

ORIGINAL ARTICLE OPEN ACCESS

Sound Guides Object Size Choices in African Penguins Through Pitch–Size Association

Francesca Terranova¹  | Francesca D'Orazio^{1,2}  | Cristina Pilenga³  | Luigi Baciadonna¹  | David Reby^{4,5}  | Livio Favaro¹ 

¹Department of Life Sciences and Systems Biology, University of Turin, Turin, Italy | ²Department of Human Neurosciences, Faculty of Medicine and Dentistry, Sapienza University of Rome, Rome, Italy | ³Zoomarine Openature Rome, Torvaianica, Italy | ⁴ENES Bioacoustics Research Laboratory, Lyon Neuroscience Research Centre (CRNL), Jean Monnet University, Saint-Etienne, France | ⁵Institut universitaire de France, Paris, France

Correspondence: Francesca Terranova (francesca.terranova@unito.it)

Received: 23 December 2025 | **Revised:** 15 April 2026 | **Accepted:** 29 April 2026

Keywords: bioacoustics | bird cognition | multimodal association | *Spheniscus demersus*

ABSTRACT

The brain's ability to associate information from different sensory modalities, known as cross-modal integration, is crucial for forming a coherent perception of objects and events. When such associations follow systematic and predictable patterns, they are referred to as cross-modal correspondences. Humans, for example, tend to associate large objects with low-pitched sounds and small objects with high-pitched sounds, a phenomenon known as pitch–size cross-modal correspondence. This phenomenon has been documented in several vertebrate species, yet direct experimental evidence in avian species remains comparatively limited. Here, we tested African penguins using a preferential choice paradigm in which individuals chose between two cubes of different sizes after hearing either a high- or low-pitched sound. Following passive exposure to congruent pitch–size pairings, penguins chose larger objects following low-pitched sounds and smaller objects following high-pitched sounds, indicating that auditory pitch guided their visual size choices. Given that individuals were pre-exposed to a limited set of pitch–size pairings, this effect may reflect either a preexisting bias consistent with cross-modal correspondences or the generalization of a learned association. Nevertheless, these findings demonstrate that penguins can integrate auditory and visual information to guide their behavior and generalize a systematic pitch–size relationship across sensory dimensions.

1 | Introduction

Animals constantly receive information from multiple sensory channels simultaneously. The ability to integrate this information is essential for constructing a unified and coherent perception of the surrounding environment [1–3]. This cross-modal integration reduces perceptual uncertainty and supports adaptive decision-making in complex natural settings [4, 5]. When specific features from different sensory modalities are systematically associated in

a predictable manner, these links are termed cross-modal correspondences [6]. These correspondences typically occur between polarized dimensions, where extreme values on one continuum (e.g., high or low) are matched with corresponding extremes on another.

In humans, one of the most robust and well-studied examples is the pitch–size correspondence: people reliably associate low-pitched sounds with large objects and high-pitched sounds

Francesca Terranova and Francesca D'Orazio are co-first authors.

David Reby and Livio Favaro are co-senior authors.

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2026 The Author(s). *Annals of the New York Academy of Sciences* published by Wiley Periodicals LLC on behalf of The New York Academy of Sciences.

with small objects [7–10]. This association is part of a broader network of cross-modal correspondences that includes associations between pitch and visual brightness, spatial height, and luminance, as well as links between other sensory qualities such as texture and color, or shape and taste [11–19]. These correspondences influence a wide range of perceptual and cognitive processes, from speeded classification tasks to everyday object recognition and aesthetic judgments.

Importantly, cross-modal correspondences are not restricted to humans. Similar patterns have been observed across a growing number of vertebrate species. Rhesus monkeys associate rising frequency sounds with looming visual stimuli [20], while domestic chicks demonstrate spontaneous associations between luminance and spatial position [21], visual and tactile cues [22], and sound–shape correspondences analogous to the human Bouba–Kiki effect [23]. Reptiles also exhibit cross-modal abilities: Hermann’s tortoises, for instance, show pitch–luminance [24] and pitch–size associations [4]. Pitch–size correspondences specifically have been experimentally demonstrated in chimpanzees [25], domestic dogs [26], and tortoises. In birds, indirect evidence suggests that vocal pitch is used to assess or signal body size: willow warbler males appear to use song pitch to assess rival body size [27], and purple-crowned fairy-wren males produce advertising songs whose low-frequency components vary with body size, consistent with body-size signaling [28]. However, direct experimental evidence for pitch–size associations in birds from controlled choice paradigms remains comparatively limited.

A key question concerns how these correspondences arise. From a theoretical perspective, they may reflect innate perceptual biases or emerge through multisensory statistical learning—the internalization of statistical regularities present in the natural environment. Through repeated exposure to consistent sensory relationships, individuals build expectations about how cues covary, such that responses extend to novel situations, thereby reducing uncertainty. Over time, these learned associations might become internalized as perceptual biases, as suggested by Bayesian coupling-prior frameworks [29], which posit that the brain encodes environmental co-occurrence statistics to bias multisensory perception. In the physical world, these statistical regularities are often grounded in biomechanical and environmental constraints: larger objects tend to produce lower-frequency sounds, and larger animals typically vocalize at lower fundamental frequencies due to larger vibrating tissues (e.g., vocal folds or syringeal membranes) and longer vocal tracts with lower resonant frequencies [30, 31]. Whether these regularities are internalized through individual development or reflect evolved perceptual biases shaped over evolutionary time remains an open question.

African penguins (*Spheniscus demersus*) provide a particularly valuable model for investigating the evolutionary origins of such mechanisms. As members of the order Sphenisciformes, they diverged from other avian lineages approximately 65 million years ago [32], offering a key comparative perspective on multisensory processing in an early-diverging bird group. Within this species, acoustic cues to body size are biologically relevant. The fundamental frequency of begging calls in chicks is negatively correlated with body mass and age [33]. In adults of African penguins’ closely related species, Humboldt penguins (*Spheniscus*

humboldti) and Magellanic penguins (*Spheniscus magellanicus*), both vocalization duration and fundamental frequency reliably encode body size and mass [34]. Furthermore, African penguins are capable of perceiving variations in fundamental frequency and formant frequencies within species-specific vocalizations [35] and successfully integrate auditory and visual information to recognize individual conspecifics [36].

Despite growing evidence that African penguins rely on acoustic parameters to infer information about signalers and integrate these cues cross-modally, it remains unknown whether auditory pitch systematically shapes visual size selection. Here, we investigated whether auditory pitch influences visual size selection in a controlled two-choice experimental setting. We predicted that if African penguins possess a pitch–size association, they should preferentially approach larger objects when hearing low-pitched sounds and smaller objects when hearing high-pitched sounds.

2 | Methods

African penguins were housed at Zoomarine (Italy). The colony comprised 21 African penguins, ranging from 6 months to 20 years old (Table S1). Each penguin was uniquely identifiable by a colored plastic ring attached to one wing, and the distinctive dot pattern on their ventral side further facilitated individual recognition [37]. The colony was housed in an outdoor exhibit of 70 m², with sand and pebbles and a 228-m³ saltwater pool. Zookeepers hand-fed the penguins three times daily.

2.1 | Audio and Visual Stimuli

We prepared controlled audio–visual stimuli for two experimental phases: an exposure phase and a subsequent testing phase.

For the exposure phase, five black cardboard cubes of different sizes were constructed (Figure S1), following evidence that birds can discriminate among cubic objects [38]. Black was chosen to maximize visual salience against the background [39]. Each cube had one removable side allowing the inside placement of a Bose SoundLink Mini II speaker connected to a Redmi Note 10 Lite smartphone. Object size was manipulated by varying cube volume. Starting from a medium cube (T0; 21 cm per side), four additional cubes were created by increasing or decreasing the volume by 5% or 10%, considering that birds encode size logarithmically [40]. Acoustic stimuli were pure tones designed to fall within the auditory sensitivity range reported for African penguins [41] and to reflect pitch contrasts comparable to those used in studies of pitch–size correspondences in other vertebrates [26]. Pure tones were generated in Praat v. 6.3.09 [42] with the following features: 2 s duration, 48 kHz sampling frequency, 0.2 Pa amplitude, and 0.1 s fade-in/out. For the medium cube (T0), we used a 430-Hz pure tone; the remaining tones were created by increasing or decreasing frequency by 5% and 10% (Figure S1). We employed pure tones to isolate the fundamental frequency as the relevant auditory cue and to avoid introducing other acoustic features (e.g., amplitude modulation or harmonic structure).

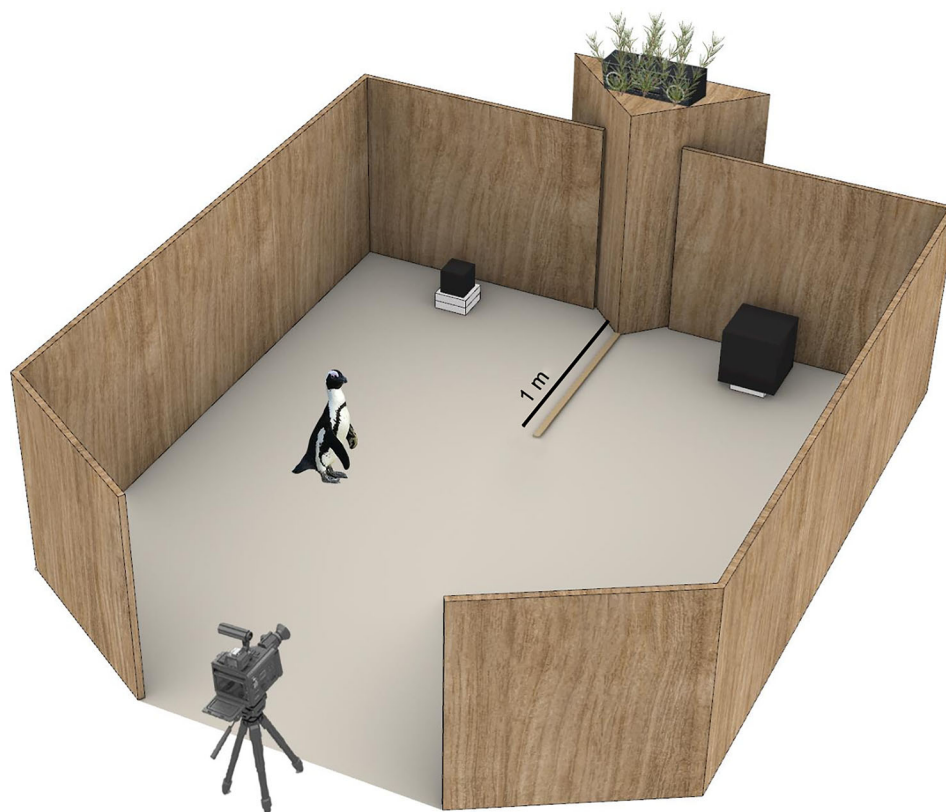


FIGURE 1 | Setup of the arena used in the testing phase. The cubes of different sizes were placed at 170 cm from each other. The penguins were exposed to acoustic stimuli from a hidden speaker positioned elevated between the stimuli. See the Supporting Information (Figure S2) for more details.

For testing, we used two new cube sizes and two novel pitches that were not presented during exposure. The cubes were a large one (28 cm per side) and a small one (13 cm per side; Figure 1), each presented with either a low-pitch (150 Hz) or high-pitch (900 Hz) tone.

All sounds were played at a sound pressure level of 65–70 dB, measured at the start of each session using a Monacor SM-2 sound level meter. We used the C-weighted scale, which emphasizes low-frequency sounds, making it suitable for assessing the acoustic characteristics relevant to the study. Measurements were conducted at a height of 70 cm and a distance of 1 m from the speaker, mimicking the position of the penguins' head during the tests.

2.2 | Exposure Phase

The exposure phase was designed to ensure that behavioral responses during testing reflected attention to the audiovisual stimuli rather than first-time novelty reactions, which are common in neophobic, nondomesticated birds such as penguins [43, 44]. This phase comprised 60 short sessions distributed across 30 nonconsecutive days and was conducted during regular feeding times; the last 24 sessions were held inside the wooden-panel arena later used for testing (Figure 1) to minimize its novelty effects. Importantly, this exposure phase cannot be construed as a training procedure, as penguins were not required to make choices, and all penguins were fed routinely regardless of whether or how they interacted with the stimuli.

During each session, while a keeper hand-fed the penguins, a single cube was placed in front of the group and paired with a pure tone. Cube size and tone frequency varied slightly across days and covaried in a congruent manner, with larger cubes paired with lower frequencies and smaller cubes with higher frequencies. Exposure occurred in a group setting, penguins were free to approach the stimuli without being separated from their social context, and food delivery followed routine feeding regardless of the penguins' behavior (Video S1). Individuals that oriented toward or pecked the cube at least once following sound playback were noted and classified as attentive to the stimuli and were subsequently included in the testing phase.

2.3 | Testing Phase

Testing began immediately after the exposure phase and lasted for 8 days. Of the initial 21 penguins, 19 were selected for the testing phase, with two individuals excluded: one juvenile, which was excluded due to being too young, and one adult, which was absent from the exposure phase because it was incubating the eggs. Each penguin completed four trials in total (no more than one trial per day): two congruent and two incongruent trials, with the target side randomized across trials (Table 1). The order of the four trial types was fully randomized.

Trials took place in a temporary arena built from wooden panels and plywood (Figure 1). The large and small cubes were placed on the ground ~170 cm apart and at matched height. A speaker was positioned centrally between them, elevated and visually con-

TABLE 1 | Audio-visual stimuli combinations. Each penguin was presented with four unique testing conditions—two congruent and two incongruent—created by a combination of two sounds (high pitch = 900 Hz; low pitch = 150 Hz) with two cubes (small size = 13 cm per side; large size = 28 cm per side).

Pitch	Object size	Object position	Condition type
High	Small	Right	Congruent
		Left	
	Large	Right	Incongruent
		Left	
Low	Small	Right	Incongruent
		Left	
	Large	Right	Congruent
		Left	

cealed by vegetation. Trials were conducted in the outdoor area of the colony enclosure, minimizing the possibility of echo and reverberation effects. Penguins were guided into the arena from an adjacent holding area by keepers. Once a penguin entered, the exit panel was closed, and the experimenter left the arena to minimize interference. The penguin was given ~1 min to settle, during which he could move freely within the arena, then the tone for that trial (high or low) was played. From audio onset, the bird's behavior was video recorded for 1 min using a Sony Handycam mounted ~30 cm above ground and oriented toward the arena entrance. A single audio stimulus was presented per trial. After the trial, the bird was released back into the main nesting area. Example trials are shown in the Supporting Information (Videos S2). We presented both cubes simultaneously because pitch-size association in other vertebrates appears to be relative rather than absolute [26, 45]. Thus, the task required the penguin to orient, look, or approach in the presence of two alternative size options under a single pitch context.

2.4 | Behavioral Analyses

Behavioral variables were coded using BORIS v.8.21.8 [46]. Video recordings were analyzed at 0.5x speed with the audio track enabled and the spectrogram was used to precisely identify stimulus onset. Coding began at the onset of the auditory stimulus for each trial, and behaviors were scored according to predefined, objective operational criteria as follows:

- *Gaze duration*: in seconds, how long each penguin observed each cube after tone onset. Gaze direction was assessed based on the orientation of the penguin's beak and scored only if the penguin clearly directed its gaze at one of the two cubes, as African penguins use forward-facing binocular vision when focusing on objects [37, 47]. Instances in which the penguin looked toward the center of the arena or toward the exit (away from the cubes) were not scored. Gaze duration was measured as a continuous variable, for example, "Penguin looked at the large object on the left for 4.5 s" or "Penguin did not gaze (0 s)" (Figure 2a). Additionally, to measure gaze preference, we calculated the preferential gaze index adapted from Albuquerque et al. [48]. The index is calculated as $(C - I)/T$, where C is the total time the penguin gazed at the congruent stimulus,

I is the total time spent gazing at the incongruent stimulus, and T is the total gaze time (congruent gaze + incongruent gaze). Hereafter, "congruent" refers to audio-visual pairs that align with cross-modal association (e.g., a large object with low pitch, a small object with high pitch), while "incongruent" refers to the reverse combinations.

- *First approach*: we recorded which cube, if any, the penguin first approached after tone onset. An approach was defined as the bird crossing into the half of the arena containing that cube (delineated by the central wooden divider). Movements to the center of the arena or near the exit were not scored. Only the first approach was recorded. This behavior was coded as a binary variable: either the penguin approached a specific object (e.g., a small object positioned on the right after a high-pitched sound) or did not approach (Figure 2b). Based on this behavior, we considered two aspects: object size choice indicating whether the penguin approached the large or small cube and selected stimuli combination, which refers to the pairing of the chosen cube with the specific audio condition in each trial (e.g., a large cube paired with a low-pitched sound).

To ensure reliability, 10% of the videos were scored by a second observer. Blind scoring was not feasible, as identifying behavioral onset required access to the audio track, which made the pitch condition visible to scorers. Interrater reliability was strong, with a kappa coefficient of 1 for the first approach, indicating perfect agreement. For the preferential gaze index, the Pearson correlation between raters was also high, with a correlation coefficient of 0.84.

2.5 | Statistical Analysis

Statistical analyses were conducted in RStudio (v. 4.1.3). We analyzed the first approach behavior to test whether object size choice varied with the pitch played. We used a chi-square test with Yates' correction to evaluate whether penguins' choices of cube size (large vs. small) were associated with the audio condition (high vs. low). We then fit a generalized linear mixed model (GLMM; package *glmmTMB* [49]) to explore how the audio condition influenced object size choice. Object size (large or small) was treated as the binary response variable, with

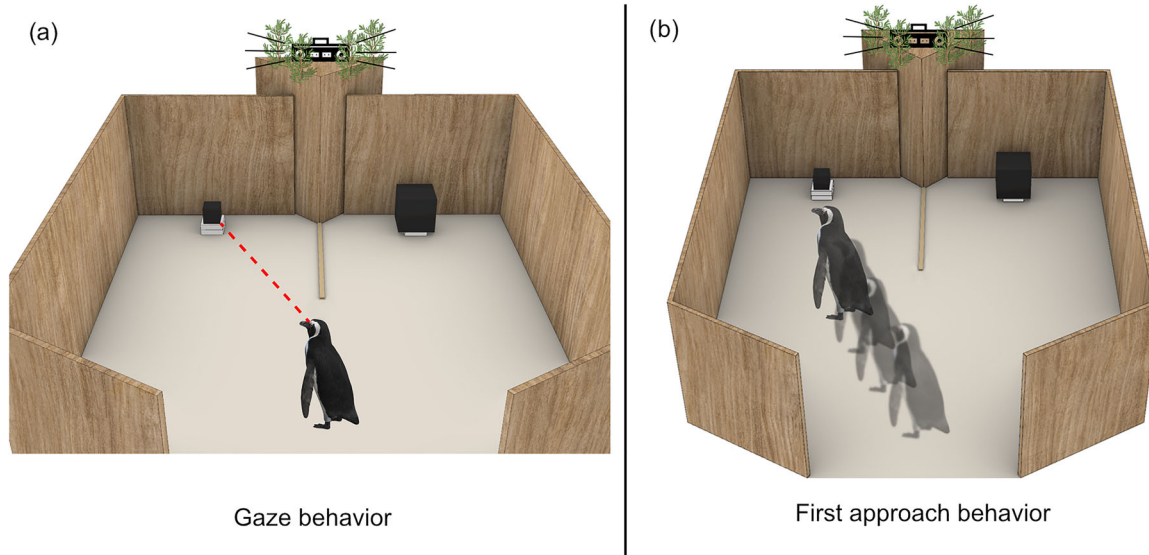


FIGURE 2 | Schematic representation of the two coded behaviors: (a) gaze behavior, showing the penguin's beak orientation toward an object, and (b) first approach behavior, illustrating the penguin's initial movement toward the small object.

audio condition (high or low) and large object position (right or left) as fixed factors. The inclusion of large object position aimed to evaluate whether penguins exhibited an attentional bias toward the large object, potentially due to its visual salience [50]. Individual ID and session were included as random effects. We verified statistical assumptions for using *glmmTMB*, confirming no collinearity among fixed factors. The significance of the full model was evaluated using a chi-square test [51], and a *drop1* analysis was performed to identify the fixed factors influencing the response [52].

To evaluate the penguins' preferential gaze behavior toward objects under different audio conditions, we first aggregated individual subjects' preferential gaze behavior data (Table S2). We then conducted the one-sample *t*-test on the aggregated scores to determine whether the mean preferential gaze index (ranging from -1 to 1) was significantly different from zero (absence of preference). This test assessed if penguins displayed an overall preference for congruent audio-visual pairings (e.g., large objects with low-pitched sound). Additionally, we calculated each subject's mean preferential gaze behavior separately for the low and high audio conditions (Table S3). This aggregation allowed us to obtain a single average preferential gaze score per audio condition for each penguin. Subsequently, one-sample *t*-tests were conducted separately for each audio condition (high-pitched and low-pitched) to investigate the preference for a specific audio condition. All assumptions for the *t*-tests were verified prior to analysis. Normality was assessed using the Shapiro-Wilk test.

3 | Results

We analyzed penguins' first-approach behavior across 72 trials to assess whether their responses varied with the pitch of the auditory cue. Penguins approached one of the two stimuli in 65 trials, while the remaining trials showed no approach. Among the 65 approach responses, penguins directed their first approach

toward the size-pitch combination consistent with the congruent pitch-size pairings (large object with low pitch, or small object with high pitch) in 80% of cases (Figure 3).

The chi-square test confirmed a statistically significant association between auditory pitch and object size choice ($\chi^2 = 21.28$, $p < 0.001$), with high-pitched sounds biasing choices toward smaller objects and low-pitched sounds toward larger objects.

Results from the GLMM analysis corroborated this pattern: approach direction was significantly influenced by the audio condition (GLMM full vs. null: $\chi^2 = 15.29$, $df = 2$, $p < 0.001$). The *drop1* function confirmed that audio condition was a significant predictor of object size choice while the large object position was not significant. The summary of the model is presented in Table 2.

Penguins approached the object whose size was consistent with the pitch-size association. Specifically, the estimated probability of approaching the smaller object was significantly higher with the high-pitched auditory cue (0.87; 95% CI: 0.57–0.97) compared to the low-pitched auditory cue (0.08; 95% CI: 0.01–0.37). Conversely, the estimated probability of choosing the larger object was higher with the low-pitched cue (0.91; 95% CI: 0.63–0.99) than with the high-pitched cue (0.13; 95% CI: 0.03–0.43; Figure S3).

The analysis of gaze behavior showed a similar pattern. The mean preferential gaze index across individuals was $M = 0.12 \pm 0.19$ SD (95% CI: 0.03–0.22, $n = 19$). This value was significantly greater than zero (one-sample *t*-test: $t = 2.74$, $df = 18$, $p = 0.0067$), indicating a general gazing preference for congruent audio-visual pairings (e.g., larger objects with low-pitched sounds and smaller objects with high-pitched sounds) (Figure 4). The one-sample *t*-test for low-pitched sounds indicated a significant positive gaze preference, with a mean preferential gaze index significantly greater than 0 ($t = 2.91$, $df = 16$, $p = 0.0051$). In contrast, the one-sample *t*-test for high-pitched sounds did not reach statistical significance, suggesting no strong preference in this condition ($t = 1.55$, $df = 18$, $p = 0.069$). This indicates that African penguins

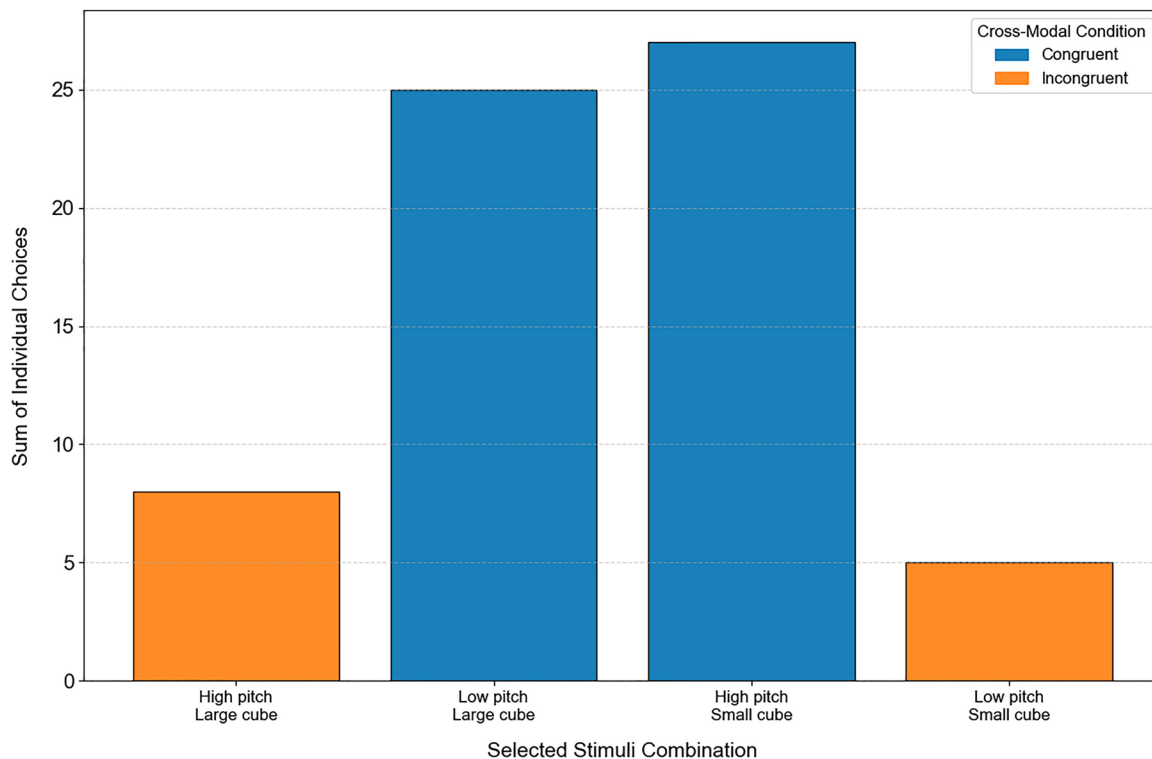


FIGURE 3 | Sum of individual choices of African penguins for different stimuli combinations under congruent and incongruent audio-visual conditions.

TABLE 2 | Summary of the generalized linear mixed model examining the influence of *audio condition* and *large object position* on *object size choice*.

	Estimate	SE	Z-value	p-value
(Intercept)	1.32	0.78	1.69	0.09
Audio condition (low)	−4.33	1.45	−2.99	0.003**
Large object position	1.18	0.82	1.44	0.15

Note: ** indicate statistical significance at $p < 0.01$.

exhibit a significant gaze preference for low-pitched sounds, whereas high-pitched sounds do not elicit a similar response.

4 | Discussion

In this study, we investigated whether African penguins systematically associate auditory pitch with visual object size. Following passive exposure to congruent pitch-size pairings, the penguins were tested with novel combinations of object size and sound pitch. During testing, individuals preferentially approached and gazed at larger objects when hearing low-pitched sounds and smaller objects when hearing high-pitched sounds. This pattern demonstrates that African penguins successfully generalized a pitch-size relationship across sensory modalities and applied it to previously unseen stimuli.

Such dimensional relationships between auditory pitch and visual size are widely documented within the framework of cross-modal correspondences [6]. Our findings also highlight that pitch-size congruent behavior extended to novel stimuli fol-

lowing passive, unsupervised exposure, whereby features in one sensory modality are systematically linked to features in another. Because the exposure phase introduced pitch-size covariation, the present design does not allow us to determine whether the observed pattern reflects a preexisting perceptual correspondence or the acquisition of a learned association during the experiment. Regardless of its developmental origin, pitch-size congruent behavior extended to novel stimuli, demonstrating sophisticated multisensory processing in this species.

Our results extend previous demonstrations that birds can form sound-shape associations [23]. The Bayesian coupling-prior framework [29] provides a principled account of such findings, proposing that pitch-size mappings may emerge from the statistical co-occurrence of lower frequencies with larger objects and larger animals in the natural environment, even without explicit training [53].

For African penguins, such regularities are particularly relevant in social contexts. Across life stages, vocal parameters encode body-size-related information: in chicks, fundamental

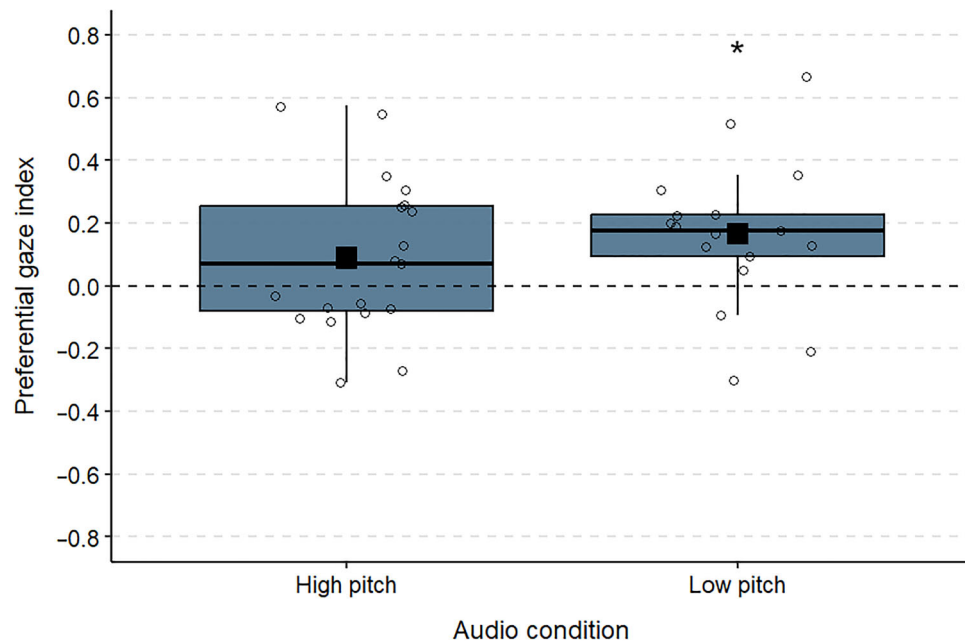


FIGURE 4 | Preferential gaze index of penguins across the two audio conditions. Higher values indicate a stronger preference for congruent audio-visual stimuli combinations. The box plots display the median (horizontal line) and interquartile range (box height) for each condition, with individual points representing penguin averages and black squares indicating the mean. The dashed line at zero represents a neutral gaze index. The asterisk (*) indicates a significant difference from zero (one-sample t -test, $p < 0.05$).

frequency covaries negatively with body mass and age [33], while in adults of African penguins' closely related species, both vocalization duration and fundamental frequency reliably signal body size and mass [34, 54]. African penguins can perceive variations in fundamental frequency and formant frequencies within species-specific vocalizations [35] and integrate auditory and visual cues during individual recognition [36]. Consequently, the ability to extract and align pitch with visual size may support coherent expectations about the physical characteristics of signalers. In densely populated colonies, where visual access to signalers may be limited or delayed, cross-modal consistency could facilitate efficient recognition of partners, offspring, and other conspecifics.

The behavioral responses that we observed were consistently aligned with the expected low-pitch/large-object and high-pitch/small-object pairings. Although approach and gaze data showed the same overall pattern, gaze behavior revealed an asymmetry: penguins exhibited a significant preference for congruent stimuli under low-pitch conditions, but not under high-pitch conditions. This asymmetry cannot be attributed to perceptual limitations, as African penguins are sensitive across the frequency range used in the study [41]. Instead, the stronger effect for low-pitched stimuli may reflect the socioecological salience of lower fundamental frequencies, which signal larger conspecifics and likely carry greater weight in intraspecific interactions such as agonistic encounters or mate choice.

The pitch-size congruent responses observed here extended to novel cube sizes and pitch values not encountered during exposure, indicating that the effect was not limited to memory of specific trained pairings. Whether this reflects learning during the exposure phase, a preexisting predisposition, or lifelong experience with natural correlations between object size and sound pitch in the environment remains unresolved [43, 53, 55].

Even without the experimental exposure phase, lifelong experience of natural correlations between object size and sound pitch could have contributed. Future studies could more directly address the developmental origins of these associations in African penguins by experimentally reversing or violating these associations during exposure, or by testing newly hatched chicks prior to substantial multisensory experience.

Exploring cross-modal associations in non-human animals offers valuable insight into the evolutionary origins of multisensory integration and the early adaptations that may have shaped its development in humans. Broadening this research across taxa will help clarify whether such associations stem from convergent evolution, where similar perceptual traits emerge independently under comparable socioecological pressures, or from a shared ancestral mechanism, given the evolutionary divergence of mammals, reptiles, and birds around 300–320 million years ago [56].

5 | Conclusion

Taken together, our findings provide the first experimental evidence that African penguins systematically associate auditory pitch with visual object size. This ability to align information across sensory modalities highlights a flexible form of multisensory processing that may support efficient perception in complex environments. Whether this pattern reflects preexisting predispositions or emerges through experience remains an open question. Nonetheless, the observed behavior demonstrates that the capacity to form and apply abstract multimodal associations

is present in an early-diverging avian lineage, underscoring the potential importance of these cognitive mechanisms in vertebrate evolution.

Author Contributions

Francesca Terranova: writing – original draft, visualization, validation, methodology, investigation, formal analysis, data curation, conceptualization. Francesca D'Orazio: writing – original draft, methodology, investigation, formal analysis, data curation. Luigi Baciadonna: writing – original draft, validation, methodology, investigation. Cristina Pilenga: writing – review and editing, supervision, resources. David Reby: writing – review and editing, supervision, conceptualization. Livio Favaro: writing – original draft, supervision, methodology, investigation, funding acquisition, conceptualization.

Acknowledgments

The authors are grateful to all the personnel of Zoomarine Italia who helped in the building of stimuli as well as with the interaction with African penguins. In particular, the authors thank all the penguin keepers of Zoomarine: Daniela de la Cadena, Francesca Cenci, Federica Alessandri, Andrea Spinelli, and Andrea Serranti. The authors also thank the pinniped keepers who helped in the initial phase of the study, in particular: Vanja Marini, Patricia Lopez, Mattia La Ragione, and Andrea Velluti. Finally, the authors are grateful to Virginia Menegatti for helping with the creation of Figures 1, 2, and S2. F.T. was supported by a PON PhD scholarship (REACT-EU FSE DM 1061). L.B. was supported by a Marie Skłodowska-Curie Postdoctoral Fellowship (FEAR-BEES-101065873) during writing. D.R. was partially supported by the Centre National de la Recherche Scientifique (CNRS) of the University of Saint-Etienne, the Institut Universitaire de France, and the Labex CeLyA.

Open access publishing facilitated by Università degli Studi di Torino, as part of the Wiley - CRUI-CARE agreement.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The data utilized in this research article are archived on Zenodo (<https://doi.org/10.5281/zenodo.14736533>).

Peer Review

For transparency, the peer review documents associated with this article are available at <https://doi.org/10.1111/nyas.70313>.

References

1. D. J. Lewkowicz and A. A. Ghazanfar, "The Emergence of Multisensory Systems Through Perceptual Narrowing," *Trends in Cognitive Sciences* 13, no. 11 (2009): 470–478, <https://doi.org/10.1016/j.tics.2009.08.004>.
2. C. Parise and C. Spence, "Audiovisual Cross-Modal Correspondences in the General Population," in *The Oxford Handbook of Synesthesia*, ed. J. Simner and E. M. Hubbard (Oxford University Press, 2013), 790–815.
3. C. Spence and O. Deroy, "How Automatic Are Crossmodal Correspondences?," *Consciousness and Cognition* 22, no. 1 (2013): 245–260, <https://doi.org/10.1016/j.concog.2012.12.006>.
4. M. Loconsole, G. Stancher, and E. Versace, "Crossmodal Association Between Visual and Acoustic Cues in a Tortoise (*Testudo hermanni*)," *Biology Letters* 19, no. 7 (2023): 20230265, <https://doi.org/10.1098/rsbl.2023.0265>.
5. V. Ratcliffe, A. Taylor, and D. Reby, "Cross-Modal Correspondences in Non-Human Mammal Communication," *Multisensory Research* 74 (2015), <https://doi.org/10.1163/22134808-00002509>.

6. C. Spence, "Crossmodal Correspondences: A Tutorial Review," *Attention, Perception, & Psychophysics* 73, no. 4 (2011): 971–995, <https://doi.org/10.3758/s13414-010-0073-7>.
7. C. Mondloch and D. Maurer, "Do Small White Balls Squeak? Pitch-Object Correspondences in Young Children," *Cognitive, Affective & Behavioral Neuroscience* 4 (2004): 133–136, <https://doi.org/10.3758/CABN.4.2.133>.
8. A. Gallace and C. Spence, "Multisensory Synthetic Interactions in the Speeded Classification of Visual Size," *Perception & Psychophysics* 68 (2006): 1191–1203, <https://doi.org/10.3758/BF031937200>.
9. I. Fernández-Prieto, J. Navarra, and F. Pons, "How Big Is This Sound? Crossmodal Association Between Pitch and Size in Infants," *Infant Behavior & Development* 38 (2015): 77–81, <https://doi.org/10.1016/j.infbeh.2014.12.008>.
10. C. Parise, "Crossmodal Correspondences: Standing Issues and Experimental Guidelines," *Multisensory Research* 29, no. 1-3 (2016): 7–28, <https://doi.org/10.1163/22134808-00002502>.
11. L. E. Marks, "On Associations of Light and Sound: The Mediation of Brightness, Pitch, and Loudness," *The American Journal of Psychology* 87, no. 1/2 (1974): 173–188, <https://doi.org/10.2307/1422011>.
12. L. E. Marks, "On Cross-Modal Similarity: Auditory-Visual Interactions in Speeded Discrimination," *Journal of Experimental Psychology: Human Perception and Performance* 13, no. 3 (1987): 384–394, <https://doi.org/10.1037/0096-1523.13.3.384>.
13. R. D. Melara and T. P. O'Brien, "Interaction Between Synesthetically Corresponding Dimensions," *Journal of Experimental Psychology: General* 116, no. 4 (1987): 323–336, <https://doi.org/10.1037/0096-3445.116.4.323>.
14. C. V. Parise, K. Knorre, and M. O. Ernst, "Natural Auditory Scene Statistics Shapes Human Spatial Hearing," *Proceedings of the National Academy of Sciences* 111, no. 16 (2014): 6104–6108, <https://doi.org/10.1073/pnas.1322705111>.
15. P. Walker, J. Bremner, U. Mason, et al., "Preverbal Infants' Sensitivity to Synaesthetic Cross-Modality Correspondences," *Psychological Science* 21 (2010): 21–25, <https://doi.org/10.1177/0956797609354734>.
16. G. Martino and L. E. Marks, "Cross-Modal Interaction Between Vision and Touch: The Role of Synesthetic Correspondence," *Perception* 29, no. 6 (2000): 745–754, <https://doi.org/10.1068/p2984>.
17. V. U. Ludwig and J. Simner, "What Colour Does That Feel? Tactile-Visual Mapping and the Development of Cross-Modality," *Cortex; A Journal Devoted to the Study of the Nervous System and Behavior* 49, no. 4 (2013): 1089–1099, <https://doi.org/10.1016/j.cortex.2012.04.004>.
18. B. G. Slocombe, D. A. Carmichael, and J. Simner, "Cross-Modal Tactile-Taste Interactions in Food Evaluations," *Neuropsychologia* 88 (2016): 58–64, <https://doi.org/10.1016/j.neuropsychologia.2015.07.011>.
19. C. Velasco, A. T. Woods, O. Deroy, and C. Spence, "Hedonic Mediation of the Crossmodal Correspondence Between Taste and Shape," *Food Quality and Preference* 41 (2015): 151–158, <https://doi.org/10.1016/j.foodqual.2014.11.010>.
20. A. A. Ghazanfar and J. X. Maier, "Rhesus Monkeys (*Macaca mulatta*) Hear Rising Frequency Sounds as Looming," *Behavioral Neuroscience* 123, no. 4 (2009): 822–827, <https://doi.org/10.1037/a0016391>.
21. M. Loconsole, M. S. Pasculli, and L. Regolin, "Space-Luminance Crossmodal Correspondences in Domestic Chicks," *Vision Research* 188 (2021): 26–31, <https://doi.org/10.1016/j.visres.2021.07.001>.
22. E. Versace, L. Freeland, and M. G. Emmerson, "First-Sight Recognition of Touched Objects Shows That Chicks Can Solve Molyneux's Problem," *Biology Letters* 20, no. 4 (2024): 20240025, <https://doi.org/10.1098/rsbl.2024.0025>.
23. M. Loconsole, S. Benavides-Varela, and L. Regolin, "Matching Sounds to Shapes: Evidence of the Bouba-Kiki Effect in Naïve Baby Chicks," *Science* 391, no. 6787 (2026): 836–839, <https://doi.org/10.1126/science.adq7188>.

24. M. Loconsole, B. Malaman, G. Stancher, and E. Versace, "Crossmodal Pitch-Luminance Association in Tortoises," *bioRxiv* (2025), <https://doi.org/10.1101/2025.01.24.634651>.
25. A. K. Kalan, R. Mundry, and C. Boesch, "Wild Chimpanzees Modify Food Call Structure With Respect to Tree Size for a Particular Fruit Species," *Animal Behaviour* 101 (2015): 1–9, <https://doi.org/10.1016/j.anbehav.2014.12.011>.
26. A. T. Korzeniowska, J. Simner, H. Root-Gutteridge, and D. Reby, "High-Pitch Sounds Small for Domestic Dogs: Abstract Crossmodal Correspondences Between Auditory Pitch and Visual Size," *Royal Society Open Science* 9, no. 2 (2022): 211647.
27. P. Linhart and R. Fuchs, "Song Pitch Indicates Body Size and Correlates With Males' Response to Playback in a Songbird," *Animal Behaviour* 103 (2015): 91–98, <https://doi.org/10.1016/j.anbehav.2015.01.038>.
28. M. L. Hall, S. A. Kingma, and A. Peters, "Male Songbird Indicates Body Size With Low-Pitched Advertising Songs," *PLoS ONE* 8, no. 2 (2013): e56717, <https://doi.org/10.1371/journal.pone.0056717>.
29. M. O. Ernst, "A Bayesian View on Multimodal Cue Integration," in *Human Body Perception From the Inside Out: Advances in Visual Cognition*, ed. G. Knoblich, I. M. Thornton, M. Grosjean, and M. Shiffrar (Oxford University Press, 2006), 105–131.
30. G. Fant, *Acoustic Theory of Speech Production* (The Hague, 1960), <https://www.semanticscholar.org/paper/Acoustic-Theory-Of-Speech-Production-Fant/5d79d86d4501b4a5381a7a8cc3562bf66cf975e6>.
31. I. R. Titze, *Principles of Voice Production* (Prentice Hall, 1994).
32. J. A. Vianna, F. A. N. Fernandes, M. J. Frugone, et al., "Genome-Wide Analyses Reveal Drivers of Penguin Diversification," *Proceedings of the National Academy of Sciences* 117, no. 36 (2020): 22303–22310, <https://doi.org/10.1073/pnas.2006659117>.
33. A. Zanoli, C. Tenneriello, I. Morandi, et al., "Acoustic Cues to Development of African Penguins (*Spheniscus demersus*) Begging Calls," *Ibis* 167, no. 1 (2025): 286–294, <https://doi.org/10.1111/ibi.13364>.
34. L. Favaro, M. Gamba, C. Gili, and D. Pessani, "Acoustic Correlates of Body Size and Individual Identity in Banded Penguins," *PLoS ONE* 12 (2017): e0170001, <https://doi.org/10.1371/journal.pone.0170001>.
35. F. Terranova, L. Baciadonna, C. Maccarone, V. Isaja, M. Gamba, and L. Favaro, "Penguins Perceive Variations of Source- and Filter-Related Vocal Parameters of Species-Specific Vocalisations," *Animal Cognition* 26 (2023): 1613–1622, <https://doi.org/10.1007/s10071-023-01806-w>.
36. L. Baciadonna, C. Solvi, S. La Cava, C. Pilenga, M. Gamba, and L. Favaro, "Cross-Modal Individual Recognition in the African Penguin and the Effect of Partnership," *Proceedings Biological sciences* 288 (2021): 20211463, <https://doi.org/10.1098/rspb.2021.1463>.
37. L. Baciadonna, C. Solvi, F. Terranova, C. Godi, C. Pilenga, and L. Favaro, "African Penguins Utilize Their Ventral Dot Patterns for Individual Recognition," *Animal Behaviour* 207 (2024): 13–21, <https://doi.org/10.1016/j.anbehav.2023.10.005>.
38. R. Cook, A. Wright, and D. Kendrick, "Visual Categorization by Pigeons," in *Behavioral Approaches to Pattern Recognition and Concept Formation*, ed. M. L. Commons, R. J. Herrnstein, S. M. Kosslyn, and D. B. Mumford (Psychology Press, 2019): 187–214, <https://doi.org/10.4324/9781315788999-10>.
39. K. Delhey, M. Valcu, J. Dale, and B. Kempenaers, "Evolutionary Predictors of the Specific Colors of Birds," *Proceedings of the National Academy of Sciences* 120 (2023): e2217692120, <https://doi.org/10.1073/pnas.2217692120>.
40. J. J. Peissig, K. Kirkpatrick, M. E. Young, E. E. Wasserman, and I. Biederman, "Effects of Varying Stimulus Size on Object Recognition in Pigeons," *Journal of Experimental Psychology: Animal Behavior Processes* 32, no. 4 (2006): 419–430, <https://doi.org/10.1037/0097-7403.32.4.419>.
41. E. G. Wever, P. N. Herman, J. A. Simmons, and D. R. Hertzler, "Hearing in the Blackfooted Penguin, *Spheniscus demersus*, as Represented by the Cochlear Potentials," *Proceedings of the National Academy of Sciences* 63, no. 3 (1969): 676–680, <https://doi.org/10.1073/pnas.63.3.676>.
42. P. Boersma, *The Use of Praat in Corpus Research* (Oxford University Press, 2013), <https://doi.org/10.1093/oxfordhb/9780199571932.013.016>.
43. K. Manabe, M. Murata, T. Kawashima, K. Asahina, and K. Okutsu, "Transposition of Line-Length Discrimination in African Penguins (*Spheniscus demersus*)," *Japanese Psychological Research* 51, no. 3 (2009): 115–121, <https://doi.org/10.1111/j.1468-5884.2009.00394.x>.
44. R. Miller, V. Slipogor, and K. Caspar, *Evolutionary Drivers of Neophobia Across the Avian Clade* (OSF, 2024), <https://doi.org/10.31219/osf.io/qhy8m>.
45. C. Spence, "On the Relative Nature of (Pitch-Based) Crossmodal Correspondences," *Multisensory Research* 32 (2019): 235–265, <https://doi.org/10.1163/22134808-20191407>.
46. O. Friard and M. Gamba, "BORIS: A Free, Versatile Open-Source Event-Logging Software for Video/Audio Coding and Live Observations," *Methods in Ecology and Evolution* 7 (2016), <https://doi.org/10.1111/2041-210X.12584>.
47. P. Hadden and J. Zhang, "An Overview of the Penguin Visual System," *Vision* 7 (2023): 6, <https://doi.org/10.3390/vision7010006>.
48. N. Albuquerque, K. Guo, A. Wilkinson, C. Savalli, E. Otta, and D. Mills, "Dogs Recognize Dog and Human Emotions," *Biology Letters* 12, no. 1 (2016): 20150883, <https://doi.org/10.1098/rsbl.2015.0883>.
49. M. Brooks, K. Kristensen, K. van Benthem, et al., "glmmTMB Balances Speed and Flexibility Among Packages for Zero-Inflated Generalized Linear Mixed Modeling," *R Journal* 9 (2017): 378–400, <https://doi.org/10.32614/RJ-2017-066>.
50. A. Ghazizadeh, W. Griggs, and O. Hikosaka, "Ecological Origins of Object Salience: Reward, Uncertainty, Aversiveness, and Novelty," *Frontiers in Neuroscience* 10 (2016): 378, <https://doi.org/10.3389/fnins.2016.00378>.
51. L. Stähle and S. Wold, "Analysis of Variance (ANOVA)," *Chemometrics and Intelligent Laboratory Systems* 6, no. 4 (1989): 259–272, [https://doi.org/10.1016/0169-7439\(89\)80095-4](https://doi.org/10.1016/0169-7439(89)80095-4).
52. D. J. Barr, R. Levy, C. Scheepers, and H. J. Tily, "Random Effects Structure for Confirmatory Hypothesis Testing: Keep It Maximal," *Journal of Memory and Language* 68, no. 3 (2013): 255–278, <https://doi.org/10.1016/j.jml.2012.11.001>.
53. C. Spence and O. Deroy, "Crossmodal Correspondences: Innate or Learned?," *i-Perception* 3, no. 5 (2012): 316–318, <https://doi.org/10.1068/i0526ic>.
54. L. Favaro, A. Zanoli, K. Ludynia, et al., "Vocal Tract Shape Variation Contributes to Individual Vocal Identity in African Penguins," *Proceedings of the Royal Society B: Biological Sciences* 290, no. 2008 (2023): 20231029, <https://doi.org/10.1098/rspb.2023.1029>.
55. H. Röfler, A. May, M. Dähne, and M. Beaulieu, "Long and Winding Road: Training Progress and Trainability Variation Across a Psychoacoustic Experiment in Penguins," *Applied Animal Behaviour Science* 256 (2022): 105764, <https://doi.org/10.1016/j.applanim.2022.105764>.
56. E. Jarvis, O. Güntürkün, L. Bruce, et al., "Opinion: Avian Brains and a New Understanding of Vertebrate Brain Evolution," *Nature Reviews Neuroscience* 6 (2005): 151–159, <https://doi.org/10.1038/nrn1606>.

Supporting Information

Additional supporting information can be found online in the Supporting Information section.

Supplementary materials: nyas70313-sup-0001-

SuppMat.docx **Supplementary video:** nyas70313-sup-0002-

video.mp4 **Supplementary video:** nyas70313-sup-0003-video.mp4