The Company of Biologists

METHODS & TECHNIQUES

Design of a robotic zebra finch for experimental studies on developmental song learning

Alice Araguas[1], Bahia Guellaï[1,2], Philippe Gauthier[3,*], Florian Richer[3,*], Guglielmo Montone[4], Adrien Chopin[5] and Sébastien Derégnaucourt[1,2,‡]

ABSTRACT

Birdsong learning has been consolidated as the model system of choice for exploring the biological substrates of vocal learning. In the zebra finch (*Taeniopygia guttata*), only males sing and they develop their song during a sensitive period in early life. Different experimental procedures have been used in the laboratory to train a young finch to learn a song. So far, the best method to achieve a faithful imitation is to keep a young bird singly with an adult male. Here, we present the different characteristics of a robotic zebra finch that was developed with the goal to be used as a song tutor. The robot is morphologically similar to a life-sized finch: it can produce movements and sounds contingently to the behaviours of a live bird. We present preliminary results on song imitation, and other possible applications beyond the scope of developmental song learning.

KEY WORDS: Animal behaviour, Robotic bird, Vocal contingency, Multimodal communication, Birdsong, Developmental learning, *Taeniopygia guttata*

INTRODUCTION

Vocal production learning is defined as 'instances where the signals themselves are modified in form as a result of experience with those of other individuals' (Janik and Slater, 2000). Humans share this ability with other animal species from few taxonomic groups including oscine songbirds. Birdsong has been established as a model of choice to study the behavioural, cellular and molecular aspects of vocal production learning (Aamodt et al., 2020). The zebra finch (*Taeniopygia guttata*) is considered as the 'lab mouse' of birdsong research. This tiny bird is very easy to breed in captivity; only males sing and they develop their short and stereotyped song (1–1.5 s) during a sensitive period of early life (25–90 days after hatching). They learn to sing by memorizing and imitating the songs of surrounding conspecifics, mainly adult males (see Derégnaucourt, 2011 for a review; Derégnaucourt and Gahr, 2013). Since the pioneering works of Immelmann (1969), thousands of zebra finches have been reared in controlled environments to examine how their experience affects the development of their song.

Birds raised in social isolation during the sensitive phase of learning develop abnormal songs that still contain sounds characteristic of the species (Price, 1979). The importance of visual and physical contacts on song learning has also been evidenced; young birds separated from their tutor by an opaque partition or a grid produce a worse imitation than birds that can physically interact with it (Derégnaucourt, 2011). Different experimental procedures were used in the laboratory to train young zebra finches to learn a song. Unlike other species of songbirds, passive exposure of zebra finch song recordings results in a poor imitation of song models (Derégnaucourt et al., 2013). Self-eliciting exposure to the song model, using an operant conditioning procedure, induces significant learning but with a high inter-individual variability: if some birds learn significantly (some even produce a very faithful copy of the song model broadcast), others show no sign of learning (Derégnaucourt et al., 2005, 2013). Inter-individual variability may have advantages for the experimenter. For example, we have shown a relationship between song learning and vocal plasticity during sleep: birds that showed stronger post-sleep deterioration during development achieved a better final imitation (Derégnaucourt et al., 2005). But it also has disadvantages. Some experimental procedures are complex (e.g. pinealectomy: Derégnaucourt et al., 2012; gene transfection: Haesler et al., 2007), and regulations regarding animal experimentation require the reduction of sample sizes. In the zebra finch, the best way to obtain a close-to-perfect copy of the song is to raise a young male with an adult male (Derégnaucourt et al., 2013). This type of learning is of course exceptional; in natural conditions, young males learn by imitating song mainly from their father, but also by copying parts of songs of other singing males living in their close environment (Derégnaucourt and Gahr, 2013). This young–adult dyadic situation used to obtain a close-to-perfect imitation of the song is a reference for birdsong research (Derégnaucourt, 2011). However, this method does not allow precise control of the different variables involved in song learning (e.g. control of the behaviour of the tutor, number of songs heard, etc.). If this method stresses the importance of social factors on song learning in the zebra finch, one cannot answer questions about the nature of these factors: are they visual or multimodal, or is this result due to attentional phenomena?

One solution is to broadcast videos of singing adult males to young finches, but so far, this does not seem reliable enough to obtain a good song copy in pupils (Deshpande et al., 2014; Ljubičić et al., 2016). Over the years, the use of robotic models has increased in animal behaviour studies with promising results in several animal species including birds. For example, in the field, male satin bowerbirds (*Ptilonorhynchus violaceus*) adjusted their courtship displays in response to different postures of robotic females (Patricelli et al., 2002). In the laboratory, a robot was used to study social attachment in young quail chicks (Jolly et al., 2016). So far, these robots were used as pre-programmed devices without contingency

[1]Laboratoire Ethologie Cognition Développement, Université Paris Nanterre, Université Paris Lumières, 92000 Nanterre, France. [2]Institut Universitaire de France, 75005 Paris, France. [3]Institut des Systèmes Intelligents et de Robotique, CNRS UMR 7222, Sorbonne Université, 75005 Paris, France. [4]Laboratoire Psychologie de la Perception, CNRS UMR 8242, Université Paris Descartes, 75006 Paris, France. [5]Sorbonne Université, Institut de la Vision, 75012 Paris, France.
*These authors contributed equally to this work

‡Author for correspondence (sebastien.deregnaucourt@parisnanterre.fr)

P.G., 0000-0002-5039-6985; A.C., 0000-0002-9289-7964; S.D., 0000-0003-0160-1986

with the behaviour of the interacting live birds. The importance of such contingency in the context of birdsong learning has already been evidenced in oscine songbirds including zebra finches. In a recent study, Carouso-Peck and Goldstein (2019) demonstrated that song imitation is influenced by a non-vocal, visual feedback given by female finches and known as a response to attractive songs. When a contingent feedback was experimentally given to young males while they were singing (the broadcast of a video of a female producing the posture), they further develop better imitations of the tutor song heard during their early life than their brother exposed to the same video non-contingently (Carouso-Peck and Goldstein, 2019). Such contingency might be necessary for the robot to be accepted as a valid tutor in the context of birdsong learning.

In order to investigate these aspects, we developed a robotic zebra finch that sings, calls, moves and reacts to the behaviour of a live zebra finch. We present also potential applications beyond the scope of birdsong learning.

MATERIALS AND METHODS
Design of the robot
The robot (named 'MANDABOT') was designed to allow for the following movements: rotation of the body around a vertical axis, rotation of the head, tail wagging and beak opening. We used taxidermic mounts of adult male zebra finches to design both the size and the shape of the different body parts.

We printed them with Multi Jet Printing on a Projet 3510SD from 3DSystems, which uses UV lights to cure parts from a liquid resin (Visijet M3 Black). MANDABOT was hand painted with mixes of Prince August paints (France) to match the colours of the plumage of an adult male zebra finch. The body is actuated by a DC motor with an encoder from Maxon. The coils and the body motor are controlled via a TB6612FNG Dual Motor Driver Carrier board from Pololu. The head is actuated by an AM1020 stepper motor from Faulhaber, which is controlled via a TMC2130 SilentStepStick driver from Trinamic. Movements of the beak and the tail are controlled by coils and magnets obtained from commercial versions of DigiBirds (Silverlit Toys Manufactory, Hong Kong, China). A three-axis accelerometer (ADXL362) permits the detection of contacts between the bird and the robot. The robot is fixed on a grey plastic box (16×8.5×8 cm). On each side of the robot, there is a perch (5 cm) connected to load cells (CZL639HD from Phidgets). The robot is equipped with a speaker (AS01808MR from PUI Audio) to broadcast pre-recorded zebra finch songs and calls. We designed the robot control using a client/server approach over Wi-Fi. We developed a custom server program running on a Beaglebone Black Wireless (physically close to the robot) and a custom client program running a laptop Dell Latitude 5300 computer running on Microsoft Windows 10. The server program handles the low-level control – move physically every actuator of the robot, read positions and rotations, detect physical events and play sounds – and is listening for commands using HTTP protocol and JSON format. The client program connects to the server and handles the high-level control: ask for positions and rotations, perch states, react, send commands, and record events, movements and sound. This approach allows more modularity and complex control over the robot (one laptop could control several robots), while using the more powerful CPU of the laptop than the Beaglebone embedded with the physical robot. We programmed the following behaviours: when the bird lands on the perch, the robot turns in his direction and randomly moves its head or broadcasts a song; when the bird leaves the perch, the body of the robot turns back to the center in a neutral position. Both closed-loop and open-loop programs can be launched for a precise period of time (e.g. 60 min). At the end of the loop, the computer generates a script containing all the sequences of movements and vocalizations produced by the robot. A script recorded in a closed-loop interaction context between the robot and a bird can then be run in a non-interactive mode with another bird. For the preliminary experiments presented below, the robot was placed in an experimental cage (46×23×27 cm) in a sound-proof chamber (103×56.5×65 cm).

Interactive vocal loops between the bird and the robot
We developed a closed-loop program to mimic the vocal interactions between two finches. Such bird–machine vocal interactions have already been successfully applied in oscine songbirds including zebra finches (Lerch et al., 2011; Benichov et al., 2016). Sounds from the experimental cage are continuously recorded using a Behringer C-2 microphone connected to a PreSonus Audio Box 1818VSL (16 bits, 44.1 kHz). The audio stream is stored in a buffer of 180 ms, which corresponds roughly to the maximum duration of a contact call produced by a zebra finch. In parallel, signal amplitude is calculated. If it exceeds the detection threshold, features of the input signal are computed. The value of this threshold must be set experimentally because it depends on the acoustic environment of the cage and also to the loudness of the vocalizing bird interacting with the robot. Because contact calls produced by finches are very brief (on average 110 ms; Ter Maat et al., 2014), the descriptor was therefore built on time windows of 32 samples (0.7 ms). Every 0.7 s, the signal received is convolved to a Hamming window, before calculating the logarithm of the corresponding spectrum as follows:

$$\text{feature}(t,:) = log|f\,ft(\text{signal})|, \tag{1}$$

where t is time and f is frequency. Before being introduced into the network, these features must be standardized. To do this, we resize the features at the minimum duration of the database calls (83 ms) and normalize each feature (115×17 matrix) between 0 and 1. The 2D image is then reorganized into a column vector (1×1955). To train the artificial neural network, we used 390 *.wav audio files, containing different calls of four different birds and noises produced by birds during movements (e.g. cage noises, wing flaps).

Following Rassak et al. (2016), the neural network used is a multi-layer perceptron with one hidden layer composed of 300 neurons with two neurons in the output layer corresponding to one of two possible categories: bird call or cage noise. If the sound belongs to the first category, then the system broadcasts a pre-recorded call from the speaker placed inside the robot. In a preliminary experiment based on 397 sounds, we obtained a correct recognition score of 0.9747 for bird calls. When a vocal sound produced by the bird (either a call or a song syllable) is detected by the system, a pre-recorded zebra finch call is broadcast through the speaker. The vocal response of the robot occurs with a minimum latency of 280 ms: 180 ms of features extraction followed by 100 ms of computation by the neural network. The duration of this latency, similar to those used by other research groups (e.g. Benichov et al., 2016), could also be experimentally increased if needed. The recorded sound (either a cage noise or a bird call) is saved as a *.wav file in a corresponding folder for eventual offline analysis or later use. An open-loop control is also possible: the system can send a call chosen randomly in the call database, at either a regular or a random interval.

Song learning experiment

Birds and housing

We used 36 males zebra finches from our colony maintained in the animal facilities of the University Paris Nanterre (14 h:10 h light: dark, at 20°C): 12 adult males (age >1 year) and 24 young males. Chicks were raised by their mother alone from 14 days post-hatching (dph) and kept in acoustic isolation from adult males to avoid any exposure to songs. Starting from 35 dph, young males were socially isolated in an individual cage (46×23×27 cm) inside a sound-proof chamber (85×65×60 cm) until 100 dph. Each cage contained three perches, and we placed a round mirror (diameter: 10 cm) above one of them to reduce the impact of social isolation. Birds had access to water, food, sand and cuttle bones *ad libitum*. Once a week this diet was supplemented with vegetables. After 2 days of habituation to the sound-proof chamber, each young finch was transferred to a cage containing a song tutor (either a live bird or a robot, see below) for 1 h, 5 days a week for five consecutive weeks. After exposure, each young finch was transferred to his housing cage. At day 1 of exposure, a grid placed in the middle of the cage separated the tutor from the pupil. Based on the results of previous experiments with a live tutor (Chen et al., 2016), we used this procedure to prevent eventual agonistic behaviours. At day 2, the grid was left only for the first 30 min of the session. Then we removed it to permit physical interactions between the young finch and the song tutor for the following 30 min. From day 3 to the end of the experiment, the tutor and the pupil could physically interact for the whole hour. At 100 dph, when song is usually crystallized, the experiment ended. All birds were then transferred to aviaries with conspecifics.

Twelve finches were trained with a live tutor (live group) and 12 other finches were trained with a robot (robot group). Regarding the live group, we used a different tutor for each bird.

In the robot group, the robot called, sang and moved in reaction to the behaviour of the pupil (see above). Songs and calls produced by live tutors in presence of the young finch (live group) were used as song models for the robot group. Sound files were prepared using Avisoft SASLab Pro; after segmenting the sound in the original *.wav file, we applied a high pass filter at 420 Hz and a volume maximization of 90%. Therefore, each young bird of the robot group was exposed to songs and calls produced by a different tutor of the live group. When the pupil landed on the perch, the robot turned in his direction and randomly moved its head or broadcast a song among the 20 of his specific tutor's repertoire. When the pupil produced a call, the robot replied with one out of 50 contact calls of the tutor's repertoire. When the pupil left the perch, the body of the robot turned back to the centre in a neutral position. Following the same experimental schedule as in the live group, we launched a closed-loop program for 60 min for each pupil of the robot group.

Song recording and analysis

During the whole experiment, the vocal activity of the birds was continuously recorded using Sound Analysis Pro software (SAP). Each cage was equipped with a Behringer C2-microphone connected to a Presonus Audio Box 1818 VSL (24 bits, 96 kHz) controlled by a DELL Optilex GX620 PC on Windows 7. Sounds were saved as *.wav files.

To measure song imitation, we selected sound files containing songs produced at 100 dph for the pupils and during the last days of the experiment for the tutors of the live group. Using Goldwave software (v6.36), we applied a high pass filter at 420 Hz (Lachlan et al., 2016) and a volume maximization of 90% to sound files. Song similarity was quantified using an automated procedure implemented in SAP that parametrically quantifies the similarity between songs. First, song segments were selected after a visual inspection of the spectrograms, by cutting the sound at the onset and offset of the selected sequence, using Sound Explorer (René Jansen, Amsterdam). Second, regions of high similarity between the segments of the pupil and the model songs were identified, and the results were aggregated into a global measure of acoustic similarity and sequence similarity. In asymmetric comparisons, the most similar sound elements of two sequences (tutor song and pupil song) are compared, independent of their position within the sequence. The smallest unit of comparison is the 9.26 ms sound interval (fast Fourier transform windows). Each interval is characterized by measures of five acoustic features: pitch, frequency modulation (FM), amplitude modulation (AM), Wiener entropy and pitch goodness. SAP (Sound Analysis Pro.; http://soundanalysispro.com/) calculates the Euclidean distance between all interval pairs from two songs, over the course of the motif, and determines a P-value for each interval pair. This P-value is based on P-value estimates derived from the cumulative distribution of Euclidean distances across 250,000 sound interval pairs, obtained from 25 random pairs of zebra finch songs. Neighbouring intervals that pass the threshold P-value (P=0.05 in this study, default value of SAP) form larger similarity segments (70 ms). In asymmetric similarity measurements, the aim is to judge how good the copy is in reference to the song model. The song model is loaded as 'sound 1' and the copies are loaded as 'sound 2' in the batch module of the SAP software. Therefore, in our study, song models (tutor songs) were loaded as 'sound 1' and pupil songs loaded as 'sound 2'. In summary, the amount of sound from the tutor song that is included into the similarity segments represents the similarity score; it thus reflects how much of the tutor's song material is found in the pupil song. This procedure was repeated 100 times, comparing 10 different exemplars of the tutor song with 10 different exemplars of each pupil's song. The mean value of these 100 comparisons was used for statistical analysis. Usually, the song segments chosen for such a comparison are song motifs of both the pupil and tutor. We also noticed that many young birds did not copy only the so-called song motif but also surrounding sounds, usually called introductory notes or connectors (Hyland Bruno and Tchernichovski, 2019). Therefore, we selected for each pupil the longest sequence of similar sounds with the tutor to calculate similarity scores.

Video recording and analysis

Physical interactions with the robot were video-recorded using a Logitech C920 webcam connected to ContaCam 4.9.9 software on an HP ProBook 650 G1 computer running on Windows 7. Days 1 and 2 of exposure were not analysed as a grid separated the pupil and the robot, preventing physical interactions between the bird and the robot for the whole session and for half of the session, respectively. We coded videos for day 2 (30 min session when physical interactions with the robot were possible), day 3 (the first time the bird could interact freely with the robot for 60 min), the first day of each following week (days 6, 11, 16, 21) and the last day of the experiment (day 25). We used an ethogram developed with the software BORIS v.7.9.16 (Table S1). We focused our analysis on the behaviours of the pupil towards the robot: physical interactions and reactions to song broadcast. Several behaviours were recorded: time spent in close vicinity of the robot (less than 10 cm) and number of pecks to different parts of the robot (beak, head, body, tail). Following Chen et al. (2016), we analysed pupils' reactions to song broadcast. We monitored five behavioural reactions: (1) stays still, (2) stands up, (3) moves on the platform, (4) flies away and

orientates his body towards the robot and (5) flies away and does not orientate his body towards the robot (see Movie 1).

Statistical analysis

Statistical analyses were conducted in R Studio v. 1.4.1103 and MATLAB R2017A. We checked distributions of data using Shapiro–Wilk tests. We used nonparametric tests or generalized linear models (GLMs) with non-normal distributions. To predict song similarity, we used a generalized linear mixed-effects models (GLMEs) with the tutor ID as random effects, two components of the number of songs heard – a linear and a quadratic one – and an interaction between the linear component and the number of songs heard, as fixed effects (reciprocal link). As a reminder, songs from each live tutor were used as song models for birds trained with MANDABOT, each bird being trained with a set of songs from a different tutor. In addition, we used song dissimilarity (100 – song

similarity, in %) as the dependent variable instead of song similarity, because song dissimilarity followed a more typical gamma distribution.

As durations on the platform encompassed durations in close vicinity that also included those when the bird was in physical contact with the robot ('clumping'), we preferred to use durations in an exclusive way. Exclusive platform durations were obtained by subtracting close vicinity durations from platform durations; exclusive close durations were obtained by subtracting clumping durations from close durations. All durations followed gamma distributions. To assess the evolution of the durations across days, we conducted a GLM with the day of experiment as a predictor. We assumed that durations followed a gamma distribution (adding 1 to all durations to avoid zero), and we modelled the relationship with the predictors using a reciprocal link. Including the tutor source in a GLME did not increase the Akaike's information criterion (AIC).
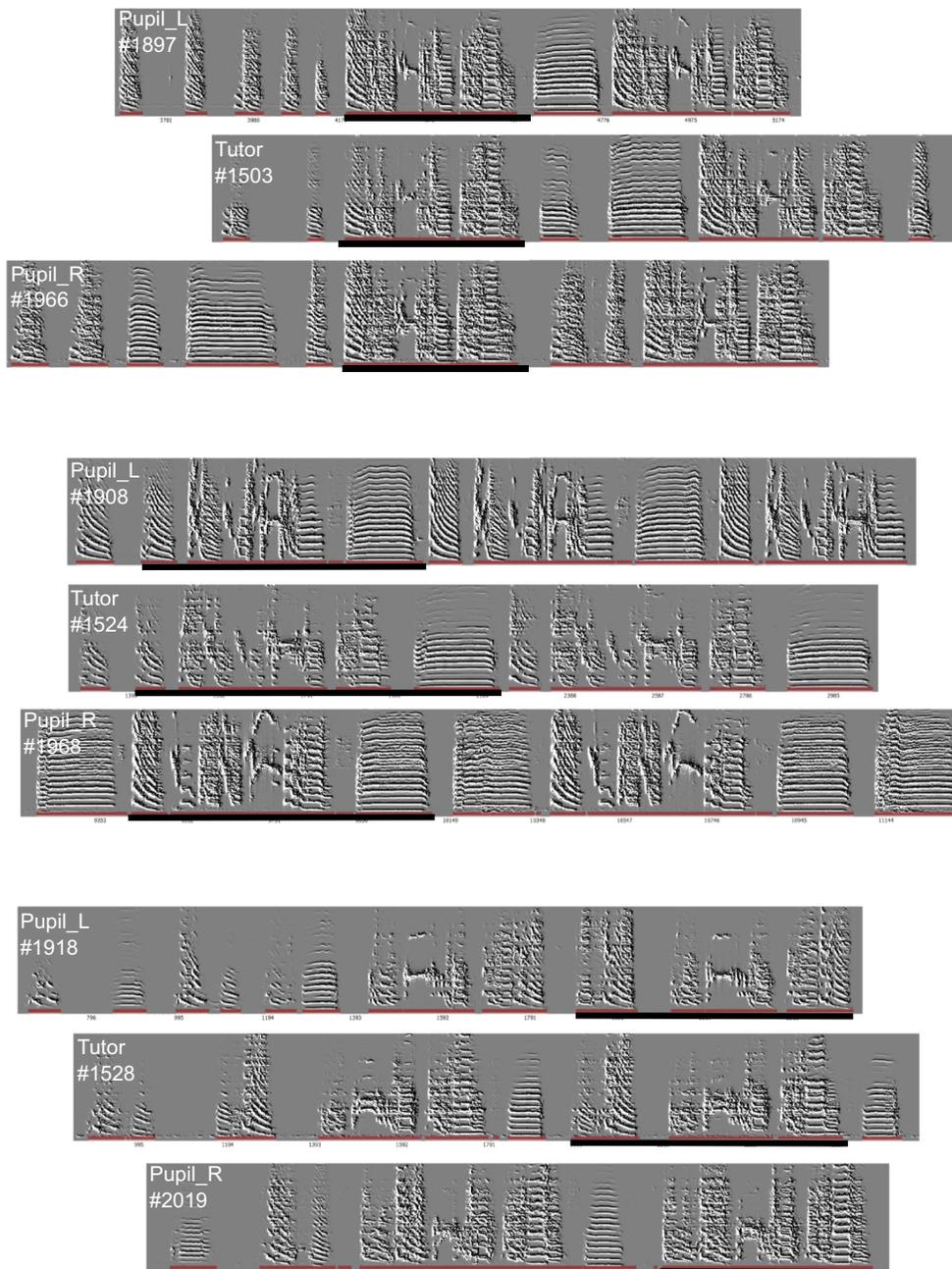

Fig. 1. Spectrograms of songs produced by three triads of birds, each triad being composed of a pupil trained with a live tutor (L, top) or MANDABOT (R, bottom). For each triad, the song model (tutor song) is presented in the middle.
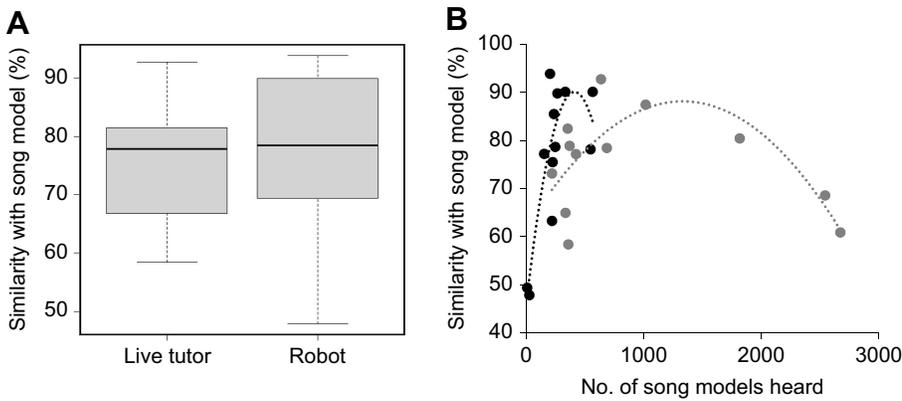
Fig. 2. MANDABOT was as effective as a live tutor for song imitation. (A) Boxplots of similarity to the song model for the live group (left) and the robot group (right). N=12 birds per group. There was no significant difference between the two groups regarding song imitation (Mann–Whitney, P=0.62). (B) Number of songs broadcast by MANDABOT (black rounds, second-order polynomial trend line, N=12 birds) or produced by the tutor (live group, grey rounds, second-order polynomial trend line, N=12 birds) in function to the similarity to the song model.

Regarding pecks on MANDABOT, we compiled pecks on the robot's beak with pecks on the head, and pecks on the robot's body with pecks on the tail. We estimated the mean number of pecks using separate GLMEs with the pupil as a random effect and the day (through a linear and a quadratic component) as a fixed effect. We assumed a Poisson distribution (because pecks are events) and a log link. We used a GLME to test whether the decreasing of the number of pecks is significant with the tutor source as a random effect.

For the analysis of behaviours expressed by the finches to song broadcast, we compiled 'stays still' and 'stands up' in a category called 'attentive reactions' as proposed by Chen et al. (2016). We used a GLME to test whether the number of attentive reactions decreased over the experiment. To characterize the evolution of the proportion of attentive reactions across days, we conducted a GLME with pupil as a random effect and day (both with a linear and a quadratic component) as a fixed effect. The proportion of attentive reactions did not follow a normal (Shapiro–Wilk test: W=0.84, P<0.001) or a clear gamma distribution. Therefore, we transformed it to the proportion of inattentive reactions (1–proportion of attentive reactions) plus one (to avoid zero values), and used a gamma distribution with a reciprocal link to model its evolution across days.

To characterize the evolution of the number of calls broadcast by MANDABOT in response to the finches' calls across days of experiment, we conducted a GLME with pupil as random effects and day (both with a linear and a quadratic component) as fixed effects. The number of calls did not follow a normal (Shapiro–Wilk test: W=0.65, P<0.001) but a clear gamma distribution. Therefore, we added one (to avoid zero values), and used a gamma distribution with a reciprocal link to model its evolution across days.

To check whether the different behavioural measures could predict song similarity to the model broadcast, we applied a GLM using a linear and quadratic predictor (normal distribution and identity link).

Ethical statement
All procedures followed the European regulations on animal experimentation and were approved by the Darwin Ethic Committee of the French Ministry for National Education, Higher Education and Research (authorization no. 20641201905141523153).

RESULTS AND DISCUSSION
There was no significant difference between the two groups regarding song imitation (Mann–Whitney, P=0.62; Figs 1, 2A). Similarity scores for the live tutor group were similar to previous published results, and scores obtained for the robot group were higher than those obtained with artificial methods used so far, such

as passive or active playback or the use of videos (e.g. Derégnaucourt et al., 2013, Varkevisser et al., 2021). Because the broadcast of tutor songs depended on the young's behaviour in the robot condition, we observed inter-individual variability in the number of songs heard (mean±s.e.m.=251.5±49). As previously explained, we used song dissimilarity, which depended on the number of songs heard, with a linear component (t_{20}=3.57, P=0.0019) and a quadratic component (t_{20}=−3.54, P=0.002), and the linear component interaction with the group (t_{20}=2.31, P=0.031; Fig. 2B). More specifically, the song similarity varied with the number of songs heard following an inverted U, and the top of the inverted U was reached faster in birds trained with MANDABOT than in the live tutor group. The two birds that produced a worse imitation were exposed to 7 and 26 exemplars, respectively, of the song models during the whole experiment. Nevertheless, as highlighted in previous studies (e.g. Deshpande et al., 2014), this
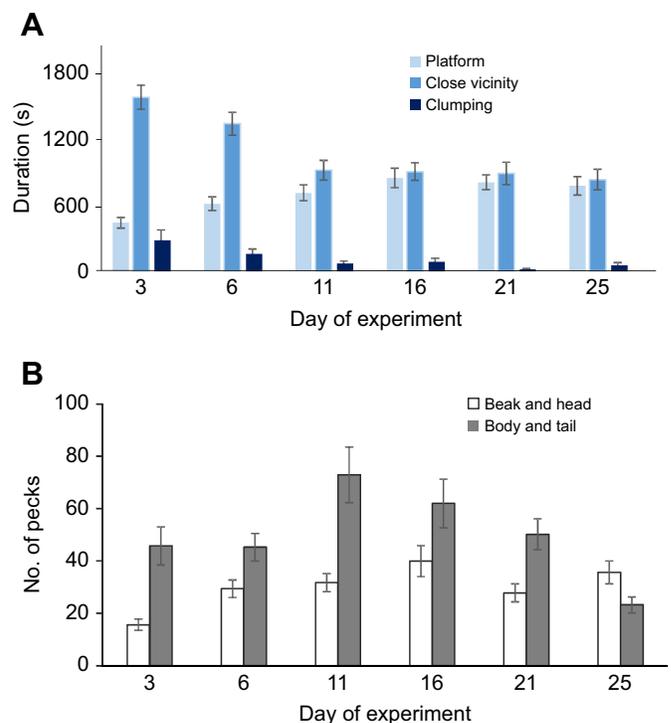

Fig. 3. Young finches exhibited an interest for MANDABOT during the whole experiment. (A) Time spent in different parts of the experimental setup: on the platform, in close vicinity to MANDABOT and in clumping (mean±s.e.m.). (B) Number of pecks given to different parts of MANDABOT: head–beak and body–tail (mean±s.e.m.).

low exposure was sufficient to trigger significant song learning, and model overabundance could have an inhibitory effect on song imitation (Tchernichovski et al., 1999).

During day 2 (the first day when physical interactions were possible with MANDABOT), pupils spent 70.65±4.92% (mean±s.e.m.) of the 30 min session on the platform and 64.57±5.39% in close vicinity. Regarding physical interactions, four birds out of 12 exhibited clumping behaviours with MANDABOT, and one of them spent 73.96% of the session in physical contact with the robot. Moreover, nine birds out of 12 pecked the robot and 57.19±2.38% of the total number of pecks were on the robot's body. These results suggest that most finches exhibited a strong interest for the robot from the first day when they could physically interact with it.

Over the experiment, the time spent on the platform increased with the experimental day ($t_{70}=-2.00$, $P=0.049$). However, the time spent exclusively in close vicinity and the time spent clumping each decreased ($t_{70}=3.08$, $P=0.0029$ and $t_{70}=-2.14$, $P=0.03$, respectively; Fig. 3A).

The number of pecks on the MANDABOT's head significantly increased until day 11 and then decreased over the experiment, in the shape of an inverted U (linear component: $t_{69}=4.38$, $P<0.001$, quadratic component: $t_{69}=-3.55$, $P<0.001$; Fig. 3B). Similarly, the number of pecks on the robot's body significantly increased until day 16 then decreased over the experiment, in an inverted U (linear component: $t_{69}=18$, $P<0.001$, quadratic component: $t_{69}=-19$, $P<0.001$).

Following a song produced by the robot, 'stays still' reactions were the most frequent ones (means±s.e.m.: 'stays still': 3.86±0.22; 'stands up': 1.95±0.15; 'moves': 0.53±0.11; 'flies away orientated': 1.83±0.18; 'flies away': 0.27±0.03).

The proportion of attentive reactions ('stays still' and 'stands up' reactions) did not change significantly with time (linear component: $t_{59}=-1.59$, $P=0.12$; quadratic component: $t_{59}=1.51$, $P=0.13$).

During the whole experiment, we also measured the number of calls broadcast by MANDABOT in response to calls produced by the young finches. We observed large inter-individual variability [mean±s.e.m. (min.–max.): 18,636±2546 (7386–36,506)]. The number of calls broadcast did not change significantly with time (linear component: $t_{297}=-0.98$, $P=0.32$; quadratic component: $t_{297}=0.87$, $P=0.38$).

Finally, we determined whether our different measures could predict success of imitation. Song similarity increased linearly with time on the platform (t-test: $t_9=2.52$, $P=0.03$), but not quadratically ($t_9=-1.91$, $P=0.088$). We reproduced this analysis with the time spent in close vicinity on the platform as a predictor, and independently, with the time spent in clumping. Neither the time spent in close vicinity (linear component: $t_9=-1.73$, $P=0.11$; quadratic component: $t_9=1.61$, $P=0.14$) nor the time spent in clumping (linear component: $t_9=1.22$, $P=0.25$; quadratic component: $t_9=-1.16$, $P=0.28$) were related to song similarity.

We conducted a similar analysis with the number of pecks on the beak or head of the robot as a predictor, and independently, with the number of pecks on the body or tail. Song similarity increased with the number of pecks on beak or head and then decreased in an inverted U (linear component: $t_9=3.10$, $P=0.013$; quadratic component: $t_9=-2.35$, $P=0.043$), but did not vary with the number of pecks on body or tail (linear component: $t_9=1.08$, $P=0.31$; quadratic component: $t_9=-0.46$, $P=0.65$). Song similarity could not be predicted by the proportion of attentive reactions exhibited by the finch following song broadcast by MANDABOT (linear component: $t_7=-0.26$, $P=0.80$; quadratic component:

$t_7=0.068$, $P=0.95$). Success of song imitation could not be predicted by the number of calls broadcast by MANDABOT (linear component: $t_9=-0.07$, $P=0.94$; quadratic component: $t_9=0.27$, $P=0.79$).

This study shows that a robot can be accepted as a valid song tutor for young zebra finches: there was no significant difference between control birds and birds tutored by MANDABOT in song imitation. Further analyses are required to better understand the role of social interactions in this result. In particular, potential experiments could investigate the importance of multimodal aspects and contingency on developmental song learning. For example, young finches could be exposed to a motionless robot, or to a robot with its beak opening synchronously or asynchronously with sound production. We could also evaluate the importance of behavioural and vocal contingencies by increasing the latency of the robot's responses. In male zebra finches, song quality depends on the social context; for example, when they sing to females, songs are faster and more stereotyped (directed songs) than when they sing alone without any audience (undirected song) (Sossinka and Böhner, 1980). Painted as a female, the robotic zebra finch could be used in this courtship context. The robotic zebra finch could also be used to investigate some aspects of social cognition such as gaze following (Butler and Fernández-Juricic, 2014). Overall, these experiments could shed light on multimodal aspects of communication in birds.

References
Aamodt, C. M., Farias-Virgens, M. and White, S. A. (2020). Birdsong as a window into language origins and evolutionary neuroscience. *Philos. Trans. R. Soc. B* **375**, 20190060.

Benichov, J. I., Benezra, S. E., Vallentin, D., Globerson, E., Long, M. A. and Tchernichovski, O. (2016). The forebrain song system mediates predictive call timing in female and male zebra finches. *Curr. Biol.* **26**, 309-318. doi:10.1016/j.cub.2015.12.037

Butler, S. R. and Fernández-Juricic, E. (2014). European starlings recognize the location of robotic conspecific attention. *Biol. Lett.* **10**, 20140665. doi:10.1098/rsbl.2014.0665

Carouso-Peck, S. and Goldstein, M. H. (2019). Female social feedback reveals non-imitative mechanisms of vocal learning in zebra finches. *Curr. Biol.* **29**, 631-636.e3. doi:10.1016/j.cub.2018.12.026

Chen, Y., Matheson, L. E. and Sakata, J. T. (2016). Mechanisms underlying the social enhancement of vocal learning in songbirds. *Proc. Natl Acad. Sci. USA* **113**, 6641-6646. doi:10.1073/pnas.1522306113

Derégnaucourt, S. (2011). Birdsong learning in the laboratory, with especial reference to the song of the zebra finch (*Taeniopygia guttata*). *Interact. Stud.* **12**, 324-350. doi:10.1075/is.12.2.07der

Derégnaucourt, S. and Gahr, M. (2013). Horizontal transmission of the father's song in the zebra finch (*Taeniopygia guttata*). *Biol. Lett.* **9**, 20130247. doi:10.1098/rsbl.2013.0247

Derégnaucourt, S., Mitra, P. P., Fehér, O., Pytte, C. and Tchernichovski, O. (2005). How sleep affects the developmental learning of bird song. *Nature* **433**, 710-716.

Derégnaucourt, S., Saar, S. and Gahr, M. (2012). Melatonin affects the temporal pattern of vocal signatures in birds. *J. Pineal Res.* **53**, 245-258. doi:10.1111/j.1600-079X.2012.00993.x

Derégnaucourt, S., Poirier, C., Van der Kant, A., Van der Linden, A. and Gahr, M. (2013). Comparisons of different methods to train a young zebra finch (*Taeniopygia guttata*) to learn a song. *J. Physiol.* **107**, 210-218. doi:10.1016/j.jphysparis.2012.08.003

Deshpande, M., Pirlepesov, F. and Lints, T. (2014). Rapid encoding of an internal model for imitative learning. *Proc. R. Soc. B* **281**, 20132630. doi:10.1098/rspb.2013.2630

Haesler, S., Rochefort, C., Georgi, B., Licznerski, P., Osten, P. and Scharff, C. (2007). Incomplete and inaccurate vocal imitation after knockdown of FoxP2 in songbird basal ganglia nucleus Area X. *PLoS Biol.* **5**, e321. doi:10.1371/journal.pbio.0050321

Hyland Bruno, J. and Tchernichovski, O. (2019). Regularities in zebra finch song beyond the repeated motif. *Behav. Process.* **163**, 53-59. doi:10.1016/j.beproc.2017.11.001

Immelmann, K. (1969). Song development in the zebra finch and other estrildid finches. In *Bird Vocalizations* (ed. R. A. Hinde), pp. 61-77. Cambridge: Cambridge University Press.

Janik, V. M. and Slater, P. J. B. (2000). The different roles of social learning in vocal communication. *Anim. Behav.* **60**, 1-11. doi:10.1006/anbe.2000.1410

Jolly, L., Pittet, F., Caudal, J.-P., Mouret, J.-B., Houdelier, C., Lumineau, S. and de Margerie, E. (2016). Animal-to-robot social attachment: initial requisites in a gallinaceous bird. *Bioinspir. Biomim.* **11**, 016007. doi:10.1088/1748-3190/11/1/016007

Lachlan, R. F., Van Heijningen, C. A., Ter Haar, S. M. and Ten Cate, C. (2016). Zebra finch song phonology and syntactical structure across populations and continents – a computational comparison. *Front. Psychol.* **7**, 980.

Lerch, A., Roy, P., Pachet, F. and Nagle, L. (2011). Closed-loop bird–computer interactions: a new method to study the role of bird calls. *Anim. Cogn.* **14**, 203-211. doi:10.1007/s10071-010-0353-6

Ljubičić, I., Bruno, J. H. and Tchernichovski, O. (2016). Social influences on song learning. *Curr. Opin. Behav. Sci.* **7**, 101-107. doi:10.1016/j.cobeha.2015.12.006

Patricelli, G. L., Uy, J. A. C., Walsh, G. and Borgia, G. (2002). Male displays adjusted to female's response. *Nature* **415**, 279-280. doi:10.1038/415279a

Price, P. H. (1979). Developmental determinants of structure in zebra finch song. *J. Comp. Physiol. Psychol.* **93**, 260. doi:10.1037/h0077553

Rassak, S., Nachamai, M. and Krishna, M. A. (2016). Survey study on the methods of bird vocalization classification. In *IEEE International Conference on Current Trends in Advanced Computing (ICCTAC)*, pp. 1-8.

Sossinka, R. and Böhner, J. (1980). Song types in the zebra finch *Poephila guttata castanotis* 1. *Z. Tierpsychol.* **53**, 123-132. doi:10.1111/j.1439-0310.1980.tb01044.x

Tchernichovski, O., Lints, T., Mitra, P. P. and Nottebohm, F. (1999). Vocal imitation in zebra finches is inversely related to model abundance. *Proc. Natl Acad. Sci. USA* **96**, 12901-12904. doi:10.1073/pnas.96.22.12901

Ter Maat, A., Trost, L., Sagunsky, H., Seltmann, S. and Gahr, M. (2014). Zebra finch mates use their forebrain song system in unlearned call communication. *PLoS ONE* **9**, e109334. doi:10.1371/journal.pone.0109334

Varkevisser, J. M., Simon, R., Mendoza, E., How, M., van Hijlkema, I., Jin, R., Liang, Q., Scharff, C., Halfwerk, W. H. and Riebel, K. (2021). Adding colour-realistic video images to audio playbacks increases stimulus engagement but does not enhance vocal learning in zebra finches. *Anim. Cogn.*, 1-26. doi:10.1007/s10071-021-01547-8