.K., Sarko, D.K., Altieri, N.A., Nidiffer, A.R., Kurela, L.R.,
977 Siemann, J.K., James, T.W., Wallace, M.T., 2014. Identifying and Quantifying Multisensory
978 Integration: A Tutorial Review. *Brain Topogr* 27, 707–730. <https://doi.org/10.1007/s10548-014-0365-7>
- 979
- 980 Sumbly, W.H., Pollack, I., 1954. Visual contribution to speech intelligibility in noise. *The Journal of*
981 *the Acoustical Society of America* 26, 212–215. <https://doi.org/10.1121/1.1907309>
- 982 Suthers, R.A., Goller, F., Wild, J.M., 2002. Somatosensory feedback modulates the respiratory motor
983 program of crystallized birdsong. *Proc Natl Acad Sci USA* 99, 5680–5685.
984 <https://doi.org/10.1073/pnas.042103199>
- 985 Talsma, D., Senkowski, D., Soto-Faraco, S., Woldorff, M.G., 2010. The multifaceted interplay
986 between attention and multisensory integration. *Trends in Cognitive Sciences* 14, 400–410.
987 <https://doi.org/10.1016/j.tics.2010.06.008>
- 988 Todt, D., Hultsch, H., Heike, D., 1979. Conditions affecting song acquisition in nightingales (*Luscinia*
989 *megarhynchos* L.). *Zeitschrift für Tierpsychologie* 51, 23–35. <https://doi.org/10.1111/j.1439-0310.1979.tb00668.x>
- 990
- 991 Town, S.M., McCabe, B.J., 2011. Neuronal Plasticity and Multisensory Integration in Filial
992 Imprinting. *PLOS ONE* 6, e17777. <https://doi.org/10.1371/journal.pone.0017777>
- 993 Tremblay, S., Shiller, D.M., Ostry, D.J., 2003. Somatosensory basis of speech production. *Nature* 423,
994 866–869. <https://doi.org/10.1038/nature01710>
- 995 Tumer, E.C., Brainard, M.S., 2007. Performance variability enables adaptive plasticity of
996 “crystallized” adult birdsong. *Nature* 450, 1240–1244. <https://doi.org/10.1038/nature06390>
- 997 Ullrich, R., Norton, P., Scharff, C., 2016. Waltzing *Taeniopygia*: integration of courtship song and
998 dance in the domesticated Australian zebra finch. *Animal Behaviour* 112, 285–300.
999 <https://doi.org/10.1016/j.anbehav.2015.11.012>
- 1000 van Atteveldt, N., Murray, M.M., Thut, G., Schroeder, C.E., 2014. Multisensory integration: flexible
1001 use of general operations. *Neuron* 81, 1240–1253.
1002 <https://doi.org/10.1016/j.neuron.2014.02.044>
- 1003 Van Kampen, H.S., Bolhuis, J.J., 1993. Interaction between auditory and visual learning during filial
1004 imprinting. *Anim Behav* 45, 623–625. <https://doi.org/10.1006/anbe.1993.1074>
- 1005 Van Ruijssevelt, L., Chen, Y., von Eugen, K., Hamaide, J., De Groof, G., Verhoye, M., Güntürkün,
1006 O., Woolley, S.C., Van der Linden, A., 2018. fMRI Reveals a Novel Region for Evaluating
1007 Acoustic Information for Mate Choice in a Female Songbird. *Current Biology*.
1008 <https://doi.org/10.1016/j.cub.2018.01.048>
- 1009 Varkevisser, J.M., Mendoza, E., Simon, R., Manet, M., Halfwerk, W., Scharff, C., Riebel, K., 2022.
1010 Multimodality during live tutoring is relevant for vocal learning in zebra finches. *Anim Behav*
1011 187, 263–280. <https://doi.org/10.1016/j.anbehav.2022.03.013>

- 1012 Veit, L., Hartmann, K., Nieder, A., 2017. Spatially Tuned Neurons in Corvid Nidopallium
1013 Caudolaterale Signal Target Position During Visual Search. *Cereb Cortex* 27, 1103–1112.
1014 <https://doi.org/10.1093/cercor/bhv299>
- 1015 Versace, E., Spierings, M.J., Caffini, M., ten Cate, C., Vallortigara, G., 2017. Spontaneous
1016 generalization of abstract multimodal patterns in young domestic chicks. *Anim Cogn* 20, 521–
1017 529. <https://doi.org/10.1007/s10071-017-1079-5>
- 1018 Wagner, H., Kettler, L., Orłowski, J., Tellers, P., 2013. Neuroethology of prey capture in the barn owl
1019 (*Tyto alba* L.). *Journal of Physiology-Paris, Neuroethology* 107, 51–61.
1020 <https://doi.org/10.1016/j.jphysparis.2012.03.004>
- 1021 Wagner, H., Takahashi, T., Konishi, M., 1987. Representation of interaural time difference in the
1022 central nucleus of the barn owl's inferior colliculus. *J Neurosci* 7, 3105–3116.
- 1023 Wallace, M.T., Meredith, M.A., Stein, B.E., 1998. Multisensory Integration in the Superior Colliculus
1024 of the Alert Cat. *Journal of Neurophysiology* 80, 1006–1010.
1025 <https://doi.org/10.1152/jn.1998.80.2.1006>
- 1026 Wallace, M.T., Stein, B.E., 1996. Sensory organization of the superior colliculus in cat and monkey.
1027 *Prog Brain Res* 112, 301–311. [https://doi.org/10.1016/s0079-6123\(08\)63337-3](https://doi.org/10.1016/s0079-6123(08)63337-3)
- 1028 Wallace, M.T., Wilkinson, L.K., Stein, B.E., 1996. Representation and integration of multiple sensory
1029 inputs in primate superior colliculus. *J Neurophysiol* 76, 1246–1266.
1030 <https://doi.org/10.1152/jn.1996.76.2.1246>
- 1031 Warren, T.L., Tumer, E.C., Charlesworth, J.D., Brainard, M.S., 2011. Mechanisms and time course of
1032 vocal learning and consolidation in the adult songbird. *J Neurophysiol* 106, 1806–1821.
1033 <https://doi.org/10.1152/jn.00311.2011>
- 1034 West, M.J., King, A.P., 1988. Female visual displays affect the development of male song in the
1035 cowbird. *Nature* 334, 244–246. <https://doi.org/10.1038/334244a0>
- 1036 Whitchurch, E.A., Takahashi, T.T., 2006. Combined auditory and visual stimuli facilitate head
1037 saccades in the barn owl (*Tyto alba*). *J Neurophysiol* 96, 730–745.
1038 <https://doi.org/10.1152/jn.00072.2006>
- 1039 Wild, J.M., 1994. Visual and somatosensory inputs to the avian song system via nucleus uvaeformis
1040 (Uva) and a comparison with the projections of a similar thalamic nucleus in a nonsongbird,
1041 *Columbia livia*. *J. Comp. Neurol.* 349, 512–535. <https://doi.org/10.1002/cne.903490403>
- 1042 Wild, J.M., Botelho, J.F., 2015. Involvement of the avian song system in reproductive behaviour.
1043 *Biology Letters* 11, 20150773. <https://doi.org/10.1098/rsbl.2015.0773>
- 1044 Wild, J.M., Gaede, A.H., 2016. Second tectofugal pathway in a songbird (*Taeniopygia guttata*)
1045 revisited: Tectal and lateral pontine projections to the posterior thalamus, thence to the
1046 intermediate nidopallium. *J. Comp. Neurol.* 524, 963–985. <https://doi.org/10.1002/cne.23886>
- 1047 Wild, J.M., Williams, M.N., 1999. Rostral wulst of passerine birds: II. Intratelencephalic projections
1048 to nuclei associated with the auditory and song systems. *J Comp Neurol* 413, 520–534.
1049 [https://doi.org/10.1002/\(SICI\)1096-9861\(19991101\)413:4<520::AID-CNE3>3.0.CO;2-B](https://doi.org/10.1002/(SICI)1096-9861(19991101)413:4<520::AID-CNE3>3.0.CO;2-B)
- 1050 Winkowski, D.E., Knudsen, E.I., 2006. Top-down gain control of the auditory space map by gaze
1051 control circuitry in the barn owl. *Nature* 439, 336–339. <https://doi.org/10.1038/nature04411>
- 1052 Woolley, S.C., Doupe, A.J., 2008. Social context-induced song variation affects female behavior and
1053 gene expression. *PLoS Biol.* 6, e62. <https://doi.org/10.1371/journal.pbio.0060062>
- 1054 Zahar, Y., Reches, A., Gutfreund, Y., 2009. Multisensory enhancement in the optic tectum of the barn
1055 owl: spike count and spike timing. *J Neurophysiol* 101, 2380–2394.
1056 <https://doi.org/10.1152/jn.91193.2008>
- 1057 Zai, A.T., Cavé-Lopez, S., Rolland, M., Giret, N., Hahnloser, R.H.R., 2020. Sensory substitution
1058 reveals a manipulation bias. *Nat Commun* 11, 5940. <https://doi.org/10.1038/s41467-020-19686-w>
- 1059
1060

1062 **Figure captions**

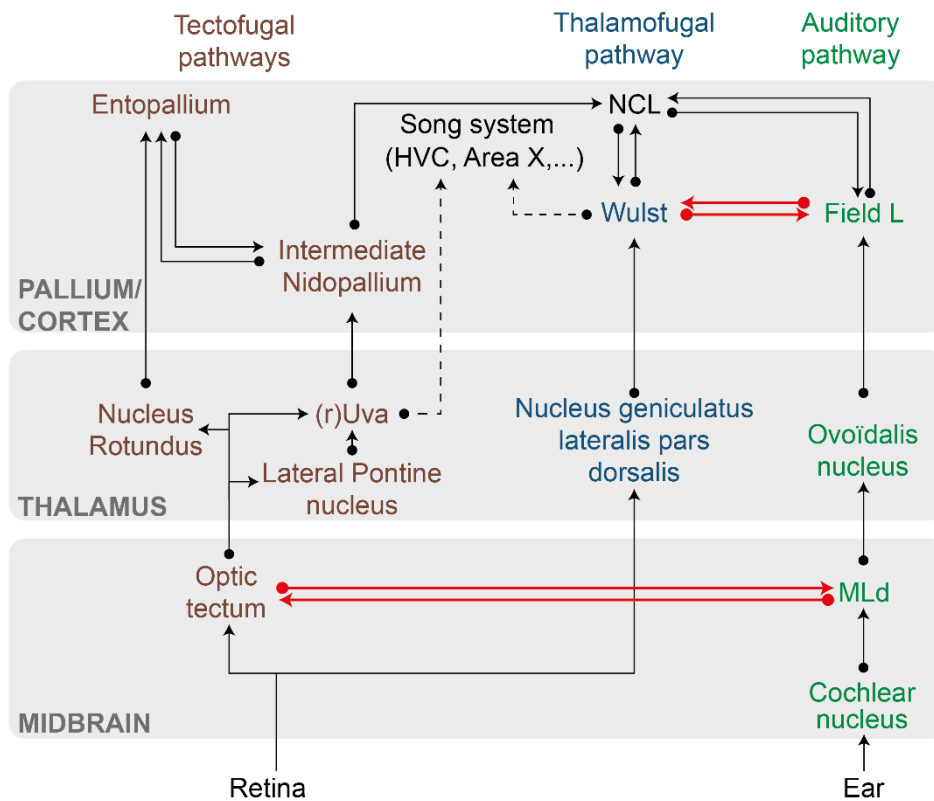


1063

1064 **Figure 1: Expected neuronal responses to unisensory and multisensory stimuli**

1065 In response to a multisensory stimulus (modalities a – green diamond - and b – black circle), neuronal
 1066 responses may result from linear properties (summation of the firing rate, FR) or non-linear properties
 1067 (FR decrease, FR increase, and/or different temporal spike patterns).

1068



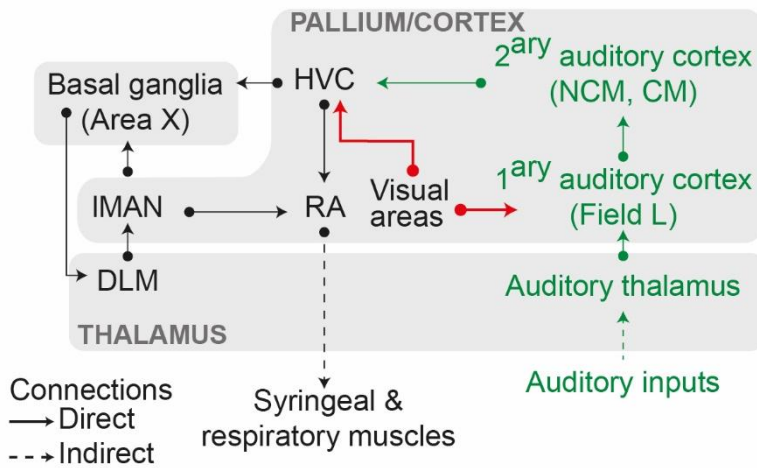
1069

1070 **Figure 2: Interconnections between visual and auditory pathways in the bird brain.**

1071 Birds have one auditory pathway (green) and two main visual pathways: the tectofugal (left, brown)
 1072 and the thalamofugal (middle, blue) pathways. Interconnections between visual and auditory areas are
 1073 reported in the midbrain (optic tectum – MLd, red arrows). In the pallium, direct connections have
 1074 been identified between visual and auditory centres (Wulst – Field L, red arrows). The Wulst is
 1075 considered as analogous to the mammalian primary visual cortex (V1). Visual and auditory
 1076 information converge in the NCL (nidopallium caudolaterale, considered as analogous to the
 1077 mammalian prefrontal cortex). In songbirds, there are indirect projections to the song system (dashed
 1078 line) originating from visual relay (auditory areas projections to the song system are not shown). MLd:
 1079 nucleus mesencephalic lateralis, pars dorsalis (indirect connections between the cochlear nucleus and
 1080 MLd are not shown); (r)UVA: (rostral) nucleus Uvaeformis. For more details on the Wulst and
 1081 (r)UVA circuits, see (Stacho et al., 2020; Wild and Gaede, 2016, respectively).

1082

1083

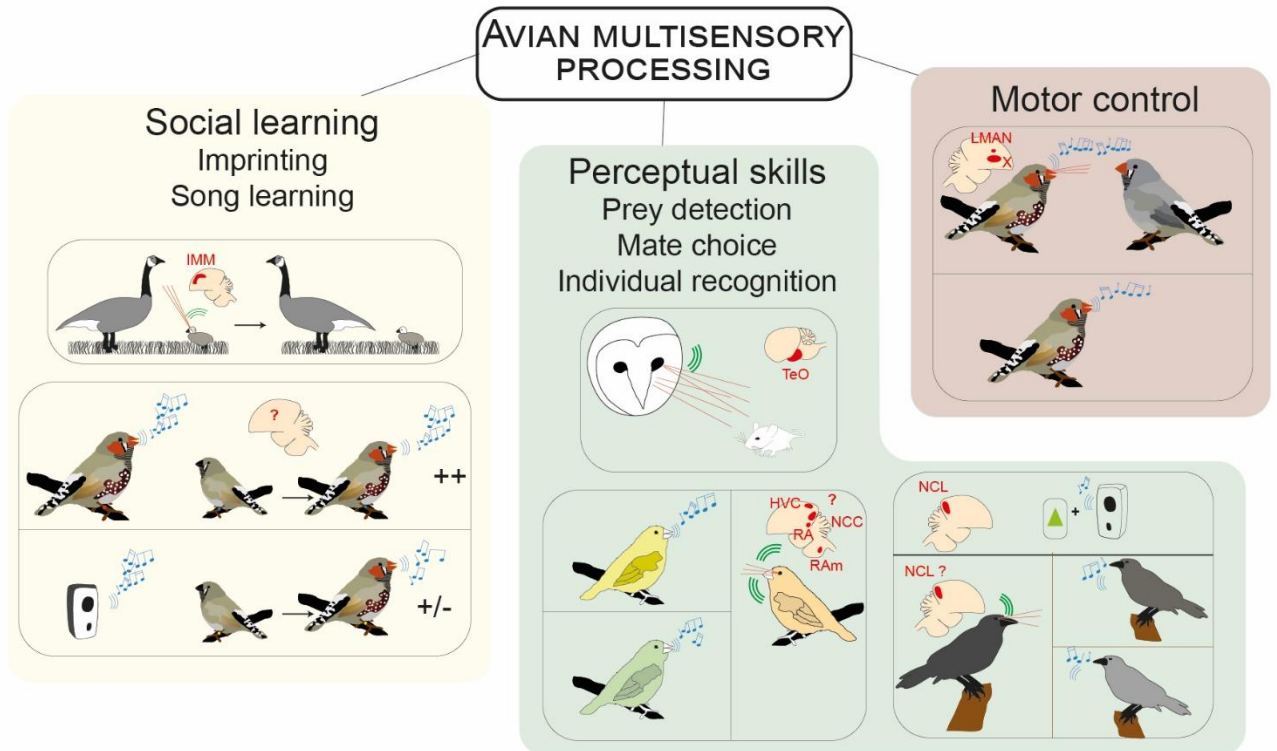


1084

1085 **Figure 3: The auditory pathway and the songbird song system**

1086 The song system includes the song motor pathway (HVC: proper name; RA: Robust nucleus of the
1087 arcopallium, analogous to the human laryngeal motor cortex) and the anterior forebrain pathway
1088 (IMAN: lateral magnocellular nucleus of the anterior nidopallium; Area X, analogous to mammalian
1089 basal ganglia; DLM: dorsolateral nucleus of the anterior thalamus). It receives inputs from the
1090 auditory pathway (green), visual areas (red line, see Figure 1 for more details) and the ventral
1091 tegmental area (not shown). CM: Caudal mesopallium; NCM: Nucleus of the caudal mesopallium.

1092



1093

1094 **Figure 4: Knowns and unknowns in the field of avian multisensory processing**

1095 Avian multisensory processing has been mostly studied in the context of social learning, perception,
 1096 and motor control. In precocial birds (left, top), filial imprinting highlighted the involvement of the
 1097 IMM, a multisensory region that receives afferents from all sensory systems and projects to motor
 1098 areas. In songbirds (left, bottom), it remains to identify which brain network yield young songbirds to
 1099 learn better copies of their tutor songs when they can directly socially interact. In barn owls (middle,
 1100 top), the multisensory based detection of preys requires the TeO, a visual area. In songbirds, female
 1101 preferences for certain song types involve song nuclei (HVC, RA) and nuclei outside of the song
 1102 system (NCC, RAm). In corvids (right, bottom), songbird species used to assess their higher cognitive
 1103 abilities, learning of arbitrary audio-visual associations recruits neurons of the NCL, avian analogue of
 1104 the mammalian prefrontal cortex, which might also be at play for multimodal recognition of
 1105 conspecifics. In male songbirds (right, top), the cortico-basal ganglia network (LMAN and Area X) is
 1106 involved in the socially-dependent modulation of the song motor control. See main text for references.

1107

1108 **Box 1: What is multisensory integration at the neuronal level?**

1109 Multisensory integration generally refers to the set of processes by which redundant and/or
1110 complementary information arriving from different sensory modalities (e.g., vision, audition, touch)
1111 are combined to yield an enriched perceptual experience of multimodal events. This form of
1112 multisensory processing reflects what is commonly reported as cue integration. Yet, multisensory
1113 processing includes other forms of integration such as sensory selection, multisensory based object
1114 recognition and segregation of cross-modal stimuli (Meijer et al., 2019). Sensory selection reflects the
1115 ability to neglect the information from one modality. Multisensory based object recognition involves a
1116 transfer of information from one modality to another during a discrimination task: animals are trained
1117 to identify an object based on a single modality (e.g. touch) and are then tested to identify the same
1118 object using another modality (e.g. sight). Modality segregation occurs when information from one
1119 modality overshadows information from another modality.

1120 While multisensory cue integration relies mostly on bottom-up neuronal processes with a feedforward
1121 transmission from low-level to associative areas, learning and cognitive skills might also modulate
1122 multisensory integration through top-down modulations exerted by higher-level areas. In the present
1123 review, multisensory processes refer not only to multisensory cue integration but also to the whole set
1124 of processes by which information of one modality influences or interacts with the processing of the
1125 specific information of another modality.

1126 Multisensory processes and integration have been quantified using a wide range of experimental
1127 approaches (Stevenson et al., 2014). At the neuronal level, the most commonly implemented method
1128 focuses on single- or multiunit responses, driven by multimodal signals consisting of two or more
1129 modality-specific components presented in a spatiotemporally congruent manner (for more details and
1130 metrics used, see Stevenson et al., 2014). These neuronal responses are often evaluated using spike
1131 counts (sometimes in combination with a spike train analysis) and are compared to responses evoked
1132 by a modality-specific component presented in a temporally and/or spatially incongruent way or to
1133 components presented separately. This comparison presumes a convergence of inputs from multiple
1134 sensory sources onto individual neurons in the brain region of interest. Neuronal responses to

1135 multimodal signals can rely on linear response properties, in which case the responses approximatively
1136 correspond to the sum of all responses to unimodal signals (Figure 1). When these responses cannot be
1137 explained by responses to unimodal signals because they are higher or lower than expected based on
1138 their summation, then this non-linearity is commonly considered as reflecting a multisensory
1139 integration. Using spike train analyses, one typically estimates of the mutual information between
1140 unimodal or multimodal stimuli and uses the temporal spike patterns to examine whether neurons
1141 transmit more information from multisensory stimuli than from a unisensory stimulus (Bizley et al.,
1142 2007). Yet, the benefit provided by multisensory integration at the behavioural level may be due to
1143 linear properties (Angelaki et al., 2009). Also, certain computational models suggest that both linear
1144 and non-linear operations could be achieved by multisensory neurons (Morgan et al., 2008; Ohshiro et
1145 al., 2011).

1146 Neurons exhibiting multisensory integration have been studied in the superior colliculus, a midbrain
1147 structure that contains multisensory convergence zones (Stein and Stanford, 2008a; Wallace et al.,
1148 1998). These neurons guide orienting behavior towards salient stimuli, a form of cue integration that
1149 relies on bottom-up processes. Experimental evidences also suggest that these neurons can drive
1150 attention through an interplay between bottom-up and top-down mechanisms (Talsma et al., 2010). In
1151 this brain structure, responses of multisensory neurons to bimodal congruent stimuli consisting of
1152 unimodal stimuli that are aligned in space and temporally synchronized are typically stronger than
1153 responses to the unimodal components. Interestingly, bimodal incongruent stimuli that are misaligned
1154 in space do not induce this enhancement in responses when compared with the unimodal responses. A
1155 depression was also observed in some responses evoked by bimodal incongruent stimuli. This
1156 differential effect suggests that congruent bimodal stimuli arise from the same object while
1157 incongruent information arise from competing objects. The multisensory integration could thus help
1158 bind the different unisensory components into a single unified object.

1159 Enhanced responses to bimodal stimuli that are suitable for target detection have been described in
1160 other regions than the superior colliculus, including primary sensory cortices (e.g. Bizley et al., 2007;
1161 Kayser et al., 2008; Knöpfel et al., 2019). However, combining stimuli from two different modalities

1162 can manifest differently in different brain regions, including a suppression of responses in primary
1163 sensory cortices (Angelaki et al., 2009; Dehner et al., 2004; Garner and Keller, 2022; Ibrahim et al.,
1164 2016; Iurilli et al., 2012; Lohse et al., 2021b). This effect may allow a sharpening of functional
1165 properties, such as orientation selectivity in V1 or reducing the impact of information from other
1166 modalities, including distracting or self-generated signals, to increase the salience of information from
1167 one specific modality (Ibrahim et al., 2016; Iurilli et al., 2012; Lohse et al., 2021b). It has also been
1168 observed in the context of task learning (Garner and Keller, 2022).

1169 Another methodological approach, measuring local field potentials, is also based on
1170 electrophysiological recordings in the brain region of interest. This has provided complementary
1171 information to the information obtained through single- and multiunit recordings, from subthreshold
1172 mechanisms at the synaptic level and provided insight into how neural populations may encode
1173 multisensory signals. A resetting of ongoing oscillations has been proposed as the possible mechanism
1174 used to amplify sensory signals by optimizing neuronal excitability (Lakatos et al., 2007; van
1175 Atteveldt et al., 2014).