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

Summary

While animals navigating the real world face a barrage of complex sensory input, their brains have evolved to perceptually compress multidimensional information by selectively extracting the features relevant for survival. For instance, 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 evaluate 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 male songs over rhythmically irregular versions of the songs. In contrast, female behaviour was invariant to manipulations affecting the songs' sequential organization, or the spectrotemporal 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.

Highlights

- Natural behaviour is used to probe how mouse listeners encode vocal sequences
- Listeners are highly sensitive to disruptions of the songs' rhythmic regularity
- Female behaviour is invariant to changes in song sequence
- Approach behaviour is robust to the removal of syllable spectrotemporal dynamics

Introduction

Animals behaving in the real world need to efficiently discriminate and monitor informative 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 compresses the acoustical complexity of vocal patterns during social goal-directed behaviour remains unclear. We asked which acoustic cues from communication sequences were *behaviourally* informative to female mouse listeners when responding to ultrasonic courtship displays from males.

Mice, like most other mammals, use complex acoustic patterns to communicate with each other in various social contexts (Portfors 2007). In particular, adult males produce rhythmic sequences of discrete, frequency-modulated pure tone "whistle-like" vocal elements in response to sensing the recent presence of a fertile female (Holy and Guo 2005, Chabout et al. 2012, Chabout et al. 2017). These courtship "songs" are attractive to sexually receptive females (Pomerantz et al. 1983), who respond to song playback with approach behaviour (Hammerschmidt et al. 2009, Shepard and Liu 2011, Asaba et al. 2014, Chabout et al. 2015, Musolf et al. 2015). Male songs facilitate copulatory success (Nomoto et al. 2018) and are thought to serve as fitness displays (Hoffmann et al. 2012, Egnor and Seagraves 2016).

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 spectrotemporal differences between individual call elements (Neilans et al. 2014), indicating that mice are *perceptually* sensitive to a wide range of acoustic features of vocalizations. On the other hand, experiments using unrewarded, ethologically-relevant place preference paradigms 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 (Chabout et al. 2015), singer species (Musolf et al. 2015), strain (Asaba et al. 2014) and kin (Asaba et al. 2014). However, which acoustic cues are used by motivated female listeners during natural behaviour in response to courtship songs remains somewhat contradictory, in part due to differences between experimental paradigms, and heterogeneity in the acoustic manipulations used in previous studies (Pomerantz et al. 1983, Hammerschmidt et al. 2009).

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 quantify which of several candidate acoustic features are monitored by the listeners. We found that females were highly sensitive to disruptions of song temporal regularity, but that their approach behaviour was not affected by manipulations affecting the songs' sequential organization or the fine spectrotemporal structure of individual syllables. 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 songs

In order to study female listeners' behavioural response to vocal sequences, we first collected a database of male C67Bl/6 mice vocalizations for acoustic stimulation in playback experiments. Following a previously published protocol (Chabout et al. 2017), 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. We chose this paradigm as 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 "courtship" acoustic displays with the purpose of attracting a female (Portfors and Perkel 2014, Matsumoto and Okanoya 2016, Nomoto et al. 2018). The vocal patterns emitted in this context consisted of sequences of individual frequency-modulated calls (see example Fig. 1A and (Holy and Guo 2005, Hage et al. 2013, Chabout et al. 2015, Tschida et al. 2019). 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 duration shorter than 63ms (local minimum)) and a minority of long syllables (local maximum = 88 ms, 14% (334/2373) of recorded syllables with duration shorter than 63ms, Fig. 1B). These male calls were restricted to the ultrasonic frequency range, with the maximal energy in individual syllables occurring around 76kHz and spanning a bandwidth of 16 kHz ([68-84 kHz, median peak frequency ± std, 76 ± 8 kHz, n = 2373 syllables; Fig. 1C). Individual syllables were organized in temporally regular sequences and often started at a constant time delay from each other (Holy and Guo 2005), as evidenced by the narrow concentration of inter-syllable interval durations around 104 ± 8 ms (median inter-syllable interval and 95% confidence interval, n = 2322 syllables; Fig. 1D). From this set of ultrasonic vocalizations, we selected a smaller, representative sample for use during subsequent playback experiments (Fig. S1). The stimulus set was composed of 7 distinct songs, totalling 957 syllables, 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]).

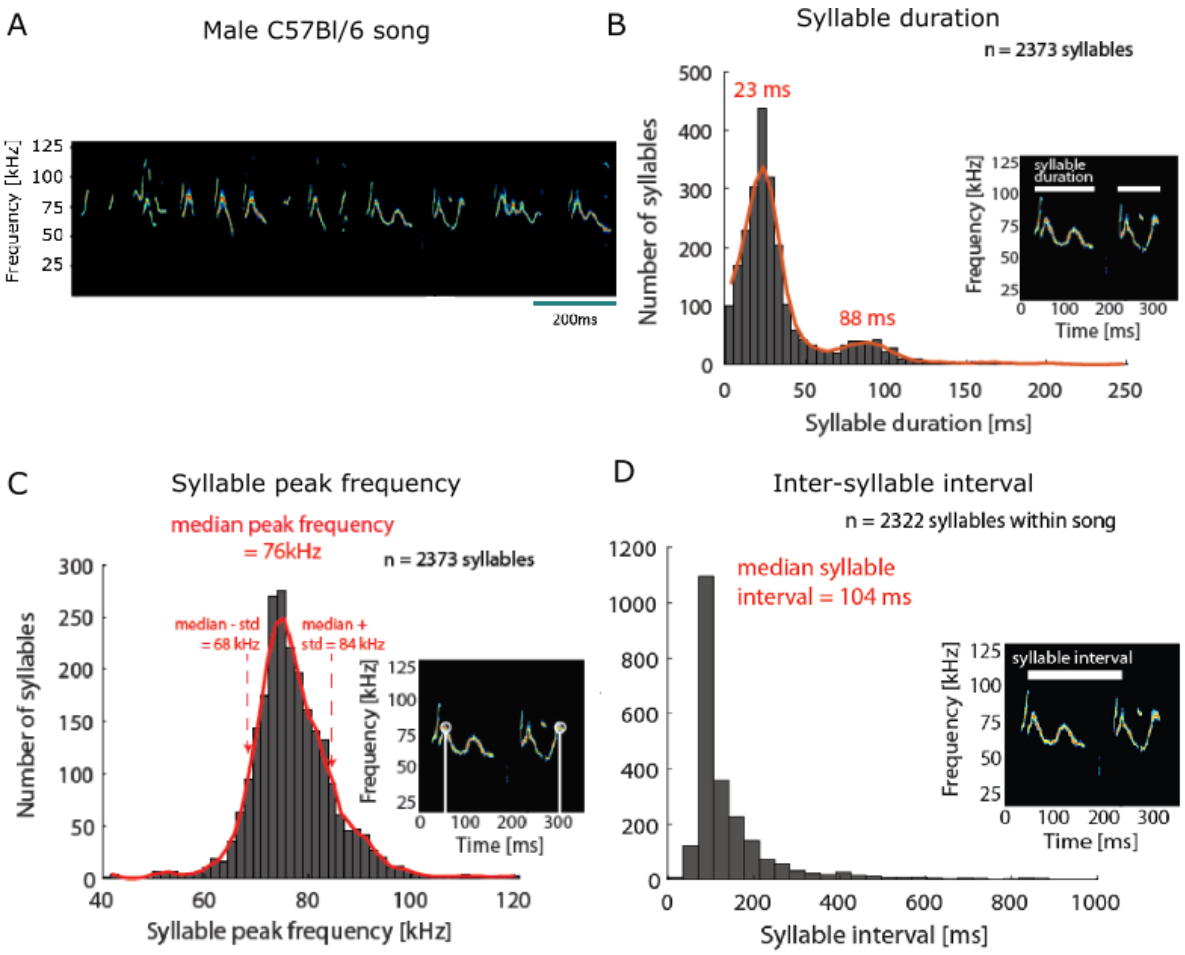

Figure 1: Acoustic features of C57Bl/6 mouse courtship songs

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

B. Distribution of individual ultrasonic syllable duration across a larger set of recorded male vocalizations 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).

D. Distribution of inter-syllable interval durations

Female mice approach playbacks of male songs over silence

Using our pre-recorded subset of male vocal sequences for stimulation within an ethologically-relevant behavioural paradigm, we then aimed to confirm that female mice displayed a preferential approach response to male songs. 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-chamber behavioural chamber (Fig. 2). Importantly, olfactory stimulation from mixed male bedding samples placed under each of the loudspeakers (Fig. 2B) was present throughout the testing session, in order to increase sexual arousal and motivation in the female listeners (Hammerschmidt et al. 2009, Asaba et al. 2014) 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 (10mn; Fig. 2A, C), one-sided playback of male vocalizations was initiated while the second speaker remained inactive (Fig. 2D, left panel). The pre-recorded set of 7 songs was repeated 4 times during one behavioural session, with the order of individual songs randomized, and the playback side alternating between trials

(Fig. 2D). A female mouse's approach response to song playback was quantified using a preference index (Fig. 2F and Methods) comparing the time an animal spent within 10cm of the sound playback source vs the silent speaker across a session (green (grey) outlines in Fig. 2B (D), respectively and Fig. 2E). This was shown by a number of previous studies to be the most sensitive readout for similar assays of female approach to male song playback (Hammerschmidt et al. 2009, Asaba et al. 2014, Musolf et al. 2015) Young female C57Bl/6 mice (aged 6-12 weeks) 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 spent more time at the source of male song playback than at the silent speaker in ¾ trials in the session, irrespective of the playback side, resulting in a positive preference index (25.3%,median across 4 trials; black dot in Fig. 2F). Female listeners tested in this paradigm demonstrated significant approach response to the playback of our subset of male songs (mean preference index significantly larger than 0, one-sample *t*-test, $t(28) = 2.83$, $p = 0.0086$; Fig. 3A), indicating that the females' behaviour discriminated between the sides of song playback and silence. This preferential approach to songs over silence did not significantly differ across the 4 sound presentation trials within a session (no significant effect of trial number on preference index value, one-way repeated measures ANOVA, $f(3, 84) = 0.40$, $p = 0.75$, $\eta_p^2 = 1.40$; Fig. 3B). Together, these results replicate previous work by other laboratories (Hammerschmidt et al. 2009, Shepard and Liu 2011, Asaba et al. 2014, Chabout et al. 2015, Musolf et al. 2015) and confirm the feasibility of recreating an ethologically-driven behavioural assay of female approach response to song playback in the controlled laboratory environment.

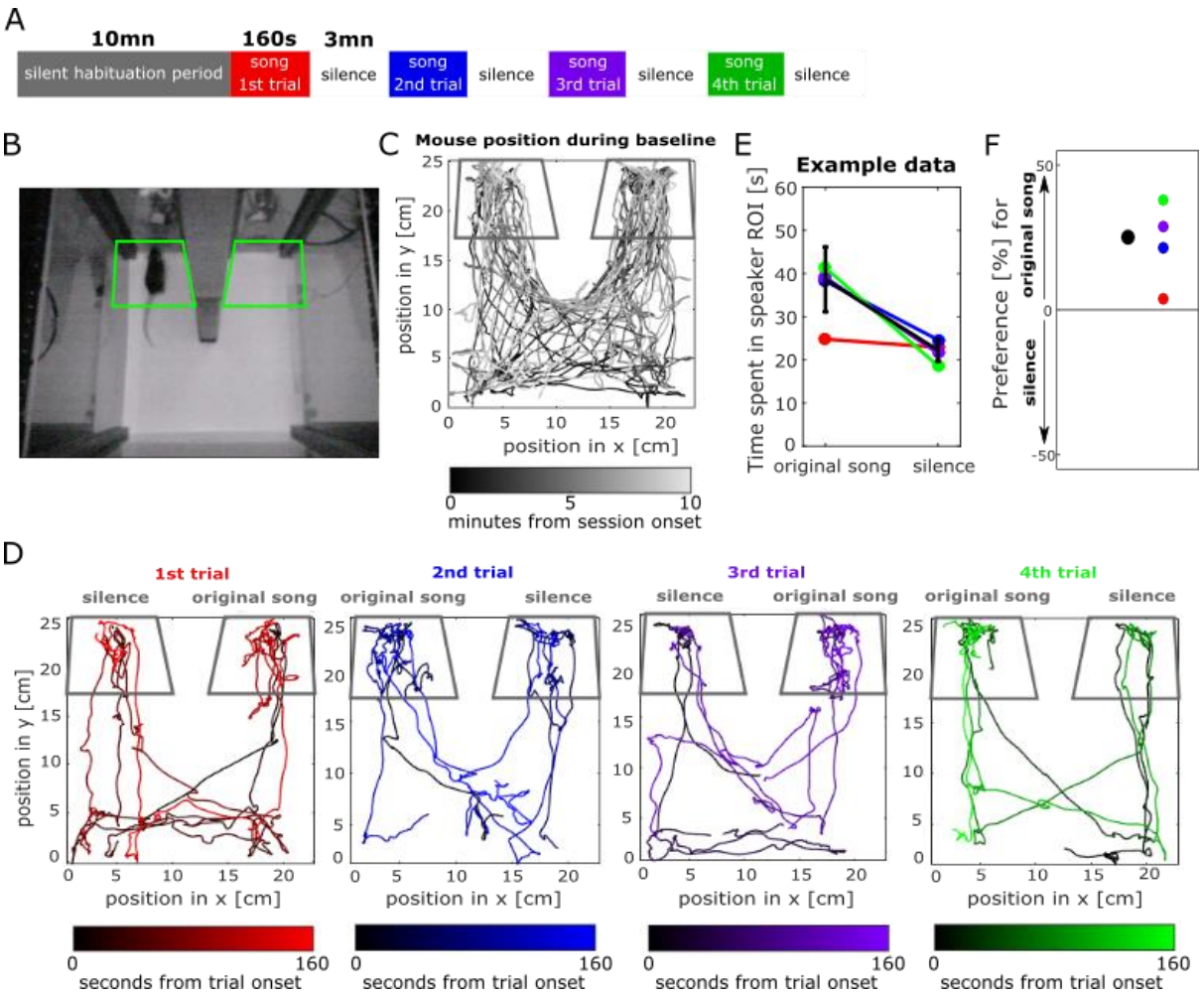

Figure 2: Example behavioural session

A. Experimental timeline for one behavioural session (40mn duration).

B. One frame of a behavioural session, recorded under infrared illumination, showing the two-compartment behavioural box with a soundproof partition (middle) positioned between two ultrasonic loudspeakers (top). Green outlines show the two "speaker zones" used as regions of interest for quantifying the animal's position.

C. Tracking of the animal's position during the 10mn silent habituation period at the start of a behavioural session. The colour code indicates time since the start of the behavioural session.

D. Tracking of the animal's position during each of the 4 sound presentation trials of a behavioural session, in this case contrasting playback of the intact songs from one side with silence from the other side. Playback sides are swapped between successive trial, such that the first and third trial (red and purple), and the second and fourth trials (blue and green) represent identical stimulation conditions, respectively. The colour code indicates time since the onset of song playback for each sound presentation trial. Dark grey outlines indicate the two speaker zones.

E. Total time the animal spent in the speaker zone corresponding to the side of song playback or silence during playback of one instance of the stimulus set. Black dots indicate the session average (median ± standard error of the mean (SEM) across the 4 trials)

F. Preference index quantifying the animal's relative place preference in response to song playback compared to silence. A preference index was computed as the difference between the time spent in the speaker zone corresponding to the song playback side, and silence, and normalized by the sum of the time spent in either speaker zones. Coloured dots (right) indicate trial-based preference indices. Black dot (left) is the median of the 4 trial-based preference indices, used to summarize the animal's preferential approach behaviour in one session. A positive preference index value reflects the animal's preferential approach of intact song playback sources over silence.

Natural behaviour for testing the behavioural sensitivity of female listeners to acoustic features in male songs

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 simultaneous playback of the pre-recorded set of male songs from one speaker, and an acoustically modified version of the songs from a 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 listeners' ability to perceive a change along the manipulated acoustic dimension. In contrast, equal approach behaviour to both intact and manipulated song playback could be interpreted as females not perceiving, or not using, the tested acoustic dimension in their natural behavioural response to courtship song.

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

In the first instance we evaluated whether listeners would be sensitive to two types of manipulations affecting changes in 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 (Chabout et al. 2015), we tested whether randomizing the order of the syllables in the song would affect their approach behaviour (Fig. 3C, left). 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. 3C, right). 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 for, such as reversing the direction of frequency trajectories (Fuzessery 1994, Tian and Rauschecker 1998, Lui and Mendelson 2003, Razak and Fuzessery 2006, Geis and Borst 2013, Issa et al. 2017, Sollini et al. 2018) (Fig. 3D, left). Temporal reversal of the song 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. 3D, right). 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.

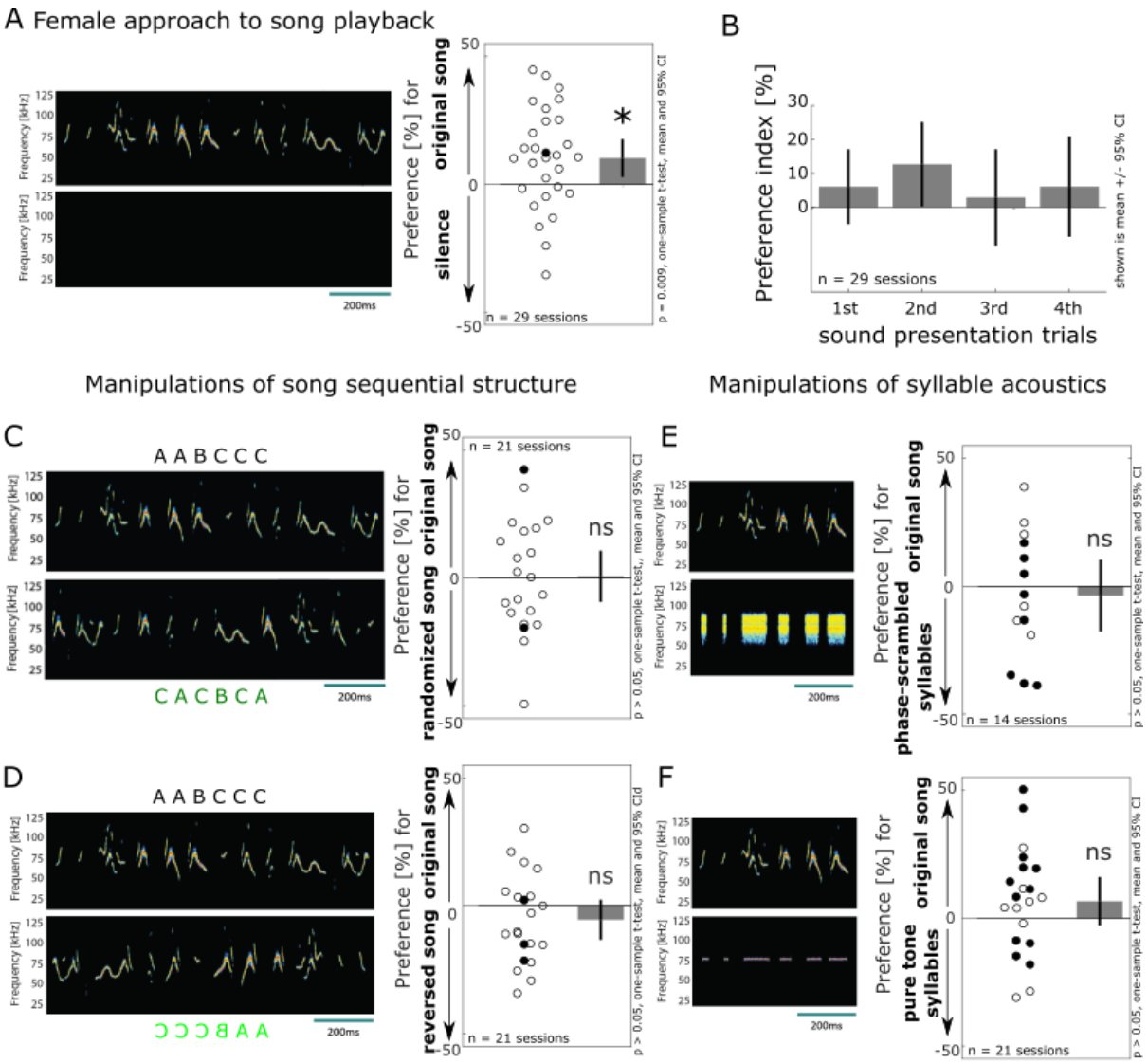

Figure 3: Female mice preferentially approach male courtship song playback over silence, but are not sensitive to sequential and spectral manipulations of courtship songs

A. Female approach behaviour during playback of intact male songs (top) vs silence (bottom). Circles indicate the preference index displayed by individual animals in one behavioural session, computed as the median value of 4 sound presentation trials (see Fig. 2). Empty circles indicate sound playback at 58 dB SPL, filled circles at 68 dB SPL. The population mean preference index in response to song playback relative to silence significantly differed from zero ($t(28) = 2.83$, $p = 0.0086$), indicating that the behaviour of female mouse listeners discriminates between song playback and silence.

B. The preference for song over silence is maintained across the 4 sound presentation trials in a session

C Female approach behaviour during simultaneous playback of intact male songs (top) and corresponding randomized syllable sequences (bottom, $t(20) = 0.126$, $p = 0.90$).

D Playback of intact songs (top) contrasted with temporally reversed songs (bottom, $t(20) = -1.46$, $p = 0.16$)

E. Playback of intact songs (top) and sequences of phase-scrambled syllables (bottom, $t(13) = -0.56$, $p = 0.59$).

F. Playback of intact songs (top) contrasted with sequences of pure tones (bottom, $t(20) = 1.46$, $p = 0.16$)

Histograms show means and error bars show 95% confidence intervals. One-sample two-tailed *t*-test, **: $p < 0.01$, ns: non significant, $p > 0.05$

Female approach behaviour is robust to removal of syllable spectrotemporal dynamics

Since auditory neurons are typically tuned to specific spectro-emporal features in the song syllables, we next wondered whether two different 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 spectrotemporal information of individual syllables by shaping each waveform with a random noise envelope (Fig. 3E, left). 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(13) = -0.56, p = 0.59; Fig. 3E, right). Further, 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) = 1.46, p = 0.16; Fig. 3F). These findings show that female behaviour was robust to the removal of fast spectrotemporal structure in the syllables, at least when sound energy was present at the natural syllable peak frequency.

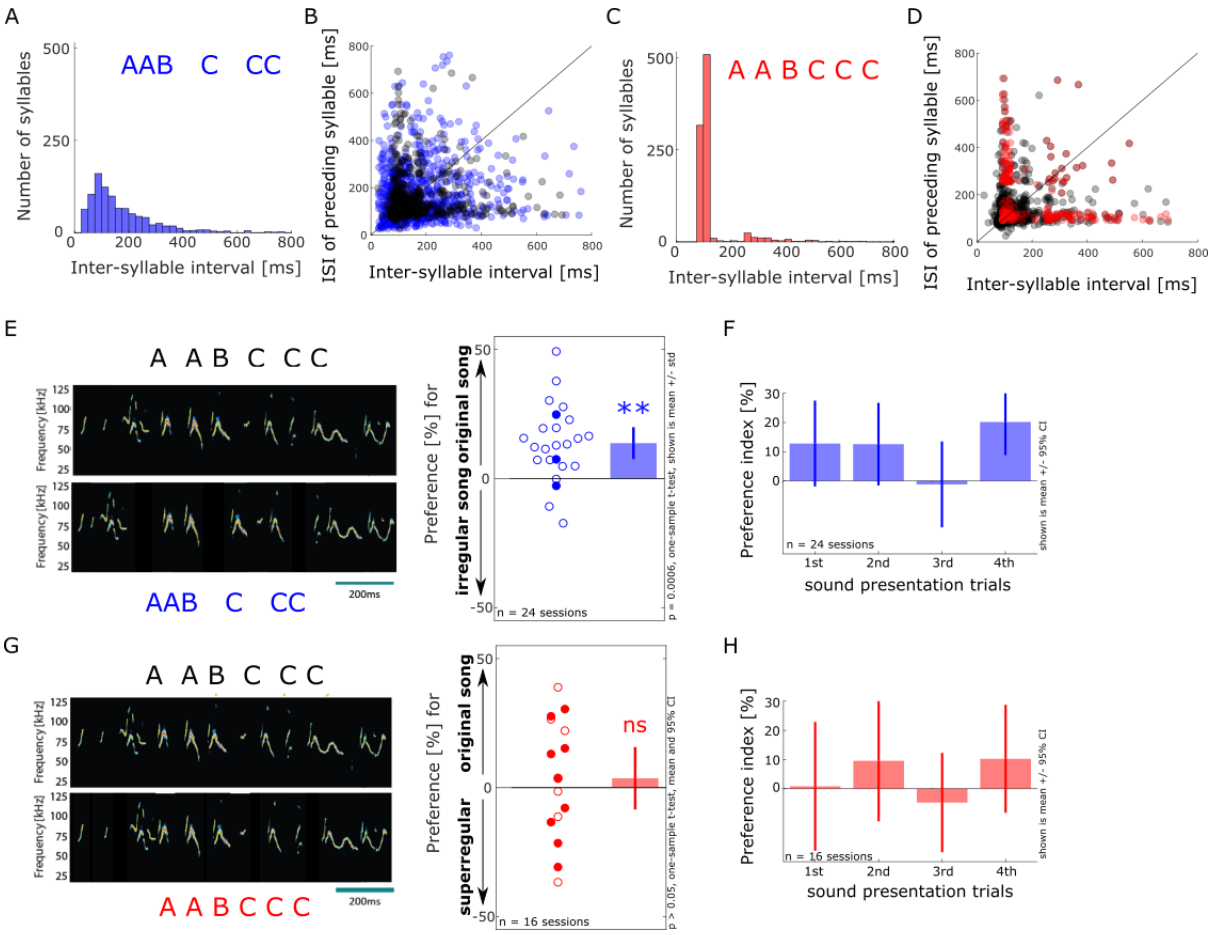

Figure 4: Female mice are sensitive to temporal manipulation of courtship songs, and preferentially approach intact over temporally irregular songs.

A. Distribution of inter-syllable interval durations across the temporally irregular song set (n = 957 syllables)

B. Sequential relationships of inter-syllable interval durations in the intact (shaded grey dots) and temporally irregular (shaded blue dots, n = 957 syllables).
C. Distribution of inter-syllable interval durations across the super-regular song set (n = 957 syllables).
D. Sequential relationship of inter-syllable interval durations in the intact (shaded grey dots) and super-regular (shaded red dots, n = 957 syllables).
E. Female approach behaviour during simultaneous playback of intact male songs (top) and temporally irregular songs (bottom). Circles indicate the preference index displayed by individual animals in one behavioural session, computed as the median value of 4 sound presentation trials (see Fig. 2 and 3). Empty circles indicate sound playback at 58 dB SPL, filled circles at 68 dB SPL. Across the population, preference indices are significantly different from zero ($t(23) = 4.60$, $p = 0.00013$), indicating that the behaviour of female mouse listeners shows a significant preference for song playback over silence.
F. Timecourse of preference for intact over irregular song across the 4 sound presentation trials in a session.
G. Female approach behaviour during simultaneous playback of intact songs (top) contrasted with temporally super-regular songs (bottom, $t(15) = 0.64$, $p = 0.53$).
H. Timecourse of lack of preference for intact over superregular song across the 4 sound presentation trials.
E,G. One-sample two-tailed *t*-test. **: $p < 0.01$, ns: non significant, $p > 0.05$. Histograms show means, and error bars show 95% confidence interval.

Female approach behaviour is reduced by 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 (Fig. 1D), we directly probed whether female listeners would be sensitive to manipulations of the temporal regularity in the song. To do this we generated a set of temporally irregular songs that consisted of the same syllable sequence with a broader distribution of silent pauses compared to intact songs (Fig. 4A, B, E left panel). 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, $t(23) = 4.60$, $p = 0.00013$, Fig. 4E, right). This preferential approach to regular over irregular songs did not significantly differ across the 4 sound presentation trials within a session (no significant effect of trial number on preference index value, one-way repeated measures ANOVA, $f(3, 69) = 1.66$, $p = 0.18$, $\eta p2 = 6.73$, Fig. 4F). This contrasts with the lack of noticeable effect of previous song manipulations on female approach behaviour. Finally, we asked whether female mice, being obviously attracted by rhythmic regularity in the song, would preferentially approach an artificial set of “super-regular” songs over intact ones. To do this 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 (Fig. 4C, D, G left panel). Yet females did not show a behavioural preference when the playback of intact songs was contrasted with that of super-regular songs (one-sample *t*-test, $t(15) = 0.64$, $p = 0.53$; Fig. 4G, right). This behavioural invariance to intact contrasted with super-regular songs did not vary as a function of sound presentation trial within a session (no significant effect of trial number on preference index value, one-way repeated measures ANOVA, $f(3, 45) = 0.63$, $p = 0.60$, $\eta p2= 4.04$, Fig. 4H). Together, these results suggest that the rhythmicity of natural male songs already reaches 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 ethologically-relevant behavioural assay to probe the behaviourally-relevant use of acoustic features from vocal sequences in mice. By utilizing the natural response of females to male-produced courtship songs, we directly tested the degree of invariance in listeners' representation of communication sounds along several acoustic dimensions. The results show that female mice selectively monitor the temporal regularity in male songs, while their behaviour was unaffected by changes in the song sequential organization, or the acoustics 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.

Perceptual vs behavioural invariance

Neurons across the mouse auditory system are tuned to different spectrotemporal features of song syllables, and mouse listeners have been shown to be perceptually sensitive to, or able to detect, the acoustic manipulations used in the current study. The upper limit of this perceptual ability has been determined using go/no-go operant conditioning to extensively train mice to discriminate between syllables from spectrotemporally distinct categories, as well as between intact and temporally reversed or pure tone versions (Neilans et al. 2014) used. However, such externally rewarded reinforcement learning and the associated sharpening of sensory acuity and neuronal tuning likely amplify the perceptual ability of a naive animal (Fritz et al. 2003, Fritz et al. 2005). Alternatively, social behaviour in naturalistic settings likely require 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 behaviour (Wachowitz and Ewert 1996). Our work complements studies quantifying the upper bounds of mouse hearing sensitivity, and shows that, in the context of an ethologically-valid behaviour, female listeners appear to ignore, or show behavioural invariance to, several acoustic features that they likely are, or can be trained to become, perceptually sensitive to.

Insensitivity to the global song structure during natural behaviour

We found the female approach to male song playback was not affected by changes to the global song structure, such as the temporal reversal of the song and the randomization of the syllable sequence in the song. Both of these manipulations preserved the relative spectrotemporal relationships within each syllable, and tested whether the order of the syllables was informative to the listener. Indeed, male songs have been shown to demonstrate some characteristic sequential features, with short syllables typically dominating the beginning of a vocal phrase (Chabout et al. 2015, Matsumoto and Okanoya 2016), which in turn would be violated by the temporal reversal of the song.While mouse listeners are known to perceptually discriminate between syllable types differing in their spectrotemporal features (Neilans et al. 2014), 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, independent of the syllable order. Our result also disambiguates previous findings on mouse sensitivity to syntactic information in male song: Chabout and colleagues (2015) previously showed, in a similar natural behavioural paradigm, that female listeners discriminated between simple and complex songs produced by males in different social contexts. Songs produced in the two

tested social contexts, however, 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. Together these 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 are in agreement with previous work by (Pomerantz et al. 1983) showing that female mice approach playbacks of synthetic 70 kHz ultrasounds over silence, when they were presented from behind one of two devocalized males. Similarly, (Ehret and Haack 1982) found that mothers do not differentiate between narrowband noise models and natural calls during pup retrieval behaviour. Hammerschmidt et al (2009) used a comparable place preference paradigm, and an artificial sequence of irregularly timed short ultrasounds that matched neither the duration or syllable rate of mouse songs, and showed that, in this case, female mice preferentially approached male songs over pure tones. Our results on female's behavioural invariance to pure tone approximations of syllables, combined with our finding that females are highly sensitive to temporal regularity, clarify their finding by suggesting that the preference for intact over ultrasound sequence observed in that study is driven by the disruption in syllable temporal regularity. This, then, suggests that sound energy in and around a critical band around the syllable range, displayed at the natural temporal structure of male songs, is a sufficient approximation of male courtship songs. Comparable robustness to spectral degradation in vocal communication, as long as temporal patterns are preserved, has been demonstrated behaviourally other animal species, including humans: the intelligibility of speech degraded in the spectral domain remains partially preserved if it is shaped with the correct amplitude envelopes (Van Tasell et al. 1987, Shannon et al. 1995).

Behavioural sensitivity to disruption of song temporal regularity

Of the subset of acoustic features tested in this study, temporal regularity in the song was the only feature whose disruption was selectively 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 females preferentially approach intact over irregular songs, suggesting that regularity is attractive to females in the context of vocal-based mate selection. Song temporal regularity is a consequence of breathing patterns regulating the production of syllables by the emitter, with individual calls being generated following the onset of the exhalation (Tschida et al. 2019). 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 regular song in the listener. Thus, song temporal regularity may represent an acoustic cue signalling the fitness or suitability of the potential mating partner to the listener (Egnor & Seagraves, 2016), to which we show that socially motivated listeners are highly attuned to. Interestingly, female behaviour did not appear to discriminate between intact songs

and a “super-regular” version of the songs, perhaps suggesting that natural mouse songs are already at ceiling in terms of the saliency of their temporal regularity: making the song even more regular did not seem to make it more attractive to female listeners. One possible explanation for this counter-intuitive finding, that remains to be directly tested, is that, as previously shown in a different sensory modality (Lak et al. 2010), some light temporal jitter on otherwise highly temporal regular sensory sequences may result in higher signal-to-noise ratio neuronal processing, and ultimately lead to increased perceived intensity of the input sequence.

Interestingly, the syllable rate in mouse songs (~ 7Hz, (Holy and Guo 2005)) broadly overlaps with that of other mammalian vocal patterns such as human speech and monkey vocalizations (Ghazanfar and Takahashi 2014). This stimulus presentation rate, resulting from the breathing rhythm of the singer, also corresponds to the theta frequency range of neuronal oscillations that have been shown to be susceptible to entrainment by streams of regularly presented stimuli. Given that active sensation during whisking and sniffing operates within the frequency range of theta (Kleinfeld et al. 2006, Kepecs et al. 2007), similar constraints may be placed on the cortical mechanisms of hearing, especially during courtship behaviour. Future work will clarify the neuronal substrates underlying the extraction, processing and representation of behaviourally relevant acoustic cues in mouse communication.

Conclusion

In summary, our results show that female mice selectively monitor the temporal regularity in male songs, while showing invariance to changes in the song sequential organization, or the acoustics 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.

Acknowledgments

We thank M. Silvestre for help with testing initial experimental designs, and C. Suess for contributing to data collection. We thank J. Linden for feedback and support. 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).

Declaration of Interests

The authors declare no competing financial interest.

Methods

Animals

90 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 6-12 weeks. 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 (Gouveia and Hurst 2013) was used exclusively. The mice had no sexual experience, but were exposed to two different groups of males through a mesh division for 5mn 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 (Shepard and Liu 2011).

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 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.

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 (Caligioni 2009, Byers et al. 2012). 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 (Byers et al. 2012). In a subset of animals tested between October 2017 and March 2018 estrous stage was identified based on visual examination of the vagina (Caligioni 2009, Byers et al. 2012). 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.

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 40mn, and contained a 10mn silent habituation period, and 4 sound presentation trials (160s of song playback) each followed by a 3mn break (Fig. 2A). The session started with the mouse being placed in the middle of the neutral zone using the handling cylinder. After the 10mn 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, 2mn 40s total sound duration), played in random order. Each trial (playback of one full stimulus set) was followed by a 3mn 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.

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 (Lopes et al. 2015) 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 estimated 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 (0-8cm). 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 in one min-intervals, starting 2mn after onset of silent habituation period and resulting in 35 intervals for one session. Sessions in which the animal displayed a significant side bias (consistent preference for the left or right side of the behavioural box, paired *t*-test, $p < 0.05$,) were excluded from the analysis.

All statistical analyses were performed using MATLAB (version R2018A). Data were tested for deviation from normality (Lilliefors’ composite goodness-of-fit test; Matlab function lillietest) 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.

References

Asaba, A., T. Hattori, K. Mogi and T. Kikusui (2014). "Sexual attractiveness of male chemicals and vocalizations in mice." Front Neurosci **8**: 231.

Asaba, A., S. Okabe, M. Nagasawa, M. Kato, N. Koshida, T. Osakada, K. Mogi and T. Kikusui (2014). "Developmental social environment imprints female preference for male song in mice." PLoS One **9**(2): e87186.

Byers, S. L., M. V. Wiles, S. L. Dunn and R. A. Taft (2012). "Mouse estrous cycle identification tool and images." PLoS One **7**(4): e35538.

Caligioni, C. S. (2009). "Assessing reproductive status/stages in mice." Curr Protoc Neurosci **Appendix 4**: Appendix 4I.

Chabout, J., J. Jones-Macopson and E. D. Jarvis (2017). "Eliciting and Analyzing Male Mouse Ultrasonic Vocalization (USV) Songs." J Vis Exp(123).

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

Chabout, J., P. Serreau, E. Ey, L. Bellier, T. Aubin, T. Bourgeron and S. Granon (2012). "Adult Male Mice Emit Context-Specific Ultrasonic Vocalizations That Are Modulated by Prior Isolation or Group Rearing Environment." Plos One **7**(1).

Egnor, S. R. and K. M. Seagraves (2016). "The contribution of ultrasonic vocalizations to mouse courtship." Curr Opin Neurobiol **38**: 1-5.

Ehret, G. and B. J. J. o. c. p. Haack (1982). "Ultrasound recognition in house mice: Key-Stimulus configuration and recognition mechanism." **148**(2): 245-251.

Fritz, J., S. Shamma, M. Elhilali and D. Klein (2003). "Rapid task-related plasticity of spectrotemporal receptive fields in primary auditory cortex." Nat Neurosci **6**(11): 1216-1223.

Fritz, J. B., M. Elhilali and S. A. Shamma (2005). "Differential dynamic plasticity of A1 receptive fields during multiple spectral tasks." J Neurosci **25**(33): 7623-7635.

Fuzessery, Z. M. (1994). "Response selectivity for multiple dimensions of frequency sweeps in the pallid bat inferior colliculus." J Neurophysiol **72**(3): 1061-1079.

Geis, H. R. and J. G. Borst (2013). "Intracellular responses to frequency modulated tones in the dorsal cortex of the mouse inferior colliculus." Front Neural Circuits **7**: 7.

Ghazanfar, A. A. and D. Y. Takahashi (2014). "The evolution of speech: vision, rhythm, cooperation." Trends Cogn Sci **18**(10): 543-553.

Gouveia, K. and J. L. Hurst (2013). "Reducing mouse anxiety during handling: effect of experience with handling tunnels." PLoS One **8**(6): e66401.

Hage, S. R., N. Gavrilov, F. Salomon and A. M. Stein (2013). "Temporal vocal features suggest different call-pattern generating mechanisms in mice and bats." BMC Neurosci **14**: 99.

Hammerschmidt, K., K. Radyushkin, H. Ehrenreich and J. Fischer (2009). "Female mice respond to male ultrasonic 'songs' with approach behaviour." Biol Lett **5**(5): 589-592.

Hoffmann, F., K. Musolf and D. J. Penn (2012). "Spectrographic analyses reveal signals of individuality and kinship in the ultrasonic courtship vocalizations of wild house mice." Physiol Behav **105**(3): 766-771.

Holy, T. E. and Z. Guo (2005). "Ultrasonic songs of male mice." PLoS Biol **3**(12): e386.

Issa, J. B., B. D. Haeffele, E. D. Young and D. T. Yue (2017). "Multiscale mapping of frequency sweep rate in mouse auditory cortex." Hear Res **344**: 207-222.

Kepecs, A., N. Uchida and Z. F. Mainen (2007). "Rapid and precise control of sniffing during olfactory discrimination in rats." J Neurophysiol **98**(1): 205-213.

Kleinfeld, D., E. Ahissar and M. E. Diamond (2006). "Active sensation: insights from the rodent vibrissa sensorimotor system." Curr Opin Neurobiol **16**(4): 435-444.

Lak, A., E. Arabzadeh, J. A. Harris and M. E. Diamond (2010). "Correlated physiological and perceptual effects of noise in a tactile stimulus." Proc Natl Acad Sci U S A **107**(17): 7981-7986.
Lopes, G., N. Bonacchi, J. Frazao, J. P. Neto, B. V. Atallah, S. Soares, L. Moreira, S. Matias, P. M. Itskov, P. A. Correia, R. E. Medina, L. Calcaterra, E. Dreosti, J. J. Paton and A. R. Kampff (2015). "Bonsai: an event-based framework for processing and controlling data streams." Front Neuroinform **9**: 7.
Lui, B. and J. R. Mendelson (2003). "Frequency modulated sweep responses in the medial geniculate nucleus." Exp Brain Res **153**(4): 550-553.
Matsumoto, Y. K. and K. Okanoya (2016). "Phase-Specific Vocalizations of Male Mice at the Initial Encounter during the Courtship Sequence." PLoS One **11**(2): e0147102.
Musolf, K., S. Meindl, A. L. Larsen, M. C. Kalcounis-Rueppell and D. J. Penn (2015). "Ultrasonic Vocalizations of Male Mice Differ among Species and Females Show Assortative Preferences for Male Calls." PLoS One **10**(8): e0134123.
Neilans, E. G., D. P. Holfoth, K. E. Radziwon, C. V. Portfors and M. L. Dent (2014). "Discrimination of ultrasonic vocalizations by CBA/CaJ mice (Mus musculus) is related to spectrotemporal dissimilarity of vocalizations." PLoS One **9**(1): e85405.
Nomoto, K., M. Ikumi, M. Otsuka, A. Asaba, M. Kato, N. Koshida, K. Mogi and T. Kikusui (2018). "Female mice exhibit both sexual and social partner preferences for vocalizing males." Integr Zool **13**(6): 735-744.
Pomerantz, S. M., A. A. Nunez and N. J. Bean (1983). "Female behavior is affected by male ultrasonic vocalizations in house mice." Physiol Behav **31**(1): 91-96.
Portfors, C. V. (2007). "Types and functions of ultrasonic vocalizations in laboratory rats and mice." J Am Assoc Lab Anim Sci **46**(1): 28-34.
Portfors, C. V. and D. J. Perkel (2014). "The role of ultrasonic vocalizations in mouse communication." Curr Opin Neurobiol **28**: 115-120.
Razak, K. A. and Z. M. Fuzessery (2006). "Neural mechanisms underlying selectivity for the rate and direction of frequency-modulated sweeps in the auditory cortex of the pallid bat." J Neurophysiol **96**(3): 1303-1319.
Shannon, R. V., F. G. Zeng, V. Kamath, J. Wygonski and M. Ekelid (1995). "Speech recognition with primarily temporal cues." Science **270**(5234): 303-304.
Shepard, K. N. and R. C. Liu (2011). "Experience restores innate female preference for male ultrasonic vocalizations." Genes Brain Behav **10**(1): 28-34.
Sollini, J., G. A. Chapuis, C. Clopath and P. Chadderton (2018). "ON-OFF receptive fields in auditory cortex diverge during development and contribute to directional sweep selectivity." Nat Commun **9**(1): 2084.
Tian, B. and J. P. Rauschecker (1998). "Processing of frequency-modulated sounds in the cat's posterior auditory field." J Neurophysiol **79**(5): 2629-2642.
Tschida, K., V. Michael, J. Takatoh, B. X. Han, S. Zhao, K. Sakurai, R. Mooney and F. Wang (2019). "A Specialized Neural Circuit Gates Social Vocalizations in the Mouse." Neuron **103**(3): 459-472 e454.
Van Tasell, D. J., S. D. Soli, V. M. Kirby and G. P. Widin (1987). "Speech waveform envelope cues for consonant recognition." J Acoust Soc Am **82**(4): 1152-1161.
Wachowitz, S. and J. P. Ewert (1996). "A key by which the toad's visual system gets access to the domain of prey." Physiol Behav **60**(3): 877-887.