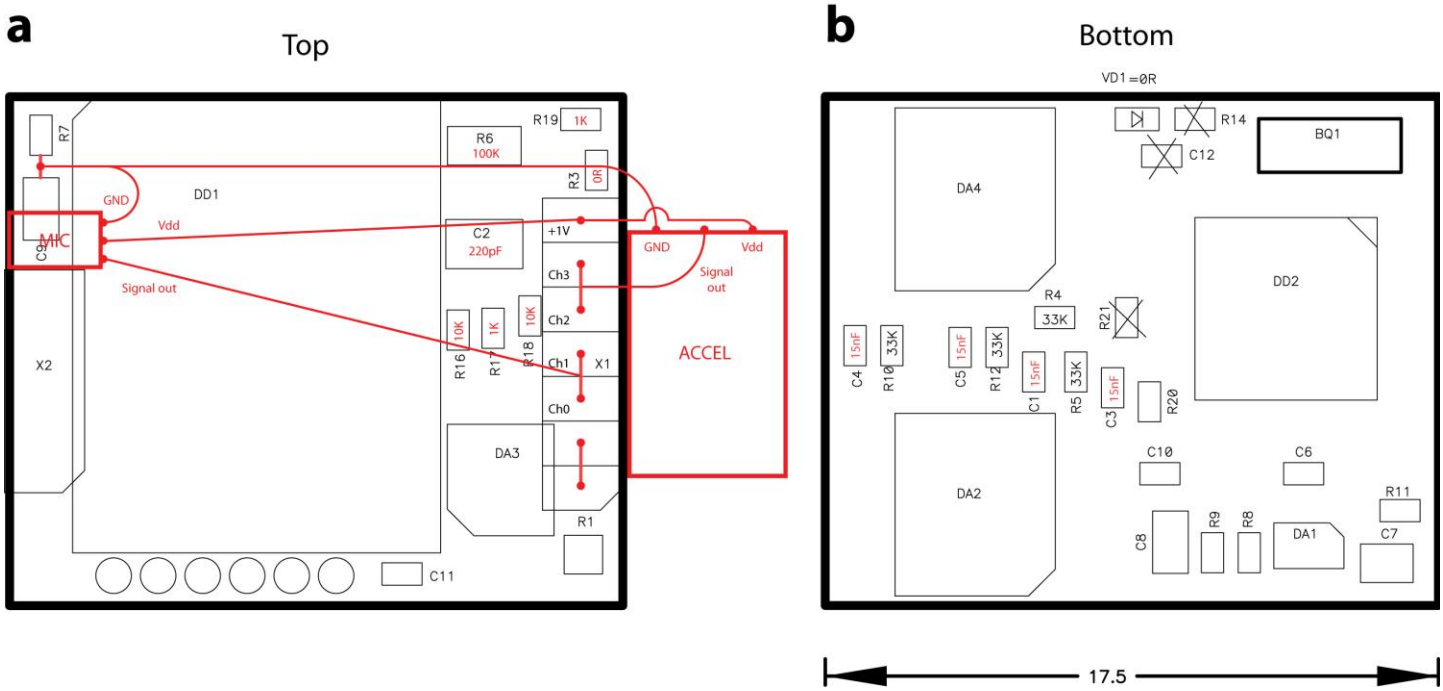
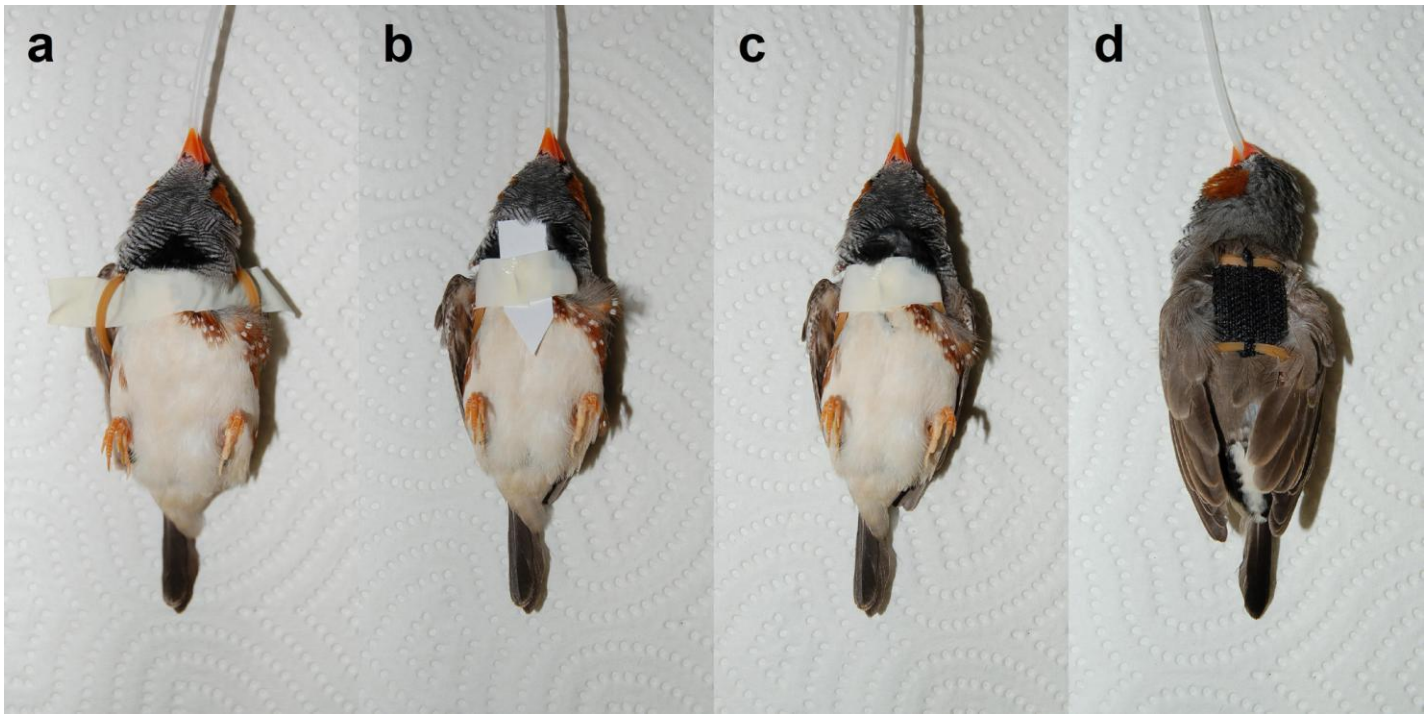


Supplementary Figure 1

Adaptation of Neurologger 2A for sound and acceleration recordings

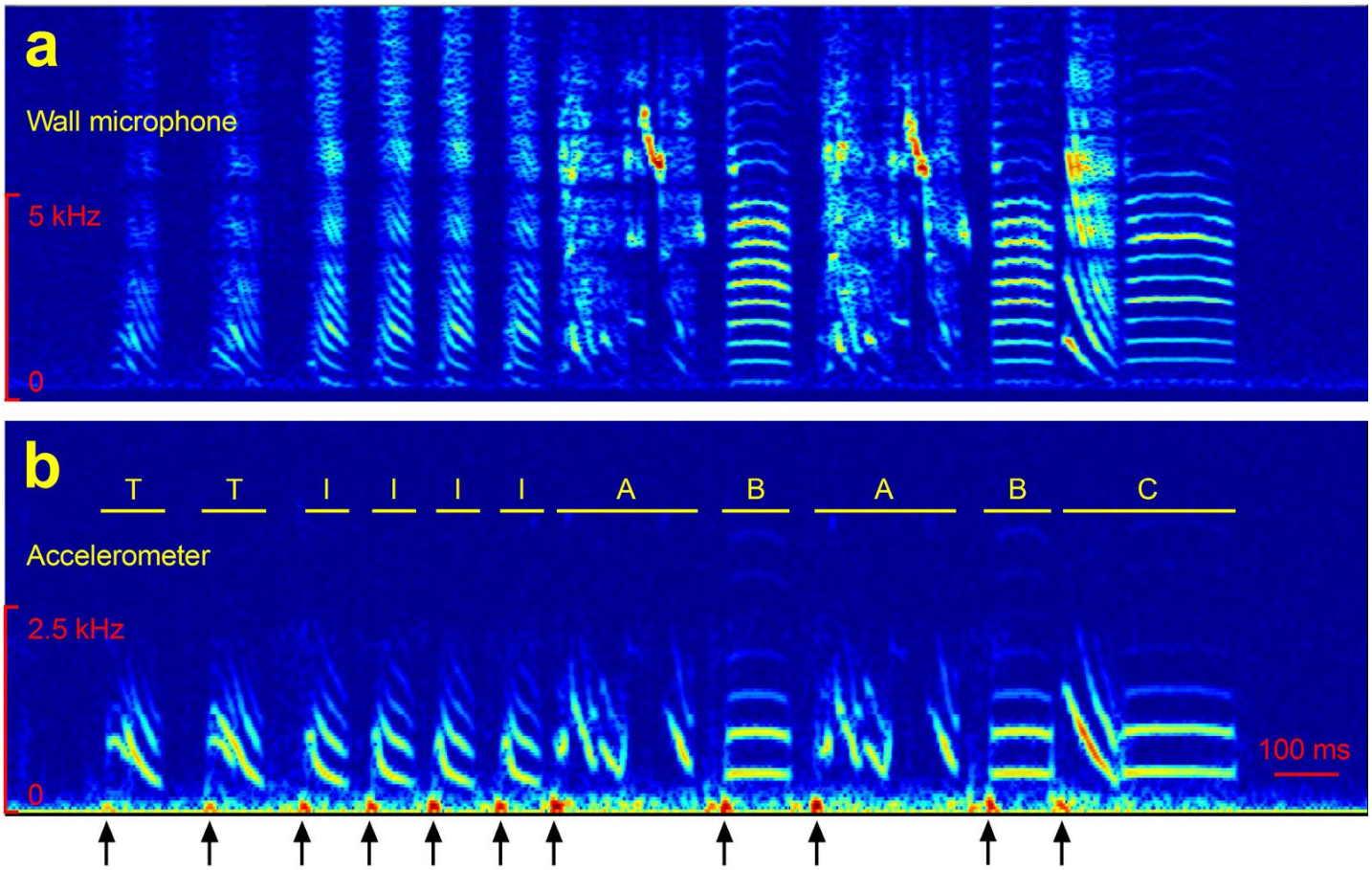
(**a**) Top view of the neurologger printed circuit board (PCB) with microphone (MIC) and accelerometer (ACCEL) attached. (**b**) Bottom view of Neurologger 2A PCB. Components that need replacement are marked in red. Wires (red lines) connect microphone and accelerometer with the board. Wire connects (soldering points) are marked by red dots. Input pads Ch0 and Ch1 are connected with the microphone signal output, and Ch2 and Ch3 with the accelerometer signal output; +1V is the power supply for the microphone and the accelerometer. Microphone and accelerometer are glued to the board by two component epoxy adhesive (DOUBLE/BUBBLE® Epoxy, www.hardmanadhesives.com).



Supplementary Figure 2

Attachment of the harness to a zebra finch

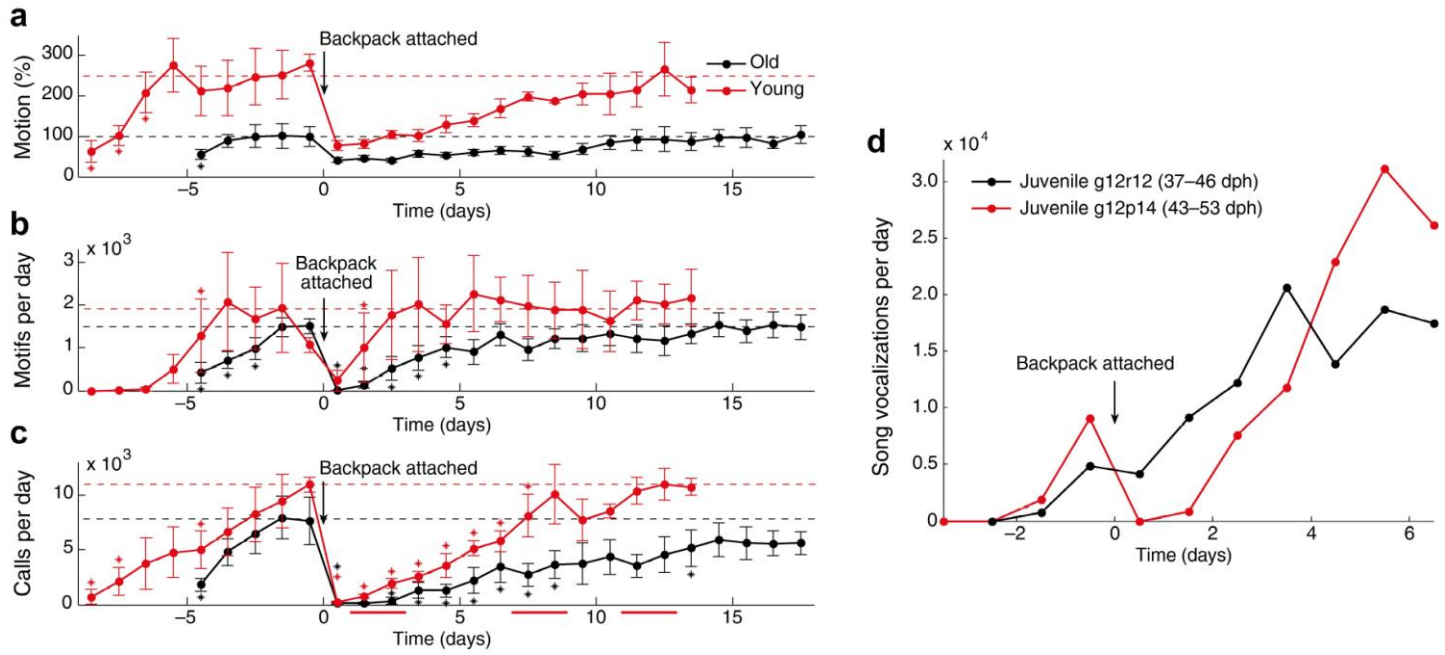
The bird is anesthetized with Isoflurane (3% induction, 1.5% support) delivered through a silicon tube going to the beak. The harness (Fig. 1a) was fabricated in advance from a piece of Velcro strip of 15x22 mm and a BüroLine rubber band of 40x1.3 mm (www.bueroline.com, Cat. #155012). The rubber band is sewn to the Velcro strip at two places, the knots of the thread are fixed by cyanoacrylate glue. (**a**) The harness is placed on the bird; the loops of the rubber band on the chest are attached to each other with a piece of thin laboratory/surgical rubber glove. (**b**) The rubber endings are glued together with a drop of cyanoacrylate, a small piece of paper is placed underneath to prevent gluing the feathers. (**c**) Bottom view of the attached harness. (**d**) Top view of the attached harness. The tightness of the harness is adjusted via the length of the rubber ring on the chest. It was found that tension of the harness has little impact on the quality of sound recordings. The tension should be such as to prevent any unnecessary pressure on the animal and to prevent sliding of the backpack on the bird's back; a loose harness can disturb the animal as much as an overly tight harness. To attach the logger onto the Velcro we recommend placing forceps between the animal's back and the harness to push against the forceps and not against the back. To remove the logger we recommend holding the Velcro strip of the harness in one hand and the backpack in the other hand.



Supplementary Figure 3

Spectrograms of a song motif recorded with the wall microphone and the accelerometer (bird g2k8)

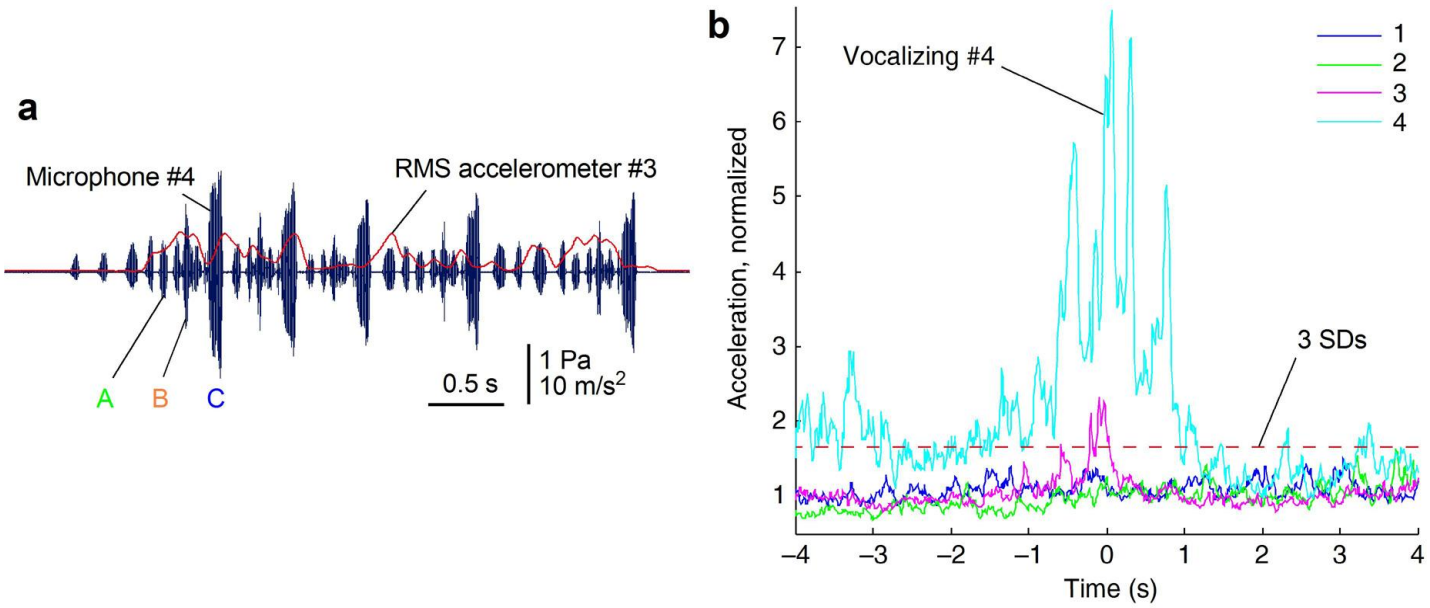
(**a**) Spectrogram of a song motif recorded with the wall microphone. (**b**) Spectrogram of the same song motif recorded with the accelerometer. Abbreviations: “T” - tet calls, “I” - introductory notes, “A”, “B” and “C” – song syllables. The low-frequency bumps in acceleration that tend to occur before syllable onsets (black arrows) are presumably associated with rapid changes in air sac pressure (respiration-related thoracic movement).



Supplementary Figure 4

Habituation of old, young and juvenile animals to the chamber and the backpack

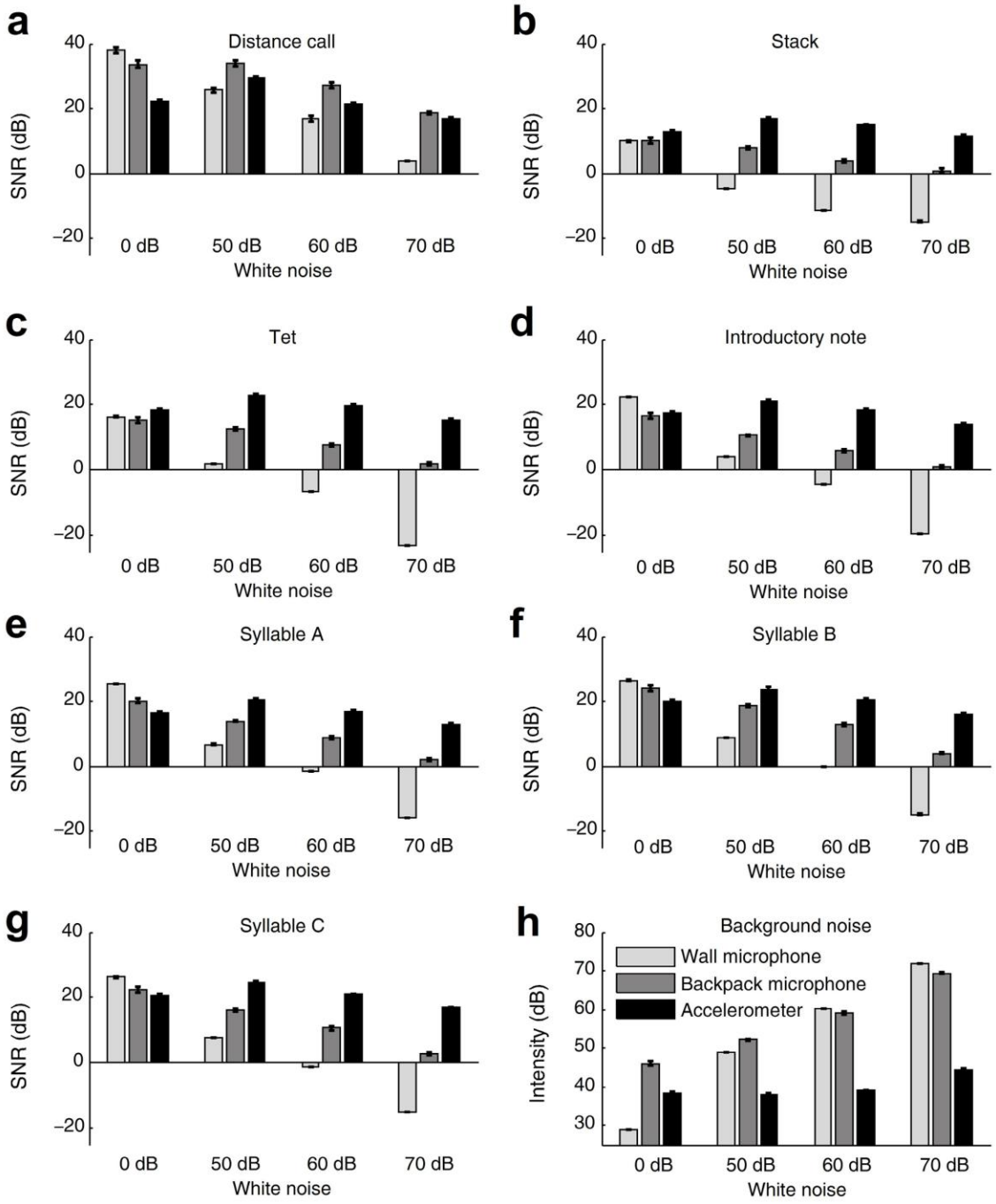
Old (n = 6) and young (n = 3) animals were habituated to the sound proof individual isolation chambers until stationary amounts of movement and singing were observed. Then, the backpacks were attached (day zero), backpacks were attached to old animals after 5 days and in young animals after 9 days. (**a**) Shown is the daily amount of movement measured by infrared video camera (motion, expressed in percent of the baseline motion (black dashed line) in old animals achieved on days [-3 -1]). The baseline value (2.27%) is an output of "Motion Detector" (range [0 100], %) averaged over light phases of days. For young animals the baseline motion (red dashed line) was given by the average motion on days [-6 -1] (5.63%). Error bars indicate s.e.m. Asterisks label days when motion significantly deviated from the corresponding baseline level (paired two-sided t-test, $P < 0.05$). Locomotor activity of young animals exceeded activity of old animals on all days ($P < 0.05$, non-paired two-sided t-test). (**b**) Number of song motifs produced per day. Baselines were computed by averaging number of motifs on days [-2 -1] in old birds and on days [-4 -2] in young birds. Asterisks mark significant deviations from baselines as before. Number of motifs produced by old and young animals differed on some days. (**c**) Number of calls per day. Number of calls in young group increased monotonically during 9 days of habituation to the chamber and did not reach saturation. For this reason we took the value on day -1 as estimate of baseline in young animals. Baseline calling in old animals was estimated as average on days [-2 -1]. Red horizontal bars along time axis indicates significant difference between old and young animals ($P < 0.05$, non-paired two-sided t-test). Notably, the rate of calling increased in young birds during habituation to the backpacks (days [1 9]) in roughly similar manner as during habituation to the chambers (days [-9 -1]). This means that the severity of calls disturbance in young birds caused by the backpacks is similar to the disturbance caused by the isolation chambers. (**d**) On two juveniles we attached 2.0 g backpacks on post-hatch-days 39 and 46 (time zero on x-axis). In the first juvenile (black) the backpack hardly decreased singing rate; in the second juvenile (red) the singing rate was transiently suppressed and picked up two days later. Both birds consistently produced more than 10'000 song syllables and introductory notes per day just 3 days after backpack attachment, suggesting that our recording technique is permissive of song learning studies in juveniles.



Supplementary Figure 5

Non-vocal response to song

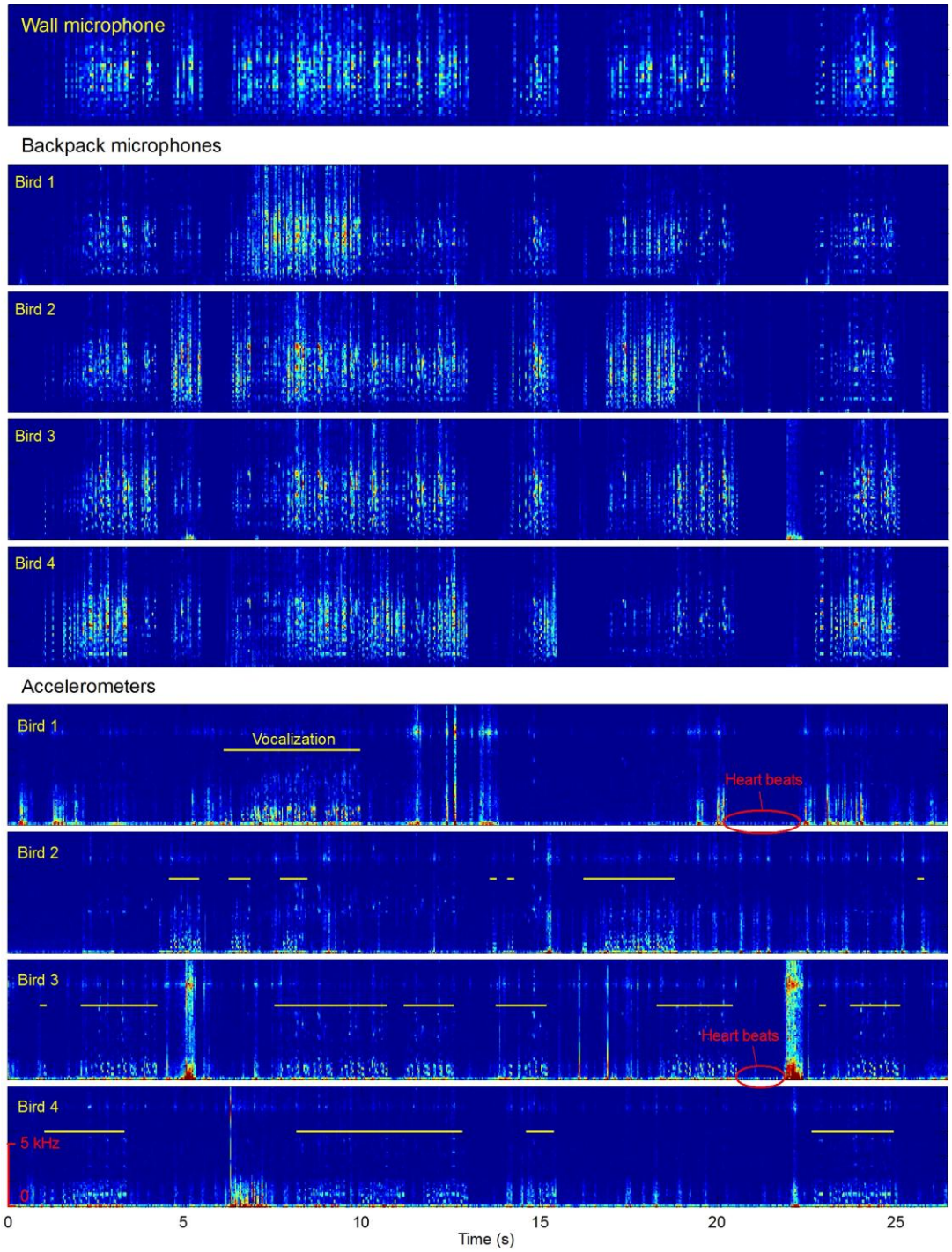
(**a**) An example of non-vocal response in co-singing birds 3 and 4. The song oscillogram (blue) in bird 4 was recorded with the wearable microphone. The non-vocal response (body acceleration, red) in bird 3 is reflected in the root-mean-square (RMS) accelerometer signal (sliding window) in the non-vocal low-frequency band < 300 Hz. Bird 3 did not vocalize during this episode (not shown). (**b**) Normalized median RMS traces of body acceleration in all four birds aligned to the beginning of the song motif in bird 4 (onset of syllable A). Shown are the median filtered RMS traces of 72 motifs. We selected only episodes in which no vocalizations in birds 1 - 3 were detected within ± 2 s from motif onset in bird 4. The 3 SDs level (red dashed line) in bird 3 was estimated using bootstrapping. Note that significant (> 3 SDs) movements in bird 3 took place in a narrow interval -217 to 40 ms around the onset of the motif. Possibly, bird 3 reacted to the preceding introductory notes or to syllable C in the previous motif. By contrast, bird 4 did not show any detectable non-vocal response to the songs of bird 3, indicating asymmetry in movement responses in these two animals.



Supplementary Figure 6

Accelerometers achieve high–signal-to-noise recordings of vocalizations irrespective of background noise level

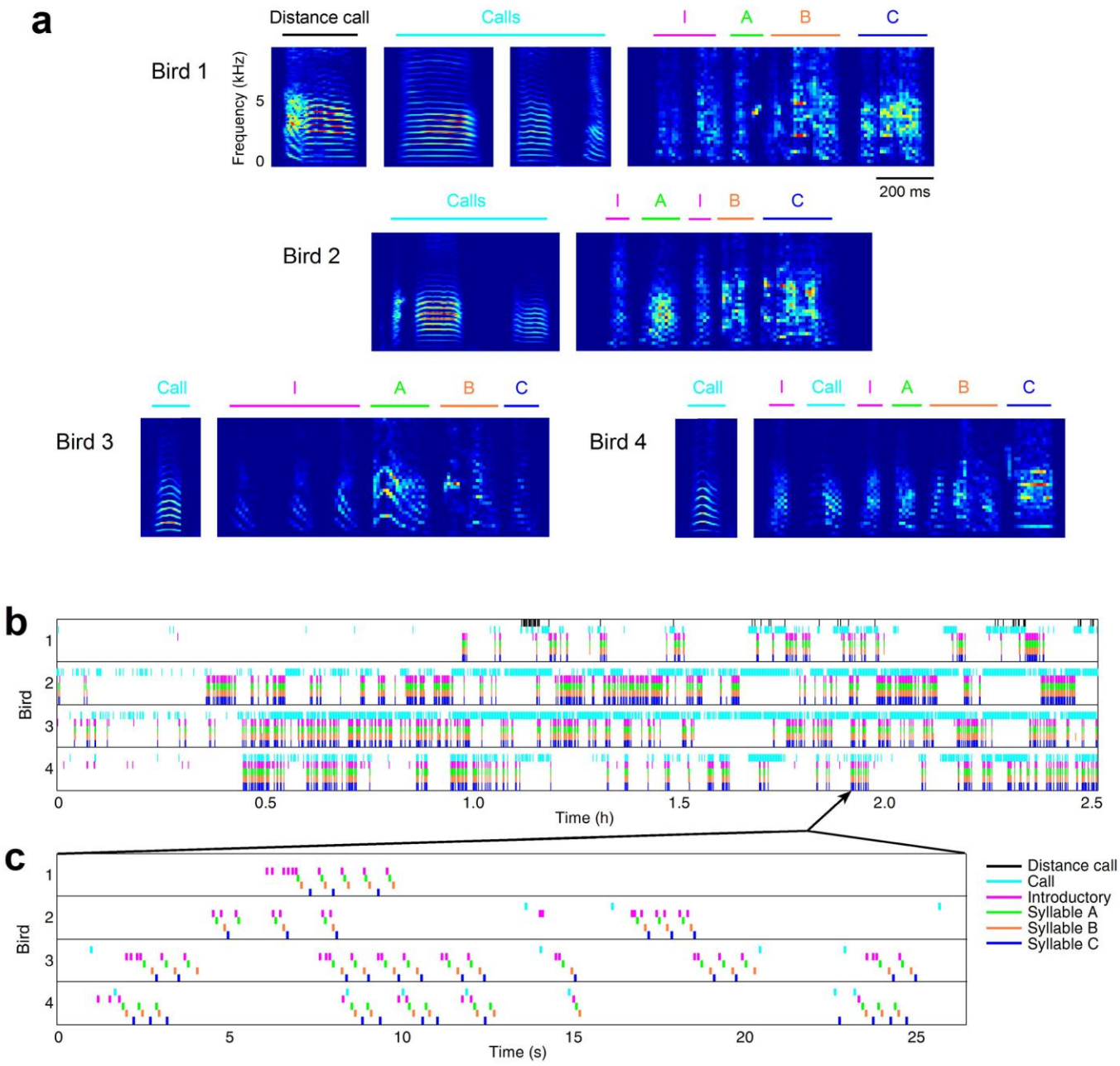
(**a-g**) SNR Sound pressure levels (SPL) achieved with the wall microphone (light gray bars), backpack microphone (medium gray bars), and acceleration SNR (dark gray bars) during production of different calls, introductory notes, and song syllables in bird g2k8. The four levels of white background noise were: 0, 50, 60 and 70 dB. SNRs vary in the range [-23.1 37.9] dB for the wall microphone, in the range [0.92 33.9] dB for the backpack microphone, and in the range [11.6 29.3] dB for the accelerometer, showing that accelerometers provide the only feasible approach to unconstrained recordings of vocalizations under noisy conditions. SPLs are shown relative to 2×10^{-5} Pa (international standard). Acceleration is shown relative to 10^{-3} m/s^2. (**h**) Intensity of background noise recorded with microphones and accelerometers. SNRs in **a-g** were computed relative to these levels.



Supplementary Figure 7

Wearable accelerometers allow discrimination of vocalizations from four males

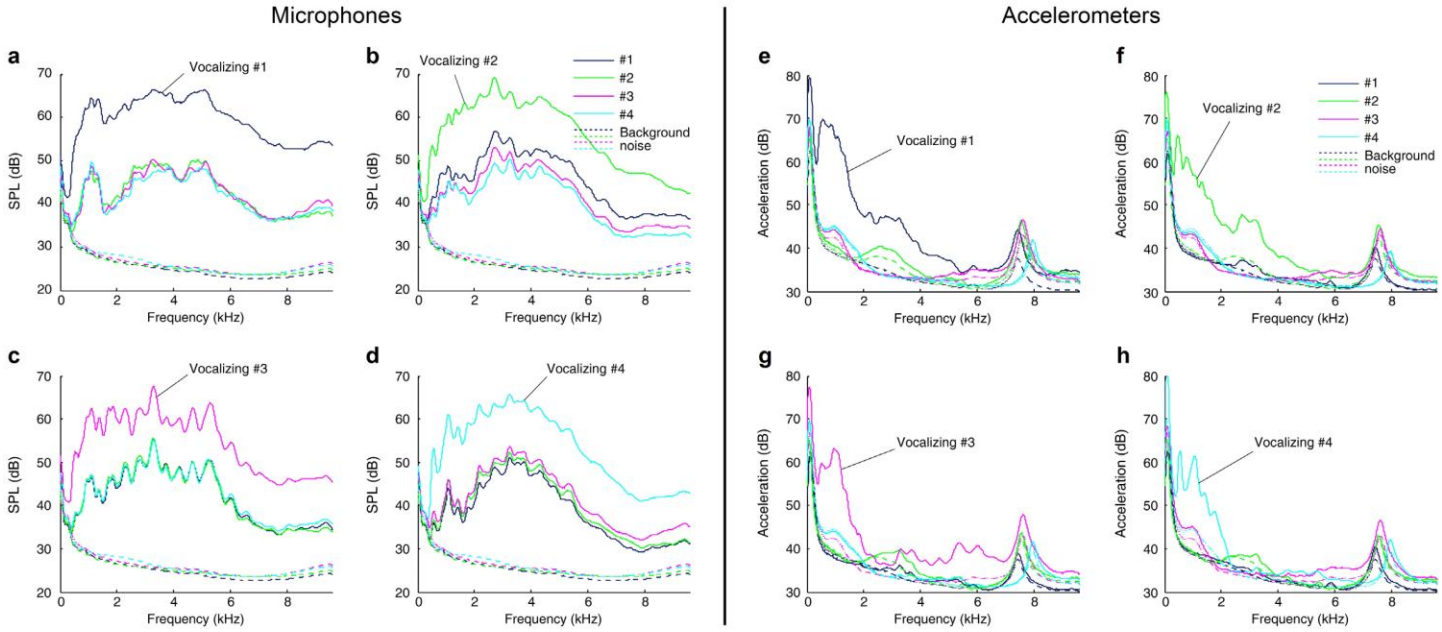
The upper panel shows a sound spectrogram of vocalization of four males to be separated and analyzed (the sound was recorded with a wall-attached microphone, see also the **Supplementary Video 1** for the birds' behavior). The spectrograms in the middle show that the wearable microphones pick up mixed vocalizations from several birds, making signal discrimination difficult at least. By contrast, accelerometers spectra (bottom) recorded vocalization of just their hosts, allowing for simple threshold-based vocal discrimination. Episodes of individual vocalizations are marked by yellow horizontal bars. The accelerometers were sensitive to rapid bird movements. During episodes of inactivity accelerometers were often able to pick up heart beats (in animals 1 and 3, red ellipses).



Supplementary Figure 8

Segmentation and classification of all vocalizations in a bird group

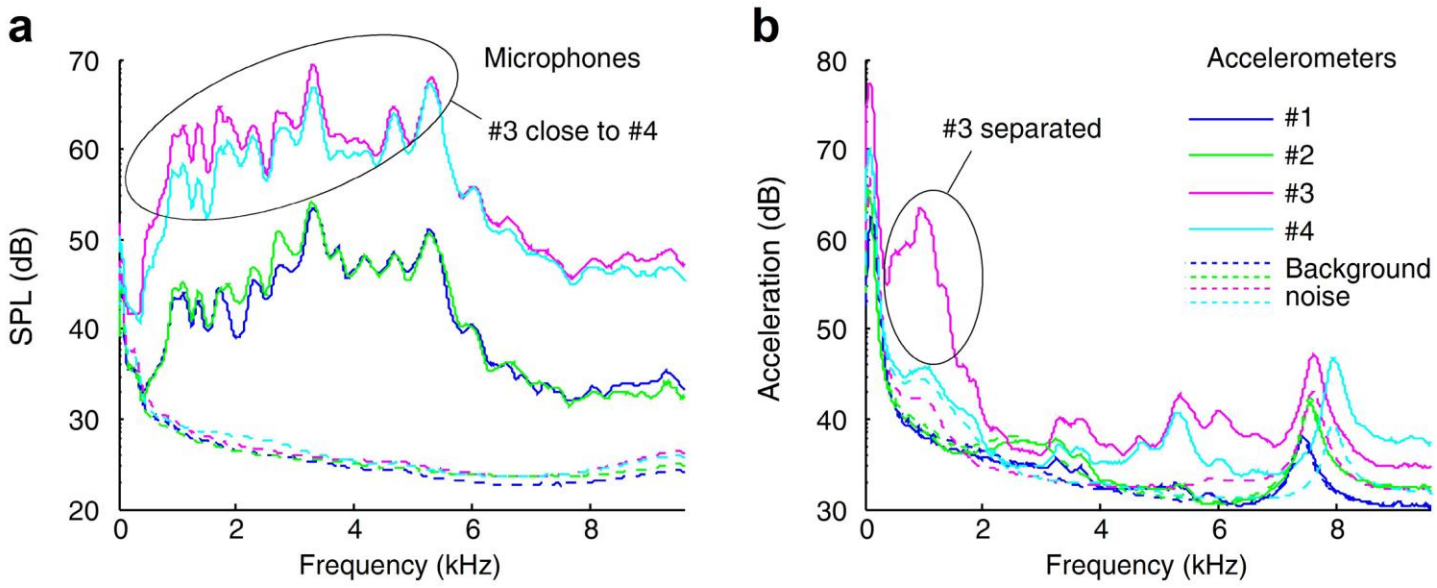
(**a**) Shown are spectrograms of example vocalizations in individual birds recorded with back-pack microphones. Song motifs consist of syllables labeled “A”, “B”, and “C”, and introductory notes labeled “I”. Only bird 1 produced distance calls (such calls are relatively rare in domesticated populations of zebra finches). (**b**) A 2.5-h temporal sequence of morning vocalizations. The zero time point marks the lights-on event at 7 a.m. Each vocalization (distance-call, call, introductory note, syllables: A, B, C) is represented by a vertical line of corresponding color. (**c**) A magnified 25-s episode taken from (b) during which all four animals sung. This interval corresponds to the spectrogram shown in **Supplementary Figure 7**.



Supplementary Figure 9

Vocalization spectra measured with backpack microphones and accelerometers

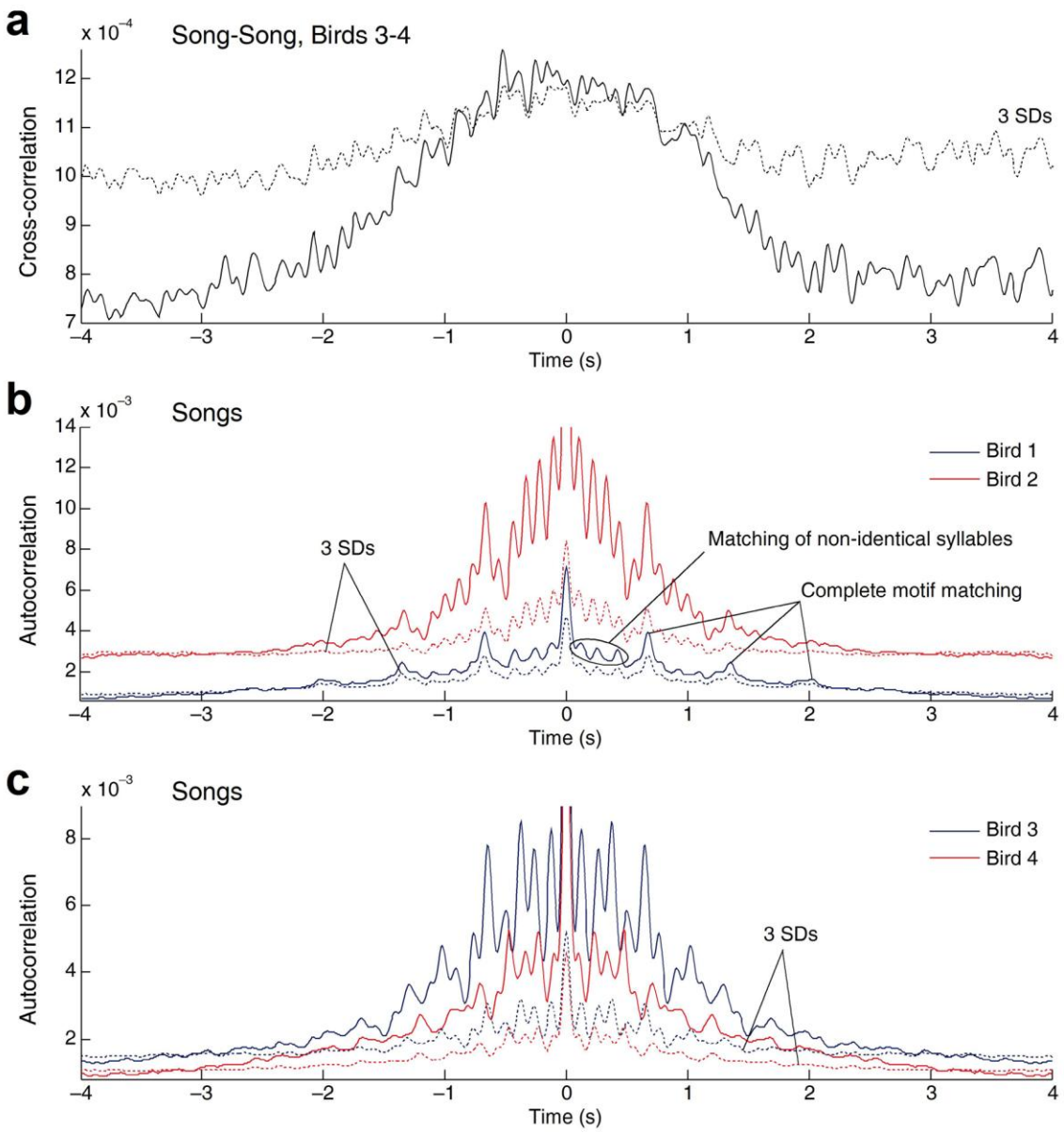
(**a-d**) show median vocalization spectra (solid lines) of birds 1-4 recorded with microphones #1 to #4. Spectra were computed from 166-ms fragments starting with vocalization onsets (calls, introductory notes and syllables). Only vocalizations that did not overlapped with vocalizations or noises (e.g. wing flaps) produced by other animals were selected. In birds 1-4 the numbers of such vocalizations were 972, 4535, 3344, and 1468, respectively. Dotted lines represent background noise recorded by microphones when no bird vocalized. To compute spectra of background noise, 10000 166-ms sound fragments during quiescence were randomly taken. Note that sound intensity at the microphone of the vocalizing bird exceeded background noise by about 30 dB. Sound intensity at microphones of listeners exceeded background noise by about 20 dB. Spectra at the microphones of listeners were attenuated but similar to spectra at carrier microphones. SPL values are given relatively to $10^{-6}\ \mathrm{Pa}/\sqrt{\mathrm{Hz}}$. (**e-h**) show median vocalization spectra (solid lines) of birds 1-4 recorded with accelerometers #1 to #4. Dotted lines represent spectra of background noise when no bird vocalized. The peak in the vicinity of 8 kHz is caused most probably by resonance of the accelerometer sensor (luckily, the resonance frequency lies above the frequency range of interest). Note that the spectra of the listeners essentially does not differ from their background noise spectrum. The spectra of the vocalizing birds exceed the background noise spectra by about 20 dB at 1 kHz. Acceleration values are provided relative to $10^{-5}\ \mathrm{m/s^2}/\sqrt{\mathrm{Hz}}$.



Supplementary Figure 10

Vocalization spectra of bird 3 when it is close to bird 4

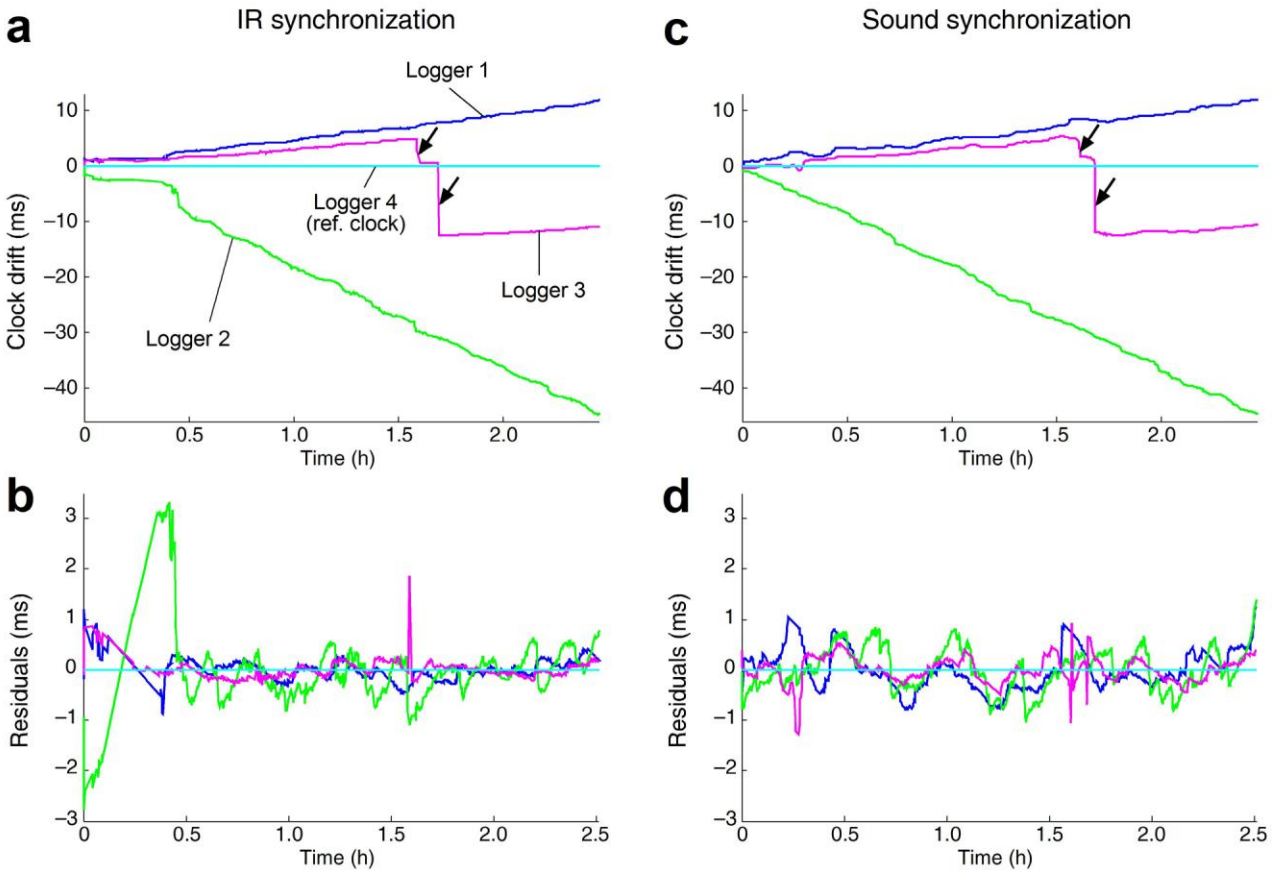
(**a**) Vocalization spectrum recorded with microphones. Note that microphones practically do not discriminate vocalizations in proximal birds 3 and 4 (red and blue lines almost coincide). Dotted lines denote background noise. Shown are spectra of 274 vocalizations in bird 3 during which the estimated distance between the two birds was ≤ 5 cm (the distance estimate was based on relative sound amplitudes between singer and listener ≤ 3 dB). (**b**) Vocalization spectra of the same vocalizations recorded with accelerometers. Note that the spectrum on the vocalizer accelerometer exceeding that of all listeners by about 20 dB (at 1 kHz).



Supplementary Figure 11

Cross-correlation (CC) and autocorrelations of songs

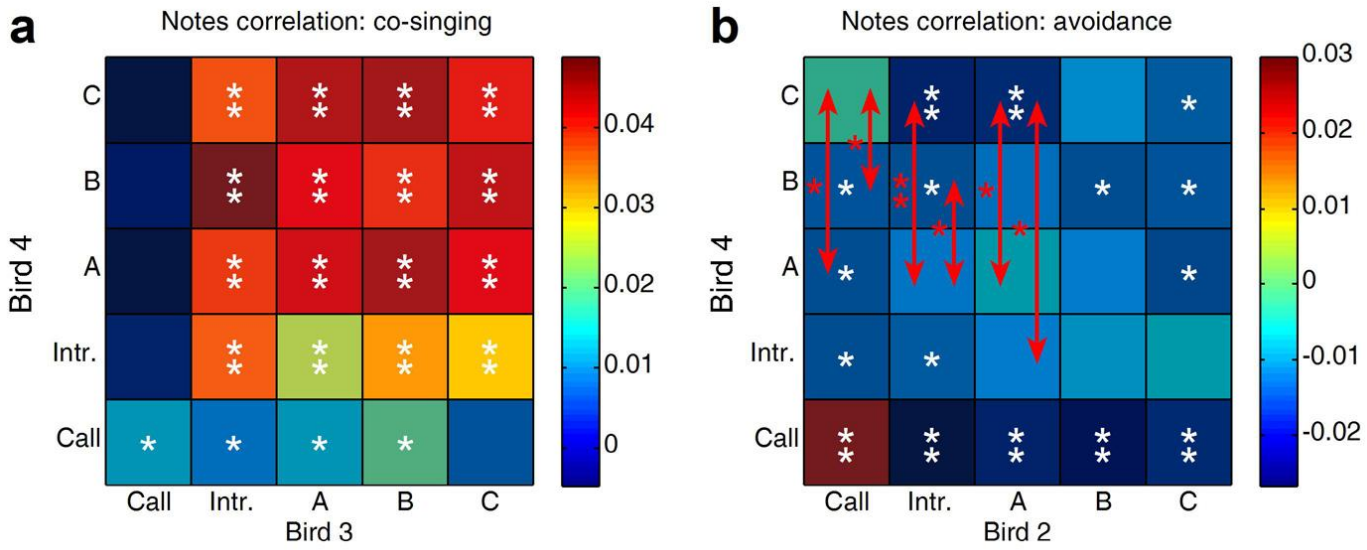
Correlations were computed after summing the number of song vocalization onsets in 50-ms sliding windows and subsequent convolution with a Gaussian kernel of width σ = 20 ms. (**a**) The CC function of song vocalizations of co-singing birds 3-4 has a wide peak in excess of 3 STDs. Shown is the same curve as in Figure 3e but without smoothing. (**b, c**) Autocorrelations of calls in birds 1 and 2 (**b**) and 3 and 4 (**c**). The autocorrelations exceeded 3 standard deviations (SDs, dotted lines of corresponding colors) in the relatively wide interval [-3 3] s. In some birds, high autocorrelation peaks were produced by complete motif matching and lower peaks by matching of non-identical syllables (**b**). The lower CC values in **a** compared to autocorrelations in **c** reveal that co-singing happened less frequently than production of two consecutive motifs by a single bird.



Supplementary Figure 12

Temporal alignment of backpack records

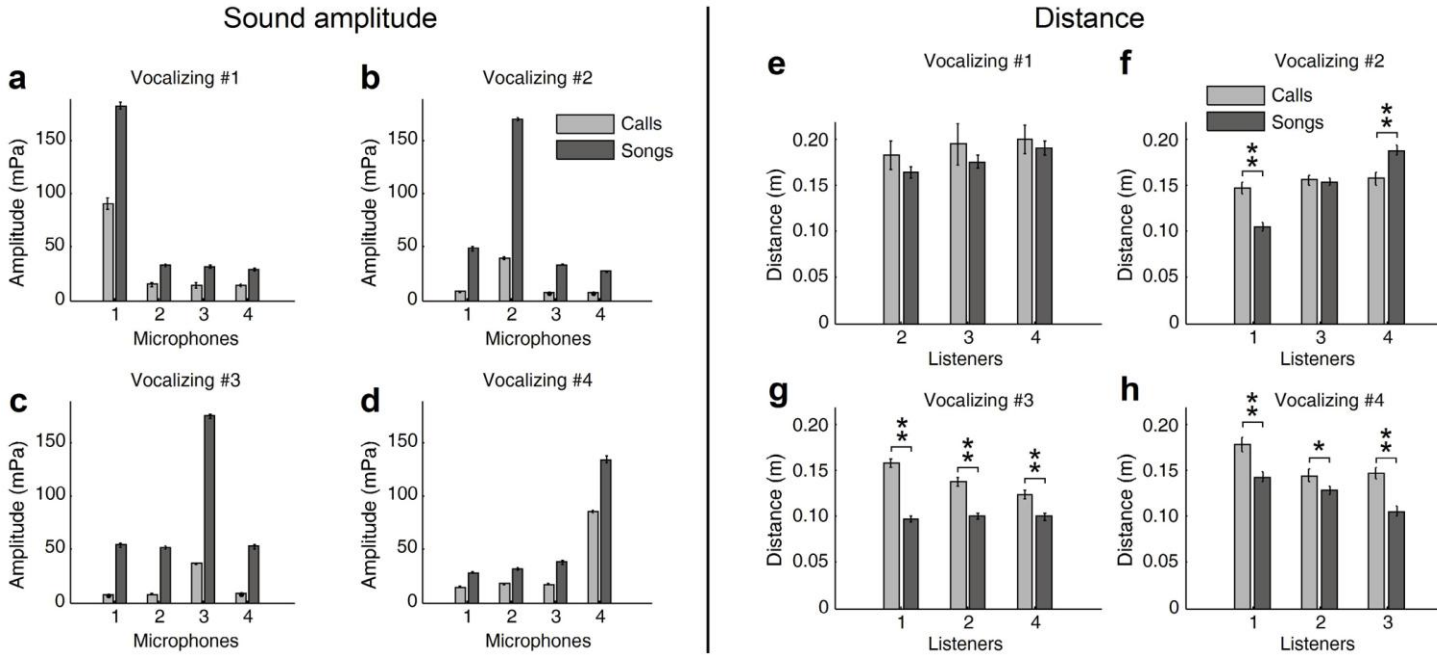
(**a**) Drifts of internal clocks on three loggers are shown by blue, green and magenta lines relatively to the fourth logger taken as a reference (turquoise horizontal line). Drift was measured by comparing patterns of IR pulses stored on loggers. A Labview program monitored sound recorded with a wall-attached microphone, every 4 ms it assessed whether a song was present, in which case a 0.8-ms IR pulse was sent to the loggers for synchronization. Temporal drift was always linear except in one logger (shown in green) in which a couple of sudden interruptions were observed for a dozen of milliseconds (black arrows). Logger records were aligned by zeroing the lines: we stretched or compressed the records by adding (linear interpolation) or removing data points. (**b**) Shown are residuals after linear regression of clock drift curves in **a** (because of the two interruptions we performed linear regression on 3 segments in logger 3). Most of the time residuals did not exceed 1 ms. Only during the first 0.5 h residuals were relatively large (< 3 ms) because of low density of IR pulses caused by infrequent singing. More frequent emission of randomized IR pulse sequences could eliminate this problem. The root-mean-square deviation was 0.46 ms for the whole record and 0.25 ms after omission of the first 0.5 h. (**c**) Drifts of internal clocks like in **a** but measured by temporal matching of sound envelopes recorded with the logger microphones. See Online Methods for details. (**d**) Residuals of sound-based synchronization similar to **b**. The root-mean-square deviation was 0.37 ms for the entire record and 0.35 ms after omission of the first 0.5 h. Thus, synchronization based on sound works as well as synchronization based on IR pulses and may be useful outdoors where IR synchronization can be difficult.



Supplementary Figure 13

Correlations between calls, introductory notes and identified syllables

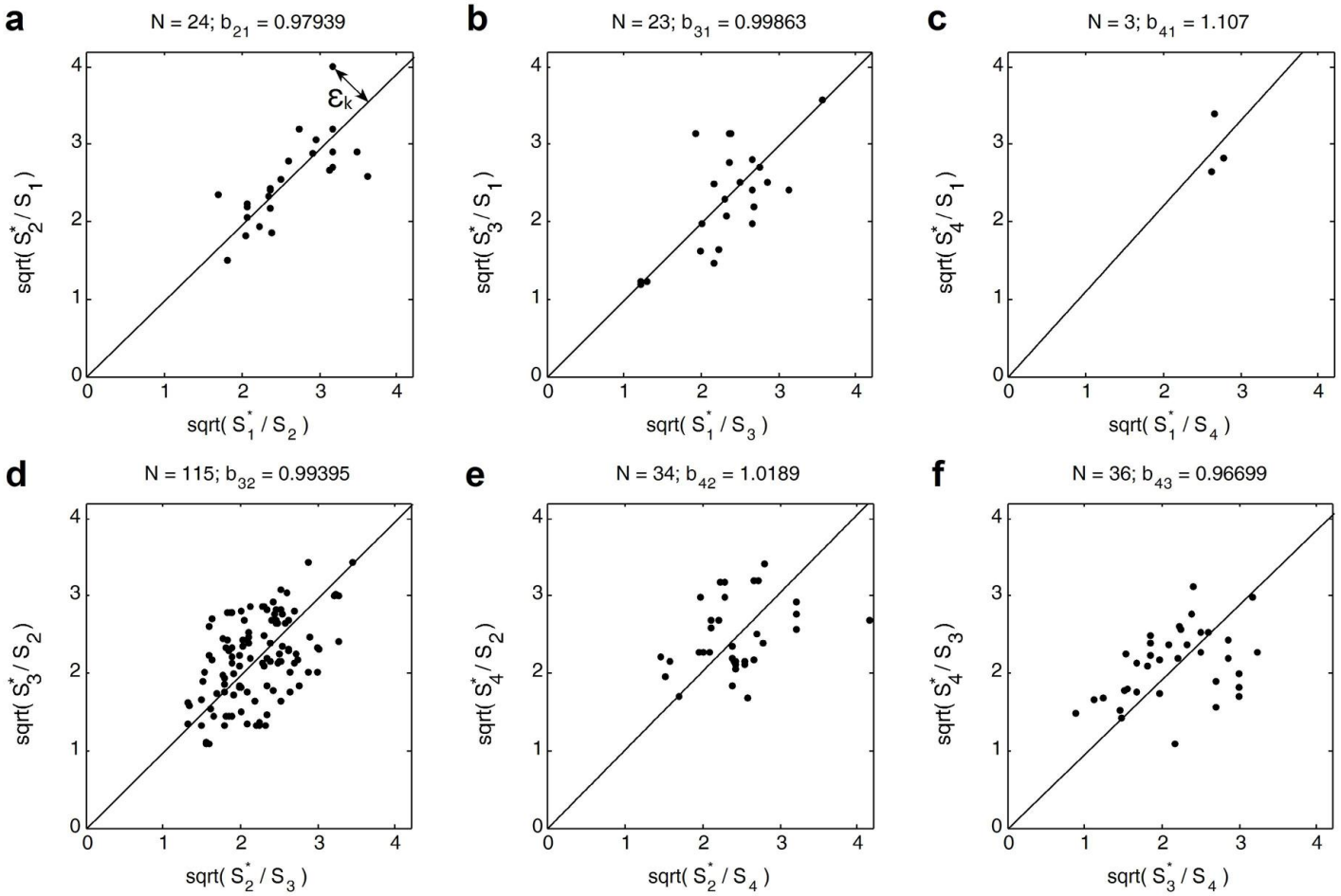
(**a**) Correlation diagram in the pair 3-4 that tended to co-sing. The color encodes the PCCs and white asterisks indicate their significance (*P < 0.05, **P<0.001). Note that all song vocalizations (syllables and introductory notes) are positively correlated to an approximately equal extent. (**b**) Correlation diagram between different vocalizations (calls, introductory notes and syllables) in pair 2-4 that avoided singing together. Red arrows and red asterisks indicate significant difference2 between PCCs. Animal 2 is able to discriminate syllables A-B and A-C in the song of animal 4: it begins to sing (an introductory note) much less frequently during syllable C than during syllable A.



Supplementary Figure 14

Sound amplitudes of vocalizations measured with near and far backpack microphones and distances between animals during songs and calls

Four panels (**a-d**) show median sound amplitudes during vocalizations of birds 1-4 recorded with microphones 1-4. Standard errors are estimated by bootstrapping. Sound amplitudes are reported in units of air pressure mPa (1 mPa = 10^{-3} H/m^2). Note that on average song vocalizations were 3.16 ± 0.80 times louder than calls (amplitude, mean ± s.e.m). The ratio of amplitudes on vocalizing and listening birds' microphones was used to estimate the distance between animals, assuming a fixed distance between the beak and the backpack microphone in the vocalizing bird (~3 cm). The four panels (**e-h**) show median distances between vocalizing birds 1 - 4 and all listeners when the vocalizing animal was either calling or singing. The shown standard errors were estimated by bootstrapping. Note that bird 1 kept a large and similar distance with all other birds during calling and singing (**e**). Bird 2 was closer to other birds during calls, but when singing it was closer to bird 1 and further away from bird 4 (**f**). The co-singing birds 3 (**g**) and 4 (**h**) were closer to other animals during singing than during calling. *P < 0.05, **P < 0.001, bootstrap.



Supplementary Figure 15

Self consistency of sound amplitude ratios validates distance estimation approach

Six panels (**a – f**) represent sound amplitude ratios on calibrated microphones in all bird pairs: 1-2, 1-3, 1-4, 2-3, 2-4, and 3-4. Each dot represents two notes (produced by two different birds) separated by a small time interval (<0.6 s between note onsets). During these quasi-synchronous notes we considered the positions of animals as fixed. The horizontal axis reports the square root of the ratio S^*_i/S_j of microphone amplitudes, where the amplitude S^*_i on the vocalizing animal's back-attached microphone is marked by an asterisk. The vertical axis displays the corresponding ratio when the other animal is vocalizing. The coefficient b_{ij} is the regression coefficient estimated by minimizing the sum of squares of residuals ε_k, k = 1,…,*N* (see panel **a**), where *N* indicates the number of quasi-synchronous notes. Coefficients b_{ij} are close to 1, indicating validity of our calibration approach, but significant spread of points reveals low accuracy of one-shot measurements. For distance estimation we assumed that both backpacks had equal distances to the beaks of carrying birds and that these distances were much smaller than the distances between animals.

Supplementary Table 1. **Number of vocalizations per type in four animals g2k8, r17y20, o6r11 and r14y4 in six sessions "Quiet", "Playback 1x", "Playback 10x", "Noise 50 dB", "Noise 60 dB" and "Noise 70 dB"**

Bird	Vocalization	Session					
		Quiet	Playback 1x	Playback 10x	Noise 50dB	Noise 60dB	Noise 70dB
g2k8	Distance call	11	6	0	9	4	14
	Stack	91	36	16	65	37	29
	Tet	395	170	89	249	165	174
	Introductory	1497	399	273	910	587	610
	Syllable A	508	133	97	273	198	242
	Syllable B	481	121	96	262	192	236
	Syllable C	264	59	45	148	103	116
	Total	**3247**	**924**	**616**	**1916**	**1286**	**1421**
	RMS misdetections	427	377	497	800	1269	2559
	RMS playback det.		0	22			
	Misclassifications (Calls/Syllables)	**10 / 2**					
r17y20	Distance call	0	0	0	0	0	0
	Stack	0	5	12	0	0	0
	Tet	4	5	5	1	2	0
	Introductory	41	72	19	5	12	0
	Syllable A	108	123	17	12	172	0
	Syllable B	108	121	17	11	176	0
	Syllable C	91	110	15	9	169	0
	Syllable D	85	105	14	8	166	0
	Total	**352**	**436**	**85**	**38**	**531**	**0**
	RMS misdetections	111	384	2540	1645	4101	3698
	RMS playback det.		0	56			
	Misclassifications (Calls/Syllables)	**0 / 2**					
o6r11	Distance call	25	0	0	0	0	0
	Stack	32	10	1	1	2	1
	Tet	58	1	0	0	0	0
	Introductory	335	0	0	0	0	0
	Syllable A	283	0	0	0	0	0
	Syllable B	283	0	0	0	0	0
	Syllable C	283	0	0	0	0	0
	Total	**1299**	**11**	**1**	**1**	**2**	**1**
	RMS misdetections	151	85	1143	128	770	936
	RMS playback det.		0	87			
	Misclassifications (Calls/Syllables)	**5 / 0**					
r19y4	Distance call	0	0	0	0	0	0
	Stack	45	26	18	0	5	4
	Tet	0	1	4	0	0	0
	Introductory	34	0	5	8	0	0
	Syllable A	54	0	5	23	0	0
	Syllable B	69	0	6	28	0	0
	Syllable C	14	0	0	10	0	0
	Total	**216**	**27**	**38**	**69**	**5**	**4**
	RMS misdetections	307	396	930	728	2164	1513
	RMS playback det.		0	166			
	Misclassifications (Calls/Syllables)	**12 / 1**					

Numbers of misclassifications were computed only in "Quiet" session because we estimated reliability of accelerometer-based classification relatively to the wall microphone that we used as "gold standard". No mixing of playback with animal vocalizations was detected; and, all counted mistakes were false classification within birds (no erroneous assignment was made of a vocalization to a different animal). During playback complete sessions were played back (including background noise). RMS playback detections took place only in the "Playback 10x" session. They constitute a small fraction of the total number of RMS misdetections indicated in the table.

Supplementary Table 2. **Pearson correlation coefficients *r* between calls or songs of different birds; and corresponding probabilities**

(**a**)

	Bird 2	Bird 3	Bird 4
Bird 1	**0.11 ± 0.01**	**0.041 ± 0.007**	**0.020 ± 0.008**
Bird 2		**0.106 ± 0.008**	**0.030 ± 0.007**
Bird 3			**0.017 ± 0.007**

(**b**)

	Bird 2	Bird 3	Bird 4
Bird 1	**4.8×10^{-30}**	**7.0×10^{-9}**	**0.019**
Bird 2		**4.7×10^{-38}**	**2.3×10^{-5}**
Bird 3			**0.014**

(**c**)

	Bird 2	Bird 3	Bird 4
Bird 1	**- 0.018 ± 0.008**	0.0026 ± 0.0112	- 0.0093 ± 0.0097
Bird 2		**- 0.020 ± 0.008**	**- 0.023 ± 0.008**
Bird 3			**0.104 ± 0.015**

(**d**)

	Bird 2	Bird 3	Bird 4
Bird 1	**0.035**	0.86	0.35
Bird 2		**0.022**	**0.0072**
Bird 3			**1.3×10^{-11}**

(**a**) Pearson correlation coefficients between calls. (**b**) Significance of correlation between calls. (**c**) Pearson correlation coefficients between song syllables. (**d**) Significance of correlation between song syllables. Values for bird pairs with significant ($P < 0.05$) Pearson correlation are highlighted in bold face.

Supplementary Table 3. **Number and percentage of calls responded with other calls**

(**a**)

Bird calling	Number of calls of calling bird	Number of answered calls in birds 1 - 4			
		Bird 1	Bird 2	Bird 3	Bird 4
Bird 1	633	56	406	239	76
Bird 2	2909	281	574	925	251
Bird 3	2963	160	871	468	209
Bird 4	879	54	211	233	72

(**b**)

Bird calling	% of answered calls in birds 1-4			
	Bird 1	Bird 2	Bird 3	Bird 4
Bird 1	8.8468	**64.1390**	**37.7567**	12.0063
Bird 2	9.6597	19.7319	**31.7979**	8.6284
Bird 3	5.3999	**29.3959**	15.7948	7.0537
Bird 4	6.1433	24.0046	**26.5074**	8.1911
% of random coincidences	4.2200	19.3933	19.7533	5.8600

(**c**)

Bird calling	% of answered calls in birds 1-4 above random level			
	Bird 1	Bird 2	Bird 3	Bird 4
Bird 1	4.6268	**44.7457**	**18.0034**	6.1463
Bird 2	5.4397	0.3385	**12.0445**	2.7684
Bird 3	1.1799	**10.0025**	-3.9585	1.1937
Bird 4	1.9233	4.6112	**6.7541**	2.3311

(**a**) Number of calls produced and answered in 2.5 morning hours excluding the first 25 minutes when vocal activity was low. A call is considered an answer when it follows a transmitter call (of another animal) within less than 0.5 s. Diagonal elements represent paired calls in a bird. (**b**) Reported are the percentage of answered calls and the expected percentage for each responder resulting from random calling in the first bird ($P(\%) = 0.5 \cdot N_2/T_{session} \cdot 100$, where N_2 - number of call of answering animal, $T_{session}$ - duration of session in seconds). Values in bird pairs with significant (>3 SD) CC peaks in **Figure 3** are highlighted in bold face. Note that the largest values are associated with significant CC peaks. (**c**) Excess percentage of calls answered (beyond random coincidences). Here again the largest values are associated with significant CC peaks in **Figure 3**.

Nature Methods: doi:10.1038/nmeth.3114

Supplementary Note

Syllable recognition

Having revealed that some birds tend to sing together whereas others avoid it (**Fig. 3d**), we were interested whether co-singing and avoidance happens at the level of the song as a whole, or at the level of the song motif. In 2 bird pairs we analyzed song-syllable correlations. We selected the co-singing pair 3-4 and pair 2-4 showing the strongest avoidance of co-singing.

Pairwise correlations between all vocalizations (calls, introductory notes, and individual song syllables) are shown by color coding in the **Supplementary Figure 13**. The co-singing bird pair 3-4 did not lock any of their song vocalizations to each other, the overlap of their vocalizations was flat (**Supplementary Fig. 13a**).

By contrast, bird 2 produced many fewer introductory notes during syllable C than syllable A of bird 4 ($P = 4 \times 10^{-5}$, **Supplementary Fig. 13b**). The difference remains significant after Bonferroni correction with 25 cells ($P = 0.001$). Thus, bird 2 can discriminate syllable A from syllable C in the motif of bird 4, suggesting it is capable of recognizing not only the song of another male as a whole, but also of discriminating syllables inside the motif, manifesting this recognition by a behavioral response (by starting to sing during one syllable more often than during another).

Distances between animals in a group

When communicating vocally with other birds, the emitter may choose not only to produce sounds, but it may also choose the spatial position relative to other birds. Recording of relative positions can help reveal the recipient to whom the vocalization was directed. We measured distances between animals by comparing sound amplitudes on the microphones of vocalizing and listening birds. We exploited the property that sound amplitudes attenuate as ~1/R with distance R from the source. We measured sound attenuation only in isolate notes that did not overlap with other vocal or non-vocal sounds. The procedure of selection of such notes is described in the **Supplementary Note** lower. To avoid possible biases caused by loggers' measurement variability we have found a method to calibrate recorders' sensitivity online during recording. That is, we selected notes produced by a couple of animals within a very small time interval (< 0.6 s) during which we assumed that relative position of the two animals was constant. This allowed us to exclude the unknown distance between animals from the system of equations and compute ratio of sensitivities of two microphones. The sensitivity of the microphone is defined as the ratio of its output (in Volts) to the air pressure on the microphone (in Pascals). The relative sensitivity is defined as the microphone sensitivity divided by the average sensitivity of the four microphones. To increase the accuracy of sensitivity estimates we selected many quasi-synchronous notes ($n = 235$) in all animal pairs ($n = 6$) and computed the least square-error solution of the non-linear equations. We have found that the four microphones used in the study had the following relative sensitivities: 0.92 ± 0.02, 1.22 ± 0.02, 0.89 ± 0.01, 0.95 ± 0.02 (value ± std, standard deviations were estimated by bootstrapping).

Online calibration of sensitivity with accuracy of ~2.3% allowed estimation of sound amplitudes during vocalization. To check whether relative positions of animals during singing differed from positions during calling we computed median amplitudes on all microphones during calls and songs (**Supplementary Fig. 14a-d**). We found that on average, amplitude on the microphone of the vocalizing animal was 118.6 ± 21.8 mPa and thus 4.98 ± 0.27 times larger than on microphones of listeners: 25.8 ± 4.1 mPa (mean ± s.e.m., $P = 1.25 \times 10^{-7}$, two-sample two-tailed *t*-test). Also, on average song vocalizations were 3.16 ± 0.80 times louder than calls (171.7 ± 11.1 mPa and 65.5 ± 15.0 mPa respectively measured on the vocalizing birds, $P = 0.0155$, non-paired two-tailed *t*-test). Thus, 2.3% sensitivity calibration was sufficient for comparison of sound amplitudes during different vocalizations.

Assuming a fixed distance between the beak and the backpack microphone in the vocalizing bird (3 cm), we computed median distances between animals during calls and songs (**Supplementary Fig. 14e-h**). Bird 1 kept a relatively large distance to other animals during calls (19.29 ± 0.51 cm) and songs (17.70 ± 0.75 cm), but these distances did not differ significantly ($P = 0.352$, bootstrap, **Supplementary Fig. 14e**), suggesting that this animal did not direct its song to another bird. Bird 2 previously shown to avoid co-singing with others kept equal distances to other animals when calling 15.39 ± 0.33 cm (paired comparisons: $P_{13} = 0.123$, $P_{14} = 0.237$ and $P_{34} = 0.968$, bootstrap, indexes here and below indicate the number of listener), but preferred to sing closely to bird 1 (10.48 ± 0.47 cm) and avoided singing near bird 4 (18.88 ± 0.62 cm, **Supplementary Fig. 14f**). The probabilities of differences in distances when bird 2 was calling and singing were $P_1 = 2.09 \times 10^{-7}$, $P_3 = 0.879$ and $P_4 = 4.59 \times 10^{-4}$, bootstrap (significant probabilities are marked by two asterisks in **Supplementary Fig. 14f**). Thus, one can suppose that the second animal directed its song to the first animals to some extent. Interestingly, the co-singing birds 3 and 4 sang closer to all other birds in comparison to their calls (**Supplementary Fig. 14g, h**). Bird 3 called on average from distances 15.80 ± 0.44, 13.80 ± 0.41 and 12.42 ± 0.42 cm to three other birds, but sang from significantly closer distances 9.70 ± 0.33, 10.10 ± 0.32 and 9.99 ± 0.40 cm ($P_1 = 5.08 \times 10^{-25}$, $P_2 = 8.48 \times 10^{-13}$ and $P_4 = 1.28 \times 10^{-5}$ respectively, bootstrap). The behavior of the bird 4 was similar. It called from distances 17.84 ± 0.81, 14.50 ± 0.69 and 14.80 ± 0.64 cm and sang from 14.31 ± 0.52, 12.89 ± 0.53 and 10.59 ± 0.53 cm. Its median distance when singing was always smaller then when calling ($P_1 = 4.21 \times 10^{-7}$, $P_2 = 0.0090$ and $P_3 = 3.03 \times 10^{-9}$ respectively, bootstrap). One can

suppose that co-singing birds were not only communicating with to each other, but were more social in general and sang to all other birds as well. Thus, distances between animals nicely complement information about vocalizations, helping to advance hypotheses about social structure of the group.

The minimal detectable distance is limited by assumption that distance between the beak and the microphone of the carrier is smaller than between this beak and microphone of another bird. As the distance between the beak and backpack microphone is ~3 cm, the minimal detectable distance can be estimated ~6 cm. The maximal distance achieved is limited by signal to background noise ratio in listeners' microphones. From the **Supplementary Figure 9a-d** one can see that sound amplitude in listeners is about 10 times lower (-20 dB) than in the vocalizing bird. Thus, distance between birds is roughly 3 cm ×10 = 0.3 m. Amplitude signal-to-noise ratio in listeners is about 10 (20 dB). Thus, maximal distance at which the vocalizing animal can be detected is roughly 0.3 m ×10 = 3 meters. Obviously, if the background noise on the listening bird is lower, the maximal distance will be larger. A 3-meter range should cover the major part of animal interactions when birds have a direct visual contact with each other. When distances between animals are larger, other methods of location determination, like GPS or triangulation with a microphone array, can be used.

Non-vocal responses to the song

As it has been shown earlier, zebra finches in a group produce significant numbers of interactive vocalizations. However, animal responses may not be limited exclusively to calls and songs. It could be that an animal does not respond vocally to another animal, but instead changes its posture or makes some movement in response to a call or a song. To investigate such responses we analyzed body vibrations recorded by the accelerometer in a non-vocal low-frequency range below 300 Hz. We computed root-mean-square (RMS) values of body acceleration in a sliding window around calls and song elements. To analyze exclusively non-vocal responses we selected fragments in which only one animal vocalized at a time. In most birds we detected no statistically significant movement responses. One of the reasons may be that such responses were masked by high background locomotor activity. However, in the pair of co-singing birds, movement responses of bird 3 to the song of bird 4 were statistically significant (**Supplementary Fig. 5**). The movements of bird 3 were synchronized with the first syllable in the motif of bird 4 (**Supplementary Fig. 5a**). The median RMS acceleration value exceeded 3 standard deviations (SDs) in the range from -217 to 40 ms from the onset of syllable A (**Supplementary Fig. 5b**). The standard deviation of the median was estimated using bootstrap. Visual inspection of records reveal that animal 3 reacted either to the beginning of the song of animal 4 (i.e. introductory notes), or to the loudest syllable C preceding syllable A in the sequence of motifs (**Supplementary Fig. 5a**). However, such responses were not detected in all renditions. As one can see in a sample record in **Supplementary Figure 5a**, responses to syllable C decreased from the first motif to the last, showing habituation of animal 3 to the song of animal 4. Analysis of video records has shown that animal 3 did not respond to the song of animal 4 by some specific behavior. Rather, its stereotyped motor behaviors such as cleaning feathers with the beak, pecking or jumping were temporarily locked to the loudest notes in the song of bird 4. Thus, it was not a specific behavioral response but a locking of ordinary movements to the external audio stimuli, as observed earlier in parrots[11,12], as well as in some other animals capable of vocal mimicry[12] and in humans[13]. Synchronization by visual cues cannot be excluded in the current experiment. The interesting finding is that movement response was observed exclusively in a pair of co-singing birds, but not in others, albeit other birds received nearly identical auditory stimulation (**Supplementary Fig. 5b**). Thus, not only the strength of auditory stimulation is important for movement responses, but the directed attention to the source of stimuli seems to be essential as well. Existence of movement response only in one of the pair of co-singing birds indicates asymmetric nature of relations between these birds. However, summarizing, the non-vocal responses were much less reliable than vocal responses and could reveal only very strong interactions in co-singing birds.

Temporal alignment of backpack records by comparing sound envelopes

Synchronization of vocalization records using IR pulses is impractical in free ranging animals in the wild. For this reason, we developed an alternative method of logger synchronization by comparing sound envelopes recorded with backpack microphones. We have found this latter method to be almost as effective as the synchronization via IR pulses.

First we high-pass filtered the original sound signal above 500 Hz with a finite impulse response (FIR) filter of order 960 in forward and reverse directions to obtain zero time shift. Then we computed the instantaneous sound amplitude by taking the absolute value of the Hilbert transform of the filtered signal. Next, we smoothed the instantaneous amplitude by convolving it with a Gaussian kernel of standard deviation σ = 20 ms (and total window span 100 ms), two times in forward and reverse directions to get zero time shift. Then we binarized the smoothed instantaneous amplitude to a sequence of rectangular pulses of height 1 and width 16 ms, each centered on a local maximum of the instantaneous amplitude exceeding 10 μV (the binary function was zero at all other time points). Then we matched these pulses similarly as we did with the IR pulses above. Such binary signals were noisier than IR pulse sequences; to obtain reliable alignment we had to increase the size of alignment frames from 20 s to 60 s (these larger frames led to 2.67 hours of computation time, i.e. 7 times longer than was needed for IR synchronization). Results are shown in **Supplementary Figure 12c, d**. The residuals of clock-drift regression were similar for IR and sound synchronization procedures (**Supplementary Figure 12** panel (**b**) versus (**d**)). In conclusion, both methods of record alignment provided sub-millisecond accuracy sufficient for precise determination of vocal responses latencies.

Measurement of perceived vocalization loudness and estimation of distances between animals

Animals can move rapidly inside the large experimental chamber of 60x60x50 cm. The perception of vocalizations may depend on proximity of the vocalizing animal to the listener. The perception of vocalization may also depend on perceived loudness of the vocalization. To reveal such dependencies we recorded vocalizations using the backpack-attached microphones. We estimated distances between animals by attenuation of sound amplitude assuming a free space in which sound amplitudes decreases with distance R as ~$1/R$. In a closed chamber this relation is approximate because of interference and multiple reflections from the walls. Thus, it is better to refer to this measurement as "vocal distance", emphasizing the degree of sound attenuation in the listeners' location as primary measurement variable.

One should note that video tracking could be used in the current conditions to record animal locations. However, here we wanted to test a method potentially applicable to naturalistic settings in which birds cannot be observed by a video camera. We used omnidirectional microphones with practically ideal spherical sensitivity (see the polar diagram in http://www.avisoft.com/usg/KnowlesFGO.htm for a similar microphone). Thus, recorded sound amplitudes should not depend on the listeners' body orientation (microphone orientation) relatively to the sound source. Presence of the animal head in the vicinity of microphone can potentially shade the signal at microphone. However, the wavelengths of typical vocalization frequencies 0.5-8 kHz lie in the range 4.125-66 cm and significantly exceed the size of the animals' heads, thus the influence of shading can be neglected. The used microphones FG-23329-D65 have factory-declared mean sensitivity -54 dB relative to 1V/0.1Pa, but may deviate from this value up to ±3 dB (-29/+41 % in amplitude). To improve the accuracy of measurements we calibrated four microphones used in the current study in the frequency range 500-7000 Hz and have found that their amplitude sensitivity varied in the range ±10%. The determined sensitivity coefficients defined as ratio of microphone sensitivity (in V/Pa) to average sensitivity of four microphones (1.0956, 1.0573, 0.9026, 0.9445) were used to improve the accuracy of the analysis. We estimated loudness of each vocalization above 500 Hz by computing the mean power in the interval -26.7/140 ms from note detection time. Only notes that did not overlap with notes of other birds were selected for loudness estimation. We considered the note as non-overlapping if no note onsets in other birds were detected in the interval -300/140 ms from the note detection time. During the 2.5-h recording session 3407 ± 556 (mean ± s.e.m.) non-overlapping vocalizations were detected per bird. These constitute 59.0 ± 2.4 % of all vocalizations. Percentages of non-overlapping introductory notes and syllables were 65.2 ± 3.2 % and 65.0 ± 3.3 %. They did not differ significantly (P = 0.93, paired two-tailed t-test applied after normalization by the Fisher transform). However, percentage of non-overlapping calls was smaller (46.4 ± 1.9 %) and differed from that of introductory notes and syllables (P = 0.053 and P = 0.043 respectively, calls against all song vocalizations, P = 0.046).

To assess possibility of distance estimation by sound attenuation we measured the power of background noise in 10000 randomly chosen 166.7-ms time intervals that did not overlap with vocalizations, and compared it with the signal power at microphones of listeners and vocalizing birds in non-overlapping notes. We have found that the amplitude of background noise was 47.2 ± 3.9 µV (median value in 10000 intervals and ± s.e.m. between birds), whereas amplitude in listeners was 547.9 ± 37.2 µV and in vocalizing birds it was 2548.5 ± 542.9 µV. These values were significantly different from each other ($P = 8.48 \times 10^{-4}$ and P = 0.0191 in listeners and vocalizing animals against background noise, P = 0.0300 in vocalizing birds against listeners, paired two-tailed t-test), suggesting that distance estimation from sound amplitude attenuation should be possible.

Note that non-stationary noises could disrupt distance estimates. The stationary background noise that consists of internal noise of the recorder and noises from the chamber ventilation and perhaps animal breathing can be easily subtracted from the spectrograms. Cancellation of such noise has been done by power subtraction (median of 10000 non-vocal intervals) from the signal power in the spectrogram.

Notes contaminated with movement artifacts (wing flaps, pecking, etc.) were removed from the analysis. We detect artifacts by measuring the power of residuals after subtraction of the scaled spectrogram of the vocalizing bird from the spectrogram of the listener. We took only the part of spectrogram that covers frequencies above 500 Hz, the dimension of this part was 244 × 24. We took the scaling coefficient that minimizes the power of residuals, i.e. provides the best match. Then we took the square root of the ratio of residual variance and power of the scaled signal of the vocalizing bird (i.e. noise to signal amplitude ratio). We normalized the distribution of such values by logarithmic transformation. After this transformation the bell-shaped histogram becomes symmetrical but still has many outliers on the positive side. We estimated the center of the distribution by the median and the standard deviation by the percentile position of 15.9% (1 STD for normal distribution). We rejected from the analysis all notes that exceeded 1.6449 STDs to the right from the distribution center (5% of normal distribution values would fall in this range). Such criterion lead only to a very moderate bias of the population mean and practically did not shift the median value that was used because of its small sensitivity to outliers. Visually detectable noisy fragments in the spectrogram were all rejected by this procedure. We found that 7.79 ± 0.47 % of notes exceeded the 1.6449 STDs threshold because of presence of outliers (17.38 ± 2.12 % calls and 4.43 ± 0.57 % song vocalizations). We selected for the analysis only notes that were not rejected in any of the three listeners. Thus, the percentage of cumulatively rejected notes was 23.66 ± 1.14 % (47.73 ± 5.01 % calls and 14.97 ± 1.73 % song vocalizations). Percentage of rejected calls was significantly larger than percentage of rejected song vocalizations (P = 0.018 and P = 0.017 for rejection in one listener and in all three listeners respectively, two-tailed paired t-test was applied after normalization

by the Fisher transform). Such bias can be explained by softer calls relatively to song vocalizations. The remaining number of vocalizations per bird was 2579 ± 828 (mean ± s.e.m.), constituting 44.9 ± 3.2 % of all vocal elements and consisting of 444 ± 151 calls and 2135 ± 697 song elements (23.9 ± 1.9 % and 55.6 ± 6.3 % respectively).

Accuracy of vocal distance estimation and correction of microphone calibration

To verify the validity of our distance estimation by sound amplitude attenuation we selected pairs of vocalizations within less than 0.6 s in every pair of animals (1-2, 1-3, 1-4, 2-3, 2-4, 3-4). We took only vocalizations that did not overlap and were not contaminated by noise. For the pairs listed above we found the following numbers of such vocalizations: 24, 23, 3, 115, 34, and 36. To verify estimated distances, we assumed that locations of animals usually do not change considerably during such a short time. And even if it does, the direction of this change can be considered random and should not bias the estimates based on a large set of samples. Thus, we assumed that the distance from singing animal 1 to listener 2 was identical to the distance from singer 2 to listener 1 at short times (<0.6 s). We define the amplitudes of vocalizations at the animals' beaks A_1 and A_2. The note of the first animal produces a response on its microphone of $S_1^* = A_1 \cdot \beta_1/L_1$, where β_1 is the sensitivity of the microphone on its backpack and L_1 the distance from the animal beak to this microphone. We denote the signal at the microphone of the vocalizing animal by an asterisk. At the same time the microphone of the second animal detects the signal $S_2 = A_1 \cdot \beta_2/L_x$, where β_2 is the sensitivity of the microphone of the second backpack and L_x the distance between animals. For simplicity we assume that $L_1 << L_x$. Dividing the first equation by the second we get the following ratio of microphone signals: $S_1^*/S_2 = L_x/L_1 \cdot \beta_1/\beta_2$. Similarly, we can write the analogous equation for the signals on two microphones when the second animal vocalizes as $S_2^* = A_2 \cdot \beta_2/L_2$ and $S_1 = A_2 \cdot \beta_1/L_x$, where S_2^* is signal on the microphone of the second vocalizing animal, S_1 the synchronous signal on the microphone of the first animal, and L_2 the distance from the animal beak to the microphone of the second animal, to obtain the ratio of signals on microphones $S_2^*/S_1 = L_x/L_2 \cdot \beta_2/\beta_1$. Dividing the second ratio by the first and taking the square root we get the following ratio of sensitivities of two microphones:

$$\frac{\beta_2}{\beta_1} = \frac{\sqrt{\frac{S_2^*}{S_1}}}{\sqrt{\frac{S_1^*}{S_2}}} \cdot \sqrt{\frac{L_2}{L_1}}. \qquad (1)$$

Because all loggers were attached in a similar way and the sizes of animals were similar, we assume that all animals maintained identical distances from their beaks to their backpack microphones, i.e. $L_1 = L_2$. In this case the ratio of microphones sensitivities is just equal to the ratio of square roots of amplitude attenuation from the vocalizer's microphone to the listener's microphone. Because we had many recordings of quasi-synchronous notes, we estimated the ratio β_2/β_1 by minimizing the sum of squares of residuals, computed as the shortest average distances ε_k to the line of the best fit (see **Supplementary Fig. 15a**), k = 1, …, N, where N is the number of quasi-synchronous notes. The equation for residuals is:

$$\varepsilon_k = \mathrm{Im}\left(Z_k \cdot \sqrt{\frac{i+b}{i-b}}\right), \qquad (2)$$

where $Z_k = \sqrt{\frac{S_1^*}{S_2}} + i \cdot \sqrt{\frac{S_2^*}{S_1}}$, $b = \frac{\beta_2}{\beta_1}$, $i = \sqrt{-1}$, k = index of the quasi-synchronous pair of notes. Multiplication by the square root above is a turn in the complex plane by the angle $-\varphi$: $\sqrt{\frac{i+b}{i-b}} = e^{-i\cdot\varphi}$, where $\varphi = \mathrm{atan}(b)$. To find microphone sensitivities we wrote similar regressions for all animal pairs, complemented them with the normalization condition $\sum_{i=1}^{4} \beta_i = 4$ and fond the solution by minimizing the sum of squares of all residuals (by an iterative procedure realized with the Matlab function *fminsearch*). We obtained the following β_i for the dataset already corrected by the microphones sensitivity coefficients determined from the static calibration described above: 0.8435, 1.1572, 0.9887, and 1.0106. Thus, the total correction coefficients were: 0.9241, 1.2235, 0.8924 and 0.9545. The obtained coefficients significantly deviated from a unit (up to 22.4%). This may be caused by slightly different protective shieldings of the microphones from direct contact with the bird feathers. Such shieldings were needed to avoid noise from touch with moving feathers. Thus, on-line calibration of sensitivity was important to increase the accuracy of distance estimates. We scaled microphone signals by these coefficients and plotted data points clouds in **Supplementary Figure 15**. As the microphone signals were scaled, the points should aggregate along the diagonal of the first quadrant. In spite of relatively large deviations of the single points from the diagonal, the clouds centers nicely on it. The deviations of single points from the diagonal most probably were caused by the interference of reflected signals from the walls. Because birds changed their positions inside the experimental chamber quite often (see **Supplementary Video 1**), all sound interferences nicely averaged over many samples. To estimate the accuracy of microphone sensitivity estimates we draw a regression line for each animal pair in **Supplementary Figure15**. As one can see, these lines displayed only very small deviations from the diagonals. The worst accuracy was achieved in the animal pair 1-4 in which only 3 quasi-synchronous notes were detected (**Supplementary Fig. 15c**). However, even in this case the regression coefficient b_{41} deviated from 1 only by 10.7%. The estimates of β_i themselves were much more accurate because they were based on all data points of all bird pairs (and not on three

points shown in **Supplementary Figure 15c**). The bootstrap procedure gave the following standard deviations σ_i for β_i: 0.018, 0.015, 0.015, and 0.023. We defined the relative accuracy ε_{ij} of our distance estimate L_x between animals i and j as $\varepsilon_{ij} = 2(L_{xi} - L_{xj})/(L_{xi} + L_{xj})$, where L_{xi} and L_{xj} are distances estimated when animals i and j vocalized quasi-synchronously, $L_{xi} = S_i^*/S_j \cdot \beta_j/\beta_i \cdot L_i$ and $L_{xj} = S_j^*/S_i \cdot \beta_i/\beta_j \cdot L_j$. Assuming as before $L_i = L_j$ we see that ε_{ij} depends on sound amplitudes S_i^*, S_j, S_j^*, S_i, and microphone sensitivities β_i, β_j. The widely distributed clouds of points in Supplementary Figure 5 reveal that uncertainty in sound amplitudes from single measurements contribute more to relative accuracy than a 2% error in microphone sensitivity. When we estimated relative accuracy ε_{ij} from single pairs of quasi-synchronous vocalization (assuming fixed microphones sensitivities β_i, β_j, **Supplementary Fig. 15**), we obtained the following averages $\langle\varepsilon_{ij}\rangle$ for bird pairs 1-2, 1-3, 1-4, 2-3, 2-4 and 3-4: 0.64, 0.45, 0.26, 0.45, 0.43 and 0.43. Taking into account that $\langle\varepsilon_{ij}\rangle$ and standard deviations of β_i are small (< 1), and assuming that measurement errors are independent, we estimated the standard error of $\langle L_x\rangle$ from N samples as $\sigma_{ij} = \sqrt{\langle\varepsilon_{ij}\rangle^2/N + \sigma_i^2 + \sigma_j^2}$, where σ_i and σ_j are estimated standard deviations of β_i and β_j, respectively. Because the typical value of $\langle\varepsilon_{ij}\rangle$ is about 0.45 and typical values of σ_i and σ_j are 0.02, the first term dominates the sum for $N < 254$. Our estimate of median distance between birds had an accuracy of $\sigma_{ij} \simeq 3\%$ (large N).

Measurement of non-vocal responses

We wanted to test whether listening to songs or calls evokes some motor response in addition to vocal responses. We took the root mean square (RMS) value of low-pass filtered (<300 Hz) accelerometer signals to estimate the strength of non-vocal responses in listening animals. First, we filtered the signal with a finite impulse response (FIR) filter with a span 50 ms in both forward and reversed directions to get zero time shift. Then we convolved the squared filtered signal with a Gaussian kernel of width $\sigma = 25$ ms in a window of 125 ms width. The square root of the resulting smoothed signal was our estimate of low-frequency acceleration of the animal.

Because vocalizations produce body vibrations in the low-frequency range, to measure non-vocal responses we selected for the analysis only episodes in which the listening animal was not vocalizing during ± 2 s from the time point of interest. To measure the reaction to the song motif we aligned all such episodes by the onset of the first syllable in the motif of the singer. As the RMS acceleration signal changes slowly we down-sampled it to 100 Hz to speed up computations. We took the median trace of a set of motifs to estimate the strength of response. The standard errors of medians were estimated by bootstrapping 1000 fragments as before. Low-frequency acceleration values that exceeded 3 STDs from the average acceleration (computed in ± 4 s time window) were considered significant. The variability in accelerometer sensitivity was eliminated by normalizing song-locked movement estimates in **Supplementary Figure 5b** by the average movement estimates in 200 randomly selected 4-s intervals without vocalizations (intervals were selected randomly without overlap). Without such normalization the colored curves in **Supplementary Figure 5b** were slightly separated in Y direction (such separation may either reflect differences in backpack attachment between animals or differences in animal baseline motor activities). The first 25 minutes after light onset were excluded from analysis because vocal activity was significantly decreased in this interval compared to the remaining part of the session. The average baseline movement estimates (RMS acceleration values without vocalizations) in four animals were 0.162, 0.222, 0.263 and 0.330 m/s^2, respectively.

References

11. Patel, A. D., Iversen, J. R., Bregman, M. R. & Schulz, I. Experimental evidence for synchronization to a musical beat in a nonhuman animal. *Current biology* **19**, 827–30 (2009).

12. Schachner, A., Brady, T. F., Pepperberg, I. M. & Hauser, M. D. Spontaneous motor entrainment to music in multiple vocal mimicking species. *Current biology* **19**, 831–6 (2009).

13. Repp, B. H. Sensorimotor synchronization: a review of the tapping literature. *Psychonomic bulletin & review* **12**, 969–92 (2005).