








RoboFinch: A versatile audio-visual synchronised robotic bird model for laboratory and field research on songbirds

Ralph Simon^{1,2}  | Judith Varkevisser³  | Ezequiel Mendoza⁴  | Klaus Hochradel⁵  |
Rogier Elsinga¹ | Peter G. Wiersma¹ | Esmee Middelburg³ | Eva Zoeter³ |
Constance Scharff⁴  | Katharina Riebel³  | Wouter Halfwerk¹ 

¹Department of Ecological Science, VU University Amsterdam, Amsterdam, The Netherlands

²Behavioral Ecology and Conservation Lab, Nuremberg Zoo, Nuremberg, Germany

³Institute of Biology Leiden, Leiden University, Leiden, The Netherlands

⁴Department of Animal Behavior, Institute of Biology, Freie Universität Berlin, Berlin, Germany

⁵Institute of Measurement and Sensor Technology, UMIT-Private University for Health Sciences, Medical Informatics and Technology GmbH, Hall in Tirol, Austria

Correspondence

Wouter Halfwerk
Email: w.h.halfwerk@vu.nl

Funding information

Human Frontier Science Program, Grant/Award Number: RGP0046/2016

Handling Editor: Camille Desjonquères

Abstract

1. Singing in birds is accompanied by beak, head and throat movements. The role of these visual cues has long been hypothesised to be an important facilitator in vocal communication, including social interactions and song acquisition, but has seen little experimental study.
2. To address whether audio-visual cues are relevant for birdsong we used high-speed video recording, 3D scanning, 3D printing technology and colour-realistic painting to create RoboFinch, an open source adult-mimicking robot which matches temporal and chromatic properties of songbird vision. We exposed several groups of juvenile zebra finches during their song developmental phase to one of six singing robots that moved their beaks synchronised to their song and compared them with birds in a non-synchronised and two control treatments.
3. Juveniles in the synchronised treatment approached the robot setup from the start of the experiment and progressively increased the time they spent singing, contra to the other treatment groups. Interestingly, birds in the synchronised group seemed to actively listen during tutor song playback, as they sung less during the actual song playback compared to the birds in the asynchronous and audio-only control treatments.
4. Our open source RoboFinch setup thus provides an unprecedented tool for systematic study of the functionality and integration of audio-visual cues associated with song behaviour. Realistic head and beak movements aligned to specific song elements may allow future studies to assess the importance of multisensory cues during song development, sexual signalling and social behaviour. All software and assembly instructions are open source, and the robot can be easily adapted to other species. Experimental manipulations of stimulus combinations and synchronisation can further elucidate how audio-visual cues are integrated by receivers and how they may enhance signal detection, recognition, learning and memory.

Ralph Simon and Judith Varkevisser shared first authorship.

Katharina Riebel and Wouter Halfwerk shared last authorship.

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial](https://creativecommons.org/licenses/by-nc/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2023 The Authors. *Methods in Ecology and Evolution* published by John Wiley & Sons Ltd on behalf of British Ecological Society.

KEYWORDS

language evolution, multimodal communication, robotic bird model, robotics, social behaviour, song learning, zebra finch

1 | INTRODUCTION

Multimodal communication is ubiquitous in nature. Plants lure pollinators with colour and fragrance, animals combine sounds with visual displays during courtship and humans gesture while talking (Fröhlich & van Schaik, 2018; Halfwerk et al., 2019; Higham & Hebets, 2013; Munoz & Blumstein, 2020; Partan & Marler, 1999; Ullrich et al., 2016). These multimodal displays presumably evolved to increase signal salience, aid species recognition or improve cognitive processing of the intended receivers, which can all contribute to increase signal attractiveness or effectiveness (Halfwerk et al., 2019; Hebets & Papaj, 2005). The mechanisms underlying improved detection and species recognition have been experimentally addressed in the past decade (Gomes et al., 2016; Halfwerk et al., 2014; Heuschele et al., 2009; Reş & Magrath, 2017; Uetz et al., 2011). However, whether and why multimodal signals improve cognitive processing and are thus easier to recognise, learn and memorise compared to signals that only target a single sensory systems is far from understood (Leavell et al., 2018), in part because these types of experiments often require high standards of stimulus control across multiple physical modalities and for long periods of time.

Human speech and bird song are learned communication systems involving primarily the auditory sensory and vocal motor modalities (Bolhuis et al., 2010; Doupe & Kuhl, 1999). Additionally, the production of speech and song is usually accompanied by concurrent visual cues (lip or beak movements). The importance of the multimodal integration of visual and auditory cues for human speech processing is demonstrated by the 'McGurk-effect': when people are experimentally exposed to a mismatch between the audio and video track of an articulated syllable (e.g. when hearing an audio track of the syllable [ba] and seeing video track of a face articulating [ga], most people will hear the syllable [da]; McGurk & Macdonald, 1976). Furthermore, infants as young as 18 to 20 weeks already recognise correspondences between specific speech sounds and the unique orofacial movements associated with them. This suggests that the integration of the auditory and visual aspects of speech are highly relevant for speech learning, even without external reinforcement (Kuhl & Meltzoff, 1982). Interestingly, in many songbird species juveniles learn to sing better when being able to hear and see their song tutors. However, whether this is due to the visual input or due to social factors remains unknown because of too many confounding variables typically associated with the experimental setups used during song tutoring experiments (Eales, 1989; Nelson, 1997; White, 2001). To disambiguate between the relevance of audio-visual and social relevance for song learning a setup is required that allows for independent experimental control of visual and auditory cues in a life-like manner.

Robotic models allow multimodal signal components to be controlled independently and to mimic animal behaviours in realistic 3D

settings (contra to using e.g. video screens; Romano et al., 2019). Robotic models also allow us to test artificial stimulus combinations to assess how receivers of multimodal signals process the different components (e.g. facilitating multisensory integration studies). Robotic lifelike models already have helped to demonstrate that multimodal signals and cues are indeed influencing receivers in non-additive ways, not predicted by their reactions to the single components (Narins et al., 2003; Patricelli & Krakauer, 2010; Reş & Magrath, 2017; Taylor & Ryan, 2013). Setups to study early sensory and/or lifetime learning and memory formation have so far not been developed or applied, but have high potential to uncover whether and how multisensory integration improves cognitive functions. Improved robotic approaches are thus needed to understand the important mechanisms and functioning involved with multimodal signalling.

Here, we describe how we designed, constructed and applied a robotic zebra finch model, which we named RoboFinch. We aimed to create a 3D printed, colour-realistically painted model of a male zebra finch whose beak and head movements during singing were deduced from high speed video recordings from singing males, thus realistically enough to function in experimental settings ranging from song acquisition studies to social behaviours to multimodal signal evolution. We based RoboFinch on the zebra finch (*Taeniopygia guttata*) because it is an important animal model for behavioural, neurobiological and genetic analyses of vocal learning. However, the methods presented here can easily be transferred to other bird species. Previous work on zebra finches uncovered important parallels to mechanisms of human speech production (Griffith et al., 2021; Pfenning et al., 2014). Traditionally, the analysis of the relevant stimuli for song learning has focused on audio playbacks, sometimes supplemented with different visual components, such as non-animated clay or plasticine models, taxidermic mounts or video presentations (Bolhuis et al., 1999; Derégnaucourt et al., 2013; Gobes & Bolhuis, 2007; Varkevisser et al., 2021). Only recently have animated robots been employed to mimic the visual presence of a bird (Araguas et al., 2022; Simon et al., 2019). These previous attempts however did not synchronise sound and movement of their robotic bird during song development. The focus of most of these earlier multimodal tutoring studies were therefore to emulate aspects of the social presence of tutors (inanimate models) or to provide contingencies (operant activation) rather than focusing on the question of underlying mechanisms associated with multimodality. These latter questions would require a more naturalistic representation of singing, and thus a high level of stimulus control to account for covariance in audio-visual cues related to the production of specific song elements.

Birds produce songs with their syrinx and modulate them by altering resonance properties of their vocal tract (Beckers et al., 2004; Elemans, 2014; Hoese et al., 2000; Nowicki, 1987; Podos et al., 2009; Riede & Goller, 2010; Westneat et al., 1993). Song modulation

involves opening and closing of the beak or altering the shape of the oropharyngeal-oesophageal cavity. Producing a specific sound will often covary with a specific posture that provides clear visual cues. Beak gape in zebra finches is for example positively related to several frequency and amplitude characteristics of different song elements (Goller et al., 2004; Ohms et al., 2010; Podos et al., 2004; Williams, 2001). Beak movements are typically fast and the development of a robotic model thus requires recording live singing males at high speeds and high levels of synchronisation between the auditory and visual channels.

To measure the speed, variation and alignment of sound and beak movements, we first used a high-speed video camera (120 frames/s) to film six adult male zebra finches during undirected singing (not directed at another bird). We used these videos as the base to create unique motion files that could drive the head and beak movements of our robots in sync with the specific acoustic song elements. Next, we 3D scanned a taxidermic zebra finch and used the 3D scan to 3D print the bodies and beaks of six different RoboFinches that we hand-painted using colours mixtures that closely matched the colours of real zebra finch plumage and beaks. We then tested the responses to these robots by presenting them to juvenile birds, either housed alone or paired with a companion bird. After confirming typical interactive behaviours (see Section 3), we used these robots during song tutoring experiments in which we exposed male tutees to a robot moving its beak and head either in or out of sync with sound playback of the songs and calls of the tutor. We compared the juveniles' behaviours during the exposure to two different RoboFinch treatments with the behaviour of juveniles exposed only to the sound of the tutor, while in social isolation, or accompanied by a female juvenile. We video-recorded these staged bird-robot encounters and analysed the behavioural acceptance of, and engagement with these artificial companions and song tutors. Finally, for a subset of the tutees we qualitatively compared their song at adulthood with that of their tutors as a proof-of-principle that our RoboFinches can function as an experimental tool to dissect the relative importance of the different multimodal components during song tutoring.

2 | MATERIALS AND METHODS

The development of the RoboFinch was carried out at VU Amsterdam. Song behaviour measurements and experiments were carried out with zebra finches from the colony at the Institute of Biology, Leiden University. Throughout, birds were housed on a 13.5/10.5 h light/dark cycle, at 20–22°C and 45%–65% humidity. Birds had ad libitum access to a commercial tropical seed mixture (Beyers, Belgium), cuttlebone, grit and drinking water. This was supplemented three times a week with hardboiled eggs and once a week with germinated tropical seeds, vegetables and fruit. Experiments were reviewed and approved by the Leiden University Committee for animal experimentation, the Leiden University Animal Welfare Body and the Centrale Commissie voor Dierproeven (CCD) of the Netherlands (permit number AVD1060020186606) and all zebra finches were housed and cared for in accordance with National (the

Experiments on Animals Act, Wod, 2014) and European legislation (EU directive no. 2010/63/EU).

2.1 | 3D scanning of the body

We scanned a taxidermic zebra finch model using a handheld 3D scanner (Eva, Artec3D). The upper and lower part of the beak were scanned at the highest resolution from a prepared skull (ATOS 5X, gom; Figure S1A,B in Appendix S1). The 3D scans were combined and adjusted in the computer aided design (CAD) program Catia V5R20 (Dassault Systèmes) to create separate shape files for the different body parts. These shape files included the upper-and lower beak, the left-and right side of the head (split along the longitudinal axis) and front-and back side of the body (split along the longitudinal axis). The body and head were split between the neck and throat, at the position of the steepest curvature (Figure S1C in Appendix S1).

2.2 | Mechanical construction

We aimed at driving the movements of various body parts separately. For the mechanics driving the head and beak we used coils from commercially available toys (see Simon et al., 2019), so called DigiBirds (Silverlit Toys Manufactory), that allowed to open and close the beak (up to 5.0mm, measured at the tip of the beak) and rotate the head (from -30° to $+30^\circ$). We designed the mechanical parts that connected the two driving coils to the different body parts also in the CAD software Catia. The coil driving the opening and closing of the beak was placed inside the head and attached via a lever to the lower beak. The upper beak was fixed to the head, as is the case in zebra finches. The coil driving the head rotations was placed inside the body. This coil moved the head from left to right. The head was connected to the body via two rods, one placed at the central axis and one placed to the front, which could be moved by the coil inside the body, resulting in the head rotating along the central axis (Figure S1C in Appendix S1). The model was mounted on a metal tube through which we ran the wiring to the coils. The metal tube was placed in a pvc stand which could rotate inside a larger tube, placed on a pvc platform. The body was rotated by a Nema 17 stepping motor (Type: 17HS3404L23P1-X1; ACT Motor GmbH) placed outside the body in the plastic platform (Figure S5 in Appendix S1). To control the stepping motor we used a Pololu Tic T825 USB Multi-Interface Stepper Motor Controller. The setup included two wooden perches placed adjacent to the RoboFinch to allow live birds to approach and inspect it from close by.

2.3 | 3D printing of the RoboFinch models

We printed the body parts and mechanical components with stereolithography 3D printing (Form 2; Formlabs), which uses a laser to cure solid isotropic parts from a liquid photopolymer resin (Grey Pro, Formlabs Resin). After assembly we checked the beak, head and

body movements and filed off parts of the beak that caused friction. Next, we painted the fully assembled models using colour patterns that closely matched light reflectance patterns of live zebra finch males (see Supplementary Methods in Appendix S1).

2.4 | Stimuli recordings from singing adult males

To deduce movements and song from live males we recorded undirected song of six adult male zebra finches. A single male was placed in a recording cage (76×45×45 cm) in the afternoon of the day before recording. The next morning, the male was recorded between 08:00 and 11:00, or until he sang three full songs. The recording cage had a clear Plexiglas window in the middle of the front side of the cage. This cage was placed on a table in a sound attenuated room. Only one cross perch was placed in the middle of the cage so that the bird would always be in focus on the camera. Audio recordings were made with a Sennheiser MKH40 microphone hanging 50 cm above the perch in the recording cage. The microphone was connected to a TASCAM DR-100MKiii recorder. Audio was recorded with a sampling rate of 96 kHz and 16-bit resolution. Video recordings were made with a Casio high speed camera (EX-ZR3600, 120 fps, 12× optical zoom) through Plexiglas in the door of the sound attenuated room. A signal bell (70027 Heidemann), which was sound attenuated to not disturb the birds was attached to the front side of the recording cage above the Plexiglas window and could be triggered from outside the sound attenuated room. The bell produced a short, impulse-like audio signal and was clearly visible on the video when the clapper touched the bell, which was later used to synchronise the audio and video recordings during stimulus preparations. At the start of each recording we triggered the bell and the camera could record 120 fps videos for up to 12 min. Audio files were filtered with a band-stop filter from 0 to 425 Hz using Praat (version 6.0.19). Audio and video were synchronised with Vegas Pro (version 14.0).

Video files were used to analyse beak and head movements using the tracking software Tracker (open source physics, physlets.org). We selected videos in which males were singing in the direction of the camera, allowing optimal tracking of the opening and closing of the beaks. We selected pixels to be tracked on the tips of the upper and lower beak (Figure S6A in Appendix S1). Head movements were tracked in a similar way, by selecting pixels on the forehead. To deduce beak opening we subtracted the *y* values of upper and lower beak and normalised them per song recording (1 was open at maximum, 0 was closed). For the head movement we deduced the *x* values from the tracker software and also normalised the displacement (1 resembled the furthest turn to the left and 0 resembled right). Data were stored in a text file with a resolution (Δt) of 8.3 ms (see Figure S6B in Appendix S1 for an example).

For each of the six males we tracked the movements of three songs with introductory notes followed by three to five motifs (mean song duration \pm SD = 4.2 \pm 1.2 s, mean number of motive repetitions \pm SD = 3.9 \pm 0.8). We also tracked the movements for two sequences each where the birds made contact calls and two sequences each where the birds just moved their head.

2.5 | Driving the RoboFinches

The RoboFinch model was controlled by a custom-built controller board, which was based on an Arduino (Adafruit 3405; Mouser electronics) and a Pololu Tic T825 USB Multi-Interface Stepper Motor Controller in order to drive the different body movements. A speaker (Blaupunkt, CB4500), connected to an amplifier (Devine MA-400), was placed behind the RoboFinch model. Playback of sound and body movements were operated via a custom-written LabVIEW program (National Instruments) that ran on a NUC PC (NUC7i3BNK; Intel Corporation). The LabVIEW program could select a soundfile and associated movement file obtained from one of the six adult singing males. To synchronise song playback and beak movement, we filmed our RoboFinches with an iPhone XR highspeed video (120 fps) and adjusted the synchronisation file with small delays in the sound file playback or the movement file playback. As song onset and onset of beak movement was not very distinct—song often started with a note that was produced with no or only minimal beak movement—we adjusted the song using the pauses in between syllables and the end of the motif for alignment. The temporal precision of each song remained accurate throughout the duration of the experiment (Figure S10 in Appendix S1). Furthermore, the temporal precision of movements associated with single song elements also remained accurate, but with this first generation model the spatial precision became gradually less accurate in the course of the experiment. This may have been related to changes in the properties of the coil, or the mechanical parts linking to the beak (see also Section 4). A subtle change in positioning could for example sometimes lead to the beak not opening or closing during a specific element. The positioning of the beak (e.g. at an intermediate beak gape size) can be influenced by changing various parameters in our software program (see detailed guideline on GitHub, link below), but at the moment requires constant readjustments, which we decided not to do during our tutoring experiment due to logistical and time constraints.

For the purpose of the experiment reported here (pairing temporally accurate movements with sound vs. the non-synchronous sound/moving robot treatment) the drifting of the spatial accuracy of specific beak movements fortunately did not affect the temporal accuracy meaning that the 3D printed RoboFinches nevertheless provided a realistic representation of the body, head and beak movements associated with calling and singing of adult male zebra finches (see Figure 1; Figures S8 and S9 in Appendix S1; Video S1 for comparison of one of our models with its live exemplar). The robots could open and close their beaks up to a rate of 100 Hz to a maximum beak gape of \pm 5 mm and rotate their heads from -30° to $+30^\circ$ and turn their bodies 180 degrees. Spectrophotoscopic readings of the beak, cheeks, flanks and back colourations matched those of a live zebra finch (Figure S2 in Appendix S1).

A detailed instruction guide containing the source code for Labview as well as the 3D printing files, sound files and motion files can be found in the Supplementary Materials in Appendix S1 and on GitHub|Zenodo (<https://github.com/SimonHubs/RoboFinch> | <https://doi.org/10.5281/zenodo.7520589>).

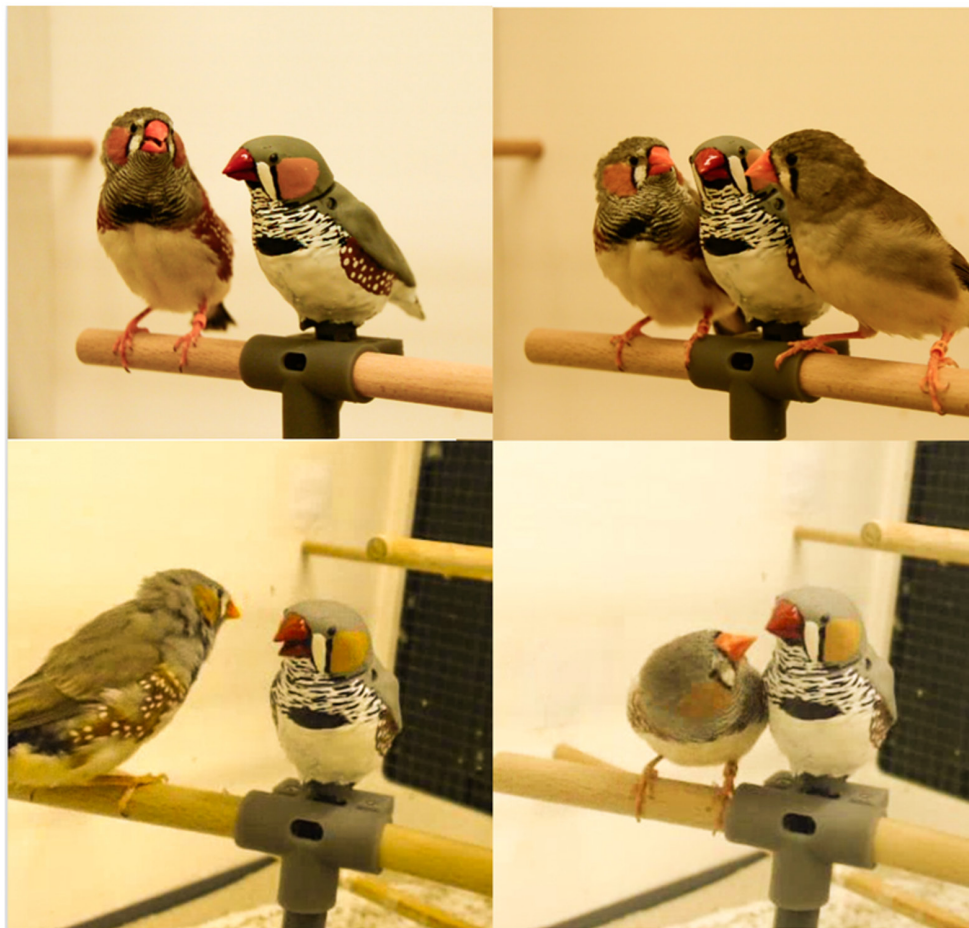


FIGURE 1 Images of young zebra finches interacting with the RoboFinch inside the cage during the first pilot behavioural experiment.

2.6 | Testing the RoboFinch with juveniles

In a pilot behavioural experiment, we tested the acceptance of the robots with two dyads of young zebra finches, ranging in age between 45 and 75 days post-hatching (dph) each dyad consisting of one male and one female. We observed the birds with a webcam for two sessions per day over the six initial days and scored their movements using frame differencing with a custom made LabView program (National Instruments; see [Figure S3](#) in Appendix [S1](#)). In a second experiment we used 45 juvenile males. Subjects stayed in their home cage with their mother and siblings from 20 to 35 dph. At 35 days they were moved to the experimental set-up, where they were tutored in one of four different conditions during the sensitive period for song learning (between 35 and 65 dph): (1) song playback only (“AudioOnly”), (2) song playback and an unrelated age-matched female housed in the same cage as the male tutee (“AudioFemale”), (3) song playback and a RoboFinch positioned next to the cage that produced beak and head movements time-aligned with the presented sound (“RoboSync”), (4) song playback and a RoboFinch positioned next to the cage that only started moving after the auditory song presentation session had finished (“RoboAsync”). The same tutor song was presented to four male tutees, each in a different tutoring treatment. For all treatments, a speaker was placed outside of the tutees cage, behind the position of the RoboFinch, or in

the corresponding position without the robot and song amplitude was set to 74 dB (Fast, A, re 20 μ Pa, Voltcraft SL-451) at 30 cm. All tutees received 6 tutoring sessions daily that lasted 30 min. During a tutoring session, three different types of files were played: songs, calls and head movements. In the RoboAsync, we created a complete mismatch by broadcasting the movement files in the half hour following the tutoring session. These movements matched movements broadcasted in the RoboSync treatment during the tutoring session (see [Figure S7](#) in Appendix [S1](#) for experimental design). Birds were observed with a webcam throughout the experiment and their behaviour scored during and in between the blocks of 10-min stimulus presentation. For more information on the setup, experimental design, data analyses and statistics, see Supplementary Methods in Appendix [S1](#).

3 | RESULTS

3.1 | Young zebra finches actively engage with the RoboFinch

To get an impression of how young zebra finches reacted to the RoboFinch we ran a pilot in which we exposed two dyads of juvenile zebra finches (aged between 45 and 75 dph), each consisting of a

male and a female, to our setup. The robot was placed adjacent to their cage behind a mesh window. The birds inside the cage could sit on a perch directly at that window. Birds in both groups spent most of their time in front of the robot, in particular when it was moving, singing or calling (Figure S3 in Appendix S1). When shortly placed inside their cages, birds would inspect the robot from close-by and we observed what seemed to be a higher incidence of juvenile singing soon after the robot sang (Figure 1, Video S2 and S3).

3.2 | Male tutees stay close to the audio-visual synchronised robot

Following up on our pilot results, we housed a total of 45 male tutees for a period of 30 days during their sensitive song acquisition phase (in this species between day 35 and day 65 post hatch) in one of four different tutor treatments. In the first—control—treatment young males were housed individually and received audio exposure to song and calls of the tutor (hereafter 'AudioOnly'). In the second—control—treatment a young male was exposed to the same audio stimuli, but was housed together with a female companion (hereafter 'AudioFemale'). As females do not sing in this species, this treatment aimed at testing a potential effect of a social companion (that was not the tutor) on song learning (Adret, 2004). The third—experimental—treatment consisted of exposure to the RoboFinch providing head and beak movement cues synchronised to the audio stimuli (hereafter 'RoboSync') and the fourth—experimental—treatment was the same as the third except that the head and beak movements were not synchronised. Instead, the RoboFinch moved only after the audio playback was finished (hereafter 'RoboAsync').

Male tutees that were housed with the RoboFinch spent most of their time on the perch or the mesh that was closest to the tutor or speaker location, while males that received the audio treatments did not spend more time in that part of the cage (LMMs; effect of treatment: $N = 24$ males, $\chi^2 = 41.1$, $df = 3$, $p < 0.001$; Figure 2a). The time spent near the location of the RoboFinches and speakers changed during the course of the experiment, when comparing the first 15 days with the last 15 days of the experiment. This effect depended on the treatment (Interaction effect of Period \times Treatment: $\chi^2 = 38.75$, $df = 3$, $p < 0.0001$; Figure 2a). Tutees in the RoboSync and the AudioFemale spent as much time close to speaker during the first half as during the second half of the experiment (both $p = 0.99$; Figure 2a). In contrast, males in the RoboAsync (z -value = 7.52, $p < 0.0001$) and the AudioOnly group (z -value = -2.61, $p = 0.036$), spent less time near the sound source during the first half compared to the second half of the experiment. Following up on these findings, we zoomed in further and compared the time spent near the robot during the first week and found that males in the RoboSync group spent 27% of their time on the perch closest to the robot setup compared to only 5% in the RoboAsync group and even less in the AudioOnly and AudioFemale groups (see also Figure S4 in Appendix S1).

Tutees also spent most of the time close to the speaker when the song was broadcast compared to the period immediately afterwards.

This effect was, however, mainly visible for the two RoboFinch treatments (interaction between song on/offset and treatment: $\chi^2 = 35.61$, $df = 3$, $p < 0.001$, Figure 3a).

3.3 | Audio-visual synchronised presentation influences tutee song behaviour

Tutees regularly practised singing during their development, but tutor treatment alone had no significant effect on the time that tutees spent singing ($\chi^2 = 2.14$, $df = 3$, $p = 0.54$). However, practicing increased over the duration of the experiment and this effect depended on tutoring treatment (Interaction effect of Period \times Treatment: $\chi^2 = 11.64$, $df = 3$, $p = 0.009$; Figure 2b). Birds in the RoboSync treatment practised their song more in the second half of the experiment compared to the first (z -value = 3.52, $p = 0.0017$), whereas in the RoboAsync treatment we did not find a statistically significant change in singing behaviour over the course of the experiment (z -value = 0.23, $p = 0.99$).

Furthermore, we found a significant interaction effect between tutor treatment and tutor song onset- and offset ($\chi^2 = 10.76$, $df = 3$, $p = 0.013$; Figure 3). In the control group that only received audio input, tutees spent about an equal amount of time singing during song playback compared to periods with no playback (AudioOnly, z -value = 0.02, $p = 1.0$), whereas tutees with a female companion sang less during tutor song playback compared to periods with no sound playback (AudioFemale, z -value = -2.66, $p = 0.03$). In the two RoboFinch groups, male tutees also sang less during sound playback when compared to periods without sound playback, but the effect was only statistically significant in the synced group (RoboSync group, z -value = -4.86, $p < 0.0001$; RoboAsync group, z -value = -2.30, $p = 0.08$), suggesting that male tutees actively paid attention to the tutor song and movements when these were in synchrony.

There was a trend for males to call more in the AudioOnly treatment which lacked a social companion, and to preen more in the two RoboFinch treatment groups, but these patterns were not statistically significant (all $p > 0.16$, Figure 2), and did not change over the course of the experiment (no effect of date, all $p > 0.13$). Visual inspection of sonograms recorded at 120 dph showed that males in all treatments developed overtly typical adult zebra finch song after 65 days of tutoring exposure, incorporating song elements of the tutor (Figure 4a–d). The analysis of copy quality in the different treatment groups is subject of future analysis.

4 | DISCUSSION

Here we report how we designed, developed and tested a life-like robotic model to study vocal communication and imitation learning in zebra finches. Based on high-quality audio and high-speed video recordings of six adult singing males, we created and validated six unique models (each for the song and associated motor gestures

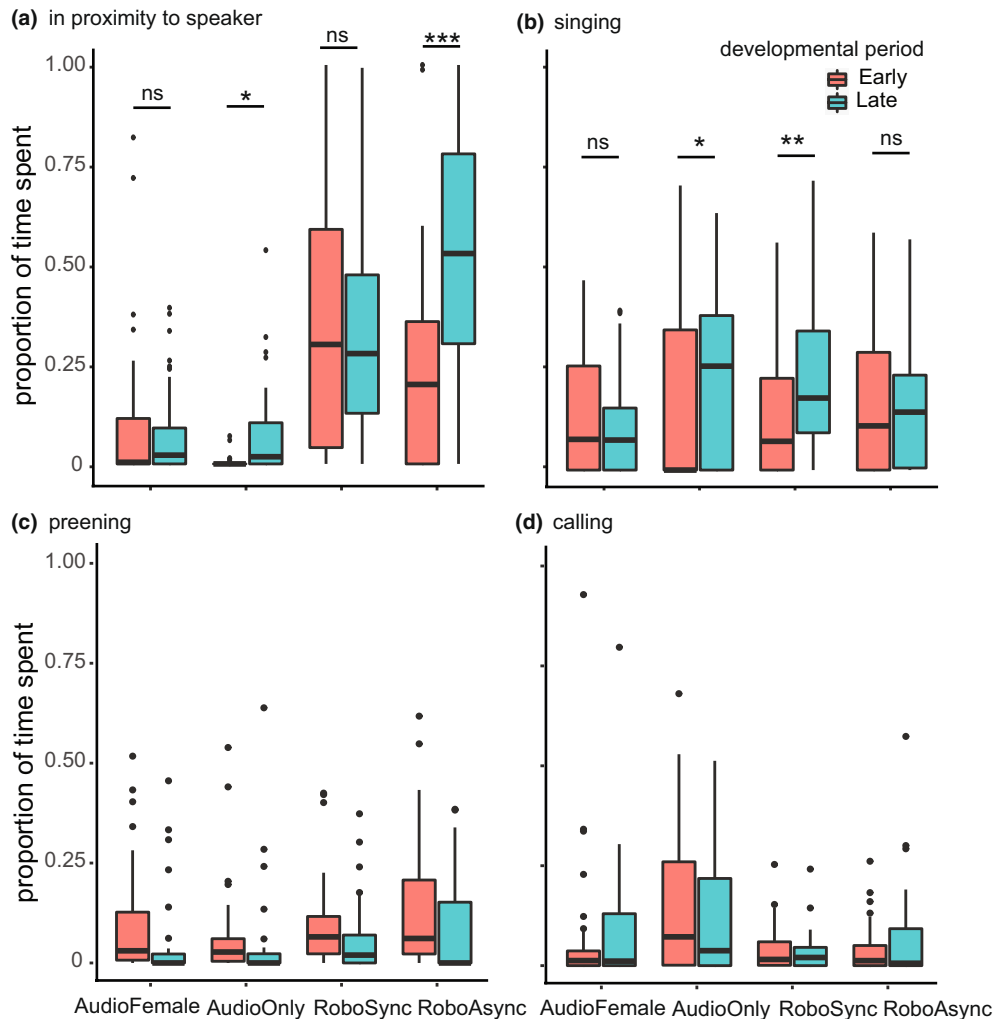


FIGURE 2 Effect of tutor treatment on tutee behaviour across their sensitive learning phase. Shown is the proportion of time juvenile males spent on different behavioural activities split over the early (<15 days) and late stage (>15 days) of the tutoring experiment, which took place from 35–65 days post hatching. (a) Proportion of time tutees spent close to the sound source and/or robots for the different treatments. Juvenile birds in the two RoboFinch treatments spent most of their time close to the tutor model, albeit in the RoboSync treatment birds approached the model from the start, whereas in the RoboAsync treatment, birds needed some more time for this. (b) Proportion of time spent singing. Juveniles in the synchronised robot treatment sang less in the beginning and more during the later stages of the tutoring experiment. (c) Proportion of time spent preening and (d) calling during the tutoring period did not differ across treatment groups nor period. Shown is raw data in boxplots (depicting the median and upper and lower 10% and 35% quartiles) and outliers. p -values ($* < 0.05$; $** < 0.01$; $*** < 0.001$) are based on post-hoc multiple comparisons between early and late stages per treatment group.

of one male) of a realistic animated setup, named RoboFinch. Our methods and setup can serve as a multimodal stimulus tool to simultaneously present zebra finches (and in the future also other species) with both visual and acoustic cues associated with vocal sound production. The RoboFinch was validated by assessing how juvenile male zebra finches interacted with it during the sensitive phase of song learning. We compared this condition to a condition where juveniles listened only to tutor song, either in social isolation or in the company of a female. Our results show that juveniles perch preferentially in the vicinity of the RoboFinch and behave towards it in similar ways as juveniles in this species normally interact with a tutor. Strikingly, the juveniles in this study often interrupted ongoing activities and actively approached the RoboFinch when it started singing, mirroring 'active listening behavior' of male juvenile zebra finches in

the presence of a live tutor (Liu et al., 2021). Furthermore, our song tutoring experiment revealed that birds can copy songs from the RoboFinch tutor, although detailed acoustic analyses will need to show whether song copy accuracy improves with multimodal stimulus presentation compared to unimodal presentation. Behavioural observations during the song acquisition and production phase showed however that birds paid attention to synchronised sounds and movements, and next to increased listening behaviour, we found more singing/motor practice, known to be beneficial to song learning (Pytte & Suthers, 1999). Interestingly, the presence of a female seemed to rescue this type of behaviour in the audio only groups. Males in the AudioOnly versus AudioFemale groups showed differences in the same parameters and in the same directions than observed between the RoboSync and RoboAsync groups: with

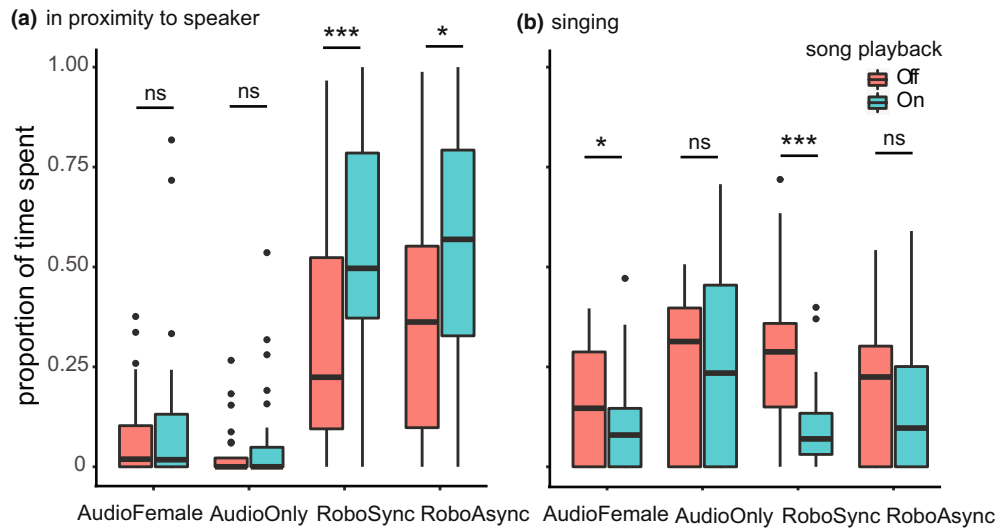


FIGURE 3 Effect of tutor treatment on tutee behaviour in relation to song playback. (a) Proportion of time tutees spent close to the sound source and/or robots during (on) and in between (off) tutor song playback. Juvenile males spent most of their time close to the robot models, especially during song playback. (b) Proportion of time spent singing. Tutees would often stop singing during tutor song exposure and resume immediately after, in particular in the tutoring treatment in which song and beak and head movements were synchronised (RoboSync). p -values ($* < 0.05$; $** < 0.01$; $*** < 0.001$) are based on post-hoc multiple comparisons between phases with and without song playback per treatment group.

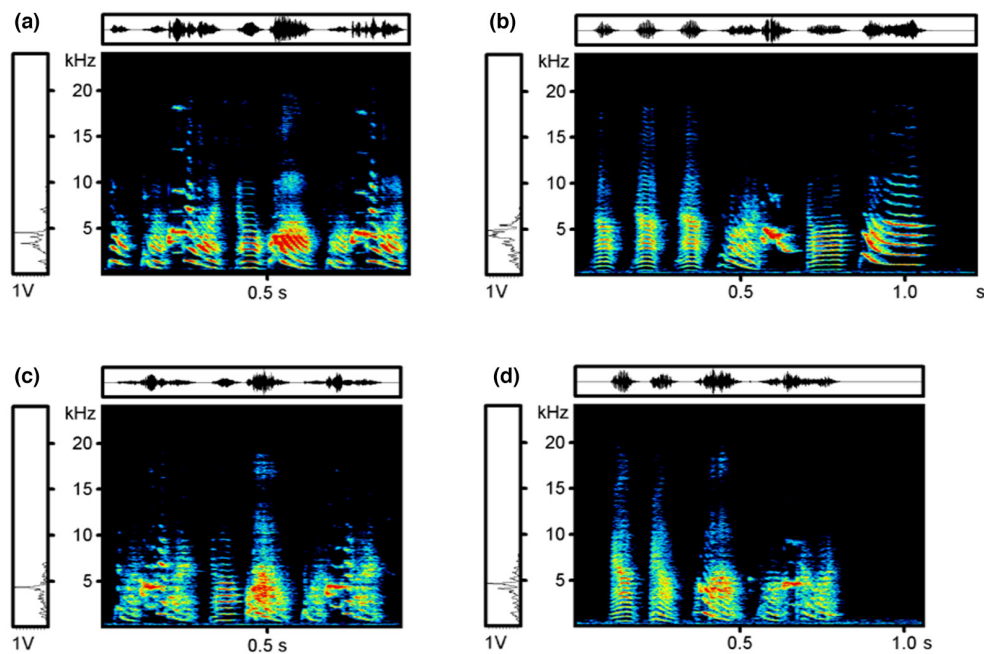


FIGURE 4 Typical song motifs of the robotic tutor and the tutees receiving different tutor regimes. (a) Motif of the robotic tutor. (b) Motif of the tutee raised in the AudioOnly condition. (c) Motif of the tutee raised in the RoboSync condition. (d) Motif of the tutee raised in the RoboAsync condition.

a female present there was consistent engagement with the song stimuli and more singing. While this suggests that there is a social effect independent of song input from the females, the comparison between the two robot groups shows that this must be yet another effect than that of the robot simply replacing a social companion. The synchronised presentation of sound and beak movements

clearly changed the salience of the RoboFinch stimulus. The observation that juveniles moved and sung differently in the synced robot treatment group compared to the non-synced and (one of the two) control groups is in line with our predictions arising from the potential importance of synchronised visual cues for song learning. The tools we developed and share open source (see Section 2) will

help to test the importance of multimodal signals for learning and memory further in the future.

Various lines of evidence support the notion that the role of audio-visual integrated cues deserves further experimental study. First, multimodal signals often generate more attention than unimodal signals, suggesting that increased stimulus salience alone might improve learning (Hebets & Papaj, 2005). Second, because sound sources are often difficult to localise, receivers may benefit from multimodal displays by integrating audio-visual information on sender identity and location (Halfwerk et al., 2019) and this might be especially true for species like the zebra finch where song serves mainly as a close-range communication signal (Loning et al., 2022). Juveniles may therefore need to stay close to their song tutors to be able to process the fine acoustic structures present in this species' song. We indeed observed juveniles to approach the speaker broadcasting tutor song only in the two robot groups, suggesting that the presence of a physical model with visual cues guides them closer to the sound source. In addition, birds in the synced group approached the robot immediately from the start of the experiment, whereas in the other groups this took a couple of days, clearly demonstrating that birds paid attention to the combined audio-visual exposure. This suggests higher salience of the audio-visual synchronised condition, which is also observed in human babies (Kuhl & Meltzoff, 1982). In future experiments our setup could allow testing the importance of specific orofacial gestures and their synchrony with production of specific song elements. For this first experiment, we only broadcast the beak and head movements during versus after song playback to test the importance of simultaneous visual and audio presentations, before moving on manipulating fine level vocal gestures. The reason behind this decision was both because logistic reasons meant a cap on the number of experimental groups and because the first generation robots, while functioning with high temporal accuracy, showed drifting in the topological accuracy of the beak movements, an aspect of the robot's performance we intend to improve in the future. Improved accuracy of fine beak movements will allow to test whether correct motor gestures improve song learning, in addition to seeing temporally synced beak movements (which might in addition increase attention).

Approach and listening behaviour during song development may reflect increased attention by the juveniles to process the relevant acoustic information encoded in adult song. Interestingly, in a study on live song tutors, a few juveniles were observed to approach adults milliseconds before the adult started to sing, a clear indication juveniles had paid attention to the tutor's gestures (Liu et al., 2021), similar to what we observed in our robot tutoring experiment. The observation that the juveniles exposed to the synchronised RoboFinch stopped singing during presentations shows that these tutees behaved most similar to live tutored birds when exposed to adult tutor song. This is different from an operant use of an animated zebra finch model (Araguas et al., 2022), where tutees that actively approached the animated model were rewarded with song by the experimenter (here the model was thus reacting to the tutees, not vice versa as in live tutoring and our setup). With these

reversed contingencies, the operant tutor group in this other robot study learned comparably well as a control live tutored group, when both groups got 1 h/day with the respective models. Other studies have used operant tutoring methods but not systematically controlled the presence/absence of visual cues of sound production or looked at contingencies between tutor (song) and tutee behaviour (Araguas et al., 2022; Carouso-Peck et al., 2020; Carouso-Peck & Goldstein, 2019; Chen et al., 2016; Derégnaucourt, 2011; Houx & ten Cate, 1999; Tchernichovski et al., 1999). These studies did therefore not address the question as to whether the multimodal nature of the song stimulus is perhaps at the base of improved vocal learning. The level of stimulus control in two modalities offered by the RoboFinch will allow future studies to investigate additive or interactive effects of passive versus active (operant) and unimodal versus multimodal (synchronised and asynchronised) exposure to song, to investigate their relative contribution to stimulus valence. Clearly, the technological advances that now allow us to have multimodally faithful representations of animal signals, as exemplified with the RoboFinch, will enable such research.

Over the past two decades, the development of robotic animal models has seen a steady increase (Romano et al., 2019), which has provided unique insights in the field of animal behaviour ranging from collective behaviour, to aggression and communication (e.g. Marras & Porfiri, 2012; Narins et al., 2003; Patricelli & Krakauer, 2010; Romano et al., 2017; Taylor & Ryan, 2013). Robotic models have been successfully applied to a wide range of taxa, including insect, fish, frogs, birds, reptiles and mammals, both in the field and in the lab (Faria et al., 2010; Fernández-Juricic et al., 2006; Martins et al., 2005; Partan et al., 2009; Taylor et al., 2008). Previous studies have so far staged short-term social interactions between robots and animals, either at the individual or the group level (Romano et al., 2019). In birds, this technique has been used to expose receivers to combined sounds and crude body movements, such as flapping the wings up and down or rotating the body or the head (Araguas et al., 2022; Fernández-Juricic et al., 2006; Reş & Magrath, 2017).

We provide our setup open-source to stimulate the use of robotics in avian research, ranging from topics as wide as social behaviour, mate choice, to learning and cognition. With some adjustments to the 3D shape files and paint mixtures, our setup can be easily adjusted for use in other similar-sized (song)bird species. As long as species-specific colouration patterns are used that match avian vision, our setup can easily function in many experiments addressing the importance of multimodal signals and cues in sexual and social behaviour, such as mate choice or staged aggression playback experiments. Our setup can even be used in the field with some adjustments to run it on battery power, although the fine-scale mechanics driving the beak movements may quickly deteriorate under harsh environmental conditions. But for many questions, a robotic model that can rotate the body and head may be suitable enough. Our setup can also be used outside the context of multimodal perception and communication in studies that require life-like moving models. For example, a life-like moving robot may increase trapping success of birds in the field, which can be particularly important

in conservation studies aimed at tracking rare, hard to catch birds, throughout their lives. Furthermore, our robot setup may be used to study the potential impact of social isolation by providing birds a life-like companion. Although speculative, we found tutees housed alone to call out more, albeit not statistically significant, compared to all other treatments, in which they were either housed with a female or a robotic companion.

We specifically aimed to develop and apply a robotic setup that allowed: (1) for fine-scale movements, and (2) for long-term experiments, with the aim to study the development of behaviour such as vocal communication through imitation learning in songbirds. The high level of (multimodal) stimulus control of our RoboFinches also allows us for example to address unresolved questions in vocal and sensory learning. It is as yet unclear whether during the sensory learning phase, juveniles only memorise the acoustic properties (i.e. the specific spectral-temporal features) of song, or also see and memorise some of the vocal gestures belonging to the associated motor programs, and if so, from which stage of the acquisition process onwards (Ohms et al., 2010; Zollinger & Suthers, 2004). The production of specific elements requires specific postures of the vocal apparatus, such as the beak gape and expanding the oropharyngeal-oesophageal cavity (Beckers et al., 2003; Suthers & Zollinger, 2004; Williams, 2001). Failing to match the correct vocal posture with the correct song element may result in suboptimal production, which can possibly be picked up by receivers, thereby reducing the singer's attractiveness (Suthers, 2004). Our tutoring experiment suggests that juvenile zebra finches, like human babies (Kuhl & Meltzoff, 1982), appear to pay specific attention to the song motor program, as they would pause their own singing behaviour when the RoboFinch simultaneously sang and moved, although we cannot rule out that they simply paid attention to any movement. Non-moving or not song aligned moving tutor models have been used before in song tutoring experiments (Araguas et al., 2022; Benichov et al., 2016; Tchernichovski et al., 2001), however, these previous setups did not allow to assess the importance of concurring audio-visual cues, whereas our setup has at least strong potential to do so.

In conclusion, creating controlled levels (in quantity and quality) of realistic stimulus representation for multiple sensory modalities has been a major issue for the systematic study of multisensory perception and multimodal communication. Studies assessing the perceptual integration of audio-visual stimuli for example often suffer from impoverished stimulus representation in one sensory domain, for example when using video screens, which lack 3D cues, and typically do not match colour or motion perception of the study system (Chouinard-Thuly et al., 2017). Here we have described how we developed a life-like robotic model of a singing zebra finch, which can be used to study all sorts of behavioural and developmental processes associated with animal communication. Our song tutoring experiment demonstrated that juveniles can readily learn to copy songs from the RoboFinch, which opens up future directions to study the importance of multimodal signals and cues in learning and memory formation, specifically in the context of song copying and vocal development. We argue that robotic setups provide unique

possibilities to create artificial stimuli that are not present in nature, but when combined with clever experimental design can provide crucial insight into the processing and functioning of a wide range of animal behaviours.

AUTHOR CONTRIBUTIONS

Wouter Halfwerk, Constance Scharff and Katharina Riebel conceived the experiment; Judith Varkevisser, Ralph Simon, Katharina Riebel, Constance Scharff and Wouter Halfwerk designed the experiment; Ralph Simon, Peter G. Wiersma, Klaus Hochradel and Rogier Elsinga designed, constructed and programmed the RoboFinch; Judith Varkevisser, Ralph Simon, Wouter Halfwerk, Esmee Middelburg and Eva Zoeter analysed the data; Judith Varkevisser, Ralph Simon, Ezequiel Mendoza, Constance Scharff, Katharina Riebel and Wouter Halfwerk discussed results and wrote the paper; All authors contributed critically to the drafts and gave final approval for publication.

ACKNOWLEDGEMENTS

We greatly acknowledge the help of Dré Kampfraath and Wesley Delmeer during the development of the RoboFinch. This work was funded by the Human Frontier Science Program (No RGP0046/2016).

CONFLICT OF INTEREST STATEMENT

The authors declare no competing interests.

PEER REVIEW

The peer review history for this article is available at <https://www.webofscience.com/api/gateway/wos/peer-review/10.1111/2041-210X.14063>.

DATA AVAILABILITY STATEMENT

Data and files to construct the RoboFinch can be found on GitHub|Zenodo (<https://github.com/SimonHubs/RoboFinch>|<https://doi.org/10.5281/zenodo.7520589>). Data on juvenile behaviour are published on Zenodo (<https://doi.org/10.5281/zenodo.7525190>).

ORCID

Ralph Simon  <https://orcid.org/0000-0002-6083-0394>

Judith Varkevisser  <https://orcid.org/0000-0001-7706-1520>

Ezequiel Mendoza  <https://orcid.org/0000-0003-4963-519X>

Klaus Hochradel  <https://orcid.org/0000-0003-4695-5809>

Constance Scharff  <https://orcid.org/0000-0002-5792-076X>

Katharina Riebel  <https://orcid.org/0000-0003-2373-8510>

Wouter Halfwerk  <https://orcid.org/0000-0002-4111-0930>

REFERENCES

- Adret, P. (2004). Vocal imitation in blindfolded zebra finches (*Taeniopygia guttata*) is facilitated in the presence of a non-singing conspecific female. *Journal of Ethology*, 22, 29–35.
- Araguas, A., Guellà, B., Gauthier, P., Richer, F., Montone, G., Chopin, A., & Derégnaucourt, S. (2022). Design of a robotic zebra finch for experimental studies on developmental song learning. *Journal of Experimental Biology*, 225(3), jeb242949. <https://doi.org/10.1242/jeb.242949>

- Beckers, G. J., Nelson, B. S., & Suthers, R. A. (2004). Vocal-tract filtering by lingual articulation in a parrot. *Current Biology*, *14*, 1592–1597.
- Beckers, G. J., Suthers, R. A., & ten Cate, C. (2003). Pure-tone birdsong by resonance filtering of harmonic overtones. *Proceedings of the National Academy of Sciences of the United States of America*, *100*, 7372–7376.
- Benichov, J. I., Benezra, S. E., Vallentin, D., Globerson, E., Long, M. A., & Tchernichovski, O. (2016). The forebrain song system mediates predictive call timing in female and male zebra finches. *Current Biology*, *26*, 309–318.
- Bolhuis, J. J., Okanoya, K., & Scharff, C. (2010). Twitter evolution: Converging mechanisms in birdsong and human speech. *Nature Reviews Neuroscience*, *11*, 747–759.
- Bolhuis, J. J., Van Mil, D. P., & Houx, B. B. (1999). Song learning with audiovisual compound stimuli in zebra finches. *Animal Behaviour*, *58*, 1285–1292.
- Carouso-Peck, S., & Goldstein, M. H. (2019). Female social feedback reveals non-imitative mechanisms of vocal learning in zebra finches. *Current Biology*, *29*, 631–636.e633.
- Carouso-Peck, S., Menyhart, O., DeVoogd, T. J., & Goldstein, M. H. (2020). Contingent parental responses are naturally associated with zebra finch song learning. *Animal Behaviour*, *165*, 123–132.
- Chen, Y., Matheson, L. E., & Sakata, J. T. (2016). Mechanisms underlying the social enhancement of vocal learning in songbirds. *Proceedings of the National Academy of Sciences of the United States of America*, *113*, 6641–6646.
- Chouinard-Thuly, L., Gierszewski, S., Rosenthal, G. G., Reader, S. M., Rieucan, G., Woo, K. L., Gerlai, R., Tedore, C., Ingley, S. J., Stowers, J. R., Frommen, J. G., Dolins, F. L., & Witte, K. (2017). Technical and conceptual considerations for using animated stimuli in studies of animal behavior. *Current Zoology*, *63*, 5–19.
- Derégnaucourt, S. (2011). Birdsong learning in the laboratory, with especial reference to the song of the zebra finch (*Taeniopygia guttata*). *Interaction Studies*, *12*, 324–350.
- Derégnaucourt, S., Poirier, C., Van der Kant, A., Van der Linden, A., & Gahr, M. (2013). Comparisons of different methods to train a young zebra finch (*Taeniopygia guttata*) to learn a song. *Journal of Physiology-Paris*, *107*, 210–218.
- Doupe, A. J., & Kuhl, P. K. (1999). Birdsong and human speech: Common themes and mechanisms. *Annual Reviews in Neurosciences*, *22*, 567–631.
- Eales, L. A. (1989). The influences of visual and vocal interaction on song learning in zebra finches. *Animal Behaviour*, *37*(507), 508.
- Elemans, C. P. H. (2014). The singer and the song: The neuromechanics of avian sound production. *Current Opinion in Neurobiology*, *28*, 172–178.
- Faria, J. J., Dyer, J. R., Clément, R. O., Couzin, I. D., Holt, N., Ward, A. J., et al. (2010). A novel method for investigating the collective behaviour of fish: Introducing 'Robofish'. *Behavioral Ecology and Sociobiology*, *64*, 1211–1218.
- Fernández-Juricic, E., Gilak, N., McDonald, J. C., Pithia, P., & Valcarcel, A. (2006). A dynamic method to study the transmission of social foraging information in flocks using robots. *Animal Behaviour*, *71*, 901–911.
- Fröhlich, M., & van Schaik, C. P. (2018). The function of primate multimodal communication. *Animal Cognition*, *21*, 619–629.
- Gobes, S. M., & Bolhuis, J. J. (2007). Birdsong memory: A neural dissociation between song recognition and production. *Current Biology*, *17*, 789–793.
- Goller, F., Mallinckrodt, M. J., & Torti, S. D. (2004). Beak gape dynamics during song in the zebra finch. *Journal of Neurobiology*, *59*, 289–303.
- Gomes, D. G. E., Page, R. A., Geipel, I., Taylor, R. C., Ryan, M. J., & Halfwerk, W. (2016). Bats perceptually weight prey cues across sensory systems when hunting in noise. *Science*, *353*(1277), 1280.
- Griffith, S. C., Ton, R., Hurley, L. L., McDiarmid, C. S., & Pacheco-Fuentes, H. (2021). The ecology of the zebra finch makes it a great laboratory model but an outlier amongst passerine birds. *Birds*, *2*, 60–76.
- Halfwerk, W., Page, R. A., Taylor, R. C., Wilson, P. S., & Ryan, M. J. (2014). Crossmodal comparisons of signal components allow for relative distance assessment. *Current Biology*, *24*, 1751–1755.
- Halfwerk, W., Varkevisser, J., Simon, R., Mendoza, E., Scharff, C., & Riebel, K. (2019). Toward testing for multimodal perception of mating signals. *Frontiers in Ecology and Evolution*, *7*, 124. <https://doi.org/10.3389/fevo.2019.00124>
- Hebets, E. A., & Papaj, D. R. (2005). Complex signal function: Developing a framework of testable hypotheses. *Behavioral Ecology and Sociobiology*, *57*, 197–214.
- Heuschele, J., Mannerla, M., Gienapp, P., & Candolin, U. (2009). Environment-dependent use of mate choice cues in sticklebacks. *Behavioral Ecology*, *20*, 1223–1227.
- Higham, J., & Hebets, E. (2013). An introduction to multimodal communication. *Behavioral Ecology and Sociobiology*, *67*, 1381–1388.
- Hoese, W. J., Podos, J., Boetticher, N. C., & Nowicki, S. (2000). Vocal tract function in birdsong production: Experimental manipulation of beak movements. *Journal of Experimental Biology*, *203*, 1845–1855.
- Houx, A. B., & ten Cate, C. (1999). Song learning from playback in zebra finches: Is there an effect of operant contingency? *Animal Behaviour*, *57*, 837–845.
- Kuhl, P. K., & Meltzoff, A. N. (1982). The bimodal perception of speech in infancy. *Science*, *218*, 1138–1141.
- Leavell, B. C., Rubin, J. J., McClure, C. J., Miner, K. A., Branham, M. A., & Barber, J. R. (2018). Fireflies thwart bat attack with multisensory warnings. *Science Advances*, *4*, eaat6601.
- Liu, W.-C., Landstrom, M., Schutt, G., Insera, M., & Fernandez, F. (2021). A memory-driven auditory program ensures selective and precise vocal imitation in zebra finches. *Communications Biology*, *4*, 1–9.
- Loning, H., Griffith, S. C., & Naguib, M. (2022). Zebra finch song is a very short-range signal in the wild: Evidence from an integrated approach. *Behavioral Ecology*, *33*, 37–46.
- Marras, S., & Porfiri, M. (2012). Fish and robots swimming together: Attraction towards the robot demands biomimetic locomotion. *Journal of the Royal Society Interface*, *9*, 1856–1868.
- Martins, E. P., Ord, T. J., & Davenport, S. W. (2005). Combining motions into complex displays: Playbacks with a robotic lizard. *Behavioral Ecology and Sociobiology*, *58*, 351–360.
- McGurk, H., & Macdonald, J. (1976). Hearing lips and seeing voices. *Nature*, *264*, 746–748.
- Munoz, N. E., & Blumstein, D. T. (2020). Optimal multisensory integration. *Behavioral Ecology*, *31*, 184–193.
- Narins, P. M., Hodl, W., & Grabul, D. S. (2003). Bimodal signal requisite for agonistic behavior in a dart-poison frog, *Epipedobates femoralis*. *Proceedings of the National Academy of Sciences of the United States of America*, *100*, 577–580.
- Nelson, D. A. (1997). Social interaction and sensitive phases for song learning: A critical review. In C. T. Snowdon & M. Hausberger (Eds.), *Social influences on vocal development* (pp. 7–22). Cambridge University Press.
- Nowicki, S. (1987). Vocal tract resonances in oscine bird sound production: Evidence from birdsongs in a helium atmosphere. *Nature*, *325*, 53–55.
- Ohms, V. R., Snelderwaard, P. C., ten Cate, C., & Beckers, G. J. L. (2010). Vocal tract articulation in zebra finches. *PLoS One*, *5*, e11923.
- Partan, S., & Marler, P. (1999). Behavior–Communication goes multimodal. *Science*, *283*, 1272–1273.
- Partan, S. R., Larco, C. P., & Owens, M. J. (2009). Wild tree squirrels respond with multisensory enhancement to conspecific robot alarm behaviour. *Animal Behaviour*, *77*, 1127–1135.
- Patricelli, G. L., & Krakauer, A. H. (2010). Tactical allocation of effort among multiple signals in sage grouse: An experiment with a robotic female. *Behavioral Ecology*, *21*, 97–106.

- Pfenning, A. R., Hara, E., Whitney, O., Rivas, M. V., Wang, R., Roulhac, P. L., Howard, J. T., Wirthlin, M., Lovell, P. V., Ganapathy, G., Mountcastle, J., Moseley, M. A., Thompson, J. W., Soderblom, E. J., Iriki, A., Kato, M., Gilbert, M. T. P., Zhang, G., Bakken, T., ... Jarvis, E. D. (2014). Convergent transcriptional specializations in the brains of humans and song-learning birds. *Science*, *346*, 1256846.
- Podos, J., Lahti, D. C., & Moseley, D. L. (2009). Chapter 5: Vocal performance and sensorimotor learning in songbirds. In M. Naguib, K. Zuberbühler, N. S. Clayton, & V. M. Janik (Eds.), *Advances in the study of behavior* (Vol. 40, pp. 159–195). Academic Press. [https://doi.org/10.1016/S0065-3454\(09\)40005-6](https://doi.org/10.1016/S0065-3454(09)40005-6)
- Podos, J., Southall, J. A., & Rossi-Santos, M. R. (2004). Vocal mechanics in Darwin's finches: Correlation of beak gape and song frequency. *Journal of Experimental Biology*, *207*, 607–619.
- Pytte, C. L., & Suthers, R. A. (1999). A bird's own song contributes to conspecific song perception. *Neuroreport*, *10*, 1773–1778.
- Reş, P., & Magrath, R. D. (2017). Deceptive vocal duets and multimodal display in a songbird. *Proceedings of the Royal Society B: Biological Sciences*, *284*, 20171774.
- Riede, T., & Goller, F. (2010). Peripheral mechanisms for vocal production in birds—Differences and similarities to human speech and singing. *Brain and Language*, *115*, 69–80.
- Romano, D., Benelli, G., Donati, E., Remorini, D., Canale, A., & Stefanini, C. (2017). Multiple cues produced by a robotic fish modulate aggressive behaviour in Siamese fighting fishes. *Scientific Reports*, *7*, 1–11.
- Romano, D., Donati, E., Benelli, G., & Stefanini, C. (2019). A review on animal-robot interaction: From bio-hybrid organisms to mixed societies. *Biological Cybernetics*, *113*, 201–225.
- Simon, R., Varkevisser, J., Mendoza, E., Hochradel, K., Scharff, C., Riebel, K., et al. (2019). Development and application of a robotic zebra finch (RoboFinch) to study multimodal cues in vocal communication. *PeerJ Preprints*, *7*, e28004v2.
- Suthers, R. A. (2004). How birds sing and why it matters. In P. R. Marler & H. Slabbekoorn (Eds.), *Nature's music: The science of birdsong* (pp. 272–295). Elsevier Academic Press.
- Suthers, R. A., & Zollinger, S. A. (2004). Producing song: The vocal apparatus. *Annals of the New York Academy of Sciences*, *1016*, 109–129.
- Taylor, R. C., Klein, B. A., Stein, J., & Ryan, M. J. (2008). Faux frogs: Multimodal signalling and the value of robotics in animal behaviour. *Animal Behaviour*, *76*, 1089–1097.
- Taylor, R. C., & Ryan, M. J. (2013). Interactions of multisensory components perceptually rescue túngara frog mating signals. *Science*, *341*, 273–274.
- Tchernichovski, O., Lints, T., Mitra, P. P., & Nottebohm, F. (1999). Vocal imitation in zebra finches is inversely related to model abundance. *Proceedings of the National Academy of Sciences of the United States of America*, *96*, 12901–12904.
- Tchernichovski, O., Mitra, P. P., Lints, T., & Nottebohm, F. (2001). Dynamics of the vocal imitation process: How a zebra finch learns its song. *Science*, *291*, 2564–2569.
- Uetz, G. W., Clark, D. L., Roberts, J. A., & Rector, M. (2011). Effect of visual background complexity and light level on the detection of visual signals of male *Schizocosa ocreata* wolf spiders by female conspecifics. *Behavioral Ecology and Sociobiology*, *65*, 753–761.
- Ullrich, R., Norton, P., & Scharff, C. (2016). Waltzing *Taeniopygia*: Integration of courtship song and dance in the domesticated Australian zebra finch. *Animal Behaviour*, *112*, 285–300.
- Varkevisser, J. M., Simon, R., Mendoza, E., How, M., van Hijlkema, I., Jin, R., et al. (2021). Adding colour-realistic video images to audio playbacks increases stimulus engagement but does not enhance vocal learning in zebra finches. *Animal Cognition*, *25*, 249–274.
- Westneat, M. W., Long, J., Hoese, W., & Nowicki, S. (1993). Kinematics of birdsong: Functional correlation of cranial movements and acoustic features in sparrows. *Journal of Experimental Biology*, *182*, 147–171.
- White, S. A. (2001). Learning to communicate. *Current Opinion in Neurobiology*, *11*, 510–520.
- Williams, H. (2001). Choreography of song, dance and beak movements in the zebra finch (*Taeniopygia guttata*). *Journal of Experimental Biology*, *204*, 3497–3506.
- Zollinger, S. A., & Suthers, R. A. (2004). Motor mechanisms of a vocal mimic: implications for birdsong production. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *271*, 483–491.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

Appendix S1. Supplementary methods, figures and tables.

Video S1. Comparison of a singing zebra finch (right) and a RoboFinch (left) playing the same song and the movements deduced from this bird.

Video S2. Juvenile male practising in front of the RoboFinch.

Video S3. Different phases of the RoboFinch project. High speed video recordings, deducing beak and head movements from videos, playing the deduced files on the robot, juvenile male listening to the singing RoboFinch and juvenile male and female interacting with the RoboFinch.

How to cite this article: Simon, R., Varkevisser, J., Mendoza, E., Hochradel, K., Elsinga, R., Wiersma, P. G., Middelburg, E., Zoeter, E., Scharff, C., Riebel, K., & Halfwerk, W. (2023). RoboFinch: A versatile audio-visual synchronised robotic bird model for laboratory and field research on songbirds. *Methods in Ecology and Evolution*, *14*, 1092–1103. <https://doi.org/10.1111/2041-210X.14063>