

## Volitional control of social vocalisations and vocal usage learning in bats

Ella Z. Lattenkamp<sup>1,2</sup>, Sonja C. Vernes<sup>1,\*</sup>, and Lutz Wiegrebe<sup>2</sup>

<sup>1</sup>Neurogenetics of Vocal Communication Group, Max Planck Institute for Psycholinguistics, Nijmegen, NDL

<sup>2</sup>Division of Neurobiology, Ludwig-Maximilians University Munich, DEU

\* corresponding author: [sonja.vernes@mpi.nl](mailto:sonja.vernes@mpi.nl)

**KEYWORDS:** bats, contextual learning, social calls, vocal conditioning, selective reinforcement, communication

### ABBREVIATIONS:

f0: fundamental frequency  
hpcf: high pass criterion frequency  
IQR: Inter quartile range  
SCF: Spectral centroid frequency  
SPL: sound pressure level

### SUMMARY STATEMENT

A novel automated training paradigm reliably elicited social calls from isolated bats and tracked their ability to adjust call features, demonstrating vocal plasticity and usage learning in bats.

## ABSTRACT

Bats are gregarious, highly vocal animals that possess a broad repertoire of social vocalisations. For in-depth studies of their vocal behaviours, including vocal flexibility and vocal learning, it is necessary to gather repeatable evidence from controlled laboratory experiments on isolated individuals. However, such studies are rare for one simple reason: eliciting social calls in isolation and under operant control is challenging and has rarely been achieved. To overcome this limitation, we designed an automated setup that allows conditioning of social vocalisations in a new context, and tracks spectro-temporal changes in the recorded calls over time. Using this setup, we were able to reliably evoke social calls from temporarily isolated lesser spear-nosed bats (*Phyllostomus discolor*). When we adjusted the call criteria that could result in food reward, bats responded by adjusting temporal and spectral call parameters. This was achieved without the help of an auditory template or social context to direct the bats. Our results demonstrate vocal flexibility and vocal usage learning in bats. Our setup provides a new paradigm that allows the controlled study of the production and learning of social vocalisations in isolated bats, overcoming limitations that have, until now, prevented in-depth studies of these behaviours.

## INTRODUCTION

Bats are highly vocal animals that possess a rich repertoire of social vocalisations including sophisticated syllable and song formations (Behr and Von Helversen, 2004; Kanwal et al., 1994; Schwartz et al., 2007; Smotherman et al., 2016; Wright et al., 2013). However, the main body of research on bat vocalisations focuses on their echolocation behaviour (Gillam and Fenton, 2016). This emphasis on echolocation calls is particularly evident when considering controlled laboratory studies of vocalisations, such as psychoacoustic measures to identify vocalisation patterns and perception thresholds in e.g. alternative forced choice experiments (Firzlaff et al., 2006). Even experiments on the Lombard effect (Luo et al., 2015), vocal plasticity (Luo and Wiegrebe, 2016), and vocal learning (Jones and Ransome, 1993) have, to a large extent, been conducted on echolocation calls. This bias has arisen due to the simple and stereotypical structure of echolocation calls and the ease with which they are measured, especially given that they are almost constantly emitted and independent of the social environment.

Meanwhile, bat social calls have received far less attention. Social calls are strongly associated with their respective behavioural context, contain a great deal of information, and show indications of high levels of flexibility (Gillam and Fenton, 2016). Thus far, studies investigating social vocalisations in bats have mainly focused on field studies (Arnold and Wilkinson, 2011; Behr and Von Helversen, 2004; Bohn et al., 2013; Boughman and Wilkinson, 1998), recordings in groups (Bohn et al., 2008; Bohn et al., 2009; Boughman, 1997; Boughman, 1998; Kanwal et al., 1994; Knörnschild et al., 2010), or ontogenetic changes in early developmental stages (Esser, 1994; Esser and Schmidt, 1989; Knörnschild et al., 2006; Prat et al., 2015). Social interactions and behavioural context have a strong impact on vocalisations as they influence the state of the emitter. Accounting for such effects is an important requirement in the study of vocal behaviours, but it is often difficult to accomplish. Disentangling changes in vocalisations triggered by social interactions and those initiated by intrinsic factors is challenging, but vital. The same is true for the precise evaluation of vocal changes due to developmental processes or learning. If changes in call parameters are to be attributed to a specific process, such as learning, detailed observations of the vocal behaviour of isolated animals are required (Siemers and Page, 2009). Yet, herein lies the problem: many animals do not spontaneously produce social calls in isolation and isolated individuals tend to fall silent (Carter et al., 2008). Since failed attempts of studying vocal behaviour in isolation usually remain unpublished, the literature is biased towards results that do not allow a separation of intrinsic variation of vocal performance and vocal changes triggered by a specific social context (Schusterman, 2008). This problem is well-known to bat researchers and as a consequence studies on vocal flexibility in isolated adult bats are rare.

In order to use bats as a model species for vocal conditioning and vocal learning experiments, this limitation needs to be overcome. Frequent social vocalisations are required for vocal conditioning experiments as they represent the working point for positive reinforcement (Schusterman, 2008). In some animals, 'easily' elicited social calls, such as food calls induced by presenting food items (Watson et al., 2015), can be brought under volitional control with comparative ease. Bats do not readily emit

food calls in isolation, making them a challenging system in which to achieve vocal conditioning. However, once volitional control of vocal output is established, social vocalisations can be studied in detail. Such vocal conditioning experiments provide the basis for the in-depth study of vocal development and vocal learning.

The use of operant conditioning paradigms involving positive reinforcement of the desired (approximate) behaviour has produced positive results in the study of vocalisations in mammalian and avian research (Koda et al., 2007; Manabe and Dooling, 1997; Manabe et al., 1997; Manabe et al., 1998; Siemers and Page, 2009). Operant control allows the in-depth investigation of call characteristics, learning behaviour, and the identification of structural constraints on vocalisations (Schusterman, 2008). Based on previous research on vocal control and flexibility in songbirds (Manabe et al., 1997; Manabe et al., 2008) and cetaceans (Richards et al., 1984), we developed an automated real-time setup and training regime, within which isolated adult bats were trained to emit social calls. Using this training regime, we aimed to (a) reliably elicit social calls from isolated bats, (b) establish an automated setup, which allows conditioning of social vocalisations in bats, and (c) track spectro-temporal changes of call parameters in response to modifications in the reward schedule. Once trained to emit social calls in isolation, we challenged the isolated bats to adjust temporal and/or spectral parameters of their calls. This was achieved by gradually increasing the lower cut-off frequency, above which the sound level for trigger reward was measured (high-pass criterion). There are a number of possible ways for the bats to adjust their vocalisations in order to overcome the added level of difficulty imposed by the high-pass criterion. To increase the energy content in frequencies above the cut-off, the bats could switch to a different type of call, increase call duration or call level, shift the energy content of the call to higher frequencies, or increase the fundamental frequency. The behavioural paradigm did not direct the bats towards any of these options and thus the choice of which strategy to use was left to the individual bats. Via digital analyses, we assessed the recorded calls before and after the activation of the high-pass criterion and, in doing so, were able to demonstrate the bats' changes in temporal and spectral call parameters. As a first step this approach establishes the stimulated, but incidental production of social vocalisations. The second step demonstrates induction of vocal modifications through selective positive reinforcement without the use of an auditory template.

## METHODS

**Animals.** Four male adult bats of the species *Phyllostomus discolor* Wagner, 1843 were used for the experiments. The animals originated from a breeding colony in Department II of the Ludwig Maximilian University (LMU) in Munich (Germany). Training and experiments were conducted at the LMU from July until November 2016, five days per week. During training sessions, the bats received food (banana pulp supplemented with infant milk powder, vitamin chalk, and honey) as a reward for successful participation in the training. On the two resting days of the week, the bats received fruits as well as meal worms (larvae of *Tenebrio molitor*). At all times, the animals had access to water *ad. lib.* This experiment was conducted under the principles of laboratory animal care and the regulations of the German Law on Animal Protection. The license to keep and breed *P. discolor* as well as all experimental protocols were approved by the German Regierung von Oberbayern (approval 55.2-1-54-2532-34-2015).

**Experimental setup.** The bats were trained in individual boxes (outside measurements: 40×48×40 cm<sup>3</sup>; w×h×d), which were lined with acoustic foam to reduce sound reflection (**Fig. 1**). All boxes were equipped with one ultrasound microphone (custom-made on basis of SPU0410LR5H, Knowles Corporation, Itasca, IL, USA) and an infrared surveillance camera (Renkforce CMOS, Conrad Electronic, Hirschau, Germany), which was transmitting a live stream from inside the boxes. A self-designed feeding device allowed remote-controlled food reward delivery. The feeding device was a metal box (outside measurements: 12×20×8 cm<sup>3</sup>; w×h×d) housing one speaker (tweeter XT19NC30-04 Peerless, Tymphany HK Ltd., Sausalito, California, USA), a flexible PVC tube (13×0.7 cm) for food reward delivery, and a drip tray. To check the bats' usage of the feeder a photoelectric barrier (EE-SX461-P11 photomicrosensor, Omron Electronics, Langenfeld, Germany) was installed in front of the feeding tube. An orange light emitting diode (LED) next to the feeding tube indicated the feeder's readiness to be activated. The microphone was fixed at 26 cm height on the wall opposite of the feeder (horizontal distance between microphone and feeding tube: 18 cm). Each microphone was connected via a microphone pre-amplifier (OctaMic II, RME, Haimhausen, Germany; level setting: -10dBV) to a multi-channel audio interface (Fireface 800, RME, Haimhausen, Germany), while the loudspeakers were connected to the Fireface via a power amplifier (Harman Kardon AVR 445, Garching, Germany).

**Data acquisition and training.** A self-written Matlab (R2007b, v7.5.0.342, MathWorks, Cambridge, MA, USA) script controlled the data acquisition. A ring buffer of 250 ms length recorded the microphone input from all eight boxes simultaneously (sampling rate: 192 kHz). A call level threshold (40 dB sound pressure level (SPL)) for feeder activation was integrated over the total buffer size. When a call exceeded this level threshold, the recording was saved and the feeder activated (for 300 ms, which equals around 0.1 ml banana pulp discharge). After each activation, the feeder was disabled for a refractory period of 5 s. Echolocation calls alone did not contain enough energy to exceed the call level threshold.

The bats were trained five days a week in a single session per day with an average length of 3.4 (max. 7.5) hours. Outside the training sessions, the bats were kept in the colony room together with 25 conspecifics. The training was split into four stages: 1) Initially all bats had a period of two sessions to get familiar with the setup, the feeding system, and the isolated condition. During this first stage, the feeder was immediately triggered when the bats broke the light barrier located directly above the tube for food delivery (**Fig. 1B**). Thus, in this stage no calls needed to be emitted in order to trigger the feeder. 2) In the second stage, the emission of social calls in isolation triggered a food reward. Whenever a vocalisations exceeded the predefined sound level of 40 dB SPL integrated over a fixed, 250 ms analysis window (call level threshold), a food reward was delivered and the recording was saved. The isolated bats were stimulated either with playbacks (random presentation of previously recorded non-aggressive social calls (from conspecifics from the same colony) in  $20 \pm 5$  s intervals throughout the training session) or constant real-time audio transmission from other boxes. If a bat reliably vocalized in isolation, it was occasionally paired with a second bat, which had not understood the task yet, in order to demonstrate the expected procedure (specifically, Bat 1 was paired four times each with Bat 3 and Bat 4; Bat 2 was paired once each with Bat 3 and Bat 4; Bat 3 was paired four times with Bat 1 and once with bat 2; Bat 4 was paired four times with Bat 1 and once with Bat 2). The pairing lasted 5-60 minutes within a training session and the procedure for food reward delivery was the same as in isolated sessions, i.e. food delivery was triggered when the call level threshold was exceeded by either of the two bats. After 15-22 sessions in the second training stage all bats reliably produced social calls in isolation. 3) In the third stage, the vocalising bats were recorded in isolation and without any auditory input for at least 16 and up to 25 sessions. All calls recorded on the last three sessions of training stage three, i.e. before the activation of the high-pass criterion, were pooled and comprised the baseline for un-stimulated vocalisations in isolation ('pre-criterion' data sets). 4) Stage four began with the activation of a spectral high-pass criterion for the feeder trigger. All social calls exceeding the call threshold continued to be saved. However, the feeder was only triggered if the threshold was exceeded in a frequency range above a high-pass cut-off frequency (high-pass criterion). The cut-off frequency was initially set to 25 kHz and then gradually increased to min. 26 kHz and max. 40 kHz. This increase was individually different and dependent on the individual call types (**Fig. 2, Table S1**). In training stage four, the bats were recorded for up to three weeks (= 9-13 sessions) with the activated high-pass criterion. During the last three recorded sessions of training stage four, i.e. with active high-pass criterion, all individuals had a constant high-pass cut-off frequency (with exception of Bat 4; see **Table S1**). Calls recorded during these last three sessions were also pooled ('post-criterion' data sets). During training stages one and two, the bats were assigned random boxes for each session. In stages three and four, the bats were recorded in the same box every training session.

**Data analysis.** A custom-written Matlab script was used for call analysis. All calls were extracted from 250 ms recordings and individually analysed. Analysed call parameters were call duration, call level, mean fundamental frequency, and spectral centroid (i.e. weighted mean of the frequencies contained in a call). Call levels are given in dB SPL, measured at the microphone. These measurements do not allow a precise statement of the intensity at the sound source, since the bats were able to move freely in the box. The maximal possible distance of the head of the bats and the microphone amounts to 44 cm (corresponding to ~30 dB call level differences). However, the bats usually stayed close to the feeder, in a distance of approximately 10-15 cm to the microphone, which corresponds to record-level variations of no more than 4 dB. We further assume constant movement patterns in all data sets and are thus able to compare relative differences in the pre- and post-criterion data sets. The fundamental frequency was tracked using the yin algorithm (de Cheveigné and Kawahara, 2002) in Matlab. Frequent ‘harmonic jumps’, i.e. falsely tracking the fundamental frequency on the wrong harmonic, were automatically detected and re-calculated to the right frequencies with the help of a custom-written Matlab script. For our analyses, we only used calls with a minimum duration of 5 ms, to conservatively exclude echolocation calls and thus only analyse social calls.

**Statistical analysis.** For the statistical evaluation of call parameter changes, we pooled data from three days before the activation of the high-pass criterion (pre-criterion; base line recordings) and from three days recorded with the high-pass criterion (post-criterion), with at least 8 sessions of adjustment in-between. The pre- and post-criterion data sets contained a total of 6,209 analysed calls. Of these, 377 calls were excluded for being of less than 5 ms length as they are likely to represent echolocation calls rather than communication calls (for exact sample sizes see **Table S1**). For one individual (Bat 4) the type of emitted call changed after the activation of the high-pass criterion (see Results), thus we split its post-criterion data into two categories (‘short’: calls longer than 5 ms, but shorter than 25 ms; ‘long’: calls of 25 ms length or longer). Note that this change in call type led to a greater cut-off frequency for Bat 4 than for the other three bats (**Table S1**). Furthermore, the adjustment of the high-pass criterion was more dynamic for Bat 4, thus its pooled post-criterion data set was recorded with a different cut-off frequency on all three analysed post-criterion days (see: **Table S1**).

Ultimately, we analysed four pre-criterion and five post-criterion data sets (**Table S1**). All data sets were evaluated separately for each bat. Due to the applied call level recording threshold, not all emitted calls were recorded, which led to a non-normal distribution of the data. The one-sample Kolmogorov-Smirnov test for continuous data confirmed that all our data sets differed significantly from a normal distribution. To examine differences between the data before and after the activation of the high-pass criterion, we thus used the Wilcoxon rank-sum test (also called Mann-Whitney U test), which is a non-parametric test for differences in distributions of continuous data. For all data sets we report the number of analysed calls, median, interquartile ranges, and *p*-values of the Wilcoxon rank-sum test (**Table S2**).

## RESULTS

### *Call types and number of emitted calls*

*P. discolor* appears to have a broad repertoire of social calls (unpublished data). However, after an initial exploration phase three of four bats consistently emitted only a single social call in order to trigger the feeder. Bat 4 represented an exception by starting to emit a second call as the training progressed to stage four. The repeatedly emitted social calls were different between the four individuals, i.e. every bat emitted one typical call (or two calls in case of Bat 4 in training stage four) (**Fig. 2**). These five different social call types could be clearly distinguished from each other as they differed in duration, fundamental frequency, and spectral centroid frequency (**Fig. 2, Table S2**).

All bats produced broadband, frequency-modulated social calls with fundamental frequencies between 10 and 25 kHz and several harmonics. The loudest component was usually the fundamental frequency, unlike echolocation calls where the 3<sup>rd</sup> or 4<sup>th</sup> harmonics are the loudest and the fundamental frequency is strongly suppressed. Moreover, social calls are much longer than echolocation calls, the former being typically between 20 and 80 ms and the latter between 1 and 3 ms. Bats 1, 2, and 3 produced rather long (median 40-54 ms), frequency modulated calls, while Bat 4 initially (i.e. pre-criterion) produced a sequence of shorter calls (median of around 6 ms) and only in training stage four (i.e. post-criterion) emitted longer social calls (median ~60 ms) (**Fig. 2D**). Nevertheless, these short social calls were still much longer than echolocation calls and dominated by their fundamental frequency. Thus, these short calls of Bat 4 were still classified as social calls.

After the activation of the high-pass criterion, three of the four bats continued to emit their typical call. Only Bat 4 changed the emitted call type. Before the activation of the high-pass criterion, Bat 4 produced sequences of 2-3 short calls, which only barely exceeded the call-level threshold. In order to exceed the sound-level threshold after the activation of the high-pass criterion, Bat 4 would have needed to emit  $\geq 4$  short calls in the 250 ms interval over which the sound level was integrated (see Methods). This difficulty may be why Bat 4 was the only bat to switch call types (**Fig. 2**). The use of an additional call type by Bat 4 does not present a gradual change of call parameters and as such, call parameters were not compared between long and short calls of Bat 4 (**Table S2**). The statistical results presented for Bat 4 result from a comparison within the short calls of this bat. The long calls of Bat 4 circumvent the need to produce calls in a sequence, which led to fewer calls being recorded in the post-criterion phase for this individual (decrease in call emission rate was -0.39 calls/min post-criterion; **Table S1**). For Bats 1, 2, and 3 the number of recorded calls increased dramatically after the activation of the high-pass criterion. The increase of total recorded calls in the post-criterion phase and the consequent increase of the call repetition rate (call rate increased between 0.29 calls/min (Bat 1) and 5.84 calls/min (Bat 2) post-criterion; **Table S1**) is an indicator for the bats' exploratory behaviour in order to meet the challenges of the high-pass criterion.



### *Changes in call duration and amplitude*

The respiratory system is used to control call on- and offset, and thus determines call duration. The respiratory system also determines call amplitude via the control of subglottal pressure. All four individuals showed significant changes in call duration and amplitude after the activation of the high-pass criterion, suggesting volitional control over their respiratory system during vocalisations.

Extensive call prolongation was observed in Bat 2, which increased its calls from a median length of 40.3 ms to 53.8 ms (IQR: 8.7 ms and 10.5 ms, respectively), while keeping all other call parameters approximately constant (**Fig. 3, Table S2**). Bats 1 and 3 also increased their call durations significantly, but to a lesser extent than Bat 2 (**Fig. 3, Table S2**). After the activation of the high-pass criterion, recorded call amplitudes showed statistically significant changes for all individuals (**Fig. 3, Table S2**). However, these changes in recorded call amplitudes are unlikely to be biologically meaningful as recorded call level changes within the same call type were on the order of  $\leq 2$  dB for all individuals (**Fig. 3, Table S2**).

### *Changes in spectral centroid and mean fundamental frequency*

The spectral parameters (i.e. spectral centroid and mean fundamental frequency) were in general much less plastic than the temporal parameters, which is in line with general findings describing the higher level of difficulty modifying the phonatory and filter systems in contrast to the respiratory system (Fitch, 2000; Janik and Slater, 1997). Although, the activation of the high-pass criterion did indeed lead to statistically significant changes of the spectral centroid for bats 1, 3, and 4, these changes were on the order of  $\leq 0.2$  kHz (difference between medians; **Table S2**).

Bats 1, 2, and 3 showed a statistically significant change in the mean fundamental frequency of their calls in response to the activation of the high-pass criterion (**Fig. 3**). For Bat 1 and 2 a significant decrease of the mean fundamental frequency was recorded (change in medians: -0.23 kHz and -0.07 kHz, respectively; **Table S2**). An increase in mean fundamental frequency was detected for Bat 3 (increase in median: 0.45 kHz; **Fig. 3, Table S2**). The mean fundamental frequency of the calls of Bat 4 (short calls) decreased by 0.03 kHz in response to the activation of the high-pass criterion, which was not a statistically significant change to the baseline calls ( $p = 0.06$ ; **Table S2**).

## DISCUSSION

We have established an automated real-time setup, which makes it possible to reliably elicit and record social calls from isolated bats. It allows the conditioning of social vocalisations and tracking of spectro-temporal changes of call parameters over time. To our knowledge, this constitutes the first report of volitional social call emission, change of spectro-temporal call features, and vocal usage learning in isolated bats in a controlled laboratory setup.

### *Setup and training regime*

Our automated setup can be used to achieve conditioning of bat vocalisations over a relatively short time scale. After only  $\leq 22$  training sessions with acoustic or conspecific stimulation, the four bats reliably produced social calls in isolation to trigger food rewards in the experimental setup. Similar to Manabe and colleagues, who also used a computer-based real-time system (Manabe and Dooling, 1997; Manabe et al., 1995; Manabe et al., 1997), we induced differentiation of vocalisations without a template sound. A maximum of just 11 sessions were needed to induce call parameter changes after initial call emission in isolation was established. This makes it a rapid training paradigm and the automated reinforcement makes this a useful method for studying more complex types of vocal learning.

The training regime used in the present study was not based on the stimulation of call emission by signal presentation, but rather the use of reinforcement of incidental vocalisations. Furthermore, no restrictions on the emitted type of call were imposed, which is known to positively support high rates of call emission and improve motivation (Adret, 1993). Studies using operant control paradigms often aim for a directed manipulation of call features (Pierce, 1985), however this was not our goal. We were interested in the vocal flexibility demonstrated by the bats when trying to overcome the difficulty imposed by the high-pass criterion. This could be achieved by a single or a combination of the following call parameter changes: using a different call type than before, increasing call duration or call level, increasing the spectral centroid frequency or the fundamental frequency.

### *Greater change in temporal than spectral call parameters*

Temporal call parameters (such as call duration and call repetition rate) and call amplitude are in general considerably more flexible and easier to adjust than frequency characteristics as they are only dependent on the respiratory system (Fitch, 2000; Janik and Slater, 1997). Changes in spectral call parameters are reliant on muscular control over the vocal folds and the exact configuration of the vocal tract (i.e. regulation of the resonance of the produced sound), which require neuromuscular control over the phonatory and filter systems, respectively (Fitch, 2000). These different levels of difficulty for the volitional change in call characteristics are reflected in our findings: while call duration is adjusted strongly and with comparative ease, spectral call features are much more static (**Fig. 3**).

Call length was extended as much as 33% (difference between medians, Bat 2, **Table S2**) within only a few experimental sessions. At the same time, the spectral call parameters showed little variation: the spectral centroid frequency increased by max. 0.3% (corresponds to 0.13 kHz difference between medians, Bat 1, **Table S2**), while the strongest registered change in mean fundamental frequency within one call type was an increase by 2.8% (corresponds to 0.45 kHz difference between medians, Bat 3, **Fig. 3, Table S2**). These very small spectral changes are unlikely to be biologically relevant, or indeed even be perceivable for other bats (Kastein et al., 2013; Krumbholz and Schmidt, 1999; Krumbholz and Schmidt, 2001; Preisler and Schmidt, 1998). Although we detected statistically significant changes in recorded call levels over time, the increase did not exceed 2 dB SPL (within one call type, **Table S2**). Since the bats were freely moving in the boxes, changes in recorded sound level due to a change of distance between bat and microphone could easily exceed the measured differences.

The presented data do not allow the unequivocal conclusion that *P. discolor* is capable of modifying frequency call parameters volitionally. The observed changes in frequency characteristics of the calls might be too small to be perceivable for the bats. In this case it is unlikely that the observed spectral changes were due to volitional adjustments of the phonatory and filter systems. Further experiments are needed to investigate their control over the complete vocal system and consequently their vocal learning capability. Vocal production learning has been indicated for *P. discolor* previously (Esser, 1994), but the ability of adult animals to match auditory templates and adjust complex spectral characteristics of their calls has not been studied thus far.

### ***Social call emission for food reward: demonstration of usage learning***

To date, only two types of social calls could reliably be elicited from isolated adult bats: distress calls (Hechavarría et al., 2016) and maternal directive calls (Esser and Schubert, 1998). As yet, a comprehensive call repertoire of social calls has not been published for *Phyllostomus discolor*. This precludes assertions about the context of the social calls recorded in the present study. However, the recorded calls are clearly non-aggressive communication calls. They do not show the characteristic spectro-temporal features of aggressive distress or “screech” calls (in *P. hastatus*: usually noisy broadband calls, which far exceed 100 ms in length (Boughman, 1997; Boughman and Wilkinson, 1998; Wilkinson and Boughman, 1998)). Instead, the calls emitted by our bats are social calls with a clear harmonic structure and approximately constant length (**Fig. 2, Table S2**). As only a single call type was repeatedly emitted by each individual, it is not possible to make statements about the general abundance of these calls in the species call repertoire.

We are certain, for several reasons, that we did not observe affective calls associated with a strongly emotional state of the animals (e.g. aggressive, stressed), but rather reliable communication calls. First, the vocalisations are dissociated from a social context, which could lead to an agitated state of the animals, as the recordings were made in isolation. Second, our experimental paradigm and the resulting recordings clearly show the bats' ability to repeatedly produce the same social call over several weeks, independent of other behavioural processes. Third, the call structure is independent of the behavioural context (e.g. 'food call' emission), since the behavioural context is the same for all individuals, while the emitted calls are dissimilar in structure, spectral features, and duration (**Fig. 2**).

The volitional emissions of social calls in a new context (e.g. labelling objects (Hage et al., 2013; Richards et al., 1984)) is considered contextual usage learning (Janik and Slater, 2000). Bearing in mind the comparable ease of call emission control, it is not surprising that contextual learning (which includes both usage and comprehension learning) can be found much more frequently in the animal kingdom than vocal production learning (Hage et al., 2013; Janik and Slater, 1997; Koda et al., 2007; Seyfarth and Cheney, 2010). We here demonstrated contextual usage learning in *P. discolor* as the bats were able to employ social calls in order to perform a contextually independent task. It is evident that the bats understood the task (i.e. emission of social calls to trigger food rewards) and further changed spectral and/or temporal call parameters when faced with the additional challenge posed by the high-pass criterion. The observed switch between context-independent call types produced by Bat 4 (single long calls vs. multiple short calls) presents an especially strong case for the demonstration of usage learning in *P. discolor*.

In summary, we succeeded in reliably eliciting social calls from isolated *Phyllostomus discolor* bats and, by demonstrating their volitional use of these communication calls out of context, captured contextual usage learning in real-time. Our results demonstrate usage learning and adjustment of call characteristics without any social feedback. Through positive reinforcement, we were able to connect social calls, which do not have an innate relation with food, with such food rewards. We thus demonstrate repeatable contextual learning. Moreover, with the help of our automated setup, we were able to track vocal adjustment in response to the implementation of a high-pass criterion without any directing auditory template. We recorded a vocal exploratory phase after changing thresholds for food reward delivery, which resulted in an adjustment of call parameters, such as change in call duration or call type. Exactly which strategy was used differed between individuals, indicating the bats' versatility when faced with the problem of the introduced high-pass criterion. This demonstration of vocal plasticity and usage learning further highlights the value of bats as a model system in the study of vocal learning in mammals. Establishing this behavioural paradigm with bats as model organism, will in future allow the in-depth investigation of the degree of motor control over the vocal system, effects of audio-vocal feedback, and consequently vocal learning.

## Acknowledgements

The authors want to thank Mirjam Knörnschild, Michael Yartsev, and Uwe Firzlaff for helpful discussion and input leading to the writing of this manuscript. We would also like to thank Lisa Hörtrich for temporarily helping with the setup maintenance in the course of a study project.

## Competing interests

The authors declare no competing interests.

## Author's contribution

EZL carried out the animal training and experimental work. EZL and LW built the setup, conducted the data processing, and analysis. LW and SCV conceived, designed, and coordinated the study and helped draft the manuscript. All authors gave final approval for publication.

## Funding

This work was funded by a Human Frontiers Science Program (HFSP) Research grant (RGP0058/2016) awarded to L.W. and S.C.V. and a Max Planck Research Group Grant awarded to S.C.V.

## Data availability

The raw data and analysis scripts used to prepare the results presented in this article are made available online in Dryad; DOI:XXXXXXX.

## REFERENCES

- Adret, P.** (1993). Vocal learning induced with operant techniques: an overview. *Netherlands J. Zool.* **43**, 125–142.
- Arnold, B. D. and Wilkinson, G. S.** (2011). Individual specific contact calls of pallid bats (*Antrozous pallidus*) attract conspecifics at roosting sites. *Behav. Ecol. Sociobiol.* **65**, 1581–1593.
- Behr, O. and Von Helversen, O.** (2004). Bat serenades - Complex courtship songs of the sac-winged bat (*Saccopteryx bilineata*). *Behav. Ecol. Sociobiol.* **56**, 106–115.
- Bohn, K. M., Schmidt-French, B., Ma, S. T. and Pollak, G. D.** (2008). Syllable acoustics, temporal patterns, and call composition vary with behavioral context in mexican free-tailed bats. *J. Acoust. Soc. Am.* **124**, 1838–1848.
- Bohn, K. M., Schmidt-French, B., Schwartz, C., Smotherman, M. and Pollak, G. D.** (2009). Versatility and stereotypy of free-tailed bat songs. *PLoS One* **4**, 1–11.
- Bohn, K. M., Smarsh, G. C. and Smotherman, M.** (2013). Social context evokes rapid changes in bat song syntax. *Anim. Behav.* **85**, 1485–1491.
- Boughman, J. W.** (1997). Greater spear-nosed bats give group-distinctive calls. *Behav. Ecol. Sociobiol.* **40**, 61–70.
- Boughman, J. W.** (1998). Vocal learning by greater spear-nosed bats. *Proc. R. Soc. B* **265**, 227–233.
- Boughman, J. W. and Wilkinson, G. S.** (1998). Greater spear-nosed bats discriminate group mates by vocalizations. *Anim. Behav.* **55**, 1717–1732.
- Briefer, E. F.** (2012). Vocal expression of emotions in mammals: mechanisms of production and evidence. *J. Zool.* **288**, 1–20.
- Carter, G. G., Skowronski, M. D., Faure, P. A. and Fenton, B.** (2008). Antiphonal calling allows individual discrimination in white-winged vampire bats. *Anim. Behav.* **76**, 1343–1355.
- de Cheveigné, A. and Kawahara, H.** (2002). YIN, a fundamental frequency estimator for speech and music. *J. Acoust. Soc. Am.* **111**, 1917–1930.
- Esser, K. H.** (1994). Audio-vocal learning in a non-human mammal: the lesser spear-nosed bat *Phyllostomus discolor*. *Neuroreport* **5**, 1718–1720.
- Esser, K. H. and Schmidt, U.** (1989). Mother-infant communication in the lesser spear-nosed bat *Phyllostomus discolor* (Chiroptera, Phyllostomidae) - evidence for acoustic learning. *Ethology* **82**, 156–168.
- Esser, K. H. and Schubert, J.** (1998). Vocal dialects in the lesser spear-nosed bat *Phyllostomus discolor*. *Naturwissenschaften* **85**, 347–349.
- Firzlaff, U., Schörnich, S., Hoffmann, S., Schuller, G. and Wiegrebe, L.** (2006). A neural correlate of stochastic echo imaging. *J. Neurosci.* **26**, 785–791.
- Fitch, W. T.** (2000). The evolution of speech: a comparative review. *Trends Cogn. Sci.* **4**, 258–267.

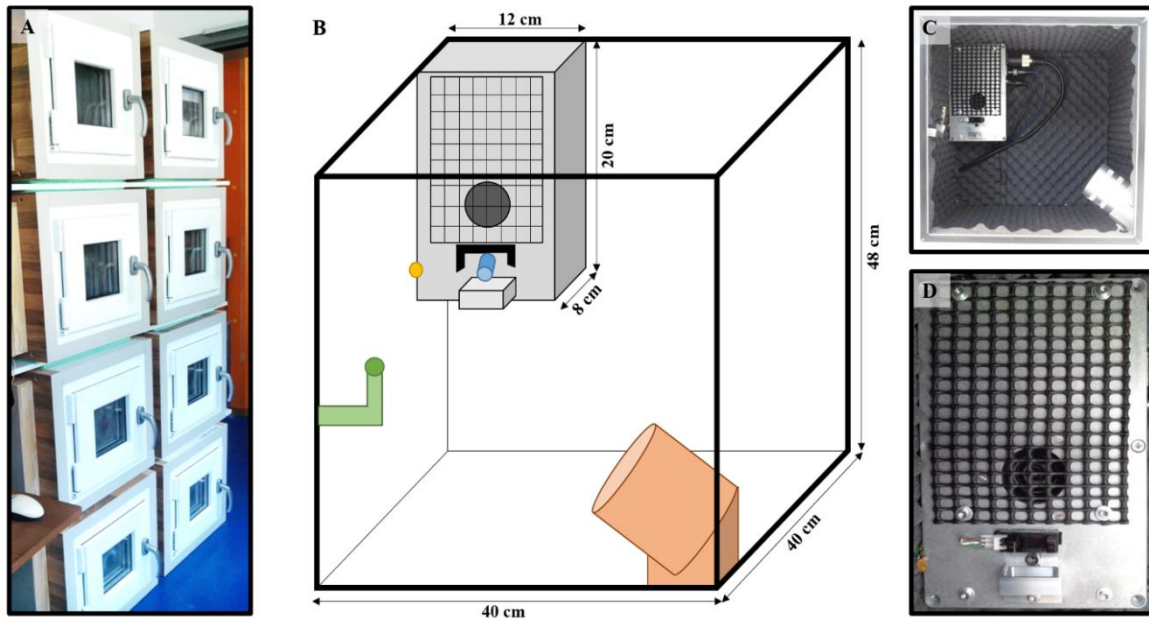
- Gillam, E. and Fenton, M. B.** (2016). Roles of acoustic social communication in the lives of bats. In *Bat Bioacoustics* (ed. Fenton, M. B.), Grinnell, A. D.), Popper, A. N.), and Fay, R. R.), pp. 117–139. Springer New York, NY.
- Hage, S. R., Gavrilov, N. and Nieder, A.** (2013). Cognitive control of distinct vocalizations in rhesus monkeys. *J. Cogn. Neurosci.* **25**, 1692–1701.
- Hechavarría, J. C., Beetz, M. J., Macias, S. and Kössl, M.** (2016). Distress vocalization sequences broadcasted by bats carry redundant information. *J. Comp. Physiol. A* **202**, 503–515.
- Janik, V. M. and Slater, P. J. B.** (1997). Vocal learning in mammals. *Adv. Study Behav.* **26**, 59–99.
- Janik, V. M. and Slater, P. J. B.** (2000). The different roles of social learning in vocal communication. *Anim. Behav.* **60**, 1–11.
- Jones, G. and Ransome, R. D.** (1993). Echolocation calls of bats are influenced by maternal effects and change over a lifetime. *Proc. R. Soc. B* **252**, 125–128.
- Kanwal, J. S., Matsumura, S., Ohlemiller, K. and Suga, N.** (1994). Analysis of acoustic elements and syntax in communication sounds emitted by mustached bats. *J. Acoust. Soc. Am.* **96**, 1229–1254.
- Kastein, H. B., Winter, R., Vinoth Kumar, A. K., Kandula, S. and Schmidt, S.** (2013). Perception of individuality in bat vocal communication: discrimination between, or recognition of, interaction partners? *Anim. Cogn.* **16**, 945–959.
- Knörnschild, M., Behr, O. and von Helversen, O.** (2006). Babbling behavior in the sac-winged bat (*Saccopteryx bilineata*). *Naturwissenschaften* **93**, 451–454.
- Knörnschild, M., Glöckner, V. and Helversen, O. Von** (2010). The vocal repertoire of two sympatric species of nectar-feeding bats (*Glossophaga soricina* and *G. commissarisi*). *Acta Chiropterologica* **12**, 205–215.
- Koda, H., Oyakawa, C., Kato, A. and Masataka, N.** (2007). Experimental evidence for the volitional control of vocal production in an immature gibbon. *Behaviour* **144**, 681–692.
- Krumbholz, K. and Schmidt, S.** (1999). Perception of complex tones and its analogy to echo spectral analysis in the bat, *Megaderma lyra*. *J. Acoust. Soc. Am.* **105**, 898–911.
- Krumbholz, K. and Schmidt, S.** (2001). Evidence for an analytic perception of multiharmonic sounds in the bat, *Megaderma lyra*, and its possible role for echo spectral analysis. *J. Acoust. Soc. Am.* **109**, 1705–1716.
- Luo, J. and Wiegrebe, L.** (2016). Biomechanical control of vocal plasticity in an echolocating bat. *J. Exp. Biol.* **219**, 878–886.
- Luo, J., Goerlitz, H. R., Brumm, H. and Wiegrebe, L.** (2015). Linking the sender to the receiver: vocal adjustments by bats to maintain signal detection in noise. *Sci. Rep.* **5**, 1–11.
- Manabe, K. and Dooling, R. J.** (1997). Control of vocal production in budgerigars (*Melopsittacus*

- undulatus*): selective reinforcement, call differentiation, and stimulus control. *Behav. Processes* **41**, 117–132.
- Manabe, K., Kawashima, T. and Staddon, J. E. R.** (1995). Differential vocalization in budgerigars: towards an experimental analysis of naming. *J. Exp. Anal. Behav.* **63**, 111–126.
- Manabe, K., Staddon, J. E. R. and Cleaveland, J. M.** (1997). Control of vocal repertoire by reward in budgerigars (*Melopsittacus undulatus*). *J. Comp. Psychol.* **111**, 50–62.
- Manabe, K., Sadr, E. I. and Dooling, R. J.** (1998). Control of vocal intensity in budgerigars (*Melopsittacus undulatus*): differential reinforcement of vocal intensity and the Lombard effect. *J. Acoust. Soc. Am.* **103**, 1190–1198.
- Manabe, K., Dooling, R. J. and Brittan-Powell, E. F.** (2008). Vocal learning in budgerigars (*Melopsittacus undulatus*): effects of an acoustic reference on vocal matching. *J. Acoust. Soc. Am.* **123**, 1729–1736.
- Mergell, P., Fitch, W. T. and Herzel, H.** (1999). Modeling the role of nonhuman vocal membranes in phonation. *J. Acoust. Soc. Am.* **105**, 2020–2028.
- Metzner, W. and Müller, R.** (2016). Ultrasound Production, Emission, and Reception. In *Bat Bioacoustics* (ed. Fenton, M. B.), Grinnell, A. D.), Popper, A. N.), and Fay, R. R.), pp. 55–92. New York, NY: Springer New York.
- Pierce, J. D.** (1985). A review of attempts to condition operantly alloprimate vocalizations. *Primates* **26**, 202–213.
- Prat, Y., Taub, M. and Yovel, Y.** (2015). Vocal learning in a social mammal: demonstrated by isolation and playback experiments in bats. *Sci. Adv.* **1**, e1500019.
- Preisler, A. and Schmidt, S.** (1998). Spontaneous classification of complex tones at high and ultrasonic frequencies in the bat, *Megaderma lyra*. *J. Acoust. Soc. Am.* **103**, 2595–2607.
- Richards, D. G., Wolz, J. P. and Herman, L. M.** (1984). Vocal mimicry of computer-generated sounds and vocal labeling of objects by a Bottlenosed dolphin, *Tursiops truncatus*. *J. Comp. Physiol.* **98**, 10–28.
- Schusterman, R. J.** (2008). Vocal learning in mammals with special emphasis on pinnipeds. In *The Evolution of Communicative Flexibility: Complexity, Creativity, and Adaptability in Human and Animal Communication* (ed. Oller, D. K.) and Gribel, U.), pp. 41–70. Cambridge, MA: MIT Press.
- Schwartz, C., Tressler, J., Keller, H., Vanzant, M., Ezell, S. and Smotherman, M.** (2007). The tiny difference between foraging and communication buzzes uttered by the mexican free-tailed bat, *Tadarida brasiliensis*. *J. Comp. Physiol. A* **193**, 853–863.
- Seyfarth, R. M. and Cheney, D. L.** (2010). Production, usage, and comprehension in animal vocalizations. *Brain Lang.* **115**, 92–100.

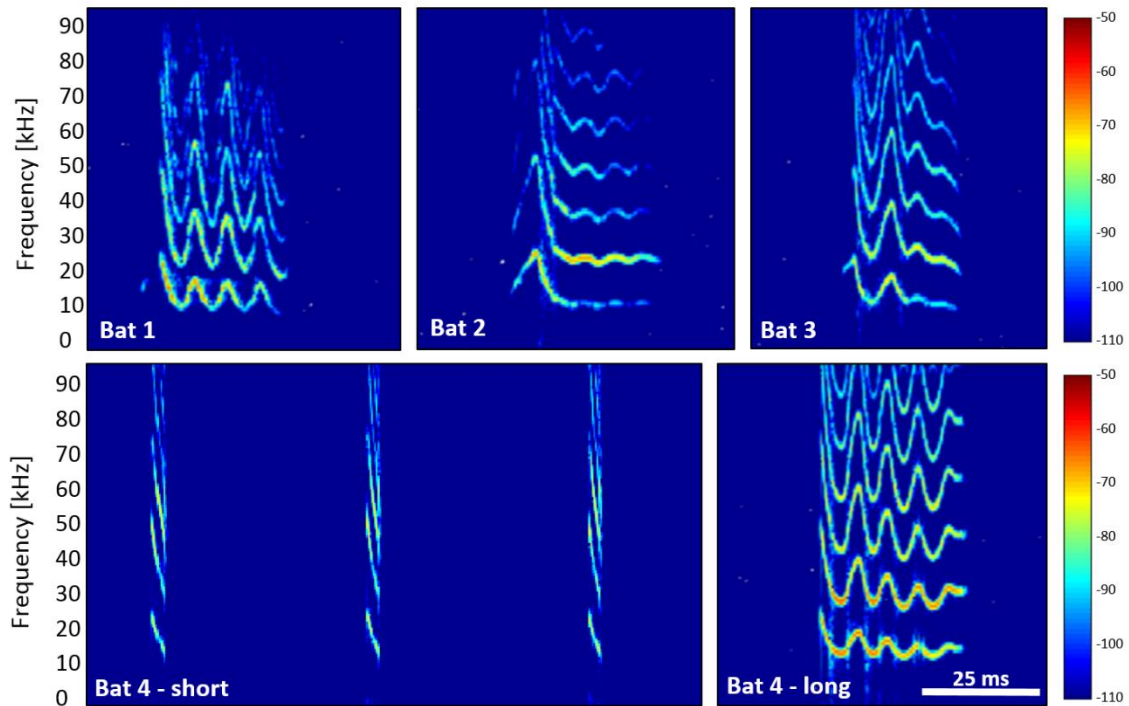


- Shapiro, A. D., Slater, P. J. B. and Janik, V. M.** (2004). Call usage learning in gray seals (*Halichoerus grypus*). *J. Comp. Psychol.* **118**, 447–454.
- Siemers, B. M. and Page, R. A.** (2009). Behavioral studies of bats in captivity: Methodology, training, and experimental design. In *Ecological and behavioural methods for the study of bats* (ed. Kunz, T. H.) and Parsons, S.), pp. 373–392. Baltimore, Maryland: John Hopkins Press.
- Smotherman, M., Knörnschild, M., Smarsh, G. and Bohn, K.** (2016). The origins and diversity of bat songs. *J. Comp. Physiol. A* **202**, 535–554.
- Sutton, D., Larson, C., Taylor, E. M. and Lindeman, R. C.** (1973). Vocalizations in rhesus monkeys: conditionability. *Brain Res.* **52**, 225–231.
- Vernes, S. C.** (2017). What bats have to say about speech and language. *Psychon. Bull. Rev.* **24**, 111–117.
- Watson, S. K., Townsend, S. W., Schel, A. M., Wilke, C., Wallace, E. K., Cheng, L., West, V. and Slocombe, K. E.** (2015). Vocal learning in the functionally referential food grunts of chimpanzees. *Curr. Biol.* **25**, 495–499.
- Wilkinson, G. S. and Boughman, J. W.** (1998). Social calls coordinate foraging in greater spear-nosed bats. *Anim. Behav.* **55**, 337–350.
- Wright, G. S., Chiu, C., Xian, W., Wilkinson, G. S. and Moss, C. F.** (2013). Social calls of flying big brown bats (*Eptesicus fuscus*). *Front. Physiol.* **4**, 1–9.

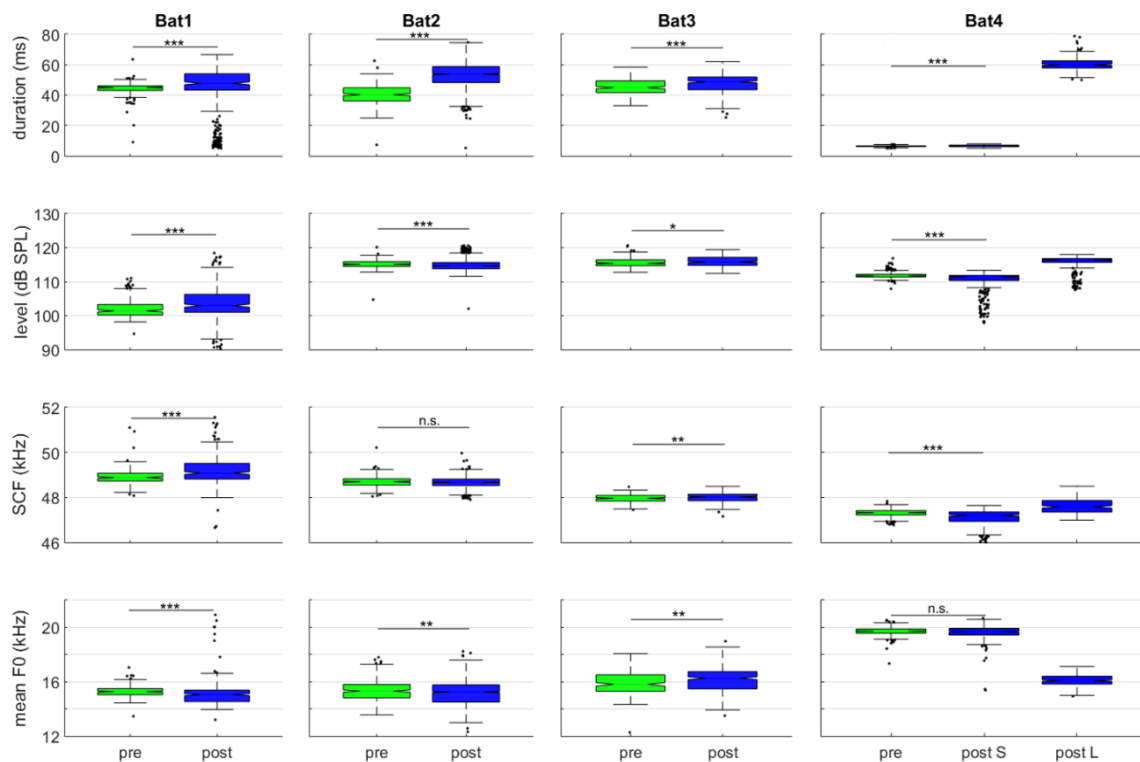
## Figures



**Figure 1: Automated behavioural training setup.** (A) Photo of eight training boxes; (B) Sketch of the inside of a training box equipped with one ultrasound microphone (green), an infrared surveillance camera (orange), an LED (yellow), and a self-designed feeding device (grey), which allowed remote controlled food reward delivery. The feeding device was a metal box housing one speaker, a flexible PVC tube (blue) for food reward delivery, and a drip tray. Furthermore, a photoelectric barrier (black) was installed in front of the feeding tube; (C) A photo of the setup inside the training box; (D) Close-up picture of the feeding device.



**Figure 2: Spectrograms of ‘typical calls’ for Bats 1-4.** The calls from Bat 1, 2, and 3 (top) were recorded from isolated individuals in sound attenuated training setup before the activation of the high-pass criterion. These calls were the same in structure of the fundamental frequency before and after the activation of the high-pass criterion. Bat 4 produced a sequence of short calls pre-criterion (lower left panel), but started to also use long calls (lower right panel) in response to the activation of the high-pass criterion.



**Figure 3: Change of temporal and spectral call parameters before (pre) and after (post) activation of the high-pass criterion.** Reported are changes of call duration, call sound level, spectral centroid frequency (SCF), and mean fundamental frequency (F0) for Bats 1-4 (from left to right). Box plot summary statistics indicated the median line and interquartile ranges for each box. Each bar contains pooled calls from three days (pre- and post-criterion). For Bat 4 the post-criterion results are split into short (S) and long (L) calls. As these long calls of Bat 4 present an additional call type and not a gradual change from the short call type, no statistical comparison between these call types was conducted. Statistical results of the Wilcoxon rank-sum test for the other call types are indicated (n.s. ( $p \geq 0.05$ ); \* ( $0.05 > p > 0.01$ ); \*\* ( $0.01 \geq p > 0.001$ ); \*\*\* ( $p \leq 0.001$ )). For detailed p-values see **Table S2**.

## SUPPLEMENTARY TABLES

**Table S1: Summary of calls used for the analyses.** Reported is the total call count within the training sessions for Bats 1-4 (split into ‘short’ and ‘long’ calls in the post-criterion phase for Bat 4) and call numbers dependent on session length. Furthermore, number of calls per session, which had a length of less than 5 ms, and were thus excluded from further analyses as they might represent echolocation calls, are reported. The applied high-pass criterion frequency (hpcf) is given in the rightmost column.

Bat	date	total call count		calls		
		(short/long)	total calls/min	< 5 ms	hpcf	
<b>1</b>		10.10.2016	126	0.53	0	0 kHz
	pre-criterion	20.10.2016	37	0.08	1	0 kHz
		21.10.2016	76	0.19	0	0 kHz
			15.11.2016	200	0.88	79
	post-criterion	16.11.2016	135	0.61	33	26 kHz
		17.11.2016	169	0.60	39	26 kHz
<b>2</b>		10.10.2016	154	1.14	1	0 kHz
	pre-criterion	19.10.2016	82	0.49	0	0 kHz
		20.10.2016	94	0.40	0	0 kHz
			03.11.2016	1107	11.41	0
	post-criterion	04.11.2016	1120	4.12	0	27 kHz
		18.11.2016	624	2.00	0	27 kHz
<b>3</b>		10.10.2016	120	0.50	2	0 kHz
	pre-criterion	19.10.2016	35	0.20	0	0 kHz
		20.10.2016	72	0.30	1	0 kHz
			04.11.2016	84	0.31	1
	post-criterion	14.11.2016	86	0.28	1	27 kHz
		15.11.2016	230	0.80	10	27 kHz
<b>4</b>		10.10.2016	386 (380/0)	1.62	6	0 kHz
	pre-criterion	19.10.2016	247 (247/0)	1.44	0	0 kHz
		20.10.2016	265 (261/0)	1.10	4	0 kHz
			16.11.2016	368 (188/107)	1.64	73
	post-criterion	17.11.2016	238 (84/74)	0.89	80	38 kHz
		18.11.2016	154 (63/45)	0.47	46	40 kHz

**Table S2: Summary of effect sizes and results of the statistical analyses.** Bat 4 uses an additional call type ('long') in the 'post-criterion' phase. As the call parameters for these long calls of Bat 4 do not experience a gradual change, but present a new call type, we did not statistically compare these calls to the 'pre-criterion' data set.

Bat		pre-criterion		post-criterion		difference in median	p-value
		median	IQR	median	IQR		
<b>1</b>	call number	238		353			
	duration [ms]	45.0	3.1	47.8	10.8	2.8	***
	level [dB SPL]	101.4	3.1	102.9	5.3	1.5	***
	F0 [kHz]	15.28	0.45	15.05	0.85	-0.23	***
	SCF [kHz]	48.88	0.35	49.08	0.70	0.20	***
<b>2</b>	call number	329		2851			
	duration [ms]	40.3	8.7	53.8	10.5	13.5	***
	level [dB SPL]	115.1	1.4	114.7	1.9	-0.4	***
	F0 [kHz]	15.32	0.99	15.25	1.28	-0.07	0.004
	SCF [kHz]	48.70	0.28	48.67	0.29	-0.03	<b>0.143</b>
<b>3</b>	call number	224		388			
	duration [ms]	44.9	7.6	48.7	8.4	3.8	***
	level [dB SPL]	115.4	1.7	115.8	2.4	0.4	0.023
	F0 [kHz]	15.81	1.22	16.26	1.27	0.45	0.003
	SCF [kHz]	47.96	0.25	48.02	0.29	0.06	0.009
<b>4</b>	call number	888 (short)		335 (short)			
	duration [ms]	6.4	0.5	6.7	1.1	0.3	***
	level [dB SPL]	111.7	0.8	111.4	1.5	-0.3	***
	F0 [kHz]	19.73	0.32	19.70	0.49	-0.03	<b>0.062</b>
	SCF [kHz]	47.32	0.19	47.20	0.42	-0.12	***
	call number			226 (long)			
	duration [ms]			59.8	4.8		
	level [dB SPL]			116	1		
	F0 [kHz]			16.08	0.57		
	SCF [kHz]			47.20	0.42		

IQR = interquartile range; F0 = mean fundamental frequency; SCF = spectral centroid frequency; \*\*\*

=  $p \leq 0.001$ ; short: calls between 5 and 25 ms length; long: calls  $\geq 25$  ms.