Reconstruction of vocal interactions in a group of small songbirds

Victor N Anisimov[1], Joshua A Herbst[2], Andrei N Abramchuk[3,5], Alexander V Latanov[1], Richard H R Hahnloser[2,4] & Alexei L Vyssotski[2]

The main obstacle for investigating vocal interactions in vertebrates is the difficulty of discriminating individual vocalizations of rapidly moving, sometimes simultaneously vocalizing individuals. We developed a method of recording and analyzing individual vocalizations in free-ranging animals using ultraminiature back-attached sound and acceleration recorders. Our method allows the separation of zebra finch vocalizations irrespective of background noise and the number of vocalizing animals nearby.

Vocal communication is an important aspect of vertebrate social behavior. However, investigations of vocal interactions have been hampered by difficulty in identifying and separating vocalizations of individual animals (**Supplementary Video 1**).

Many technologies for recording vocalizations from animal groups exist, most of which are inconvenient for one or several reasons. Microphone arrays used in studies of echolocating bats[1,2] cannot disentangle avian vocalizations because songbird calls are longer and of lower frequency than bat calls, and these song echoes interfere with the original wave from the source. Head-mounted microphones[3,4] record high-quality signals from the carrier and substantially attenuate signals from other birds, but only when birds are not too close to each other. Skull-attached piezo-accelerometers and contact microphones[5] potentially allow discrimination of vocal signals from the carrier even in close proximity of other animals, but their implantation requires surgical intervention. Also, in small birds the microphone should be attached by the tethered cable to the commutator at the top of the chamber to lead the signal to the sound recording system. Battery-powered ultraminiature radio microphones[6,7] can eliminate inconvenient cables; the lightest of them weighs only 0.6 g but is capable of transmitting signal to distances of several meters[6].

To record vocalizations from a group of free-ranging animals, we stored data locally on a logger attached to animals' backs. We designed a back-attached audio and acceleration recorder (**Fig. 1a**) based on Neurologger 2A (refs. 8,9). Every logger was equipped with both a miniature microphone and an accelerometer (or 'contact microphone'; **Supplementary Figs. 1** and **2** and **Supplementary Note**).

Carried by adult zebra finches (13–17.5 g, $n = 9$), accelerometers register vibrations that reflect the vocal output of the carrier bird (**Fig. 1b** and **Supplementary Fig. 3**), whereas the microphone records what the animal hears (auditory input).

We measured the effects of the recording backpacks (weighing ~3 g) on diverse behavioral parameters. Before and after backpack attachment, we measured the amount of motion and the number of song motifs and calls per day of old birds (age 14–45 months, average 25 months, 5 males and 1 female) and young adult birds (6 months, 3 males) housed individually. The backpacks transiently suppressed movements and vocalization rates in young and old birds, but within less than 2 weeks these parameters recovered to their pre-backpack values (**Supplementary Figs. 4a–c**). Singing rate recovered the fastest—within less than 1 week; we also found that 40-d-old and 48-d-old juvenile birds (15.2 and 15.7 g, respectively, $n = 2$) produced song vocalizations already at 1 and 2 d after attachment of backpacks that were lighter than the original version (~2 g; **Supplementary Fig. 4d**). Moreover, the behavior of the juveniles on the day after backpack attachment was hardly distinguishable from the behavior on the previous day without a backpack (**Supplementary Videos 2–4**).

To test our method, we placed loggers on two zebra finches (**Supplementary Video 5**). The accelerometers detected song-related vibrations up to 5 kHz. The accelerometer sensitivity was sufficient even for detecting heartbeats, respiratory patterns before vocalization onset and body movement (**Fig. 2a** and **Supplementary Figs. 3** and **5**).

To assess the sound-separation performance achieved with the back-attached accelerometers, we recorded vocalizations of four individually housed zebra finches exposed to different levels of acoustic white noise (0, 50, 60 and 70 dB in 2.5-h sessions; **Fig. 2b–d**). Only one bird (g2k8) produced songs and calls in all four recording sessions (**Supplementary Table 1** and **Supplementary Fig. 6**). To demonstrate the strength of our approach, we selected the softest (<40 dB; **Fig. 2c**) and most difficult vocalization to detect—'stack' calls—in bird g2k8. Even in the quiet environment, the signal-to-noise ratio (SNR) of the acceleration signal (12.9 dB) was superior to SNRs of signals recorded with wall and backpack microphones (10.0 and 10.2 dB, respectively; **Fig. 2d**). In 70-dB background noise, the microphone signals were almost completely masked by noise (SNR = 14.7 and 1.0 dB, respectively). In contrast, the acceleration signal was still high above the noise (SNR = 11.6 dB; **Fig. 2d**), thus demonstrating robustness of accelerometer-based song recordings with respect to environmental noise.

Acceleration spectrograms allowed us not only to discriminate vocalizations from four males but also to classify the vocalizations

[1]Faculty of Biology, Moscow State University, Moscow, Russian Federation. [2]Institute of Neuroinformatics, University of Zurich and ETH, Zurich, Switzerland. [3]Echelon-Geolife, Moscow, Russian Federation. [4]Neuroscience Center Zurich (ZNZ) Zurich, Switzerland. [5]Present address: Luxoft Ltd., Moscow, Russian Federation. Correspondence should be addressed to A.L.V. (alexei@vyssotski.ch).

RECEIVED 19 APRIL; ACCEPTED 24 AUGUST; PUBLISHED ONLINE 28 SEPTEMBER 2014; DOI:10.1038/NMETH.3114

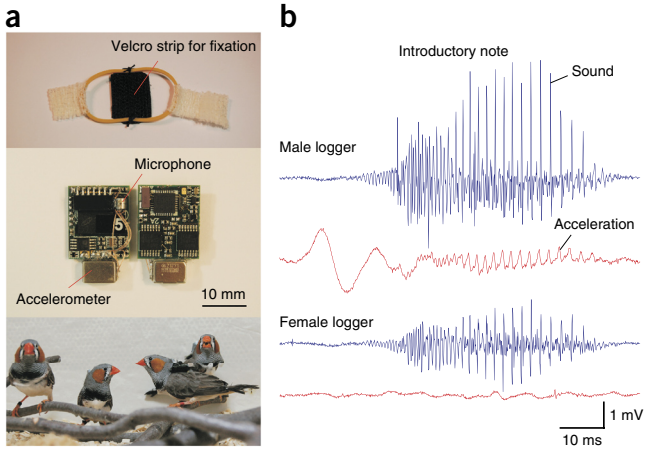

Figure 1 | A wearable sound and acceleration logger for recording individual vocalizations in a group of songbirds. (**a**) Harness with Velcro strip for rapid fixation of the logger to the back of the bird (top), top and bottom views of Neurologger 2A used for sound and acceleration recording (center), and four zebra finches with the backpacks during the experiment (bottom). (**b**) Raw records of an introductory note recorded with backpacks placed on the singing male (top) and a listening female (bottom).

into different types (**Supplementary Figs. 7** and **8**). Relative to gold-standard microphone recordings (0-dB noise), we found misclassification rates of 4.1% for calls (n = 661) and 0.19% for song syllables recorded with the accelerometer (0-dB noise; n = 2,631 vocalizations; **Supplementary Table 1**). Such rates are roughly comparable to the variability among researchers in our lab who classify songs from individually housed birds.

In two sessions, we recorded vocalizing birds (n = 4) during playback of their own vocalizations that we had recorded during a previous session. Playback amplitudes mimicked a singing bird either 40 cm away from the subject (1× amplitude) or 4 cm away from it (10× amplitude). We clustered acceleration signals and found that r.m.s. thresholding of 1× playback did not yield any false positive detections of vocalizations (**Supplementary Table 1**). In contrast, for sessions with 10× playback, one session per bird (n = 4), acceleration-r.m.s. thresholding produced false detection of 25% ± 18% of played vocalizations. However, all these penetrations of playback into the acceleration channel were easily spotted by visual inspection of spectrograms (**Fig. 2e**). Thus, these misdetections did not introduce any classification mistakes.

As a final test, we simultaneously recorded from four individuals in one chamber. We measured vocalization-related sound spectra and acceleration spectra; both were dominated by vocalizations of the carrier (**Supplementary Fig. 9**). We paid particular attention to episodes in which the sound r.m.s. on a listener was less than 3 dB below that of the vocalizer. In these cases, the sound spectra were almost identical on both loggers (**Supplementary Fig. 10a**). In contrast, the simultaneously recorded acceleration spectra of the vocalizer still exceeded that of the listener by 20 dB in the vicinity of 1 kHz, revealing that acceleration-based separation of vocalization spectra is possible even in difficult situations (**Supplementary Fig. 10b**).

To probe for directed vocal interactions in animal pairs, we computed Pearson correlation coefficients (PCCs) among onsets of either calls or song syllables (including introductory notes) in 250-ms time windows (see Online Methods). Calls in all animal pairs were weakly but positively correlated (**Fig. 3a** and **Supplementary Table 2a,b**), indicating that birds increased their call rates when other birds were

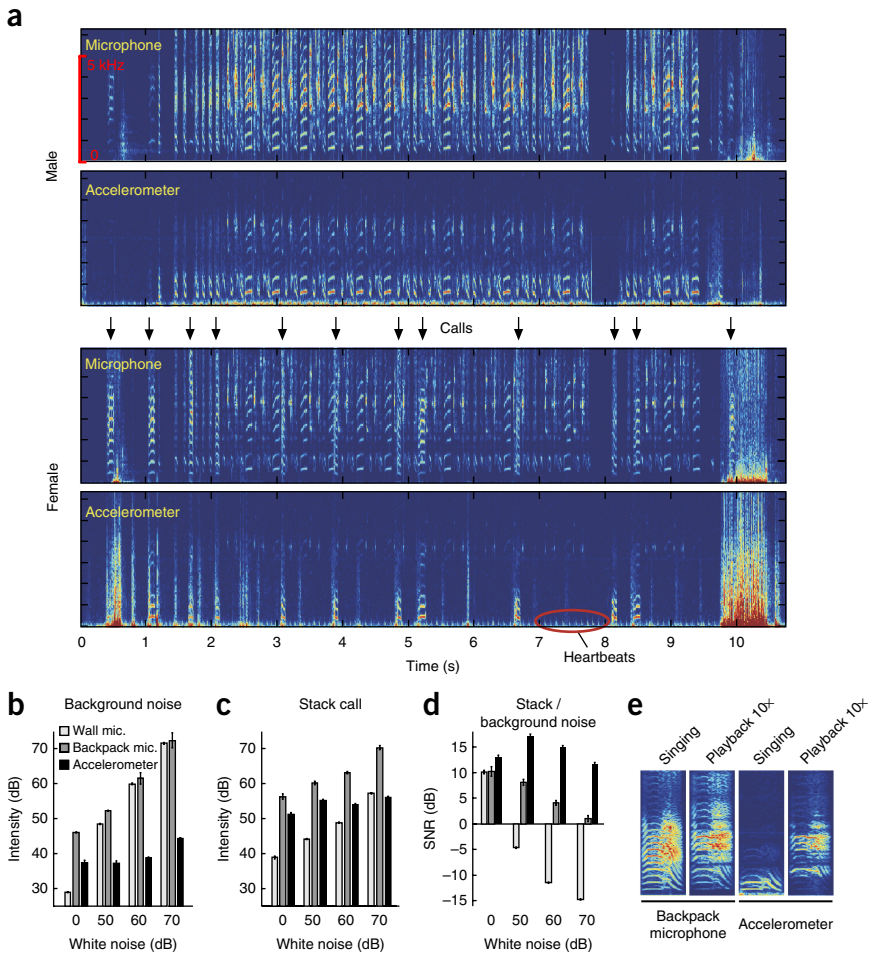

Figure 2 | Acceleration allows for reliable segmentation of vocalizations. (**a**) Sound and acceleration spectrograms of male-female vocal interactions (see also **Supplementary Video 5**). Black arrows indicate calls in the female. Heartbeats are indicated with a red oval. (**b**) Sound and acceleration intensities recorded with a wall microphone, backpack microphone and accelerometer in response to white noise (mean ± s.e.m. for median values in birds g2k8 and r20y17). Signals were high-pass filtered (>400 Hz); sound intensity is given relative to 2 × 10^{-5} Pa (international standard), and acceleration intensity relative to 10^{-3} m/s^2. (**c**) Sound and acceleration intensity of soft 'stack' calls in bird g2k8. s.e.m. is estimated by bootstrapping. (**d**) SNR of soft stack calls depending on background white-noise level. (**e**) Example spectrograms of produced and played-back distance call (10× amplitude) in bird g2k8. Spectrogram color scales have been individually normalized to match the respective signal range. Spectrogram width, 0.19 s; height, 9.6 kHz.

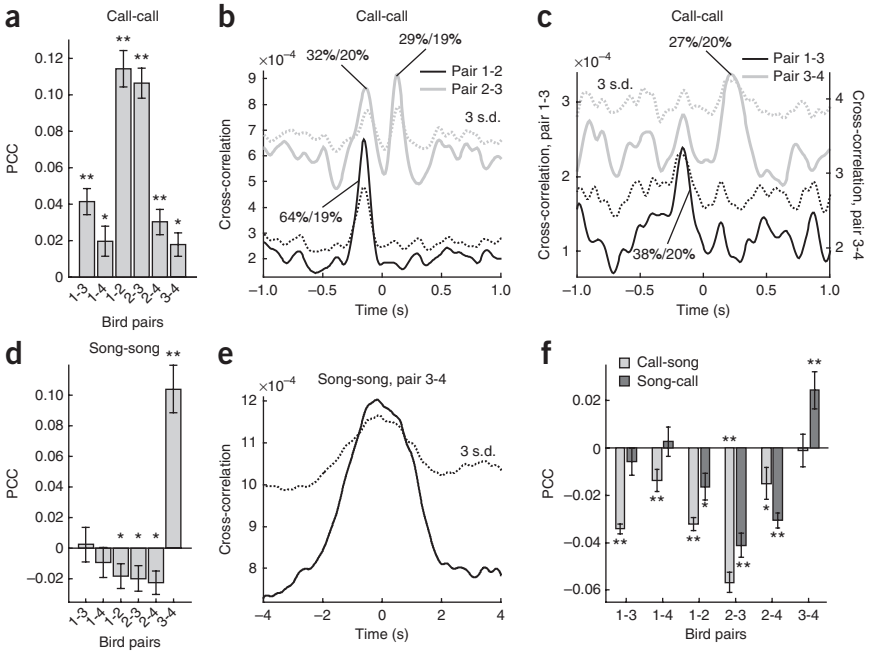

Figure 3 | Vocal interactions in a group of songbirds. (**a**) Pearson's correlation coefficients (PCCs) between calls in different birds (*$P < 0.05$, **$P < 0.001$, double-sided, estimated by bootstrap). Error bars, s.e.m. (**b**) Cross-correlation (CC) between calls in two bird pairs. Dotted lines of corresponding colors indicate 3 s.d. levels. Part of answered calls/estimated part of random call overlap are indicated (%/%). (**c**) CC function between calls in pairs 1-3 and 3-4. (**d**) PCCs between song syllables. (**e**) CC function between song syllables in pair 3-4. (**f**) PCCs between songs and calls.

calling, communally aggregating calls in 'vocal bursts'[10].

To reveal which bird called first and which responded, we computed cross-correlation (CC) functions between calls in all bird pairs; the four pairs with highest CC peaks are shown (**Fig. 3b,c**). The CC peak in pair 1-2 occurred at a negative time lag of −150 ms, indicating that bird 2 responded to calls of bird 1 (**Fig. 3b**). In contrast, bird 1 did not respond to calls of bird 2 because no significant CC (>3 s.d.) was found at a positive time lag. In pair 2-3, the call-call CC function revealed two independent peaks, one at a negative lag of −130 ms and another at a positive lag of 130 ms, indicating that these animals responded to each other's calls (the intervals of significant CCs were from −243 to −72 ms and from 68 to 186 ms, respectively). The call CCs computed in pairs 1-3 and 3-4 were weaker but exceeded 3 s.d. in narrow time intervals from −181 to −129 ms and from 190 to 283 ms, respectively (**Fig. 3c**). Thus, bird 3 responded to calls of birds 1 and 4, whereas birds 1 and 4 did not respond to the calls of bird 3. Call CCs in other pairs did not display any significant peaks. We also computed the percentage of calls answered in 0.5-s intervals following a call: the pairs with significant CCs showed the largest percentage of answered calls (ranging from 27% to 64%), larger than the estimated percentage from random call overlap (4–20%; **Supplementary Table 3**). PCCs between song vocalizations were also notably diverse. Song syllables in bird 2 were negatively correlated with song syllables in all other birds, whereas pair 3-4 showed positive song correlation (**Fig. 3d** and **Supplementary Table 2c,d**). To reveal the temporal pattern of co-singing in pair 3-4, we computed the CC function of their song vocalizations (**Fig. 3e**) and observed a wide peak centered near zero lag, suggesting that both animals responded to each other's songs. In fact, the CC peak width was similar to the peak widths of song autocorrelations in individual birds (**Supplementary Fig. 11**). PCCs between calls and song vocalizations were significantly negative in 8 of 12 directed bird pairs (**Fig. 3f**), suggesting that birds had a tendency to separate singing and calling. Exceptions from this rule may be linked with individual affinity (i.e., 'co-singing').

In summary, we have demonstrated that accelerometers reliably signal vocal output of the carrier bird and that the described technology opens possibilities for studying vocal interactions in the laboratory and possibly in the wild in much more detail than was feasible before. These diverse relationships between songs and calls that we observed point to a complex structure of the social group.

METHODS

Methods and any associated references are available in the online version of the paper.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

The study was supported by research grants from the Scientific & Technological Cooperation Programme Switzerland-Russia (A.L.V., A.V.L. and V.N.A.), the University of Zurich (A.L.V.) and the Swiss National Science Foundation 31003A_127024 (R.H.R.H.), and by the European Research Council under the European Research Community's Seventh Framework Programme (ERC grant AdG 268911 FP7/2007-2013).

AUTHOR CONTRIBUTIONS

V.N.A. and A.L.V. performed the experiments and analyzed data. J.A.H. wrote software for real-time song detection and synchronization of the loggers. A.N.A. wrote logger firmware. A.V.L. and R.H.R.H. supervised the project and wrote software for the data analysis. A.L.V. designed the study and wrote the paper together with R.H.R.H.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

1. Yartsev, M.M., Witter, M.P. & Ulanovsky, N. *Nature* **479**, 103–107 (2011).
2. Barchi, J.R., Knowles, J.M. & Simmons, J.A. *J. Exp. Biol.* **216**, 1053–1063 (2013).
3. Sober, S.J. & Brainard, M.S. *Nat. Neurosci.* **12**, 927–931 (2009).
4. Andalman, A.S. & Fee, M.S. *Proc. Natl. Acad. Sci. USA* **106**, 12518–12523 (2009).
5. Grohrock, P., Häusler, U. & Jürgens, U. *J. Neurosci. Methods* **76**, 7–13 (1997).
6. Hiryu, S., Hagino, T., Riquimaroux, H. & Watanabe, Y. *J. Acoust. Soc. Am.* **121**, 1749–1757 (2007).
7. Hiryu, S., Shiori, Y., Hosokawa, T., Riquimaroux, H. & Watanabe, Y. *J. Comp. Physiol. A Neuroethol. Sens. Neural Behav. Physiol.* **194**, 841–851 (2008).
8. Vyssotski, A.L. *et al. J. Neurophysiol.* **95**, 1263–1273 (2006).
9. Vyssotski, A.L. *et al. Curr. Biol.* **19**, 1159–1166 (2009).
10. Elie, J.E., Soula, H.A., Mathevon, N. & Vignal, C. *J. Acoust. Soc. Am.* **129**, 4037–4046 (2011).

ONLINE METHODS

Wearable sound/acceleration logger. The wearable sound/acceleration logger is based on Neurologger 2A, available from Evolocus (http://www.evolocus.com/). The original Neurologger 2A in configuration for neuronal recordings was capable of recording four channels in the frequency range 0.300–3.3 kHz (−3 dB cutoffs) at a sampling rate of 19.2 kHz per channel and 10-bit resolution (±500-μV input range). The storage capacity of the logger is 1 GB, sufficient for a 2.5-h recording. We adjusted the parameters of the amplifier to suit the audio recordings, i.e., we changed the frequency range to 320–7,200 Hz to cover the main part of the zebra finch vocalization spectrum. The first-order filters produced relatively smooth cutoffs with attenuation of −20 dB per decade. To estimate the required dynamic range, we measured both the peak amplitude of bird vocalizations and the noise level in the empty chamber. The ratio of these amplitudes was about 2^{12}. Thus, a 12-bit or better analog-to-digital conversion (ADC) system was needed to record zebra finch vocalization without loss of sound quality. To exceed the 10-bit ADC limit, we recorded each signal (sound or acceleration) on two channels with different amplification gains (11 and 101). Thus, the total dynamic range covered by such recordings was $\log_2(101/11\times1{,}024) = 13.2$ bits (input referenced ADC step 19.3 μV, range ± 181.8 mV). The necessary changes in the amplification cascades are shown in **Supplementary Figure 1**. The high-pass RC filter with cutoff frequency 320 Hz is formed by the 33-kΩ resistor and 15-nF capacitor. The proper resistors are already present on the board, but the capacitors have to be replaced. As we have four channels available, one should replace four capacitors C1, C3, C4 and C5 with C = 15 nF, 0402, ≥4V, 5% preferable (**Supplementary Fig. 1b**). The amplification ratio is determined by the ratio of a 100-kΩ resistor and a 1-kΩ (for $K_u = 101$) or a 10-kΩ (for $K_u = 11$) resistor. Thus, one should put as resistors R16 = 10 kΩ, R17 = 1 kΩ, R8 = 10 kΩ, and R19 = 1 kΩ, 0402, 1% (**Supplementary Fig. 1a**). Additionally, one should put a bridge R3 = 0, 0402, for making one common reference for all channels.

We attached the miniature microphone FG-23329-D65 and the accelerometer BU-21771-000 (Knowles Electronics) to the logger in such a way to place the microphone as close as possible to the beak (**Fig. 1a** and **Supplementary Fig. 1a**). The weight of the Neurologger 2A board with 1 GB of memory is 781 mg. The weights of the microphone and accelerometer were 36 mg and 276 mg, respectively. Additionally, we mounted an infrared (IR) receiver that we used to synchronize the multiple loggers deployed in the experiment. Its weight (including the connectors) was 488 mg. Thus, the total weight of all electronic components (without batteries) is 1.58 g. The circuitry was powered by a couple of Zn-Air DA10 1.4 V batteries (http://www.duracell.com) connected in series to provide more than 2 V required for the logger (and total current of 4.2 mA), or by a rechargeable 3.7-V Li-polymeric battery (Full River 301218, 20 mA h, 18 × 12 × 3 mm^3; http://www.fullriver.com/). The weight of Zn-Air batteries with the connecting cables was 644 mg, and that of the Li-poly battery was 750 mg. We found the rechargeable battery to be preferable because no battery replacement was needed between sessions. Contributing to the total backpack weight of 2.6 g was also fixing material.

If weight is a concern, it is in fact possible to easily decrease the weight of the backpack down to 1.4 g, which allowed us to use them on zebra finch juveniles and even smaller birds. First, one can decrease the weight by 0.41 g by using a 10-mA-h Full River battery 201013HS10C (0.34 g) instead of a 20-mA-h Full River 301218HS20C (0.75 g) battery used in the study. Second, by removing the IR receiver with the connector, an additional 0.49 g can be saved (in this case synchronization of different loggers can be done using the microphone signal; **Supplementary Fig. 12c**,**d**). Note that without the IR receiver, power consumption will drop down to 4.0 mA; thus, a 10-mA-h battery will be sufficient for a 2.5-h recording. Third, an additional 0.28 g of weight savings can be achieved by replacing the currently used accelerometer with LIS344AHH (http://www.st.com/). In combination, these savings help to reduce the backpack weight from 2.6 g down to 1.4 g.

Synchronization of recorders. In addition to four analog input channels, each logger is equipped with a digital input channel used for synchronization (the digital signal is written to memory at 19.2 kHz synchronously with the analog signals). The digital channel is connected with a miniature infrared (IR) receiver that detects flashes of two LEDs positioned at the top of the cage. These LEDs are controlled by the "Neurologger synchronizer" (http://www.evolocus.com), which is controlled by a custom script written in Matlab (MathWorks). The script also controls illumination of the chamber via a solid-state relay connected to a DLP-2232PB-G interface board (http://www.dlpdesign.com/).

At the start of the recording session, we switched on the light and marked the recording onset with 50 IR pulses of 0.8-ms duration and 0.8-ms gaps—the corresponding bit sequence of approximately 15 logical 1s followed by 15 0s was stored on the logger. Logger synchronization was precise with accuracy of one ADC count (52 μs). The same sequence of pulses was sent at the end of the recording session. Recorders have independent clocks with guaranteed precision of 30 p.p.m. Thus, the divergence of clocks in any two loggers should not exceed 0.54 s during a 2.5-h recording session. The maximal divergence we measured in our loggers was 57 ms. To compensate potential clock drift, we regularly delivered synchronizing IR pulses during the recording session. To automate record alignment, we delivered unique pulse sequences determined by vocalization detection events. The latter were detected (at a sampling rate of 4 ms) using a custom LabVIEW-based song detection software (National Instruments) connected to a wall-attached microphone. At each detected event, a digital pulse was sent to the Neurologger synchronizer, and a 0.8-ms-duration IR pulse was emitted. Because the fine temporal pattern of bird vocalizations is unique in each song bout, so was the sequence of IR pulses.

Animals and experimental schedule. We used 15 adult zebra finches (>120 d old), 2 females and 13 males, and two juvenile males (37 and 43 d old at the beginning of the experiment). Three of the 13 adult males were young (age 6 months, brothers, grew up together). Before the experiment, other adult animals had never been together in the same cage. Adult animals were raised with both parents up to age 60 d in breeding cages, and then they were transferred to communal male and female cages until the start of the experiment. Two juvenile birds were isolated from the family at the beginning of the experiment. Records from a male-female adult pairing are shown in **Figures 1b** and **2a**. Six old birds (age from 14 to 35 months, average 25 ± 4.7 months, 5 males, 1 female), three young and two juvenile birds were used in backpack habituation

doi:10.1038/nmeth.3114

experiments (**Supplementary Fig. 4**). Four males were used in playback experiments (**Fig. 2b–e** and **Supplementary Figs. 3** and **6**), and another four males were recorded in one group (**Fig. 3** and **Supplementary Figs. 5** and **7–15**). Birds were kept on a 13/11-h light/dark cycle with free access to water and food.

The behavior of animals was continuously recorded with two video cameras (Logitech C905, 800 × 600 pixels, 15 f.p.s.) attached to the top edges of the experimental chamber in a way to cover all ground area and permitting the observation of animals from two opposite sides. Cameras were controlled by "Motion Detector" software (http://sourceforge.net/projects/motiondetector/) that continuously saved video streams in 1-min fragments to a hard disk (AVI format). Motion Detector was written in Borland Delphi 7.0 (Borland) using the VisionLab 4.5 (Mitov Software) library. In addition, to simplify the analysis, when a song was detected using the wall-attached microphone, the video stream of the vocalization episode was saved to disk in two AVI files (from the left and the right video cameras, respectively). Each AVI file contained a sound track recorded with the camera built-in microphone. The recording sessions lasted 2.5 h and started at artificial sunrise at 7 a.m.

To inspect habituation behavior to the backpack weight, we attached to 11 animals a logger dummy with weight equal to that of the backpack (3.0 g for adults, 2.0 g for juveniles including the harness shown in **Fig. 1a**). The monitored animals in habituation experiments were kept individually in 23 × 40 × 30 cm^3 cages placed inside a 60 × 60 × 50 cm^3 soundproof recording chamber (**Supplementary Videos 2–5**). An infrared video camera was placed 28 cm away from the long side of the cage to monitor the entire area inside the cage (see an image of cage in the screenshot of Motion Detector (http://sourceforge.net/projects/motiondetector/). The IR camera was custom made from a Logitech C905 video camera by removing the IR light filter from inside and placing two layers of exposed and developed Kodak Gold 200 color film to stop visible light. The IR illuminator was constructed from two sets of three IR-emitting diodes (CN304, Stanley) connected in series with 13-Ω current-limiting resistors (one resistor per each series of three IR diodes) to obtain 50-mA currents per series from a 5-V voltage source. The IR illuminator was placed just near the camera. Zebra finches practically did not move in darkness. For this reason, only light periods were analyzed. Bird vocal activity was recorded with the wall-attached microphone. At the first stage of the experiment, habituation of an animal to the recording chamber was monitored. After 5 or 9 d of habituation (in 'old' and 'young' adult birds, respectively) dummy backpacks with matched weight and shape were attached to the animals. The total duration of the recording was 23 d. We studied the habituation behavior in terms of amount of locomotion, the number of song motifs and the number of calls per day. Juvenile birds were tested under similar conditions as adults, except that in juveniles the dummy backpacks were attached after 2 d of singing, and the recordings lasted for a total 10–11 d.

All animals not in habituation experiments (male-female pair and four birds used in communal recording) were habituated to the backpack weight during at least 1 week before the experiment. Four birds used in communal recording were jointly housed in the sound-proof recording chamber of size 60 × 60 × 50 cm^3. All birds were able to fly in the cage with logger dummies attached; after a few days they demonstrated normal behavior that was visually indistinguishable from the behavior of untreated birds.

Loggers were placed on the animals 1 h before the recording session: sleeping animals were grabbed in the dark, the dummies were replaced with the loggers and animals were returned back to the dark chamber. Backpacks were configured to stay in sleeping mode 1 h to preserve memory.

Data analysis. *Matching two 10-bit records to obtain 13.2-bit resolution.* Each signal was recorded with two separate amplification gains and stored to memory at 10-bit resolution. We then combined these signals to obtain a new signal with larger dynamic range. The combined signal was formed by the high-resolution (high gain) signal, in which we substituted the clipped samples by the corresponding low-resolution samples. This procedure did not allow us to smoothly extrapolate the high-resolution signal (because of variability of component values in the amplification cascades); therefore, we linearly regressed the low-resolution signal onto the high-resolution signal at points at which no clipping was observed and then used these regression coefficients to replace the clipped high-resolution samples with low-resolution predictions. Regression coefficients were computed in consecutive 15-s epochs to avoid biases potentially linked with discharging batteries.

Temporal alignment of backpack records. Because the estimated cumulative drift of backpack clocks exceeded the typical syllable durations, we performed a fine alignment of acquired data records in Matlab. The starts and ends of the recording sessions were determined on the basis of the detected patterns of 50 IR pulses. We defined as reference clock either (i) the logger that showed a total recording time closest to the mean of all others at the end of the recording session or (ii) an external clock (usually an external clock is preferable because it is noise free). Data in other loggers were then aligned to the reference clock.

The clock drifts were not constant, and for this reason a linear temporal stretching/compressing of logger records on the basis of start/stop times was inappropriate. To obtain a dynamic temporal alignment, we computed the divergence of logger records every 10 s by detecting the unique patterns of IR pulses generated in each logger record when animals vocalized. To find matching patterns, we aligned IR pulse sequences in 20-s windows by shifting one record (in increments of 50 μs) relatively to another up to an assumed maximal deviation of ±0.25 s. When fewer than 16 IR pulses were contained in such 20-s alignment windows, no alignment was performed. In total, we computed about 900 alignment points during a 2.5-h session, resulting in ~10^7 sequence comparisons done on two GTX580 GPUs using the Matlab function "filter" (the total computation time was about 23 min). The clock drift during an example 2.5-h recording session is shown in **Supplementary Figure 12a**. When sudden interruptions (delays) were observed, we separately computed the moving medians from the left and right sides of this singularity.

Logger records were aligned by zeroing the moving medians: The records were stretched or compressed by adding or removing a small number of data points. If a point was added, its value was computed by linear interpolation between adjacent points. When the clock drift was linear, we estimated the accuracy of alignment in terms of the residuals of a linear regression of clock drift (**Supplementary Fig. 12b**). The obtained temporal accuracy

was sufficient for comparing temporal patterns in bird vocalizations. Matlab code for data importing and alignment and a simple spectrogram viewer are provided in the **Supplementary Software** together with a test data set.

Separation and classification of vocal fragments. The analysis of vocal patterns of several animals in a group was done one-by-one. At first, to remove the low frequency movement artifacts, we band-pass filtered the accelerometer signal in the frequency range 400 Hz–5 kHz (−3 dB). A finite impulse response filter (FIR) of order 320 was used. Input data were processed in both forward and reverse directions (doubled the filter order). The resulting sequence had precisely zero phase distortion. To detect syllable/call onsets, we computed root mean square (r.m.s.) acceleration in a sliding window of span 512 and step size 32 (given the sampling rate of 19.2 kHz, the window amounted to 26.67-ms width and the step size to 1.67 ms). When the r.m.s. value exceeded a fixed threshold, the short fragment around it (from −26.7 to +140 ms, ~166-ms total duration) was selected for further analysis as a potential syllable or call. We adjusted the detection threshold individually for each animal to detect all vocalizations reliably and to keep the number of false positive detections due to locomotor artifacts as small as possible. Next, we computed spectrograms of these ~166-ms fragments using the fast Fourier transform (FFT) with length 512 (resulting in a spectral resolution of 37.5 Hz) and step size of 128 samples (FFT was applied to the tapered data given by a zero padded Hanning window of size 256 samples). Thus, the spectrogram dimensions were 257 × 24 (zero frequency included). We clustered the spectrograms (syllables) visually using a custom graphical user interface (GUI) in which spectrograms are routinely sorted on the basis of their Euclidean distances. We always verified the relative positions of classified syllables in the motif to avoid classification errors.

Correlation analyses. To compute correlations between vocalizations in bird pairs, we split the 2.5-h session into non-overlapping 250-ms bins for each bird and computed the number of vocalizations of interest in each bin. Correlations were quantified by the Pearson correlation coefficient (PCC) between these numbers in different bird pairs. Because of strong dependence of counts in adjacent bins, the standard way of computing significance of PCC for independent samples produced overestimated probabilities. For this reason we estimated significance of PCCs by bootstrapping. We split the recording session in 1,000 equal fragments (of 9-s duration each) and sampled (1,000 samples) with replacement from this set 10,000 times. We assumed that dependencies of birds' vocalizations beyond time spans of 9 s can be neglected. Indeed, bootstrapping using longer fragments produced similar results. Probabilities smaller than 10^{-4} were estimated using the normal approximation.

To compute autocorrelation and cross-correlation (CC) functions, we counted target vocalizations in 50-ms sliding windows with a 10-ms sliding step, and then we smoothed that sequence by convolving it with a Gaussian kernel of s.d. $\sigma = 20$ ms (and total window span 100 ms). We then computed unbiased autocorrelation or CC functions between these smoothed functions (over a range of time lags up to ±200 s). The first 25 min of the recording session were excluded from the analysis because of reduced vocal activity. The s.d. of autocorrelation or CC functions were estimated by bootstrapping 1,000 fragments as described above. Correlation values deviating by more than 3 s.d. from the average correlation function were considered significant. The CC curve in **Figure 3e** was smoothed by convolving it with a Gaussian kernel of $\sigma = 125$ ms (and total window span 625 ms). The nonsmoothed version of the plot is shown in **Supplementary Figure 11a**.

Supplementary Figures 13–15 show, respectively, correlations between calls, introductory notes and identified syllables; sound amplitudes of vocalizations measured with near and far backpack microphones; and a self-consistency test for estimation of distance between animals during songs and calls.

Reproducibility. *Sample size.* In all figures of the main text (**Figs. 1–3**), the data from individual animals or pair interactions are presented. **Supplementary Figure 4** contains data from 9 birds, sufficient for demonstration of significance at 0.05 level (two-sided *t*-test).

Randomization. All animals selected for the social interaction experiment newer had been in contact with each other before the experiment. Animals were selected from the colony randomly with the exception described in "Animals and experimental schedule."

Blinding. The experiment was completely automated, and experimenters were not in contact with the animals during recording sessions. Statistics of habituation to backpacks (**Supplementary Fig. 4**) were computed relatively to individual baselines recorded before. No untreated control group was used.

Statistical tests. For between-subject comparisons, the *t*-test was used. The distribution of data does not contradict normal distribution assumptions. Variances within groups were similar (see whiskers in **Supplementary Fig. 4**). Verification of significance of vocal interactions between animals was done by the bootstrap data shuffling.

Ethics statement. All experimental procedures were approved by the Cantonal Veterinary Office of the Canton of Zurich, Switzerland (license numbers 123/2010 and 207/2013).

doi:10.1038/nmeth.3114