Courtship behaviour reveals temporal regularity is a critical social cue in mouse communication

Catherine Perrodin[1]*, Colombine Verzat[1,2], Daniel Bendor[1]*

[1]Institute of Behavioural Neuroscience, Department of Experimental Psychology, University College London, WC1H 0AP, United Kingdom

[2]Idiap Research Institute, CH-1920, Martigny, Switzerland

*For correspondence:

c.perrodin@ucl.ac.uk (CP)

d.bendor@ucl.ac.uk (DB)

Keywords: natural behaviour, neuroethology, acoustics, social, mice, sequence processing, perception, ultrasound, vocalizations, auditory

Summary

While animals navigating the real world face a barrage of sensory input, their brains evolved to perceptually compress multidimensional information by selectively extracting the features relevant for survival. Notably, communication signals supporting social interactions in several mammalian species consist of acoustically complex sequences of vocalizations. However, little is known about what information listeners extract from such time-varying sensory streams. Here, we utilize female mice's natural behavioural response to male courtship songs to identify the relevant acoustic dimensions used in their social decisions. We found that females were highly sensitive to disruptions of song temporal regularity, and preferentially approached playbacks of intact over rhythmically irregular versions of male songs. In contrast, female behaviour was invariant to manipulations affecting the songs' sequential organization, or the spectro-temporal structure of individual syllables. The results reveal temporal regularity as a key acoustic cue extracted by mammalian listeners from complex vocal sequences during goal-directed social behaviour.

Animals behaving in the real world need to efficiently discriminate and monitor relevant sensory input, such as information communicated by conspecifics, from a constantly evolving, complex and multidimensional barrage of stimulation. In many social species, including humans, vocal communication takes the form of time-varying sequences of acoustic elements and demands a rapid, socially-appropriate behavioural response. Yet how the listener's brain effectively extracts salient information from acoustically complexity vocal patterns during social goal-directed behaviour remains unclear. Here, we asked which acoustic cues from communication sequences were informative to female mouse listeners when responding to male courtship songs.

Mice, like most other mammals, use complex acoustic patterns to communicate with each other in a number of different social contexts[1,2]. In particular, adult males produce rhythmic sequences of discrete, frequency-modulated pure tone elements in response to sensing the recent presence of a fertile female[3-5]. These courtship songs are attractive to sexually receptive females[6], who respond to song playback with approach behaviour[7-12]. Male songs facilitate copulatory success[13] and are thought to serve as fitness displays[14,15].

Female mice are able to detect and use acoustic information in male vocalizations. On the one hand, reinforcement learning experiments have shown that mouse listeners are able to detect and report spectro-temporal differences between individual ultrasonic call elements[16,17], indicating that mice are *perceptually* sensitive to a wide range of acoustic features in the vocalizations. On the other hand, experiments using ethologically-relevant place preference paradigms without external reward have shown that females use, or are *behaviourally* sensitive to, certain acoustic characteristics in male songs to guide social decisions, such as discriminating between social contexts[9], singer species[10], strain[18,19] and kin[20]. However, which of the many acoustic cues available in male courtship songs are used by motivated female listeners during natural behaviour remains unclear.

In this study, we asked how listeners process and use complex vocal sequences during natural behaviour. We exploited female mice's natural behavioural response to playbacks of male courtship songs in order to independently manipulate and identify which of several candidate acoustic features are monitored by the listeners. We found that female approach behaviour was highly sensitive to disruptions of song temporal regularity, but that it was not affected by global manipulations of the songs' sequential structure, or the local removal of syllable spectro-temporal dynamics. The results highlight temporal regularity as a key social cue monitored by female listeners during goal-directed social behaviour.

Results

Acoustic characteristics of vocal sequences emitted by male C57Bl/6 mice

In order to study female listeners' behavioural response to vocal sequences, we first collected a database of male C57Bl/6 mice vocalizations for acoustic stimulation in playback experiments. Following a previously published protocol[5], we obtained audio recordings of ultrasonic vocalizations emitted by males in response to the presentation of mixtures of urine samples collected from conspecific females in estrus. This paradigm was deemed optimal for generating stimuli to use in ethologically-inspired playback experiments, as vocalizations emitted in this context are emitted by solitary males and reflect long-range acoustic "courtship" displays with the purpose of attracting a female[13,21,22]. The vocal patterns emitted in this context consisted of sequences of individual frequency-modulated calls (see example Fig. 1A, 3A, S1A and [3,9,23-25]). Across the set of recorded male vocalizations, the duration of continuous call elements, or syllables, followed a bimodal distribution, with a majority of short syllables (local maximum = 23ms, 86% (2039/2373) of recorded syllables with durations shorter than 63ms (local minimum)) and a minority of long syllables (local maximum = 88 ms, 14% (334/2373) of recorded syllables with durations longer than 63ms, Fig. 1B). These male calls were restricted to the ultrasonic frequency range, with the maximal energy in individual syllables occurring around 76kHz (median peak frequency and 95% confidence interval (C.I.), 76 ± 0.3 kHz, n = 2373 syllables) and spanning a bandwidth of 16 kHz ([68-84 kHz]; Fig. 1C). Individual syllables were organized in temporally regular sequences and often started at an approximately constant time delay from each other[3,25], as evidenced by the narrow concentration of inter-syllable interval durations around 104 ± 2 milliseconds (median inter-syllable interval and 95% C.I.; interquartile range of syllable interval = 61ms, n = 2322 syllables followed by another syllable within 2 seconds; Fig. 1D). From this set of ultrasonic vocalizations, we selected a smaller, representative sample for use in subsequent playback experiments (n = 957 syllables, dark grey histograms in Fig. 1B-D). This stimulus set was composed of 7 distinct songs produced by C57Bl/6 male mice aged 22 weeks on average

(median age, range [10-27 weeks]) in response to the presentation of female urine. The songs varied in duration, lasting on average 33.4s (median duration, range [13-41s]), and were composed of an average of 145 syllables (median, range [95-186]).

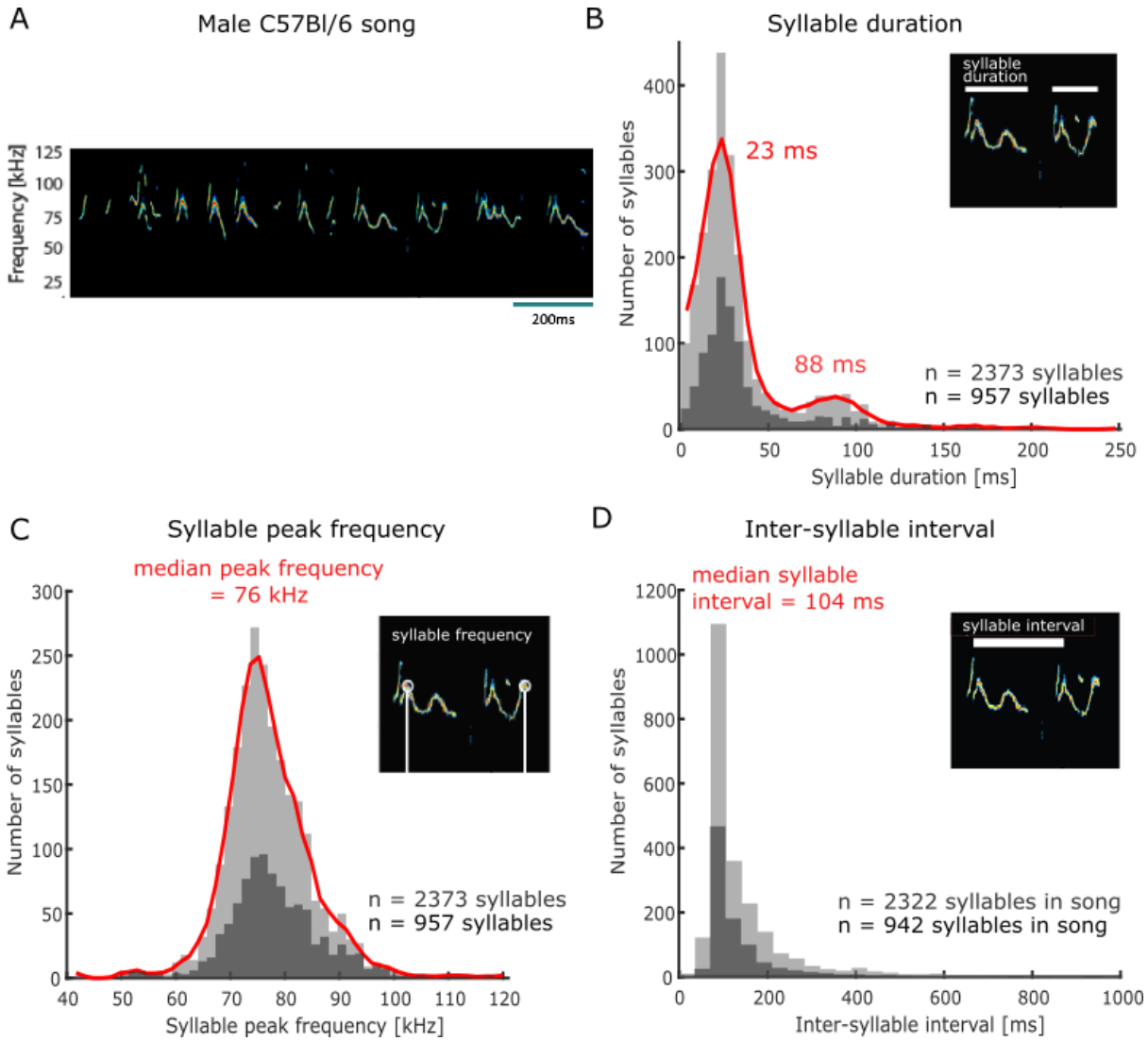

Figure 1: Acoustic features of C57Bl/6 mouse courtship songs

A. Spectrogram of a segment of ultrasonic male vocalizations used for stimulation

B. Distribution of individual ultrasonic syllable durations across a large set of male mouse vocalizations (n = 2373 syllables, light grey histogram; red line is smoothed distribution), including the stimulus set used in subsequent playback experiments (n = 957 syllables, dark grey), emitted in response to the presentation of urine samples from females in estrus.

C. Distribution of syllable peak frequency (point of maximum amplitude across the call element in kHz) across all recorded syllables (n = 2373, light grey), and the subset of syllables used for playback (n = 957, dark grey).

D. Distribution of inter-syllable interval durations. The analysis was restricted to syllables with a subsequent syllable starting within 2 seconds (all recorded syllables: n = 2322, light grey; playback stimulus set: n = 942, dark grey).

Female mice preferentially approach playbacks of male songs over silence

Using our pre-recorded subset of male vocal sequences for stimulation, we then aimed to confirm that female mice displayed a preferential approach response to male songs in our ethologically-inspired behavioural paradigm. We used a place preference assay to evaluate female listeners' behaviour in response to the playback of our pre-recorded song set in a two-compartment behavioural chamber (Fig. 2A). Importantly, olfactory stimulation from mixed male bedding samples placed under each of the loudspeakers was present throughout the testing session, in order to increase sexual arousal and motivation in the female listeners[11,18] and create a multisensory coherent approximation of a natural social situation. Following a silent habituation period during which the animal was free to explore the behavioural box (at least 10min; grey traces in Fig. 2B, C), one-sided playback of male vocalizations was initiated while the second speaker remained inactive (red traces in Fig. 2B, C). The pre-recorded set of 7 songs was repeated 4 times during one behavioural session, with the order of individual songs randomized in each trial, and the playback side alternating between trials (Fig. 2B, C). A female mouse's approach response to song playback was measured by quantifying the time the animal spent within a "speaker zone" over the course of a behavioural session (white (grey) trapezoidal outlines in Fig. 2A (C), respectively). Young female C57Bl/6 mice (aged 5-11 weeks; Fig. S3) participated once in the playback experiment, while in the fertile (proestrus or estrus) stage of their estrous cycle (as assessed by vaginal cytology, see Methods).

In the example session illustrated in Fig. 2, the mouse occupied both speaker zones equally during the silent baseline (Fig. 1D) and the first sound playback trial (red trace in Fig. 2F, I). Over the course of the 3 subsequent trials, the mouse displayed a consistent preferential approach towards the source of male song playback over the silent speaker, irrespective of the playback side (blue, purple and green traces, Fig. 2F, I and video files S2A, B). An animal's approach behaviour over the course of an entire session can be quantified using a preference index (see Methods and Fig. 2E, G), that accurately captures the cumulative preferential dwell time at the song playback vs silent speaker (see correspondence between preference index values in Fig. 2G and endpoints of trial temporal profiles in Fig. 2J). This preference index metric was also shown by a number of previous studies to be the most sensitive readout for similar assays of female approach to male song playback[10,11,18]. In this example, the mouse preferentially approached the song playback over silence in ¾ trials, resulting in a preference index value of 25.3% (median of 4 trials; black dot in Fig. 2G). The behaviour of female listeners tested in this paradigm significantly discriminated between song playback and silence (n = 29 sessions, mean preference index significantly different from 0, one-sample *t*-test, $t(28) = 2.81$, $p = 0.0088$; Fig. 2H), and mouse listeners were overwhelmingly attracted to the playback of our subset of male songs. The preference for song over silence was robust to different lengths of the speaker zone used to calculate the preference index (Fig. S4A, B). Mouse listeners demonstrated significant approach response after as little as 16.9 seconds of sound playback across the 4 sound presentation trials (normalized cumulative time in speaker zone significantly different from 0, one-sample t-test, $p<0.05$; Fig. 2K), with preference further increasing over the course of song playback. The build-up of preferential approach to song was evident in the vast majority of animals, mostly sustained over a behavioural session (Fig. 2L), and similar across the 4 sound presentation trials (Fig. S4B). Together, these results replicate previous work by other laboratories[7-11] and confirm the feasibility of recreating an ethologically-driven behavioural assay of female approach response to male song playback in the controlled laboratory environment.

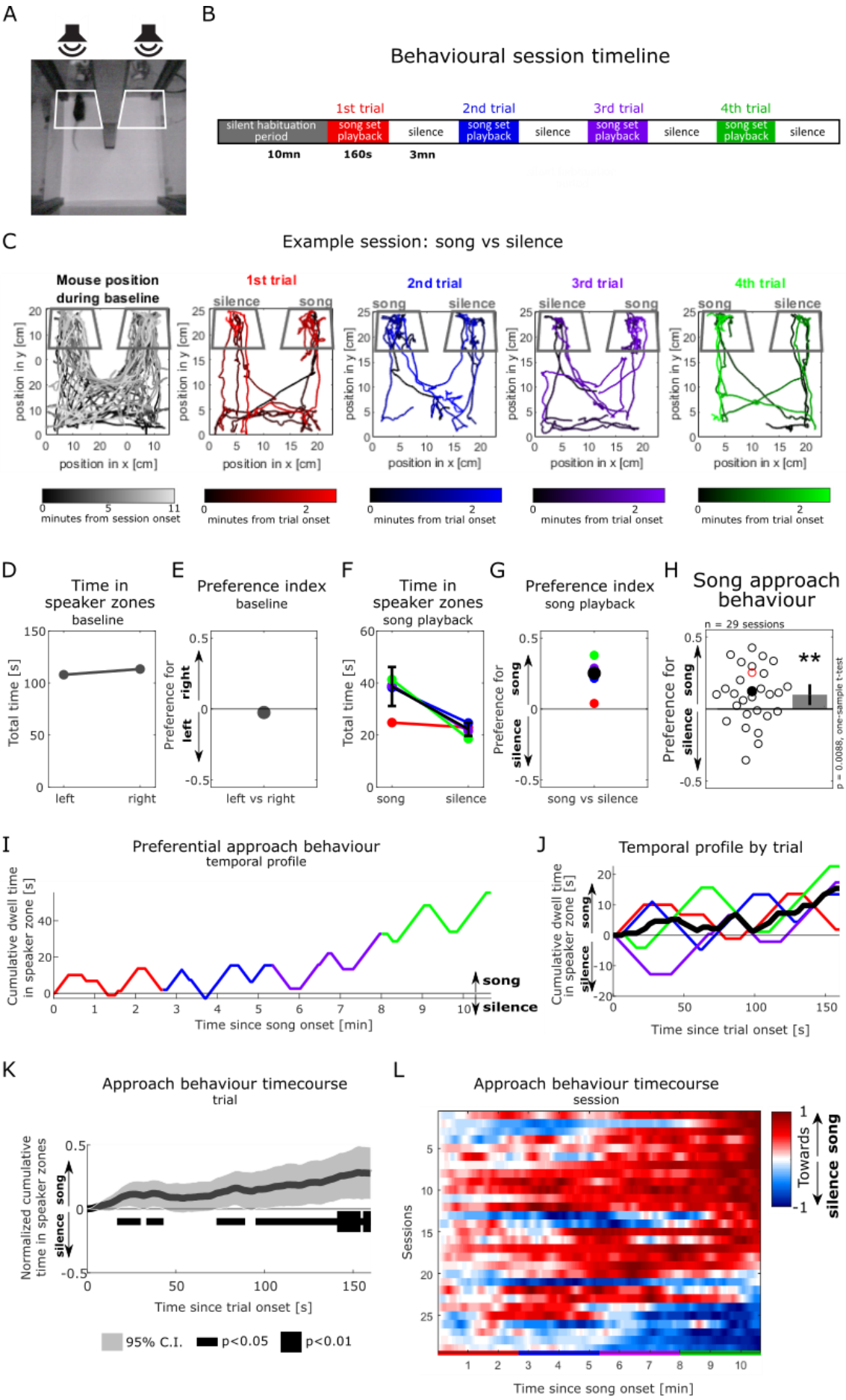
A
B
Behavioural session timeline
1st trial
2nd trial
3rd trial
4th trial
silent habituation period
song set playback
silence
song set playback
silence
song set playback
silence
song set playback
silence
10mn
160s
3mn
C
Example session: song vs silence
Mouse position during baseline
1st trial
2nd trial
3rd trial
4th trial
silence
song
position in y [cm]
position in x [cm]
minutes from session onset
minutes from trial onset
D
Time in speaker zones
baseline
Total time [s]
left
right
E
Preference index
baseline
Preference for
left
right
left vs right
F
Time in speaker zones
song playback
Total time [s]
song
silence
G
Preference index
song playback
Preference for
song
silence
song vs silence
H
Song approach behaviour
n = 29 sessions
Preference for
song
silence
**
p = 0.0088, one-sample t-test
I
Preferential approach behaviour
temporal profile
Cumulative dwell time in speaker zone [s]
song
silence
Time since song onset [min]
J
Temporal profile by trial
Cumulative dwell time in speaker zone [s]
song
silence
Time since trial onset [s]
K
Approach behaviour timecourse
trial
Normalized cumulative time in speaker zones
silence
song
Time since trial onset [s]
95% C.I.
p<0.05
p<0.01
L
Approach behaviour timecourse
session
Sessions
Towards
silence
song
Time since song onset [min]

Figure 2: Female mice preferentially approach playbacks of male songs over silence

A. Video frame showing the testing box with a soundproof partition (middle) positioned between two ultrasonic loudspeakers. White outlines show the two “speaker zones” used as regions of interest for quantifying the animal’s position.

B. Experimental timeline

C. Tracking of the animal’s position during an example behavioural session, in which a silent baseline period (leftmost panel) is followed by the playback of intact songs from one side contrasted with silence from the other side (coloured panels). Colour saturation indicates time since start of the experimental period of interest. Dark grey outlines indicate the speaker zones.

D. Time the animal spent in the speaker zones during the silent baseline period in C.

E. Index quantifying the animal’s relative place preference to the left vs right speaker zones during the silent baseline. This index was computed as the difference between the time spent in either speaker zones, normalized by the sum of the time in both speaker zones (see Methods). A preference index value close to zero indicates no preference to either side.

F. Time spent in the speaker zones corresponding to song playback and silence during each song presentation trial. Error bars: standard error of the mean.

G. Preference index in response to the playback of intact songs compared to silence over the course of the example session. The preference index is the difference between the time in the song playback and the silent speaker zones, divided by the sum of the time in both speaker zones. A positive index value reflects the animal’s preferential approach to the sources of song playback over silence.

H. Population summary of female approach response to playback of intact male songs (positive values) over silence (negative). Each circle is the preference index displayed by individual animals in one behavioural session (median of 4 sound presentation trials). The red circle corresponds to the example session in C & G. Open circles: sound playback at 58 dB SPL, filled circles: 68 dB SPL. Bar plot shows mean preference index across sessions and 95% confidence interval (C.I.). One-sample two-tailed *t*-test, **: $p < 0.01$.

I. Temporal profile of approach behaviour over the four sound presentation trials in the example session in C, calculated as the cumulative sum of time in the intact song playback (positively weighted) vs silent (negatively weighted) speaker zone.

J. Trial-averaged profile of approach behaviour to song playback in the example session, calculated as in I.

K. Population-averaged approach behaviour time course in response to intact mouse songs vs silence, calculated as in J. The dark grey trace indicates the mean of trial-based temporal profiles across all sessions (n = 29). For each session, the median of 4 sound presentation trials (e.g., black trace in J) was normalized to its maximal amplitude. The horizontal black bar indicates time bins during the course of a sound playback trial in which the cumulative approach behaviour significantly deviates from zero (one-sample two-tailed *t*-test).

L. Normalized temporal profiles of approach behaviour to mouse songs vs silence over the course of 4 sound presentation trials (x-axis, coloured bars) for each of the behavioural sessions (y-axis, n = 29), calculated as in I. Sessions (lines) are ordered by the amplitude of their last element.

F, G, I, J: Black traces indicate the session average (median) across the four sound presentation trials (coloured traces)

Female approach behaviour is not affected by changes to global song structure

Given this proof of principle that socially-motivated female listeners approach the source of male song playback, we then extended this natural behavioural paradigm to directly evaluate how listeners perceive and use acoustic features from vocal sequences. Specifically, we tested the hypothesis that, during naturalistic goal-directed behaviour in a social context, female listeners perceptually compress the high sensory dimensionality of male songs by selectively monitoring a reduced subset of meaningful acoustic features in isolation. We quantified females' place preference in the same behavioural assay in response to the near-simultaneous playback of the pre-recorded set of male songs from one speaker, and an acoustically modified version of the songs from the second speaker. Given that females preferentially approach the speaker playing the male song, differences in the relative time spent at the source of intact vs manipulated sound playback are interpreted to demonstrate the behavioural relevance of the manipulated acoustic dimension. In contrast, equal approach behaviour to both intact and manipulated song playback is interpreted as females not

perceiving, or not relying on, the tested acoustic dimension in their natural behavioural response to courtship song.

In the first instance we evaluated whether listeners would be sensitive to two types of global manipulations affecting the song structure at longer timescales. Building on previous work suggesting that female mice use syntactic information to discriminate songs produced in different social contexts[9], we tested whether randomizing the order of the syllables in the song would affect their approach behaviour (Fig. 3A, left and S1B). We found that the sequential organization of syllables in the male songs was not necessary for normal approach behaviour (mean preference index in response to intact vs randomized songs did not differ from zero, one-sample *t*-test, $t(20) = 0.126$, $p = 0.90$; Fig. 3A, right). This behavioural invariance to changes in the syllable order was apparent at all timepoints of song set playback (Fig. 3B), as well as at different speaker zone lengths (Fig. S4C), and was comparable across the 4 sound presentation trials (Fig. S5D). This confirms that the lack of sensitivity to syllable sequence structure we observe here was not caused by the selection of specific temporal or spatial parameters for analysis.

Then, we contrasted intact songs with a time-reversed version of the songs. While preserving the long-term spectra and range of spectral changes, this manipulation generated sounds with novel temporal features that different subsets of auditory neurons are tuned to, such as reversing the direction of frequency trajectories[26-32] (Fig. 3C, left and S1C). Temporal reversal of the vocal sequence did not significantly affect females' approach behaviour (mean preference index in response to intact vs reversed songs did not differ from zero, one-sample *t*-test, $t(20) = -1.46$, $p = 0.16$; Fig. 3C, right). This lack of behavioural sensitivity to song temporal reversal was present throughout sound playback in the vast majority of animals tested (Fig. 3D), across the sound presentation trials (Fig. S5E), and for different spatial regions of interest (Fig. S4F). Together, these observations suggest that female listeners are not relying on global cues reflecting the overall structure of male song sequences when selecting which sound source to approach.

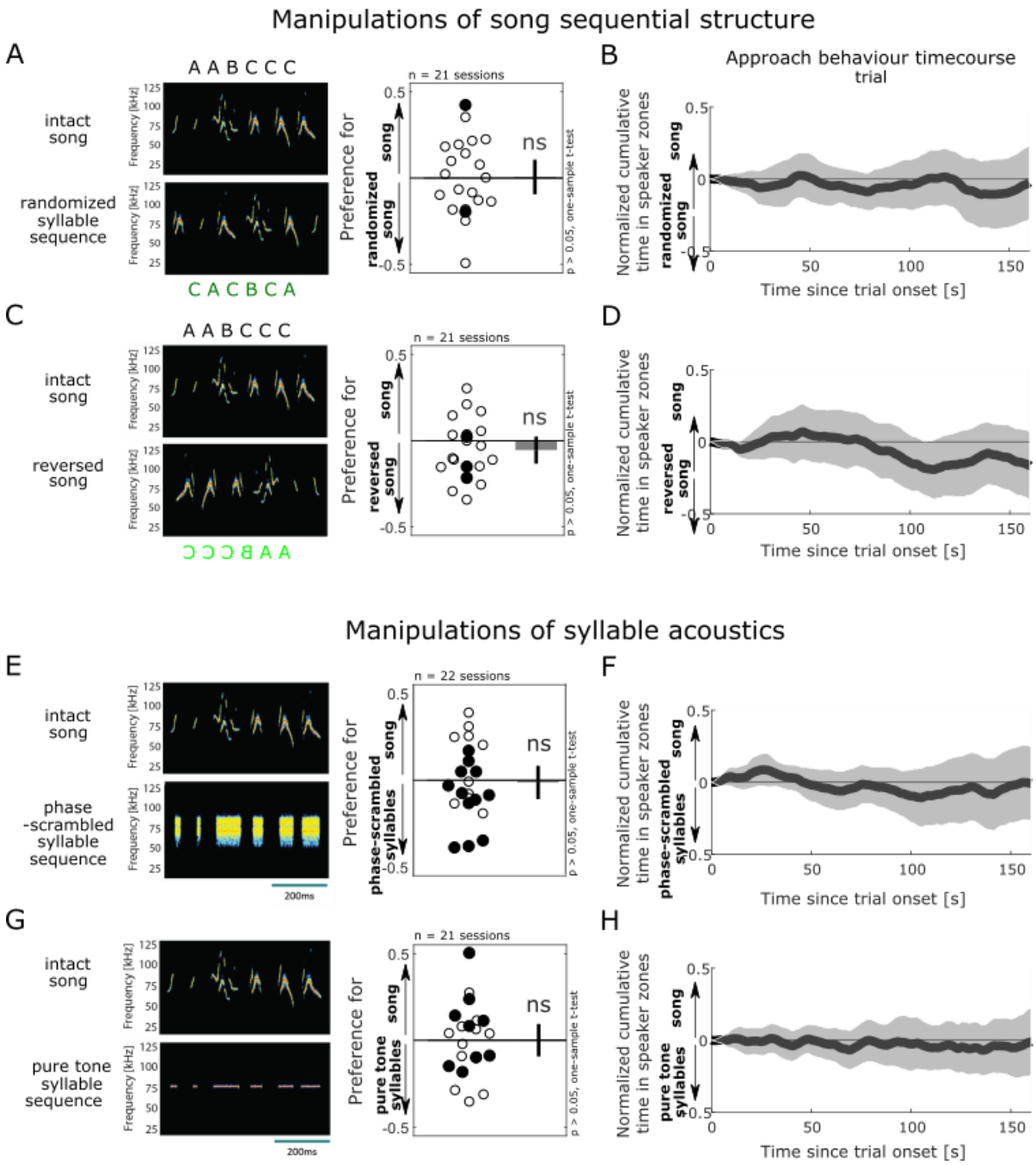

Figure 3: Female approach behaviour is not affected by changes to global song structure or the removal of syllable spectro-temporal dynamics

A. Female approach behaviour during simultaneous playback of intact male songs (top) and corresponding randomized syllable sequences (bottom,), displayed as in Fig. 2H. Each circle is the preference index displayed by individual animals in one behavioural session (median of 4 sound

presentation trials). Open circles indicate sound playback at 58 dB SPL, filled circles at 68 dB SPL. Bar plot shows mean preference index across sessions, and error bar show 95% confidence interval (C.I.). One-sample two-tailed *t*-test, ns: non-significant, $p > 0.05$.

B. Population timecourse of approach behaviour during the playback of intact (positively weighted) vs randomized songs (negatively weighted), displayed as in Fig. 2K. The dark grey trace indicates the population mean of trial-based temporal profiles across all sessions (n = 21 sessions). For each session, the median of 4 sound presentation trials (e.g., black trace in J,) was normalized to the absolute value of its maximal amplitude. At no time bins during the course of a sound playback trial did the cumulative approach behaviour significantly deviate from zero (one-sample two-tailed *t*-test, all $p>0.05$).

C Playback of intact songs (top) contrasted with temporally reversed songs (bottom)

D. Typical timecourse of approach behaviour during the playback of intact (positively weighted) vs reversed songs (negatively weighted).

E. Playback of intact songs (top) and sequences of phase-scrambled syllables (bottom).

F. Typical timecourse of approach behaviour during the playback of intact (positively weighted) vs phase-scrambled syllable sequences (negatively weighted).

G. Playback of intact songs (top) contrasted with sequences of pure tones (bottom).

H. Typical timecourse of approach behaviour during the playback of intact (positively weighted) vs pure tone sequences (negatively weighted).

Female approach behaviour is robust to the removal of syllable spectro-temporal dynamics

Since auditory neurons are typically tuned to specific spectro-temporal features in the song syllables, and mice are able to use those features to discriminate between individual syllables[16,17], we next wondered whether manipulations disrupting the fine acoustic structure of individual calls in the male vocal sequences would be more easily discriminated by female listeners. We first created a phase-scrambled version of the songs that preserved the average frequency spectrum of the syllables, but removed the spectro-temporal dynamics of individual syllables and shaped each noise burst with a flat envelope (Fig. 3E, left and S1D). Females approached the intact and phase-scrambled song playback similarly (mean preference index in response to intact vs phase-scrambled songs did not differ from zero, one-sample *t*-test,

t(22) = -0.27, p = 0.79; Fig. 3E, right). Second, comparing intact songs with more drastically simplified sound sequences that stripped down each syllable into a pure tone at the peak syllable frequency (76kHz) also failed to elicit differential approach behaviour (mean preference index in response to intact vs pure tone sequences did not differ from zero, one-sample *t*-test, t(20) = 0.034, p = 0.97; Fig. 3G and S1E). These two instances of behavioural insensitivity to changes in the syllable acoustic trajectories were robust to variations in temporal (Fig. 3F, H and Fig. S5F, G) and spatial (Fig. S4D, G) parameters used for analysis. Together, these findings show that female behaviour was robust to the removal of fast spectro-temporal structure in the syllables, at least when sound energy was still present at the natural syllable peak frequency.

Female approach behaviour is reduced by the disruption of song temporal regularity

Lastly, following the earlier observation that syllable onsets occur at consistent time intervals from each other across the song sequences (see [25] and Fig. 1D), we directly probed whether female listeners would be sensitive to manipulations of the temporal regularity in the male song. To do this we generated a set of temporally irregular songs that consisted of the same syllable sequence with a broadened distribution of silent pauses compared to intact songs (interquartile range of syllable interval in irregular songs = 134ms, compared to 61ms for intact songs; Fig. 4A, B, E left panel and S1F). Interestingly, females were highly sensitive to the disruption of song temporal regularity, and preferentially approached intact over irregular song playback (mean preference index in response to intact vs irregular songs significantly differed from zero, one-sample *t*-test, t(23) = 3.43, p = 0.00023, Fig. 4E, right). The sensitivity to temporal regularity was robust and similarly strong across all sizes of speaker zones (Fig. S4E). The temporal profile of approach behaviour of a majority of animals favoured intact, regular songs over the course of a session (Fig. 4G), and across sound presentation trials (S5A). On average, mouse listeners started showed this sustained preference for regular over irregular songs after 85.4 seconds of sound playback (normalized cumulative time in speaker

zone significantly different from 0, one-sample t-test, p<0.05; Fig. 4F). This longer latency to behavioural discrimination after song onset compared to when approaching intact song playback over silence (16.9 seconds, Fig. 3B) might reflect the need for additional accumulation of evidence and temporal integration when discriminating between the two concurrent sound streams. These striking results in response to disruptions of the song temporal regularity contrast with the lack of noticeable effect of other manipulations of the song and syllable structure on female approach behaviour.

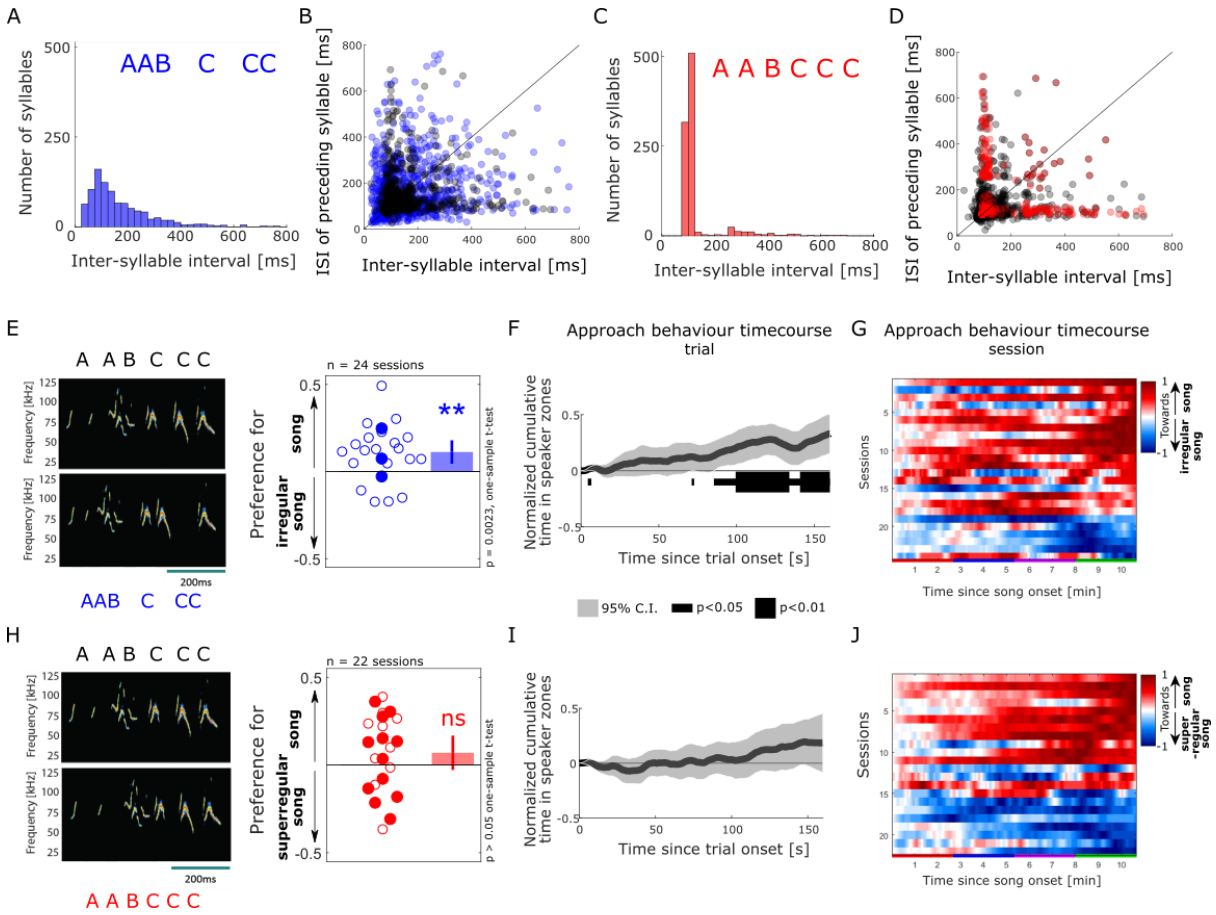

Figure 4: Female approach behaviour is sensitive to disruption of courtship temporal regularity.

A. Distribution of inter-syllable interval (ISI) durations across the set of temporally irregular songs, calculated as in Fig. 1D.

B. Sequential relationships of ISI durations in the intact (shaded grey dots) and temporally irregular (shaded blue dots).

C. Distribution of ISI durations across the super-regular song set.

D. Sequential relationship of ISI durations in the intact (shaded grey dots) and super-regular (shaded red dots).

A-D: n = 957 syllables

E. Female approach behaviour during simultaneous playback of intact male songs (top) and temporally irregular songs (bottom), displayed as in Fig. 2H. Each circle indicates the preference index displayed by individual animals in one behavioural session (median of 4 sound presentation trials). Open circles: sound playback at 58 dB SPL, filled circles: 68 dB SPL. One-sample two-tailed *t*-test. **: $p < 0.01$, ns: non significant, $p > 0.05$. Bar plots show means, and error bars show 95% C.I.

F. Population-averaged time course of approach behaviour in response to intact (positively weighted) vs temporally irregular (negatively weighted) mouse songs, displayed as in Fig. 2K. The dark grey trace indicates the mean of normalized, trial-based temporal profiles across all sessions (n = 24). The black bar indicates time bins in which the cumulative approach behaviour significantly deviates from zero (one-sample two-tailed *t*-test).

G. Temporal profiles of approach behaviour to intact vs temporally irregular songs over the course of 4 sound presentation trials (x-axis, coloured bars) for each of the behavioural sessions (y-axis. n = 24), displayed as in Fig. 2L.

H. Female approach behaviour during simultaneous playback of intact songs (top) and temporally super-regular songs (bottom), displayed as in panel E.

I. Population time course of approach behaviour to intact (positively weighted) vs temporally super-regular (negatively weighted) mouse songs, displayed as in panel F.

J. Temporal profiles of approach behaviour to intact vs temporally super-regular songs over the course of 4 sound presentation trials, displayed as in panel G.

Finally, we asked whether female mice, being obviously attracted by stronger rhythmic regularity in the songs, would preferentially approach an artificial set of “super-regular” songs over intact ones. To this end we created a set of highly rhythmic songs in which the silent breaks were modified such that the onset of the syllables in each song occurred exactly at the dominant inter-syllable interval, effectively narrowing the distribution of inter-syllable times (interquartile range of syllable interval in super-regular songs = 24ms, compared to 61ms for intact songs; Fig. 4C, D, H left panel and S1G). Female listeners did not show any consistent

behavioural preference when the playback of intact songs was contrasted with that of super-regular songs (one-sample *t*-test, $t(21) = 1.48$, $p = 0.15$; Fig. 4H, right). This was reflected in the time course of approach behaviour, which showed large variability across animals over a session with no clear emerging pattern (Fig. 4J). All 4 sound presentation trials elicited comparable response profiles (Fig. S5C). This behavioural invariance to improvements of the songs' temporal regularity remained when tested at different timepoints during song playback (Fig. 4I), or using different spatial parameters for analysis (Fig. S4H). Together, these results suggest that the rhythmicity of natural male songs is already sufficiently salient to reach the criterion for approach behaviour by female listeners, and cannot be improved by additional regularisation of syllable onsets.

Discussion

In this study, we developed an ethological assay to probe the behavioural relevance of several candidate acoustic features in mouse communication. By exploiting female mice's natural approach response to male-produced courtship songs and using competing playbacks of intact and manipulated songs, we directly tested how listeners use acoustic information from vocal sequences. The results show that female mice monitor temporal regularity in male songs, while their behaviour was unaffected by changes in the song sequential organization, or the fine acoustic structure of individual syllables. The findings highlight the selective extraction of song temporal regularity, which may be an acoustic signature of singer fitness, as an effective, potentially evolutionarily conserved behavioural strategy for the sensory compression of complex vocal sequences during goal-directed social behaviour.

On the one hand, neurons across the mouse auditory system are tuned to different spectro-temporal features in song syllables, and mouse listeners have been shown to be perceptually sensitive to, or able to detect, the acoustic manipulations of individual syllables used in the current study. The upper limit of this perceptual ability has been determined using go/no-go operant conditioning to extensively train animals to discriminate between syllables from spectro-temporally distinct categories, between intact and temporally reversed or pure tone versions[16], or between partial from whole syllables[17]. However, such externally rewarded reinforcement learning and the associated sharpening of sensory acuity and neuronal tuning likely bias and amplify the perceptual ability of a naive animal[33,34]. On the other hand, social behaviour in naturalistic settings requires the selective monitoring of specific informative sensory features, while cognitively suppressing or ignoring other dimensions that may however be perceptible, as classically demonstrated in the toad's visual system during prey capture[35]. Our work complements previous studies that quantified the upper bounds of mouse hearing sensitivity. We show that, in the context of an ethologically-valid behaviour, female listeners show behavioural invariance to several acoustic features that they are, or can be trained to become, perceptually sensitive to.

We found that females' approach to male song playback was not consistently affected by changes to the global structure of the vocal sequence, such as the temporal reversal of the song and the randomization of the syllable order in the song. Both of these manipulations preserve the relative spectro-temporal relationships within each syllable, as well as the physical complexity and acoustic characteristics of the sequence, and tested whether the order of the syllables was informative to the listener. Indeed, songs from male mice have been shown to demonstrate some characteristic sequential features, with short syllables typically dominating the beginning of a vocal phrase[9,22,25,36,37]. This structure would be violated both by the syllable order randomization, and the temporal reversal of the song. While mouse listeners are known to perceptually discriminate between syllable types differing in their spectro-temporal features[16], our finding that female approach behaviour was not affected by a shuffling of the syllable sequence indicates that, during courtship behaviour, females monitor the presence/absence or relative occurrence of specific syllable types, or other acoustic features of the songs, independent of the syllable order. Our result also disambiguates previous findings on mouse sensitivity to syntactic information in male songs: Chabout and colleagues[9] previously showed, in a similar behavioural paradigm, that female listeners discriminated between syntactically simple and complex songs produced by males in different social contexts. Songs produced in the two tested social contexts differed both in the first-order syllable sequencing and in the composition of the syllable repertoire. Our finding on females' behavioural insensitivity to syllable order thus suggests that listeners may have been primarily exploiting differences in repertoire composition to discriminate and ultimately select songs produced from a specific social context. Thus our results show that female listeners do not strongly rely on the global structure of the syllable sequence during courtship behaviour.

When manipulating the songs at the more local level of individual syllables, we found that female listeners approached sequences of noise bursts or pure tones at the syllable peak frequency similarly to intact male songs. Our finding on female mice's behavioural invariance to replacing syllables by pure tone sequences replicates previous work that showed female

mice approaching playbacks of synthetic 70 kHz ultrasounds over silence, when these were presented from behind one of two devocalized males[6]. In contrast, Hammerschmidt and colleagues[11] used a place preference paradigm comparable to the one used in this study, and an artificial sequence of irregularly timed short ultrasounds that matched neither the duration or syllable rate of mouse songs. In this case, the authors showed that female mice preferentially approached male songs over pure tones. Our results resolve the apparent contradiction between these previous studies: by showing both females' behavioural invariance to pure tone approximations of syllables and their high sensitivity to temporal regularity, our work suggests that the preference for intact over ultrasound sequence observed in the Hammerschmidt study is driven by the co-occurring disruption in sequence temporal regularity. Thus, the evidence suggests that sound energy in and around a critical band corresponding to the syllable frequency range, displayed at the natural temporal organization of male songs, is a sufficient approximation of male courtship songs to drive female approach behaviour. Our and others' results on female courtship behaviour complement the body of work on maternal behaviour by lactating mothers: as long as the temporal structure of pup calls is maintained, pup retrieval behaviour is robust to the removal of most of the spectral structure of ultrasonic pup isolation calls[38], and three-harmonic stack approximations of sonic wriggling calls elicit normal maternal responses[39]. Similar importance of temporal patterns for the perception of social communication, over the fine acoustic structure of individual elements, has been demonstrated behaviourally in other animal species, including humans: as illustrated in vocoded speech, the intelligibility of speech degraded in the spectral domain remains partially preserved if it is shaped with the correct amplitude envelopes[40,41].

Of the subset of acoustic features tested in this study, temporal regularity in the song was the only feature we observed to be extracted by female listeners from male acoustic courtship displays. Our other tested contrasts show that several other perceptually dramatic disruptions of song features were nevertheless not impacting on the female approach behaviour, as long as the temporal patterning of the songs was preserved, suggesting a high degree of saliency

for temporal regularity as a socially informative cue. We found that mouse listeners preferentially approached intact over irregular courtship songs, suggesting that regularity is attractive to females in the context of vocal-based mate selection. The temporal regularity observed in mouse songs is a consequence of breathing patterns regulating the production of syllables by the emitter, as individual calls are generated following the onset of exhalation[23,25,42]. Disruption to the temporal regularity of male song, as artificially introduced in this study, may for instance result from irregular breathing cycles and thus "stuttering" singing by the male, or by the inability to sustain bouts of vocal production of sufficient duration to elicit a salient percept of regularity in the listener. A recent study of courtship songs by Foxp2 knockout mice provided direct evidence for a link between male emitter fitness and song temporal regularity, by showing songs by mutant mice contain rhythmic irregularities[36]. From the perspective of the listener, temporal predictability has been shown to facilitate auditory detection of near-threshold stimuli, compared to an identical aperiodic sequence[43]. Additionally, in human listeners, speech intelligibility is disrupted by temporal jittering the sentences[44,45]. Thus, perhaps by improving the perceptual signal-to-noise ratio of male courtship songs, temporal regularity may represent an acoustic cue signalling the fitness or suitability of the potential mating partner to the listener[14], to which we show that socially motivated listeners are highly attuned to.

Female behaviour did not appear to discriminate between intact songs and an artificial "super-regular" version of the songs, suggesting that natural mouse songs might already at ceiling in terms of the perceptual saliency of their temporal regularity and therefore, their attractiveness to female listeners. In the somatosensory system, neuronal responses in barrel cortex responded indistinguishably to lightly temporally jittered and perfectly regular sequences of whisker deflections, when the stimulus presentation rate was slower than 20Hz[46]. In a hypothesis that remains to be directly tested in the auditory system, it is thus possible that, at the natural syllable rates of mouse songs, removing the low temporal jitter

present in intact songs to create super-regular songs does not significantly impact on perceptually relevant neuronal substrates.

The syllable rate in mouse songs (~ 7Hz[3]) broadly matches with that of other mammalian vocal patterns such as monkey vocalizations and human speech[47,48]. On the one hand, this stimulus presentation rate in mouse songs is directly determined by the breathing rhythm of the singer. Given that active sensation during whisking and sniffing operates within the frequency range of both respiratory-coupled oscillations[49] and the hippocampal theta rhythm in rodents[50-53], similar constraints may be placed on the cortical mechanisms of hearing[54], especially during natural behaviour. This intriguing hypothesis could be addressed in future work by simultaneously tracking the listener's hippocampal theta oscillations and sniffing behaviour during song playback. On the other hand, the syllable rate in mouse songs generally corresponds to the timescale of slow amplitude envelope in human speech, which is known to contribute to the identification of both prosodic and syllabic content[55]. While vocal sequences such as human speech and mouse songs are quasi-periodic rather than metronomically regular signals, they both contain sufficiently salient temporal regularity to allow listeners to make predictions about the incoming signal[55]. In particular, the comparable range of temporal regularities found in mouse and human vocal sequences corresponds to the theta frequency of cortical oscillations that have been shown to be susceptible to entrainment by streams of regularly presented stimuli[55,56]. An influential model of speech processing proposes that cortical theta oscillations in the 4-8 Hz range entrain to the slow temporal envelope modulations in speech, and may play a role in parsing continuous sensory input by coordinating neuronal excitability to optimally support the decoding of phonemes[57]. Whether similar neuronal dynamics can be observed in response to courtship song sequences in the mouse auditory system, and what neuronal substrates causally support the behaviourally-relevant extraction and representation of acoustic features identified in this study, remains to be determined.

Conclusion

In summary, we used an ethologically-driven approach to evaluate the behavioural relevance of several acoustic features in mouse communication. Our results identify a key role of temporal regularity, but invariance to global and local song structure, in the goal-directed use of vocal patterns by female listeners. The findings highlight the selective monitoring of temporal regularity in communication sounds as an effective, potentially evolutionarily conserved behavioural strategy for the sensory compression of complex vocal sequences during mammalian vocal perception.

Methods

Animals

83 female C57Bl/6J inbred mice (*M. musculus*, Charles River Laboratories, UK) participated in the experiments. All experimental procedures were carried out in accordance with a UK Home Office Project License approved under the United Kingdom Animals (Scientific Procedures) Act of 1986, and in compliance with international legislation governing the maintenance and use of laboratory animals in scientific experiments (European Communities Council Directive of November 24, 1986, 86/609/EEC). Animals were housed in same-sex groups of 3-5 per cage and under a reversed 12 hours light/dark cycle. Food pellets and water were provided *ad libitum*. Animals were tested in 8 batches of 8-12 animals between September 2017 and September 2019. Mice participated in the playback experiments between the ages of 5-11 weeks (mean = 9 weeks; Fig. S3). In order prevent any impact of handling stress from influencing behavioural testing, the animals were regularly handled and habituated to all aspects the behavioural protocol for at least a week before starting experiments. Tube handling[58] was used exclusively. The mice had no sexual experience, but were exposed to two different groups of males through a mesh division for 5 min each following each behavioural testing session. This allowed olfactory, auditory and visual contact with males while preventing physical contact and intercourse, and has been suggested to help maintain female's motivation for approach behaviour[8].

Acoustic stimuli

Ultrasonic vocalizations were obtained from 5 male C57BL/6 mice in response to urine samples from conspecific females in estrus, across 11 recording sessions. The urine stimulus consisted of the tip of one sterile cotton bud soaked in a mixture of freshly-collected urine from at least two animals. Sounds were recorded at a sampling frequency of 250 kHz (and, Avisoft Bioacoustics, Germany) using a condenser ultrasonic microphone (CM16/CMPA), recording

interface (UltrasoundGate 416H) and software (Avisoft Recorder version 5.2.09, all from Avisoft Bioacoustics, Germany).

The stimulus set used in the playback experiments ("intact/original songs") consisted of a subset of seven songs produced by 3 C57BL/6 male mice (median age 22 weeks) over 6 recording sessions. The sound files were high-pass filtered above 40 kHz, and were denoised using the frequency-domain noise reduction algorithm in Avisoft SASlab Pro (version 5.2.09, Avisoft Bioacoustics, Germany). Syllable detection and segmentation was verified manually. Sound files were up-sampled to 260420 Hz with anti-aliasing for playback.

Acoustic manipulations

An artificial set of randomized syllable sequences was created by shuffling the sequential order of the syllables within each song. Each syllable remained paired with its subsequent silent interval, in order to preserve the distribution of inter-syllable intervals across the songs.

The set of temporally-reversed songs was creating by playing each original song backwards, e.g. flipped along the time axis such that the last sample of the last syllable was played first, etc, until the first sample of the first syllable which was played last.

Phase-scrambled songs were created by replacing the phase of the original signal with a random phase value, for each original song separately. The resulting high-frequency noise syllables were then ramped with a 0.5 ms linear rise/fall time.

Pure tone sequences were created by replacing each syllable with a pure tone of the matching duration, including a 0.5 ms linear rise/fall time. Tone frequency was selected to match the median peak frequency of all recorded syllables, e.g. 76k Hz.

For the phase-scrambled and pure tone sequence sets, sound amplitude was constant across all syllables in one song, and was adjusted for each song separately in order to match the mean root mean square amplitude of syllables in the corresponding original songs.

A set of temporally irregular songs was created by replacing the silent pauses associated with each syllable (time intervals between the offset of one syllable and the onset of following syllable) with randomly generated break durations, under the condition that the duration of the new song be equal to that of the original song.

A super-rhythmic version of each of the original songs was created by scaling the silent pauses between each pair of successive syllables such that every inter-syllable interval (time between the onset of one syllable and the offset of the following syllable) matched the median inter-syllable interval of a given song. For syllables that were longer than the median inter-syllable interval, the silent pause was adjusted to “skip a cycle” such that the next syllable occurred at twice the median syllable interval. The total duration of each super-rhythmic song was matched to that of the corresponding original song by modifying the longer inter-bout interval (silent pause between regular sequences of syllables).

Sounds were played using MATLAB (Matlab version R20015a; MathWorks, Natwick, MA, USA) and a digital signal processor (RX6, Tucker-Davis Technologies, FL, USA). Acoustic stimuli were delivered via two free-field electrostatic speakers driven (ES1, Tucker-Davis Technologies, FL, USA), placed at ear level, 3 cm from the mesh window at the edge of the enclosure. The frequency response of the loudspeaker was ± 6 dB across the frequency range used for stimulation [68-84 kHz]. Sounds were played either at 53-59 dbSPL, or 63-70 dbSPL, as measured just inside the testing box (7cm from the speakers). Before the start of the first behavioural session of the day, correct playback of ultrasonic stimuli was confirmed visually using the Avisoft-RECORDER software and condenser ultrasonic microphones (Avisoft CM16/CMPA, both by Avisoft Bioacoustics, Germany) placed at the edge of the behavioural box.

Estrous staging

Estrous stage was assessed daily based on vaginal cytology[59,60]. Samples were collected by flushing warmed sterile saline over the vaginal opening. The unstained cell samples collected

via vaginal lavage were then visualized using bright-field microscopy (Leica DFC365FX). Estrus stage was estimated visually based on the relative proportions of leukocyte, nucleated and cornified epithelial cells, following the estrous cycle stage identification tool outlined in [59]. In a subset of animals tested between October 2017 and March 2018 estrous stage was identified based on visual examination of the vagina[59,60]. Both estrus staging methods required brief (10s) tail restraint, and took place daily at similar times in the morning, at least 2-3 hours before a behavioural experiment in order to prevent any handling stress from impacting the behaviour.

Behavioural testing

All behavioural tests were conducted during the dark (active) phase of the light cycle, between 12:00 and 20:00. Animals participated in the experiments once per estrous cycle, when identified as being in estrus or proestus. A minimum of 5 days separated successive test sessions for each mouse. Animals were tested at most once in each given sound contrast. The order of experimental contrasts each animal participated in was assigned pseudo-randomly, in order to balance the ages of animals taking part in each type of behavioural tests (no difference in mean listener age across all tested contrasts, one-way ANOVA, $F(6, 153) = 0.013$, $p = 1.0$; Fig. S3).

Behavioural assays took place in a two-compartment Plexiglas behavioural box (22.7 x 25.2 x 22.5cm), under infrared LED illumination (no visible light, Fig. 2B). A soundproof partition partly divided the behavioural box down the middle into two “speaker zones” joined by a larger “neutral zone”. Wall panels with a mesh insert were positioned at the end of each speaker zone, such that the mesh-covered openings were level with two speakers placed outside of the box, on either side of the partition. This allowed undistorted delivery of acoustic stimuli separately into each of the two speaker zones, while playbacks from both speakers were equally audible in the centre of the neutral zone. The behavioural box was placed inside a double-walled soundproof booth (IAC Acoustics), whose interior was covered by 4-cm thick acoustic absorption foam (E-foam, UK). Two trays containing 2 g of a mixture of soiled bedding

freshly collected from two cages of males were placed below each speaker, outside of the behavioural box. After each behavioural testing session, the wall panels were washed with soapy water and dried, and the behavioural boxed wiped down first with 70% ethanol, then with distilled water.

One behavioural session lasted for 40 min, and contained a 10 min silent habituation period, and 4 sound presentation trials (160s of song playback) each followed by a 3 min break (Fig. 2B). The session started with the mouse being placed in the middle of the neutral zone using the handling cylinder. After the 10min silent habituation period, the first playback trial was initiated manually as soon as the animal returned to the middle of the neutral zone. Each playback trial consisted of one concatenated presentation of the songs in the stimulus set (7 songs, 2min 40s total sound duration), played in random order. Each trial (playback of one full stimulus set) was followed by a 3min silent break, before the next trial could be initiated.

In a “song vs silence” contrast experiment, the song set was played back through one of the two speakers, while the other remained silent. In a “original vs manipulated song” contrast experiment, the original song set and corresponding manipulated version of the song set were played back simultaneously, each through a different speaker. The playback onsets of individual pairs of original and manipulated songs were jittered with respect to each other by a time delay randomly sampled between 20-90 ms, with the leading stimulus set randomly assigned for each trial. The speaker/playback side for each sound presentation programmes was randomly assigned at the beginning of one session, and alternated across the 4 consecutive trials within each session. This strategy allowed to distinguish, for each session, between a behavioural preference for one of the sound types despite changes in the location of the sound source, and a preference for one of the physical locations (side bias).

Video tracking

The position of the animal was recorded under infrared illumination, using an overhead camera (640x480 pixel resolution, Play Station 3 Eye, Sony) that had been customized by manual

removal of its infrared blocking filter. A custom image processing pipeline in Bonsai [61] detected the position of the animal against the while background and saved the x and y coordinates of the centroid of the body, as well as the corresponding timestamps (ca 30 frames/s). On- and onset timestamps for song playback were recorded by manual key presses.

Data analysis

Tracking data were analyzed in MATLAB using custom scripts (version R2018a; MathWorks, Natwick, MA, USA). For timestamps during which the detection of the mouse's body failed, the x-y position was interpolated by repeating the previously available position. For each playback side, a "speaker zone" was defined as the area of the behavioural box adjacent to each mesh opening in the wall panel, ranging 86 mm into each arm of the divided section of the box. The relative strength of the animal's place preference for one stimulus over the other was captured using a preference index defined as follows: $Preference\ index = \frac{t(song)-t(manipulated\ song)}{t(song)+t(manipulated\ song)}$, where *t(song)* is the time in seconds the animal spent in the speaker zone adjacent to the side of original song playback for this trial and *t(manipulated song)* is the time spent in the other speaker zone, e.g. the side of manipulated song playback or silence. The preference index was computed for each sound presentation trial. The preference index for one behavioural session was calculated as the median of the 4 trial-based sound preference indices for that session.

Side bias across a testing session was assessed by quantifying the time in seconds an animal spent in the left and right speaker zone during each of the four sound presentation trials. Sessions in which the animal displayed a clear side bias (consistent preference for the left or right side of the behavioural box despite changes in the sound playback side, evaluated as all four trials showing a positive, or all negative, difference between the times in the left and right speaker zones) were excluded from the analysis (29% ± 7% of behavioural sessions excluded due to side bias, mean and 95% C.I.).

To evaluate the temporal evolution of each session's approach behaviour, one positive (resp. negative) increment was assigned to each video frame during which the animal was inside the area (speaker zone) in front of the speaker playing back intact (resp. manipulated) songs. Frames during which the animal was outside either speaker zones were assigned a value of zero. Temporal profiles quantifying the accumulation of dwell time in either speaker zones were computed by taking the cumulative sum of these traces along the relevant trial(s), and dividing by the video frame rate to obtain times in seconds. Before comparing or pooling data across animals, each trace from an individual session or trial was normalized to the absolute value of its maximum amplitude. When displaying all traces individually, sessions were sorted by the amplitude of their last element.

All statistical analyses were performed using MATLAB (version R2018A). Data were tested for deviation from normality using Lilliefors' composite goodness-of-fit test (Matlab function lillietest) and Shapiro-Wilk's parametric hypothesis test of composite normality (Matlab function swtest) at $p<0.05$, before using parametric tests. Significant population place preference for each contrast was tested using one-sample two-tailed *t*-test against zero. Unless otherwise stated all statistical tests are two-tailed. In all figures, significance levels are indicated as follows: *: significant at the 0.05 level, **: significant at the 0.01 level.

Data and code availability

All data used to produce the figures in this study are available in figshare with the identifier [final data DOI(s) TBC ahead of publication]. The computer code used in this study is available at [final github location TBC ahead of publication].

References

1 Portfors, C. V. Types and functions of ultrasonic vocalizations in laboratory rats and mice. *J Am Assoc Lab Anim Sci* **46**, 28-34 (2007).

2 Liu, R. C., Miller, K. D., Merzenich, M. M. & Schreiner, C. E. Acoustic variability and distinguishability among mouse ultrasound vocalizations. *J Acoust Soc Am* **114**, 3412-3422, doi:10.1121/1.1623787 (2003).

3 Holy, T. E. & Guo, Z. Ultrasonic songs of male mice. *PLoS Biol* **3**, e386, doi:10.1371/journal.pbio.0030386 (2005).

4 Chabout, J. *et al.* Adult Male Mice Emit Context-Specific Ultrasonic Vocalizations That Are Modulated by Prior Isolation or Group Rearing Environment. *Plos One* **7**, doi:ARTN e29401

DOI 10.1371/journal.pone.0029401 (2012).

5 Chabout, J., Jones-Macopson, J. & Jarvis, E. D. Eliciting and Analyzing Male Mouse Ultrasonic Vocalization (USV) Songs. *J Vis Exp*, doi:10.3791/54137 (2017).

6 Pomerantz, S. M., Nunez, A. A. & Bean, N. J. Female behavior is affected by male ultrasonic vocalizations in house mice. *Physiol Behav* **31**, 91-96, doi:10.1016/0031-9384(83)90101-4 (1983).

7 Asaba, A. *et al.* Developmental social environment imprints female preference for male song in mice. *PLoS One* **9**, e87186, doi:10.1371/journal.pone.0087186 (2014).

8 Shepard, K. N. & Liu, R. C. Experience restores innate female preference for male ultrasonic vocalizations. *Genes Brain Behav* **10**, 28-34, doi:10.1111/j.1601-183X.2010.00580.x (2011).

9 Chabout, J., Sarkar, A., Dunson, D. B. & Jarvis, E. D. Male mice song syntax depends on social contexts and influences female preferences. *Front Behav Neurosci* **9**, 76, doi:10.3389/fnbeh.2015.00076 (2015).

10 Musolf, K., Meindl, S., Larsen, A. L., Kalcounis-Rueppell, M. C. & Penn, D. J. Ultrasonic Vocalizations of Male Mice Differ among Species and Females Show Assortative Preferences for Male Calls. *PLoS One* **10**, e0134123, doi:10.1371/journal.pone.0134123 (2015).

11 Hammerschmidt, K., Radyushkin, K., Ehrenreich, H. & Fischer, J. Female mice respond to male ultrasonic 'songs' with approach behaviour. *Biol Lett* **5**, 589-592, doi:10.1098/rsbl.2009.0317 (2009).

12 Asaba, A. *et al.* Male mice ultrasonic vocalizations enhance female sexual approach and hypothalamic kisspeptin neuron activity. *Horm Behav* **94**, 53-60, doi:10.1016/j.yhbeh.2017.06.006 (2017).

13 Nomoto, K. *et al.* Female mice exhibit both sexual and social partner preferences for vocalizing males. *Integr Zool* **13**, 735-744, doi:10.1111/1749-4877.12357 (2018).

14 Egnor, S. R. & Seagraves, K. M. The contribution of ultrasonic vocalizations to mouse courtship. *Curr Opin Neurobiol* **38**, 1-5, doi:10.1016/j.conb.2015.12.009 (2016).

15 Hoffmann, F., Musolf, K. & Penn, D. J. Spectrographic analyses reveal signals of individuality and kinship in the ultrasonic courtship vocalizations of wild house mice. *Physiol Behav* **105**, 766-771, doi:10.1016/j.physbeh.2011.10.011 (2012).

16 Neilans, E. G., Holfoth, D. P., Radziwon, K. E., Portfors, C. V. & Dent, M. L. Discrimination of ultrasonic vocalizations by CBA/CaJ mice (Mus musculus) is related to spectrotemporal dissimilarity of vocalizations. *PLoS One* **9**, e85405, doi:10.1371/journal.pone.0085405 (2014).

17 Holfoth, D. P., Neilans, E. G. & Dent, M. L. Discrimination of partial from whole ultrasonic vocalizations using a go/no-go task in mice. *J Acoust Soc Am* **136**, 3401, doi:10.1121/1.4900564 (2014).

18 Asaba, A., Hattori, T., Mogi, K. & Kikusui, T. Sexual attractiveness of male chemicals and vocalizations in mice. *Front Neurosci* **8**, 231, doi:10.3389/fnins.2014.00231 (2014).

19 Sugimoto, H. *et al.* A role for strain differences in waveforms of ultrasonic vocalizations during male-female interaction. *PLoS One* **6**, e22093, doi:10.1371/journal.pone.0022093 (2011).
20 Musolf, K., Hoffmann, F. & Penn, D. J. Ultrasonic courtship vocalizations in wild house mice, *Mus musculus musculus*. *Animal Behaviour* **79**, 757-764, doi:https://doi.org/10.1016/j.anbehav.2009.12.034 (2010).
21 Portfors, C. V. & Perkel, D. J. The role of ultrasonic vocalizations in mouse communication. *Curr Opin Neurobiol* **28**, 115-120, doi:10.1016/j.conb.2014.07.002 (2014).
22 Matsumoto, Y. K. & Okanoya, K. Phase-Specific Vocalizations of Male Mice at the Initial Encounter during the Courtship Sequence. *PLoS One* **11**, e0147102, doi:10.1371/journal.pone.0147102 (2016).
23 Tschida, K. *et al.* A Specialized Neural Circuit Gates Social Vocalizations in the Mouse. *Neuron* **103**, 459-472 e454, doi:10.1016/j.neuron.2019.05.025 (2019).
24 Hage, S. R., Gavrilov, N., Salomon, F. & Stein, A. M. Temporal vocal features suggest different call-pattern generating mechanisms in mice and bats. *BMC Neurosci* **14**, 99, doi:10.1186/1471-2202-14-99 (2013).
25 Castellucci, G. A., Calbick, D. & McCormick, D. The temporal organization of mouse ultrasonic vocalizations. *PLoS One* **13**, e0199929, doi:10.1371/journal.pone.0199929 (2018).
26 Issa, J. B., Haeffele, B. D., Young, E. D. & Yue, D. T. Multiscale mapping of frequency sweep rate in mouse auditory cortex. *Hear Res* **344**, 207-222, doi:10.1016/j.heares.2016.11.018 (2017).
27 Lui, B. & Mendelson, J. R. Frequency modulated sweep responses in the medial geniculate nucleus. *Exp Brain Res* **153**, 550-553, doi:10.1007/s00221-003-1618-y (2003).
28 Sollini, J., Chapuis, G. A., Clopath, C. & Chadderton, P. ON-OFF receptive fields in auditory cortex diverge during development and contribute to directional sweep selectivity. *Nat Commun* **9**, 2084, doi:10.1038/s41467-018-04548-3 (2018).
29 Tian, B. & Rauschecker, J. P. Processing of frequency-modulated sounds in the cat's posterior auditory field. *J Neurophysiol* **79**, 2629-2642, doi:10.1152/jn.1998.79.5.2629 (1998).
30 Fuzessery, Z. M. Response selectivity for multiple dimensions of frequency sweeps in the pallid bat inferior colliculus. *J Neurophysiol* **72**, 1061-1079, doi:10.1152/jn.1994.72.3.1061 (1994).
31 Razak, K. A. & Fuzessery, Z. M. Neural mechanisms underlying selectivity for the rate and direction of frequency-modulated sweeps in the auditory cortex of the pallid bat. *J Neurophysiol* **96**, 1303-1319, doi:10.1152/jn.00020.2006 (2006).
32 Geis, H. R. & Borst, J. G. Intracellular responses to frequency modulated tones in the dorsal cortex of the mouse inferior colliculus. *Front Neural Circuits* **7**, 7, doi:10.3389/fncir.2013.00007 (2013).
33 Fritz, J. B., Elhilali, M. & Shamma, S. A. Differential dynamic plasticity of A1 receptive fields during multiple spectral tasks. *J Neurosci* **25**, 7623-7635, doi:10.1523/JNEUROSCI.1318-05.2005 (2005).
34 Fritz, J., Shamma, S., Elhilali, M. & Klein, D. Rapid task-related plasticity of spectrotemporal receptive fields in primary auditory cortex. *Nat Neurosci* **6**, 1216-1223, doi:10.1038/nn1141 (2003).
35 Wachowitz, S. & Ewert, J. P. A key by which the toad's visual system gets access to the domain of prey. *Physiol Behav* **60**, 877-887, doi:10.1016/0031-9384(96)00070-4 (1996).
36 Castellucci, G. A., McGinley, M. J. & McCormick, D. A. Knockout of Foxp2 disrupts vocal development in mice. *Sci Rep* **6**, 23305, doi:10.1038/srep23305 (2016).
37 Hertz, S., Weiner, B., Perets, N. & London, M. Temporal structure of mouse courtship vocalizations facilitates syllable labeling. *Commun Biol* **3**, 333, doi:10.1038/s42003-020-1053-7 (2020).

38 Ehret, G. & Haack, B. J. J. o. c. p. Ultrasound recognition in house mice: Key-Stimulus configuration and recognition mechanism. **148**, 245-251, doi:10.1007/bf00619131 (1982).
39 Gaub, S. & Ehret, G. Grouping in auditory temporal perception and vocal production is mutually adapted: the case of wriggling calls of mice. *J Comp Physiol A Neuroethol Sens Neural Behav Physiol* **191**, 1131-1135, doi:10.1007/s00359-005-0036-y (2005).
40 Shannon, R. V., Zeng, F. G., Kamath, V., Wygonski, J. & Ekelid, M. Speech recognition with primarily temporal cues. *Science* **270**, 303-304, doi:10.1126/science.270.5234.303 (1995).
41 Van Tasell, D. J., Soli, S. D., Kirby, V. M. & Widin, G. P. Speech waveform envelope cues for consonant recognition. *J Acoust Soc Am* **82**, 1152-1161, doi:10.1121/1.395251 (1987).
42 Sirotin, Y. B., Costa, M. E. & Laplagne, D. A. Rodent ultrasonic vocalizations are bound to active sniffing behavior. *Front Behav Neurosci* **8**, 399, doi:10.3389/fnbeh.2014.00399 (2014).
43 Lawrance, E. L., Harper, N. S., Cooke, J. E. & Schnupp, J. W. Temporal predictability enhances auditory detection. *J Acoust Soc Am* **135**, EL357-363, doi:10.1121/1.4879667 (2014).
44 Ghitza, O. & Greenberg, S. On the possible role of brain rhythms in speech perception: intelligibility of time-compressed speech with periodic and aperiodic insertions of silence. *Phonetica* **66**, 113-126, doi:10.1159/000208934 (2009).
45 Pichora-Fuller, M. K., Schneider, B. A., Macdonald, E., Pass, H. E. & Brown, S. Temporal jitter disrupts speech intelligibility: a simulation of auditory aging. *Hear Res* **223**, 114-121, doi:10.1016/j.heares.2006.10.009 (2007).
46 Lak, A., Arabzadeh, E. & Diamond, M. E. Enhanced response of neurons in rat somatosensory cortex to stimuli containing temporal noise. *Cereb Cortex* **18**, 1085-1093, doi:10.1093/cercor/bhm144 (2008).
47 Ghazanfar, A. A. & Takahashi, D. Y. The evolution of speech: vision, rhythm, cooperation. *Trends Cogn Sci* **18**, 543-553, doi:10.1016/j.tics.2014.06.004 (2014).
48 Chandrasekaran, C., Trubanova, A., Stillittano, S., Caplier, A. & Ghazanfar, A. A. The natural statistics of audiovisual speech. *PLoS Comput Biol* **5**, e1000436, doi:10.1371/journal.pcbi.1000436 (2009).
49 Tort, A. B. L. *et al.* Parallel detection of theta and respiration-coupled oscillations throughout the mouse brain. *Sci Rep* **8**, 6432, doi:10.1038/s41598-018-24629-z (2018).
50 Kleinfeld, D., Ahissar, E. & Diamond, M. E. Active sensation: insights from the rodent vibrissa sensorimotor system. *Curr Opin Neurobiol* **16**, 435-444, doi:10.1016/j.conb.2006.06.009 (2006).
51 Kepecs, A., Uchida, N. & Mainen, Z. F. Rapid and precise control of sniffing during olfactory discrimination in rats. *J Neurophysiol* **98**, 205-213, doi:10.1152/jn.00071.2007 (2007).
52 Grion, N., Akrami, A., Zuo, Y., Stella, F. & Diamond, M. E. Coherence between Rat Sensorimotor System and Hippocampus Is Enhanced during Tactile Discrimination. *PLoS Biol* **14**, e1002384, doi:10.1371/journal.pbio.1002384 (2016).
53 Kleinfeld, D., Deschenes, M. & Ulanovsky, N. Whisking, Sniffing, and the Hippocampal theta-Rhythm: A Tale of Two Oscillators. *PLoS Biol* **14**, e1002385, doi:10.1371/journal.pbio.1002385 (2016).
54 Schroeder, C. E., Wilson, D. A., Radman, T., Scharfman, H. & Lakatos, P. Dynamics of Active Sensing and perceptual selection. *Curr Opin Neurobiol* **20**, 172-176, doi:10.1016/j.conb.2010.02.010 (2010).
55 Peelle, J. E. & Davis, M. H. Neural Oscillations Carry Speech Rhythm through to Comprehension. *Front Psychol* **3**, 320, doi:10.3389/fpsyg.2012.00320 (2012).
56 Doelling, K. B., Arnal, L. H., Ghitza, O. & Poeppel, D. Acoustic landmarks drive delta-theta oscillations to enable speech comprehension by facilitating perceptual parsing. *Neuroimage* **85 Pt 2**, 761-768, doi:10.1016/j.neuroimage.2013.06.035 (2014).

57 Giraud, A. L. & Poeppel, D. Cortical oscillations and speech processing: emerging computational principles and operations. *Nat Neurosci* **15**, 511-517, doi:10.1038/nn.3063 (2012).
58 Gouveia, K. & Hurst, J. L. Reducing mouse anxiety during handling: effect of experience with handling tunnels. *PLoS One* **8**, e66401, doi:10.1371/journal.pone.0066401 (2013).
59 Byers, S. L., Wiles, M. V., Dunn, S. L. & Taft, R. A. Mouse estrous cycle identification tool and images. *PLoS One* **7**, e35538, doi:10.1371/journal.pone.0035538 (2012).
60 Caligioni, C. S. Assessing reproductive status/stages in mice. *Curr Protoc Neurosci* **Appendix 4**, Appendix 4I, doi:10.1002/0471142301.nsa04is48 (2009).
61 Lopes, G. *et al.* Bonsai: an event-based framework for processing and controlling data streams. *Front Neuroinform* **9**, 7, doi:10.3389/fninf.2015.00007 (2015).

Acknowledgments

We thank Margaux Silvestre for help with testing initial experimental designs, Camille Suess for contributing to data collection, and Jennifer Linden for feedback and support. We thank Joshua Jones-Macopson and Erich Jarvis for advice on optimizing murine singing behaviour, and developing an approach behaviour paradigm. Tania Barkat, Florian Studer and Sebastian Reinartz provided helpful comments on the manuscript. This work was supported by the Swiss National Science Foundation (P2SKP3_158691, CP) the Wellcome Trust (110238/Z/15/Z, CP), the Medical Research Council (MR/M002889/1, DB), and a Royal Society Research Grant (RG130622, DB).

Author Contributions

Conceptualization, Data curation, Formal Analysis, Visualization and Writing – original draft: CP; Funding acquisition and Methodology: CP, DB; Resources: CP, CV, DB
Investigation and Software: CP, CV; Writing – review & editing: CP, DB.

Competing Interests

The authors declare no competing financial interest.