OXFORD

Article

Vocal individuality and rhythm in male and female duet contributions of a nonhuman primate

Dena J. CLINK[a,*], Johny S. TASIRIN[b], and Holger KLINCK[a]

[a]Bioacoustics Research Program, Cornell Lab of Ornithology, Cornell University, Ithaca, NY, 14850, USA and
[b]Faculty of Agriculture, Sam Ratulangi University, Manado, Indonesia

*Address correspondence to Dena J. Clink. E-mail: djc426@cornell.edu.

Abstract

Duetting, or the stereotypical, repeated and often coordinated vocalizations between 2 individuals arose independently multiple times in the Order Primates. Across primate species, there exists substantial variation in terms of timing, degree of overlap, and sex-specificity of duet contributions. There is increasing evidence that primates can modify the timing of their duet contributions relative to their partner, and this vocal flexibility may have been an important precursor to the evolution of human language. Here, we present the results of a fine-scale analysis of Gursky's spectral tarsier *Tarsius spectrumgurskyae* duet phrases recorded in North Sulawesi, Indonesia. Specifically, we aimed to investigate individual-level variation in the female and male contributions to the duet, quantify individual- and pair-level differences in duet timing, and measure temporal precision of duetting individuals relative to their partner. We were able to classify female duet phrases to the correct individual with an 80% accuracy using support vector machines, whereas our classification accuracy for males was lower at 64%. Females were more variable than males in terms of timing between notes. All tarsier phrases exhibited some degree of overlap between callers, and tarsiers exhibited high temporal precision in their note output relative to their partners. We provide evidence that duetting tarsier individuals can modify their note output relative to their duetting partner, and these results support the idea that flexibility in vocal exchanges—a precursor to human language—evolved early in the primate lineage and long before the emergence of modern humans.

Key words: primate vocal communication, rhythm, vocal plasticity, vocal individuality

Duetting, or the stereotypical coordinated vocalizations between 2 or more individuals (Langmore 2002), occurs across a diverse range of taxa including insects (Bailey 2003), frogs (Tobias et al. 1998), birds (Thorpe et al. 1972; Hall 2004), and nonhuman primates including the families Tarsiidae (tarsiers; Burton and Nietsch 2010; Nietsch 1999), Pitheciidae (titi monkeys; Müller and Anzenberger 2002; Robinson 1979, 1981), Hylobatidae (gibbons and siamangs; Geissmann and Orgeldinger 2000; Geissmann 1999, 2002; Keith et al. 2009), and in some lemur species, such as the indri (Indriidae; Tecot et al. 2016) and sportive lemur (Lepilemuridae; Méndez-Cárdenas and Zimmermann 2009). In nonhuman primates, duetting behaviors co-occur with pair-living and territoriality (Haimoff 1986). Although the function of duets in primates is still a topic of research, primate duets may mediate intergroup spacing (Mitani 1985a), moderate territorial conflicts (Mitani 1985b; Cowlishaw 1992), and serve to advertise and/or strengthen the pair-bond (Geissmann and Orgeldinger 2000). In nocturnal primates, duets may also serve an extra function of facilitating the reuniting of individuals after a night of foraging alone (the pair reunion model; MacKinnon and MacKinnon 1980; Méndez-Cárdenas and Zimmermann 2009).

As primate duets typically travel over multiple conspecific territories, it seems likely that they are providing information to conspecifics about caller or group status and identity. It is possible that in territorial animals listeners can glean information from caller identity simply from the location of the caller, but more likely that call structure also provides cues to caller identity (Ham et al. 2016; Torti et al. 2018), and this may vary according to call type (Volodin et al. 2011; Bouchet et al. 2012). The majority of analyses on individual- or pair-level variation in primate duets have focused on the gibbons, or lesser apes (Oyakawa et al. 2007; Keith et al. 2009; Terleph et al. 2015, 2018; Clink et al. 2017, 2018b; Lau et al. 2018), with some work on indris (Gamba et al. 2016; Torti et al. 2017). Vocal individuality may be simply the result of a unique combination of genetic and environmental factors on the ontogeny of individual calls in the absence of selective pressures for individual distinctiveness (McGregor 1993; Suthers 1994). For example, in ring-tailed lemurs *Lemur catta*, it was found that morphological variation in vocal tracts explains individuality in vocalizations (Gamba et al. 2017). But, the fact that certain call types are more individually distinct than others provides evidence that some calls have been under selection to provide cues to caller identity (Mitani et al. 1996).

At its most simple definition, rhythm is defined as a pattern of events in time (McAuley 2010), and the ability to perceive and follow rhythm is ubiquitous across human cultures, along with the ability for flexible turn-taking in vocal exchanges (Stivers et al. 2009). The ubiquity across cultures implies that these abilities may have been a precursor to human language, and there is some evidence for the ability for flexible turn-taking in nonhuman primates. Primate duets generally exhibit stereotyped, alternating exchanges between individual callers (Haimoff 1986). In male white-handed gibbons, it was found that they flexibly time their contribution to the duet relative to the female, and tend to interrupt females only if female calls exhibit abnormal structure (Terleph et al. 2018). In indris, the proportion of time spent cosinging was influenced by the dominance status of the caller, with dominant animals overlapping more than nondominants (Gamba et al. 2016), indicating that these animals can also flexibly time their vocalizations.

Tarsiers are small, nocturnal primates that are found only in Southeast Asia, specifically in Borneo, Sumatra, Sulawesi, the Philippines, and surrounding islands (Groves and Shekelle 2010). Tarsiers are a particularly interesting group of primates as they exhibit a range of social organization, from solitary to pair-living (Gursky 2003), with the pair-living tarsiers on Sulawesi regularly engaging in duets (MacKinnon and MacKinnon 1980). Tarsier duets tend to occur around dawn, as the animals return to their sleeping trees, but are also sometimes emitted during the night in territorial encounters with other groups (MacKinnon and MacKinnon 1980), providing evidence that duets function, at least in part, to mediate territorial conflicts. Species with high-amplitude duets are predicted to be communicating with extra-pair receivers (Dahlin and Benedict 2014), and tarsier duets can be heard up to 500 m in neighboring territories (Gursky 2015). Since tarsier duets can be heard over multiple conspecific territories, they are probably communicating with extra-pair individuals, and it is possible that some portion of their duet contains information about individual or pair identity. Sulawesi tarsier duets have previously been shown to exhibit strong patterns of geographic variation, with different species and populations across Sulawesi showing distinct duet patterns (Nietsch 1999; Shekelle et al. 2008; Burton and Nietsch 2010).

Here we report the results of a fine-scale analysis of Gursky's spectral tarsier *Tarsius spectrumgurskyae* duets recorded in Tangkoko National Park, Sulawesi. To our knowledge, there have not been analyses of individual variation in duet contributions, or of duet timing in Sulawesi tarsiers. The tarsier lineage split from all other extant primates at least 58 mya (Goodman et al. 1998), and had an exceptionally long period of independent evolution (Merker et al. 2009). Therefore, a thorough, quantitative assessment of tarsier duets may contribute to our understanding of the factors that contributed to the evolution of duetting across the Order Primates, as well as provide possible insights into the refinement of rhythm and turn-taking abilities that occurred over the course of human evolution (Gamba et al. 2016). Specifically, our objectives were to 1) investigate individual variation in the male and female contributions to the duet; 2) quantify the rhythm (Sasahara et al. 2015) of tarsier duets to understand rhythmic consistency within and between individuals; 3) quantify and compare differences in rate of note repetition between males and females; and 4) to examine the temporal precision of male and female contributions relative to each other, along with the amount of overlap or cosinging between callers.
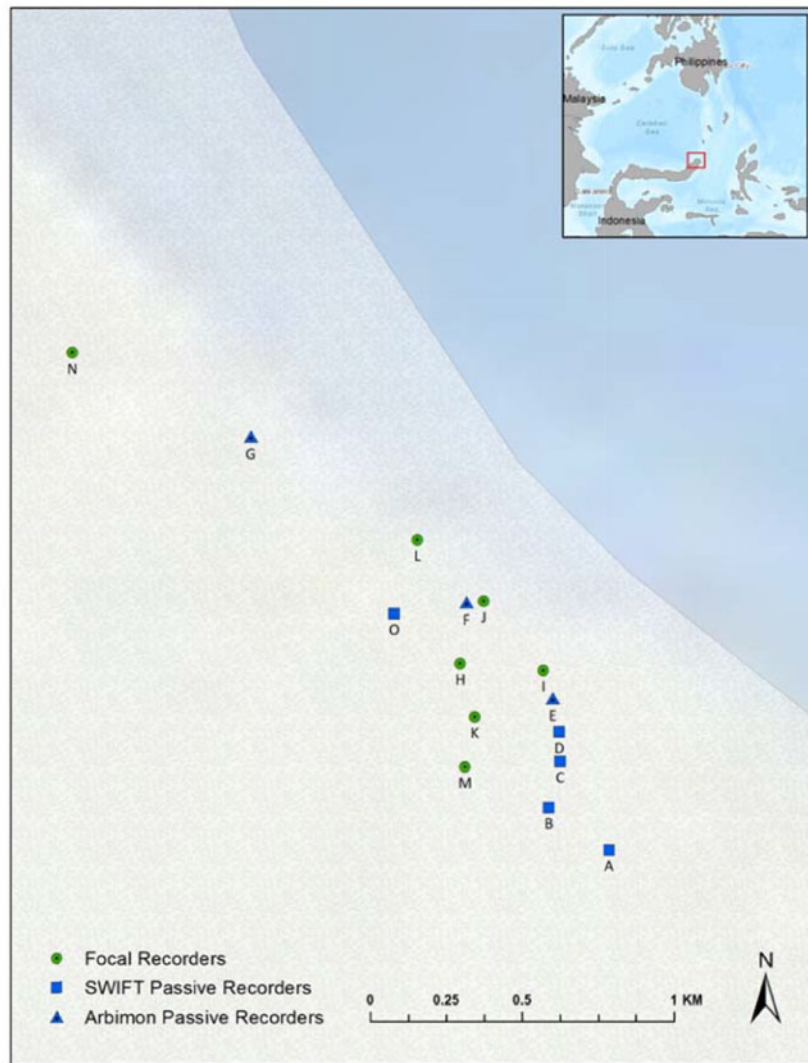
Methods

Study site and subjects

Our study focused on Gursky's spectral tarsier (referred to hereafter as spectral tarsiers, *T. spectrumgurskyae*, Shekelle et al. 2017) in Tangkoko National Park, Sulawesi, Indonesia (1°33′45.30″N, 125°10′17.53″E; Figure 1). Data for this study were collected during July and August 2018. We collected acoustic data using a combination of focal and autonomous recordings (see description below). Tarsier population density at Tangkoko National Park is ~156 individuals per km² (Gursky 1998) with tarsier territory size varying from 0.016 km² (1.6 ha) to 0.041 km² (4.1 ha; Gursky 1998). Tarsiers exhibit a preference for sleep trees in the genus *Ficus*, show a high fidelity to a small number of sleep trees in their territory (Gursky 2003), and call from their sleep tree each morning (MacKinnon and MacKinnon 1980). We distinguished between pairs based on recording location alone but given certain aspects of tarsier behavior and ecology—specifically their territoriality, nonoverlapping home ranges and sleep tree fidelity—we felt confident in distinguishing between pairs in this manner. Our criteria for distinguishing among pairs were as follows: the pair must be recorded >150 m away from another pair (for autonomous recordings) or we had to make visual and auditory contact with the pair (in the case of focal recordings). In contrast to most other nonhuman primates, unhabituated tarsiers exhibit relatively little fear in the presence of potential predators (human or otherwise; MacKinnon and MacKinnon 1980), so it seems unlikely that the presence of human observers substantially altered their duetting behavior.

Acoustic data collection

We collected data using a combination of focal recordings and strategically placed autonomous recorders near to known sleeping trees. For focal recordings, we used a RØDE NT-USB Condenser Microphone (Røde Microphones, Sydney, Australia) connected to a 32 GB Apple iPad Air (Apple Inc., Cupertino, CA) and the Voice Record Pro application at a sampling rate of 44.1 kHz and 16 bits. In addition, we used 2 different autonomous recording devices, either an ARBIMON portable recorder (Aide et al. 2013) which

Figure 1. Map of recording locations of tarsier pairs in Tangkoko National Park, Sulawesi, Indonesia. Each point denotes the recording location of a tarsier pair and the shape of the points reflects the type of recorder used (see "Methods" section for details).

recorded at 44.1 kHz and 16 bits, or a SWIFT recorder (Koch et al. 2016) recording at 48 kHz and 16 bits. ARBIMON recorders were programmed to record from 6 PM to 6 AM daily, SWIFT recorders recorded 24 h continuously, and focal recordings were taken opportunistically during the early morning hours by D.J.C. and a research assistant. All recordings were saved as Waveform Audio Files and ARBIMON recorders saved 1-h long files at a size of 317.5 MB, SWIFT recorders saved 40-min files at a size of 230.4 MB and focal recordings files were of variable duration and size. The number of recording days per autonomous recorder ranged from 2 to 7. Although tarsier duets can be heard up to 500 m away by a human observer (Gursky 2015), the detection distance of our recording devices was much lower than that—typically <50 m for high-quality recordings. Therefore, all recordings included in the present analysis were taken within ∼50 m of the calling pair. Given their relative predictability in calling location, we were able to get high-quality recordings of tarsier duets during focal recordings, and using the autonomous recorders with relative certainty of pair identity. Using the autonomous recorders, there were some cases wherein we were able to get high-quality recordings over multiple days. We include a

summary of the number of phrases for each sex by recorder type along with number of recording days in Table 1. The uneven sample size between males and females is the result of having to exclude certain male phrases because there were 2 males vocalizing at the same time, and in 1 case there was a tarsier female who called without a male.

Acoustic analysis

Male and female tarsiers exhibit a high-degree of sex-specificity in their duet contributions (Nietsch 1999; Burton and Nietsch 2010), so we could easily distinguish male and female duet contributions from the spectrograms. For our acoustic analyses, we included only duets with a high signal-to-noise ratio (≥ 10 dB) where it was clear there was only one calling individual of each sex. Tarsier duets are composed of multiple phrases wherein the male produces rapidly repeating notes, and the female produces a series of notes that increase in duration and decrease in bandwidth (see Figure 2 for a representative spectrogram of a tarsier duet and phrase). We focused our analysis on phrases within the duet, as opposed to the whole duet, for the following reasons. First, the number of phrases in the

Table 1. Summary of tarsier pairs and duet phrases analyzed in this study

Pair ID	Sex	Number of recording days	Number of phrases	Sex	Number of recording days	Number of phrases	Recorder type
A	F	1	4	M	1	4	Autonomous (SWIFT)
B	F	2	10	M	2	10	Autonomous (SWIFT)
C	F	2	12	M	2	12	Autonomous (SWIFT)
D	F	1	4	M	1	4	Autonomous (SWIFT)
E	F	3	13	M	2	9	Autonomous (ARBIMON)
F	F	2	12	M	2	11	Autonomous (ARBIMON)
G	F	1	10	M	1	3	Autonomous (ARBIMON)
H	F	1	4	M	1	4	Focal
I	F	1	1	M	1	1	Focal
J	F	1	3	M	1	3	Focal
K	F	1	1	M	1	1	Focal
L	F	1	3	M	1	3	Focal
M	F	1	3	M	1	3	Focal
N	F	1	3	M	0	0	Focal
O	F	1	9	M	1	9	Autonomous (SWIFT)
		Total female phrases	92		Total male phrases	77	

We analyzed duet phrases from 14 males and 15 females. In some cases using the autonomous recorders, we recorded high quality duets over multiple mornings, and the number of recording days is equivalent to number of duets included per pair.

duet was variable, making it difficult to decide which features from the whole duet to include for our analyses, particularly as our methods of discriminating individuals required us to use the same number of features for each phrase (see "Classification of individuals" section). Second, the phrases were highly stereotyped, making it easy to discern when they start and end, which made them a relatively easy unit of analysis for comparison. Third, previous studies of vocal individuality in nonhuman primates (Terleph et al. 2015; Clink et al. 2017; Lau et al. 2018), or on duet rhythm (Gamba et al. 2016) have focused on a particular call type or phrase within a duet, as opposed to the entire duet. And lastly, previous qualitative analyses of tarsier duets focused on the duet phrase, so our analysis will be comparable with previous analyses of tarsier duets (Shekelle 2008).

Given the limitations of data storage capabilities for the ARBIMON autonomous recorders and our desired sampling scheme, we were necessarily limited to a sampling rate of 44.1 kHz. Tarsiers have ultrasonic vocalizations (Gursky 2015), but our analyses focused on the fundamental frequency of the duet phrases, which range from 6 to 14 kHz, indicating that a sampling rate of 44.1 kHz was more than sufficient for our analysis. Before analysis and to ensure spectrogram measurements were comparable among different recorder types, we low-pass filtered the data with a 48 kHz sampling rate >22.5 kHz using a Butterworth filter, and then down-sampled to a sampling rate of 44.1 kHz using Audacity(R) 2.3.0 software (2018). We created spectrograms using Raven Pro 1.5 Sound Analysis Software (Bioacoustics Research Program, Cornell Lab of Ornithology, Ithaca, NY). Spectrograms were created with a 1600-point (33.3 ms) Hann window (3 dB bandwidth = 43.1 Hz), with 50% overlap, and a 2048-point discrete fourier transform, yielding time and frequency measurement precision of 16.7 ms and 24.1 Hz.

There were variable numbers of introductory notes for each phrase, so we defined the start of a phrase as the first instance in a duet where the female notes were emitted with a spacing of ≤2 s, and we defined the end of the phrase as the instance when the female stopped calling for at least 2 s. We created separate selection tables for males and females in Raven Pro, and for each note estimated the following features: start and stop time, note duration, along with the minimum frequency (kHz), and the maximum frequency (kHz). To increase intraobserver reliability in spectrogram measurements,

we used the robust measurements in Raven, which calculate measurements based on the energy of the particular call segment and are less sensitive to user variability (Charif et al. 2010). For our duration features, we used 90% duration, and for our frequency features, we used 5% and 95% frequencies, so actual duration and frequency values were slightly lower or higher than those reported. For each phrase, we also estimated the duration of the phrase, the number of notes, and the rate of note output (number of notes divided by duration of the phrase; Figure 3).

Classification of individuals

For most modeling and discriminative tasks, it is a requirement that the length of each feature vector for each observation is equal (in this case, the number of features included for each phrase of the tarsier duet). The fewest number of female notes in a particular tarsier phrase was 6; therefore, we were limited to including the features of 6 or fewer notes in our analyses of female duet phrases, and for consistency we only included 6 notes for male duet phrases. The commonly used method of classifying individual vocalizations—linear discriminant function analysis (Terleph et al. 2015; Clink et al. 2017)— was inappropriate for our analysis as we included multiple phrases from multiple tarsiers which were in some cases collected over multiple days, meaning that our study had a 3-factorial design, and not a 2-factorial design that is required when using linear discriminant function analysis (Mundry and Sommer 2007). If the same individual (or pair) is recorded at 2 separate times, this violates the assumptions of statistical independence for linear discriminant function analysis (Venables and Ripley 2002). If we would have included only phrases from each pair recorded on a single day in our analysis, then we could have used a linear discriminant function analysis without violating the assumptions, but we wanted to include the largest sample size possible, so chose to use a "state-of-the-art" support vector machine (SVM; Cortes and Vapnik 1995).

The use of a permuted discriminant function analysis would have been appropriate in our case (Mundry and Sommer 2007), but we opted for the use of the SVM, which is a more commonly used classification algorithm in human speech recognition (Dahake et al. 2016), and increasingly in studies of nonhuman primate
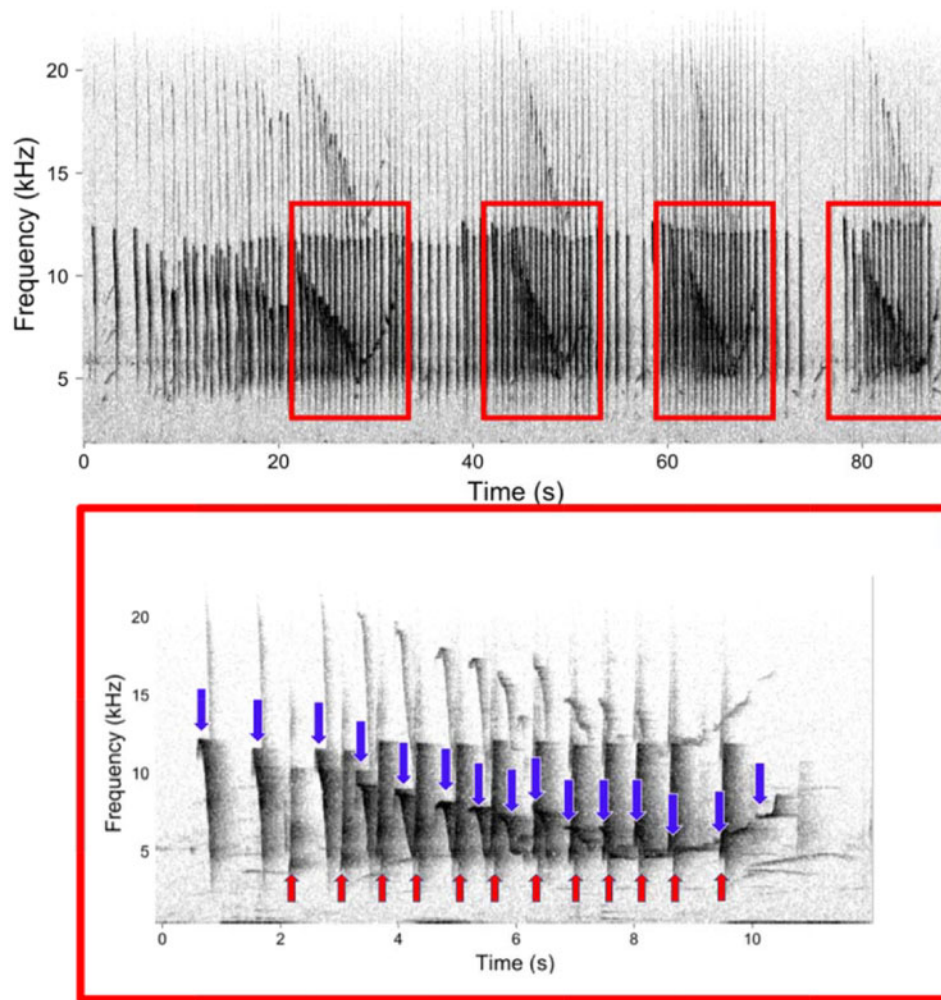
Figure 2. Representative spectrograms of a tarsier duet with multiple repeated phrases (top) and a single phrase (bottom). Female notes are denoted by blue, downward facing arrows and male notes are denoted by red, upward facing arrows. Spectrograms were created using the same settings outlined in the methods, and the single phrase spectrogram was zoomed in on the time axis.

vocalizations (Fedurek et al. 2016; Turesson et al. 2016; Clink et al. 2018a). SVMs have fewer assumptions regarding data independence (Hsu and Lin 2002) than linear discriminant function analysis and were therefore a more reasonable choice for our classification problem. To see how well different individuals could be distinguished based on features estimated from the spectrograms, we used the R package "e1071" to create multiclass SVM models (which are appropriate for classification when there are >2 classes or groups) using a radial kernel, and we used the "tune" function to estimate the best values for the gamma and cost parameters. We used leave-one-out cross-validation to determine our classification accuracy. We also used SVM recursive feature elimination (Kuhn 2008; Colby 2011) to rank the features estimated from the spectrogram in terms of their importance for classifying individuals.

Pair-level signatures

Given the subjective nature of spectrogram feature extraction, we were interested to see if there were pair-level differences in duets that we were potentially missing based on our choice of features. Estimation of Mel-Frequency (Hz) cepstral coefficients (MFCCs) provides an automated alternative to spectrogram feature extraction, and MFCCs are commonly used in human speech recognition

and animal call classification applications (Clemins et al. 2005; Lee et al. 2006; Chou et al. 2008; Mielke and Zuberbühler 2013; Clink et al. 2018a). We calculated MFCCs for the duet phrase (as opposed to estimating features for individual notes like with spectrogram feature extraction), and as tarsier duets consist of alternating notes between males and females with substantial overlap, MFCCs calculated for each phrase necessarily reflect pair-level differences, as opposed to individual differences between males and females. We calculated MFCCs for each tarsier phrase using the R package "tuneR" (Ligges et al. 2018).

As with any type of classification using SVM, we needed to create MFCC feature vectors for each tarsier phrase that were of equal length. Tarsier phrase duration varied among pairs (ranging from 4 to 36 s), so we divided each phrase into an equal number of time windows and calculated MFCCs for each time window. We found that using 4 time windows resulted in the highest classification accuracy, and including more time windows did not increase our accuracy. We divided each phrase into 4 time windows, and calculated 12 MFCCs for each time window along with the delta-cepstral coefficients, which are the first-order derivatives of the original cepstral coefficient, and are meant to capture dynamics of MFCCs over the course of the signal (Beigi 2011; Kumar et al. 2011). We omitted the
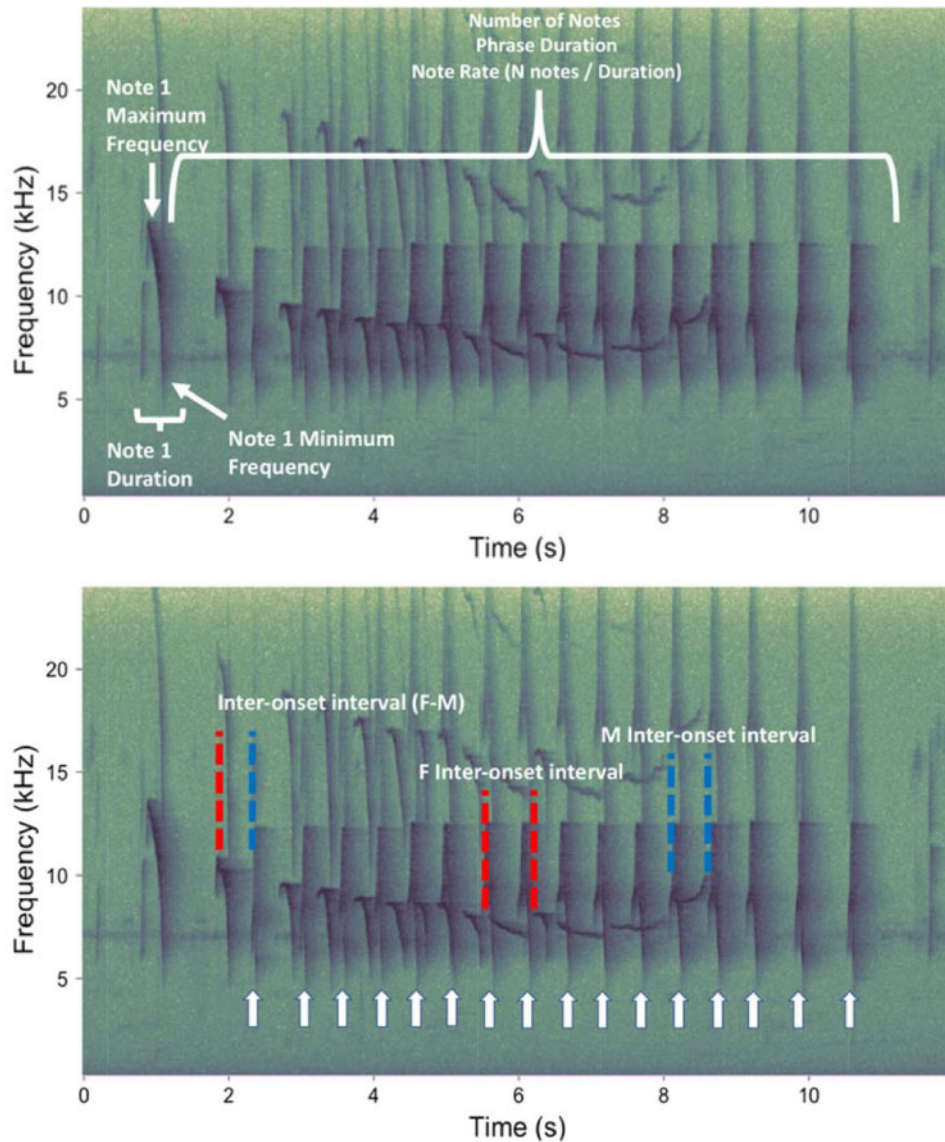
Figure 3. Representative spectrogram of a tarsier phrase outlining the features used for individual classification (top) and the temporal measures taken to estimate male and female rhythm bands, along with temporal precision of males relative to females (bottom). The white arrows indicate individual male notes. For each male and female note, we estimated the duration, minimum frequency (kHz), and maximum frequency (kHz). For each phrase, we estimated the duration, number of notes, and note rate (number of notes over total duration) for males and females separately. Spectrograms were created in Raven Pro using the spectrogram settings outlined in the methods.

first MFCC for each time window, as this MFCC is correlated with signal strength and is dependent on distance of caller from the recording device (Han et al. 2006; Muda et al. 2010). We also included the duration of the phrase which resulted in a feature vector of length 65. We then used SVMs as outlined above to classify tarsier phrases by pair.

Male and female rhythm

To quantify rhythm structure of tarsier phrases, we used an approach that was originally developed to quantify developmental trajectories of birdsong (Sasahara et al. 2015). For each of the male and female phrase contributions, we calculated the interonset intervals using Raven Pro selection tables for consecutive notes s_i and s_{i+i} where $\mathrm{IOI}(_i) = t(s_{i+1}) - t(s_i)$ for male notes and female notes

separately (Figure 3). We calculated a kernel density estimate for the interonset intervals for each phrase, using the "density" function in R and a smoothing bandwidth of 0.1 ms. We then computed the local maxima for each density PLOT using the discrete analogue to the second derivative applied to the kernel density estimate. Different rhythm peaks are the result of a varying number of "local maxima" or peaks of interonset interval densities, and following Sasahara et al. (2015) we term these peaks "rhythm bands." A representative density plot highlighting our method of estimating rhythm bands is shown in Figure 4. We estimated the number of rhythm bands for each phrase for each tarsier individual, and used the R package "ggpubr" (Kassambara 2017) to create boxplots of number of rhythm bands for males and females separately. We then used the "compare_means" function to calculate a 1-way analysis of variance to test for differences between males, females, and males
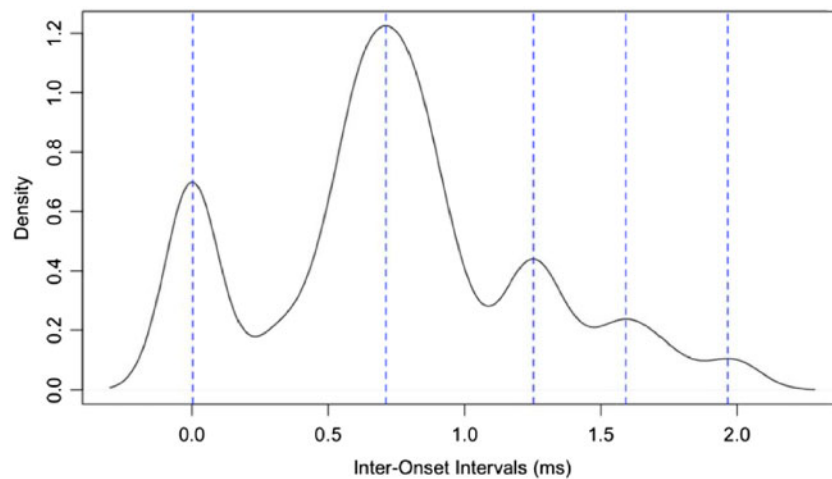
Figure 4. Density plot of the interonset interval distribution for 1 tarsier female duet phrase highlighting local maxima or rhythm bands. For this particular tarsier female duet phrase, there were 5 peaks, which correspond to 5 "rhythm bands."

relative to females in number of rhythm bands. To test for correlation between number of male and female rhythm bands, we used a Pearson's correlation test.

Cosinging, rate of note repetition and temporal precision of tarsier duets

We calculated the duration of cosinging, following Gamba et al. (2016), as the percentage of the total phrase duration that males and females were calling at the same time using Raven Pro selection tables by summing the duration during which males and females overlapped during each phrase, and calculating the percentage of the total phrase duration that males and females were calling together. To calculate rate of note repetition, we divided each tarsier phrase into 3-s bins (median call duration was ~9 s) and calculated the number of male and female notes in each 3-s bin. We used a Pearson's correlation test to investigate the correlation between male and female rate of note repetition. To calculate temporal precision of males relative to females (and females relative to males), we used selection tables created in Raven Pro, and calculated the duration from the start of the individual's note to the beginning of their partner's next note (Figure 3). To determine if female timing was predictive of male timing, and vice versa, we calculated a Granger's causality test, which is a statistical test developed to determine if 1 time series could be used to forecast another time series (Granger 1969; Brandt et al. 2008). We calculated the Granger's causality test in the R package "lmtest" (Zeileis and Hothorn 2002) for each phrase based on the interonset interval of male notes to female notes, and vice versa.

Data access and R code

All data and R code needed to recreate analyses are provided as online Supplementary material. In addition, sound files used for analysis will be made available from the corresponding author on reasonable request.

Ethical note

The research presented here adhered to all local and international laws. Institutional approval was provided by Cornell University (IACUC 2017-0098). Approval was also granted by the Indonesian

Ministry of Research, Technology and Higher Education (Permit number: 2881/FRP/E5/Dit.KI/VII/2018).

Results

Classification of individuals and pairs

We report the results of the analysis of 92 female tarsier calls and 77 male tarsier calls from 15 different pairs. Our ability to classify females based on features estimated from the spectrogram using SVM and leave-one-out cross-validation was substantially better than our ability to classify males, with a classification accuracy of 80% for females and 64% for males. Using SVM recursive feature elimination, we found that spectral features ranked highest in terms of classifying females and males (see Table 2 for feature summary and ranking along with mean, range, and standard deviation [SD] of features for females and males). Our ability to classify pairs was 70%, based on the MFCCs estimated for each phrase.

Male and female rhythm

We found that there was substantial interindividual variation in the number of rhythm bands for female phrases but not male phrases. The results of a 1-way analysis of variance were statistically significant for the number of bands among individual females ($P < 0.05$) but not for males ($P = 0.33$; Figure 5A). We found that males had fewer rhythm bands (mean = 3.2 rhythm bands, range = 2–5) than females (mean = 4.4 rhythm bands, range = 2–7; $P < 0.05$), indicating that male note output was more consistent than females, but there was a strong correlation between the number of rhythm bands a male and female exhibited in a particular phrase (R = 0.36, $P < 0.05$; Figure 5B).

Tarsier cosinging, temporal precision, and rate of note repetition

We found that tarsiers differed substantially in the amount of cosinging exhibited during duetting, with pairs overlapping anywhere from 6% to 40% of the total duration of the duet, and some pairs consistently overlapped more than others ($P < 0.05$; Figure 6). We found that males were precise in their note timing relative to females, with a median interonset interval across males of 0.14 s. There were consistent differences between individual males in their

Table 2. Description of features used for classification, along with mean and SD for females and males

Feature[a]	Female mean and range	Female SD	Male mean and range	Male SD
Note 6 maximum frequency (kHz)	9.3 (6.2–12.8)	1.8	11.8 (7.6–13.5)	1.2
Note 5 maximum frequency (kHz)	9.6 (6.2–13.4)	1.8	11.8 (7.9–13.7)	1.1
Note 4 maximum frequency (kHz)	10.2 (6.9–13.8)	1.6	11.8 (7.9–13.7)	1.2
Note 5 minimum frequency (kHz)	6.8 (5.6–8.5)	0.7	6.8 (5.1–9.8)	0.8
Note 6 minimum frequency (kHz)	6.6 (5.3–8.9)	0.6	6.7 (5.1–9.1)	0.7
Note 3 maximum frequency (kHz)	10.8 (8.2–14.2)	1.4	11.9 (8.7–13.8)	1.0
Note 2 maximum frequency (kHz)	11.5 (9.3–14.6)	1.5	11.9 (7.4–13.8)	1.2
Note 1 maximum frequency (kHz)	12.0 (9.3–15.2)	1.5	11.8 (8.7–13.5)	1.2
Number of notes in a phrase	14.78 (6–41)	6.72	18.36 (6–48)	9.48
Note 4 minimum frequency (kHz)	7.0 (5.9–8.6)	0.5	6.7 (5.0–10.4)	0.8
Note 3 minimum frequency (kHz)	7.1 (6.4–8.8)	0.6	6927 (5.0–11.0)	0.9
Note 2 minimum frequency (kHz)	7.3 (6.4–9.0)	0.7	6.6 (5.2–8.1)	0.6
Note 1 minimum frequency (kHz)	7.5 (5.9–9.8)	0.8	6.5 (4.8–8.1)	0.6
Note 6 duration (s)	0.54 (0.24–0.98)	0.19	0.28 (0.17–0.48)	0.08
Note 5 duration (s)	0.55 (0.25–0.95)	0.16	0.26 (0.15–0.40)	0.06
Note 4 duration (s)	0.51 (0.26–0.83)	0.19	0.25 (0.12–0.40)	0.07
Note 3 duration (s)	0.47 (0.30–0.88)	0.11	0.24 (0.14–0.40)	0.06
Note rate (number of notes/duration [s] of phrase)	1.36 (0.74–1.98)	0.28	1.55 (1.02–2.27)	0.25
Note 1 duration (s)	0.5 (0.304–0.744)	0.12	0.25 (0.17–0.38)	0.05
Phrase duration (s)	12.31 (4.02–36.77)	6.84	12.18 (3.09–35.93)	6.81
Note 2 duration (s)	0.46 (0.28–0.71)	0.11	0.25 (0.14–0.40)	0.06

[a]Features are ranked by their importance for classifying individual females.

relative timing ($P < 0.05$; Figure 7), and all but 3 of the males were in the range (<0.2 s) established for classifying duets as precise, at least in songbirds (Dahlin and Benedict 2014). Females were also precise, with a median interonset interval of 0.13 s, and there were consistent differences among individual females in their timing relative to males ($P < 0.05$), with 5 females exhibiting median interonset intervals >0.2 s (Figure 7). Differences in relative note timing were not influenced by male rate of note output, as there was not a significant correlation between male timing and rate of male note output ($P = 0.84$). To determine how well female timing could predict male timing, and vice versa, we applied the Granger's causality test on the timing of male and female notes between phrases, and found that the timing of female calling was useful in predicting male note output 50% of the time whereas male calling was predictive of female note output only 10% of the time ($n = 77$ phrases, $P < 0.05$), which shows that males adjust their timing relative to their partner more often than females do.

Discussion

Tarsier female duet phrases encode more information about caller identity than male duet phrases, as we were able to correctly classify 80% of female phrases, whereas our classification accuracy for male phrases was only 64%. We found that females had marked variation in the rhythm bands of their duet phrases, with some females showing relatively high variance in the spacing between notes, and others showing relatively low variance, whereas males were more consistent. We found that tarsiers tracked their partner's rhythm in the sense that if the female was more variable, so was the male. All tarsier phrases exhibited some degree of cosinging, with some pairs overlapping substantially. The median interonset intervals for males relative to females was ~0.14 s (and for females relative to males it was 0.13 s), and only 3 out of 14 males (and 5 out of 14 females) had a median interonset interval >0.2 s, indicating that most individuals were precise (Dahlin and Benedict 2014) in their note
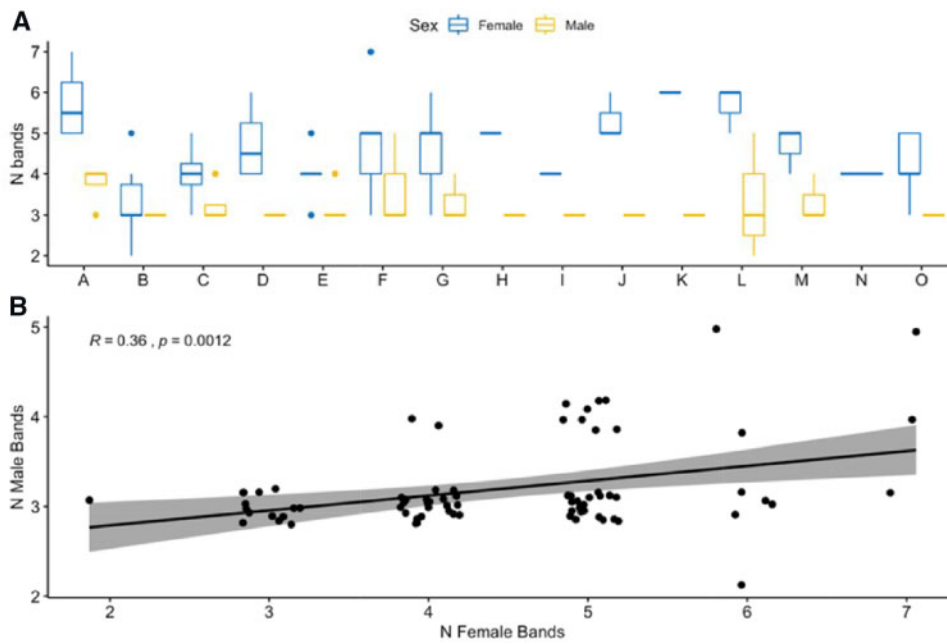
Figure 5. Rhythm bands of tarsier male and female phrases. **(A)** Boxplot of the number of bands averaged per individual male and female for each of the 15 tarsier pairs examined. There were substantial differences between males and females in the number of rhythm bands present, with females having a higher number of rhythm bands (and therefore more variable rhythm than males). In addition, females exhibited substantial interindividual differences in the number of rhythm bands. The tarsier pair "N" did not have a male vocalizing, so only the female is represented in the figure. **(B)** There was a positive correlation between the number of rhythm bands in a female tarsier phrase and the number of bands in the corresponding male tarsier phrase. Points are jittered for better visualization on the plot.
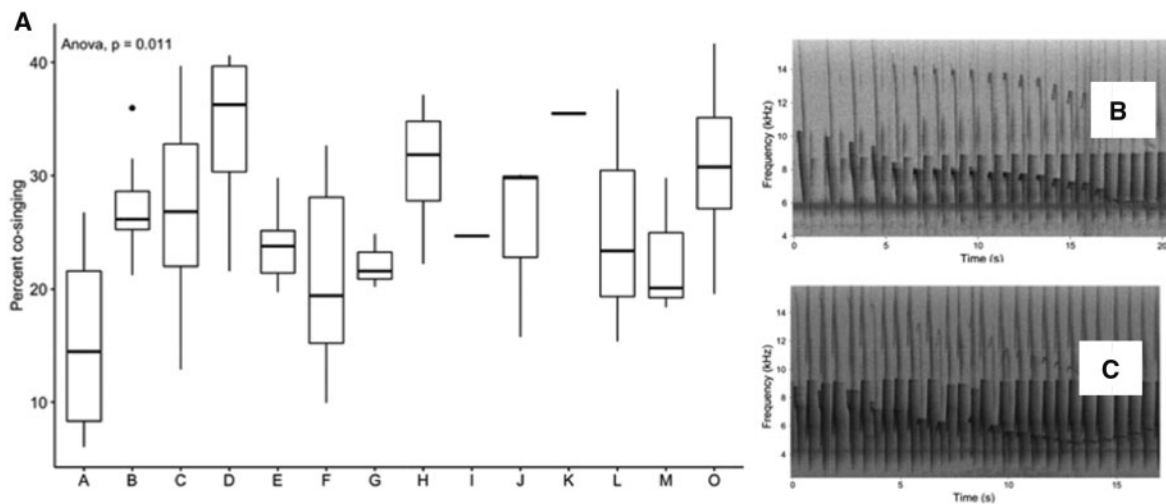

Figure 6. Tarsier pairs differ substantially in their amount of cosinging. **(A)** Tarsiers exhibit variation in the duration of cosinging (the percentage of the total phrase duration that males and females were calling at the same time), and some pairs overlap much more than others, but in every phrase there is at least some overlap between males and females. **(B)** A representative spectrogram of a tarsier phrase that exhibits a relatively low amount of cosinging from pair "A." **(C)** A representative spectrogram of a tarsier phrase that exhibits a relatively high amount of cosinging from pair "D."

output relative to their partner. We also found a correlation between male and female note rate, providing evidence that tarsiers modify the timing of their note output relative to their partner.

Potential for re-recording the same individuals

Although we aimed to reduce the possibility of categorizing 2 separate tarsier groups as the same group, the fact that many of our recordings were taken using autonomous recording units in the absence of a human observer means that there is the potential for misclassification of pairs. We aimed to reduce this by only using high-quality recordings in our analysis, in a sense using signal-to-noise ratio as a proxy for distance. We assumed that there was only one pair duetting close enough to our recorders to give a high signal-to-noise ratio, given the territorial nature of tarsiers and the fact that we never saw or heard tarsier duets emitted from 2 separate groups in close proximity, although it can happen occasionally (MacKinnon and MacKinnon 1980). Recording 2 separate pairs and classifying them as a single pair would have influenced our results
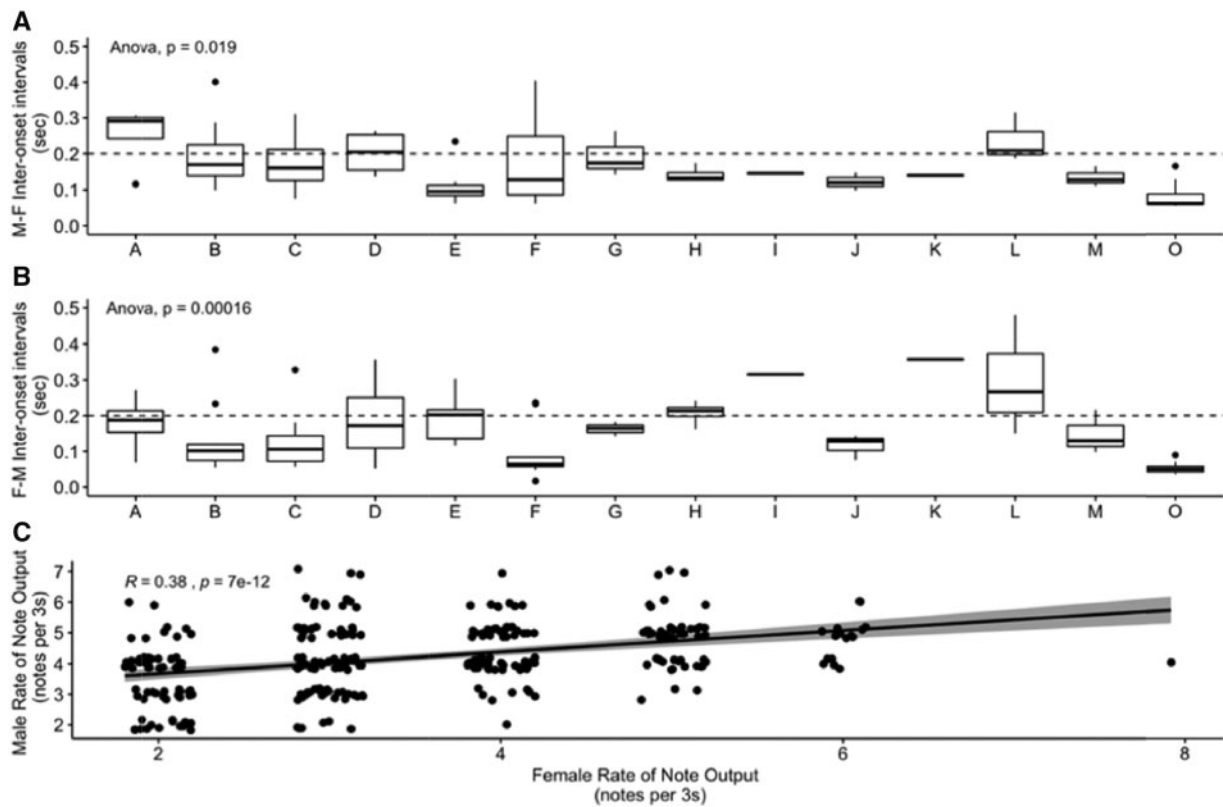
Figure 7. (A) Boxplot of interonset intervals for males relative to females averaged by tarsier pair; (B) boxplot of interonset intervals for females relative to males averaged by tarsier pair; and (C) the correlation of rate of note output (number of notes per 3 s) between males and females within a particular phrase. The dotted horizontal line indicates the 0.2 s mark, which is the proposed cutoff for temporal precision in duetting partners (Dahlin and Benedict 2014). Males and females can both be considered "precise" in their output relative to their partner. In addition, there was a correlation between the rate of note repetition between males and females, indicating that as one partner increases their note output, so does the other. Points are jittered on the plot for better visualization.

in the following ways. First, it would have led to a decrease in our classification accuracy. As our classification accuracy for females is already relatively high (80%), it seems that it is unlikely we included multiple individuals that were incorrectly classified as a single individual. Second, classifying 2 groups as one would have resulted in higher variance for our measures of rhythm, interonset intervals and cosinging duration. As our measures for these variables were relatively consistent within groups (as indicated by low variance), it seems that it was unlikely that many pairs were misclassified.

Individual classification

Spectral tarsier duets, like gibbon duets, exhibit sex-specificity in duet contributions. If the duet function was mainly for intrapair communication, then sex-specific contributions may not be necessary as the receivers (or partners) would be able to distinguish their mates call without the need for sex-specific contributions (Dahlin and Benedict 2014). Alternatively, sex-specific parts of the duets may have evolved initially for extra-pair signaling, but may also provide information to the pair-mate regarding motivation or physical condition (Seyfarth and Cheney 2003). We found that the female duet phrase was more individually distinct than the male phrase. This is consistent with findings of Bornean gibbon duets, where the female contribution was distinct between individuals (Clink et al. 2017), but the male contribution was less distinct (Lau et al. 2018). Given the similarities in sex-specificity, and degree of encoding of individual identity into respective male and female duet contributions,

between gibbon and tarsier duets, it is possible that the duets of these distantly related primates serve the same function(s).

Differences in individual identity encoding of duet contributions may be related to different functions of the male and female contribution. For example, in chimpanzees pant hoots are more stereotyped and individually distinct than pant grunts, and this may be related to different selection pressures on different call types (Mitani et al. 1996). There is evidence that primates respond differently to vocalizations from different individuals based on playback experiments, indicating that primates do pick up on acoustic cues related to differences between individuals (Cheney and Seyfarth 1980, 1999; Rendall 2003). The combination of sex-specificity and individual distinctiveness of tarsier duets are likely conveying information to conspecifics about both members of the pair, but it is unclear which acoustic cues tarsiers pick up on, and if they alter their behavior in response to variation in acoustic cues.

Pair-level signatures

Feature extraction from the spectrogram is a commonly used means of data reduction for analysis of primate acoustic signals (Mitani et al. 1996; Terleph et al. 2016; Clink et al. 2017), and is useful for hypothesis testing as it allows researchers to measure specific features of acoustic signals such as note frequency and duration. But, feature extraction from the spectrogram is subjective in the sense that researchers must make choices about which features to estimate, and this will be influenced by the particular research question. For example, Nietsch (1999) was interested in geographic

variation in Sulawesi tarsier duets, and extracted 6 features from spectrograms of tarsier duet phrases, focusing on global measures such as the maximum frequency of the highest note in a duet. In this study, we estimated 21 features from each duet phrase, focusing on values from individual notes, as we were interested in distinguishing among tarsier individuals from a single site.

In an attempt to reduce the subjectivity inherent with classification using features extracted from the spectrogram, and also to test the value of a pair-level metric of acoustic similarity, we calculated MFCCs for each tarsier duet phrase. As MFCCs are not calculated based on the linear Hertz scale normally used to describe sound, but are based on the "mel" scale, differences in MFCCs are difficult to interpret in a biologically meaningful way (Mielke and Zuberbühler 2013). But, MFCCs have been shown to reliably reflect not only identity of the caller (Clemins et al. 2005; Clink et al. 2018a) or pair (this study), but also the age, social status, and motivation state or context of the call (Deshmukh et al. 2012; Fedurek et al. 2016). Given the complexity of most animal acoustic signals, it seems unlikely that the primate auditory systems responsible for processing vocalizations focus on a single acoustic parameter (Fedurek et al. 2016), but rather a suite of reduced features analogous to MFCCs. Calculating MFCCs employs several principles with known counterparts in human auditory processing—frequency decomposition, mel-filtering of the frequency axis, and compression of amplitudes (Holmberg et al. 2006)—which means that variation in MFCCs may be more relevant to listening tarsier conspecifics than variation in features estimated from the spectrogram.

Rhythm in tarsier duets

To date, most studies of variation in rhythm have focused on the developmental trajectory of rhythm in birdsong (Saar and Mitra 2008; Sasahara et al. 2015). Here, we investigated interindividual variation in rhythm in spectral tarsier duet phrases, and found that male tarsiers were more consistent in the rhythm output of their notes relative to females, which is not particularly surprising based on visual inspection of spectrograms of the duet phrase. In indris, there is sexual dimorphism in interonset intervals, with males exhibiting longer interonset intervals than females (Gamba et al. 2016), which is opposite to the pattern seen in tarsier phrases. We also found that females differed substantially from other individuals of the same sex in terms of the number of rhythm bands, but males did not, which means that male spectral tarsiers in our population are relatively consistent in the rhythm of their duet contributions.

It is possible that variation in the number of rhythm bands seen across female individuals is related to age of calling individual (*sensu* Sasahara et al. 2015) or that temporal variation is related more to social status (dominant or subordinate) and individual identity, as is the case with indris (Gamba et al. 2016). It is also possible that interindividual variation in spectral tarsier rhythm provides cues to caller identify. For example, in wild chimpanzees (*Pan troglodytes*), there are consistent interindividual differences in buttress drumming, which may provide acoustic cues for listeners to identify drumming males (Arcadi et al. 1998). There is evidence that nonhuman primates can detect rhythmic differences, as shown in studies of captive cotton-top tamarins (*Saguinus oedipus*) that were able to discriminate between different languages only if they were characterized into different rhythmic classes (e.g., Polish and Japanese) but not if they were in the same rhythmic class (e.g., English and Dutch; Tincoff et al. 2005). Therefore, in addition to variation in temporal and spectral features of the notes of tarsier duet phrases, variation in rhythm may also provide acoustic cues to caller identity.

Temporal precision

In songbirds, duets are considered to be temporally precise when there is a response latency of ≤ 0.2 s (Dahlin and Benedict 2014). We found that tarsier individuals are precise in their timing relative to their partner. Our analyses support the observation by Nietsch (1999) that spectral tarsiers adjust their duet contributions through simultaneous acceleration or deceleration. The ability for ongoing temporal adjustment of note output exists in other tarsier species, as Togian tarsier duets exhibit a complete avoidance of temporal overlap of notes (Nietsch 1999). It is possible that variation in temporal precision, or coordination between pairs, is related to the amount of time that a pair has been together, as animals that have been paired longer may have a higher degree of temporal coordination (Hall and Magrath 2007). It is also possible that temporal coordination is an honest signal of the coalition quality (Hall and Magrath 2007). Alternatively, variation in temporal precision could simply be the result of age-related changes in tarsier duet contributions, and that partners' note output correlates because individual singing rates change with age in general. In male banded wrens (*Thryophilus pleurostictus*), age-related changes were found in the consistency of trill notes (de Kort et al. 2009) as well as rate of trill note output (Vehrencamp et al. 2013), so it is possible that changes such as these occur in tarsiers. Further lines of research investigating temporal coordination among tarsier pairs with known age and pair-bond length, along with playback experiments with more and less temporally precise duets will be informative.

Cosinging

Tarsier pairs exhibited a variable degree of overlap, ranging from 10% to 40% of the total phrase duration. At the start of the duet phrase, male and female tarsiers tend to emit notes antiphonally, but as the female notes increase in duration and decrease in bandwidth, the male notes tend to overlap the females. Duets across the Primate Order exhibit a variable degree of overlap, with titi monkey duets overlapping the majority of the time (Adret et al. 2018), whereas in many species in the genus *Hylobates* males rarely overlap with females when the female is emitting a great call (Clink et al. 2018b; Terleph et al. 2018). In the family Hylobatidae, species in which duet contributions are sexually dimorphic tend to overlap less, whereas species that are not sexually dimorphic tend to overlap more (Deputte 1982). And indris—which are not sexually dimorphic—also overlap substantially in their chorusing (Gamba et al. 2016). A meta-analysis on duetting birds did not find a relationship between duet function—either intrapair or extrapair communication and degree of overlap or temporal precision of notes (Dahlin and Benedict 2014). Within various tarsier species across Sulawesi, there is substantial variation in the amount of overlap between cosingers (Nietsch 1999). Given high levels of variation in the amount of cosinging across Sulawesi tarsiers, it seems unlikely that this variation reflects differences in the function of the duets. It is possible that variation in cosinging in tarsiers is the result of long-term reproductive isolation of tarsier populations (Driller et al. 2015) which lead to nonadaptive changes in duet structure over time.

Evidence for vocal flexibility in spectral tarsiers

Tarsier duets exhibit clear structural and temporal regularities, but little is known about how tarsiers (and nonhuman primates in general) coordinate singing (Terleph et al. 2018). We investigated the potential for vocal flexibility in tarsier duet phrases in 3 ways.

First, we found that the number of male and female rhythm bands in a particular phrase is correlated. Second, we showed that individual tarsiers were precise in the timing of their notes relative to their partner. And third, we showed that there was a correlation between the rate of female note output and the rate of male note output. The presence of antiphonal duets in a species is not sufficient to determine that animals have the ability for flexible turn-taking, and it must be shown that animals can flexibly modify their output in response to their partners (Terleph et al. 2018). Although vocalizations in nonhuman primates are generally thought to be developmentally or genetically fixed (Geissmann 1984), we show that spectral tarsiers have the ability to simultaneously modify the rate of note output of duet phrases—and in a sense the ability to track their duetting partner—which suggests that tarsier duets exhibit a degree of behavioral plasticity. Future research into the structure of "abnormal" tarsier duet phrases wherein one or the other partner abandons their respective part (*sensu* Terleph et al. 2018) will be informative. Our results are consistent with findings on chimpanzees (Fedurek et al. 2013), gibbons (Terleph et al. 2018), marmosets (Chow et al. 2015), and indris (Gamba et al. 2016) which show that nonhuman primates have the capacity to modify their vocal output relative to their partner. This flexibility in vocal interactions is a universal in human language (Stivers et al. 2009; Levinson 2016), and our results add support to the idea that this precursor to human language evolved long before the appearance of modern humans.

Supplementary Material

Supplementary material can be found at https://academic.oup.com/cz.

References

Adret P, Dingess K, Caselli C, Vermeer J, Martínez J et al., 2018. Duetting patterns of titi monkeys (Primates, Pitheciidae: Callicebinae) and relationships with phylogeny. *Animals* **8**:178.

Aide TM, Corrada-Bravo C, Campos-Cerqueira M, Milan C, Vega G et al., 2013. Real-time bioacoustics monitoring and automated species identification. *PeerJ* **1**:e103.

Arcadi AC, Robert D, Boesch C, 1998. Buttress drumming by wild chimpanzees: temporal patterning, phrase integration into loud calls, and preliminary evidence for individual distinctiveness. *Primates* **39**:505–518.

Bailey WJ, 2003. Insect duets: underlying mechanisms and their evolution. *Physiol Entomol* **28**:157–174.

Beigi H, 2011. *Fundamentals of Speaker Recognition*. New York: Springer Science & Business Media.

Bouchet H, Blois-Heulin C, Pellier A-S, Zuberbühler K, Lemasson A, 2012. Acoustic variability and individual distinctiveness in the vocal repertoire of red-capped mangabeys *Cercocebus torquatus*. *J Comp Psychol* **126**:45.

Brandt PT, Colaresi M, Freeman JR, 2008. The dynamics of reciprocity, accountability, and credibility. *J Conflict Resolut* **52**:343–374.

Burton JA, Nietsch A, 2010. Geographical variation in duet songs of Sulawesi tarsiers: evidence for new cryptic species in south and southeast Sulawesi. *Int J Primatol* **31**:1123–1146.

Charif RA, Waack AM, Strickman LM, 2010. *Raven Pro 1.4 user's manual*. Ithaca, NY: Cornell Lab of Ornithology.

Cheney DL, Seyfarth RM, 1980. Vocal recognition in free-ranging vervet monkeys. *Anim Behav* **28**:362–367.

Cheney DL, Seyfarth RM, 1999. Recognition of other individuals' social relationships by female baboons. *Anim Behav* **58**:67–75.

Chou CH, Liu PH, Cai B, 2008. On the studies of syllable segmentation and improving MFCCs for automatic birdsong recognition. Proceedings of the 3rd IEEE Asia-Pacific Services Computing Conference. Yilan, Taiwan: APSCC; 2008, 745–750.

Chow CP, Mitchell JF, Miller CT, 2015. Vocal turn-taking in a non-human primate is learned during ontogeny. *Proc R Soc B Biol Sci* **282**: 20150069–20150069.

Clemins PJ, Johnson MT, Leong KM, Savage A, 2005. Automatic classification and speaker identification of African elephant *Loxodonta africana* vocalizations. *J Acoust Soc Am* **117**:956–963.

Clink DJ, Bernard H, Crofoot MC, Marshall AJ, 2017. Investigating individual vocal signatures and small-scale patterns of geographic variation in female Bornean gibbon *Hylobates muelleri* great calls. *Int J Primatol* **38**: 656–671.

Clink DJ, Crofoot MC, Marshall AJ, 2018a. Application of a semi-automated vocal fingerprinting approach to monitor Bornean gibbon females in an experimentally fragmented landscape in Sabah, Malaysia. *Bioacoustics* **28**: 1–17.

Clink DJ, Grote MN, Crofoot MC, Marshall AJ, 2018b. Understanding sources of variance and correlation among features of Bornean gibbon *Hylobates muelleri* female calls. *J Acoust Soc Am* **144**:698–708.

Colby J, 2011. *(multiple) Support Vector Machine Recursive Feature Elimination - mSVM-rfe* [Cited 2019 March]. Available from: http://github.com/johncolby/SVM–RFE

Cortes C, Vapnik V, 1995. Support-vector networks. *Mach Learn* **20**: 273–297.

Cowlishaw G, 1992. Song function in gibbons. *Behaviour* **121**:131–153.

Dahake PP, Shaw K, Malathi P, 2016. Speaker dependent speech emotion recognition using MFCC and Support Vector Machine. In: *2016 International Conference on Automatic Control and Dynamic Optimization Techniques (ICACDOT), Pune, India*. IEEE. 1080–1084.

Dahlin CR, Benedict L, 2014. Angry birds need not apply: a perspective on the flexible form and multifunctionality of avian vocal duets. *Ethology* **120**: 1–10.

Deputte BL, 1982. Duetting in male and female songs of the white-cheeked gibbon *Hylobates concolor leucogenys*. In: Snowdon ST, Brown CH, Petersen MR, editors, *Primate Communication*. Cambridge: Cambridge University Press, 67–93.

Deshmukh O, Rajput N, Singh Y, Lathwal S, 2012. Vocalization patterns of dairy animals to detect animal state. In: *Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012), Stockholm, Sweden*. IEEE. 254–257.

Driller C, Merker S, Perwitasari-Farajallah D, Sinaga W, Anggraeni N et al., 2015. Stop and go: waves of tarsier dispersal mirror the genesis of Sulawesi Island. *PLoS ONE* **10**:e0141212.

Fedurek P, Machanda ZP, Schel AM, Slocombe KE, 2013. Pant hoot chorusing and social bonds in male chimpanzees. *Anim Behav* **86**:189–196.

Fedurek P, Zuberbühler K, Dahl CD, 2016. Sequential information in a great ape utterance. *Sci Rep* **6**:38226.

Gamba M, Favaro L, Araldi A, Matteucci V, Giacoma C et al., 2017. Modeling individual vocal differences in group-living lemurs using vocal tract morphology. *Curr Zool* **63**:467–475.

Gamba M, Torti V, Estienne V, Randrianarison RM, Valente D et al., 2016. The indris have got rhythm! Timing and Pitch variation of a primate song examined between sexes and age classes. *Front Neurosci* **10**:249.

Geissmann T, 1984. Inheritance of song parameters in the gibbon song, analysed in 2 hybrid gibbons (Hylobates pileatus × H. lar). Folia Primatol 42: 216–235.

Geissmann T, 1999. Duet songs of the siamang Hylobates syndactylus: iI. Testing the pair-bonding hypothesis during a partner exchange. Behaviour 136:1005–1039.

Geissmann T, 2002. Duet-splitting and the evolution of gibbon songs. Biol Rev 77:57–76.

Geissmann T, Orgeldinger M, 2000. The relationship between duet songs and pair bonds in siamangs Hylobates syndactylus. Anim Behav 60: 805–809.

Goodman M, Porter CA, Czelusniak J, Page SL, Schneider H et al., 1998. Toward a philogenetic classification of based on DNA evidence complemented by fossil evidence. Mol Phylogenet Evol 1:1–14.

Granger CWJ, 1969. Investigating causal relations by econometric models and cross-spectral methods. Econometrica 37:424.

Groves C, Shekelle M, 2010. The genera and species of Tarsiidae. Int J Primatol 31:1071–1082.

Gursky S, 1998. Conservation status of the spectral tarsier Tarsier spectrum: population density and home range size. Folia Primatol 69:191–203.

Gursky S, 2003. The behavioral ecology of the spectral tarsier Tarsius spectrum. Evol Anthropol Issues News Rev 11:226–234.

Gursky S, 2015. Ultrasonic vocalizations by the spectral tarsier Tarsius spectrum. Folia Primatol 86:153–163.

Haimoff EH, 1986. Convergence in the duetting of monogamous Old World primates. J Hum Evol 15:51–59.

Hall ML, 2004. A review of hypotheses for the functions of avian duetting. Behav Ecol Sociobiol 55:415–430.

Hall ML, Magrath RD, 2007. Temporal coordination signals coalition quality. Curr Biol 17:R406–R407.

Han W, Chan C-F, Choy C-S, Pun K-P, 2006. An efficient MFCC extraction method in speech recognition. 2006 IEEE International Symposium on Circuits and Systems. Kos, Greece: IEEE. 4 p.

Ham S, Hedwig D, Lappan S, Choe JC, 2016. Song functions in nonduetting gibbons: Evidence from playback experiments on Javan gibbons (Hylobates moloch). Int J Primatol 37:225–240.

Holmberg M, Gelbart D, Hemmert W, 2006. Automatic speech recognition with an adaptation model motivated by auditory processing. IEEE Trans Audio Speech Lang Processing 14:43–49.

Hsu C, Lin C, 2002. A comparison of methods for multiclass support vector machines. Neural Networks IEEE Trans 13:415–425.

Kassambara A, 2017. Ggpubr:"ggplot2" Based Publication Ready Plots. R Package version 01 6.

Keith SA, Waller MS, Geissmann T, 2009. Vocal diversity of Kloss's gibbons Hylobates klossii on the Mentawai Islands, Indonesia. In: Whittaker D, Lappan S, editors. The Gibbons- Developments in Primatology: Progress and Prospects Vol. 89. 51–71.

Koch R, Raymond M, Wrege P, Klinck H, 2016. SWIFT: a small, low-cost acoustic recorder for terrestrial wildlife monitoring applications. In: North American Ornithological Conference. Washington, DC. 619 p.

de Kort SR, Eldermire ERB, Valderrama S, Botero CA, Vehrencamp SL, 2009. Trill consistency is an age-related assessment signal in banded wrens. Proc R Soc B Biol Sci 276:2315–2321.

Kuhn M, 2008. Caret package. J Stat Softw 28:1–26.

Kumar K, Kim C, Stern RM, 2011. Delta-spectral cepstral coefficients for robust speech recognition. ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings. Prague, Czech Republic: IEEE. 4784–4787.

Langmore NE, 2002. Vocal duetting: definitions, discoveries and directions. Trends Ecol Evol 17:451–452.

Lau A, Clink DJ, Crofoot MC, Marshall AJ, 2018. Evidence for reduced individuality in male gibbon codas. Int J Primatol 39:670–684.

Lee C-H, Chou C-H, Han C-C, Huang R-Z, 2006. Automatic recognition of animal vocalizations using averaged MFCC and linear discriminant analysis. Pattern Recognit Lett 27:93–101.

Levinson SC, 2016. Turn-taking in human communication: origins and implications for language processing. Trends Cogn Sci 20:6–14.

Ligges U, Krey S, Mersmann O, Schnackenberg S, 2018. tuneR: Analysis of Music and Speech. Available from: https://CRAN.R-project.org/package=tuneR

MacKinnon J, MacKinnon K, 1980. The behavior of wild spectral tarsiers. Int J Primatol 1:361–379.

McAuley JD, 2010. Tempo and rhythm. In: Music Perception. New York, NY: Springer, 165–199.

McGregor PK, 1993. Signalling in territorial systems: a context for individual identification, ranging and eavesdropping. Philos Trans R Soc London Ser B Biol Sci 340:237–244.

Méndez-Cárdenas MG, Zimmermann E, 2009. Duetting: a mechanism to strengthen pair bonds in a dispersed pair-living primate Lepilemur edwardsi? Am J Phys Anthropol 139:523–532.

Merker S, Driller C, Perwitasari-Farajallah D, Pamungkas J, Zischler H, 2009. Elucidating geological and biological processes underlying the diversification of Sulawesi tarsiers. Proc Natl Acad Sci USA 106:8459–8464.

Mielke A, Zuberbühler K, 2013. A method for automated individual, species and call type recognition in free-ranging animals. Anim Behav 86:475–482.

Mitani JC, 1985a. Gibbon song duets and intergroup spacing. Behaviour 92: 59–96.

Mitani JC, 1985b. Responses of gibbons Hylobates muelleri to self, neighbor, and stranger song duets. Int J Primatol 6:193–200.

Mitani JC, Gros-Louis J, Macedonia JM, 1996. Selection for acoustic individuality within the vocal repertoire of wild chimpanzees. Int J Primatol 17: 569–583.

Muda L, Begam M, Elamvazuthi I, 2010. Voice recognition algorithms using Mel frequency cepstral coefficient (MFCC) and dynamic time warping (DTW) techniques. J Comput 2:2151–9617.

Müller AE, Anzenberger G, 2002. Duetting in the titi monkey Callicebus cupreus structure, pair specificity and development of duets. Folia Primatol 73:104–115.

Mundry R, Sommer C, 2007. Discriminant function analysis with nonindependent data: consequences and an alternative. Anim Behav 74:965–976.

Nietsch A, 1999. Duet vocalizations among different populations of sulawesi tarsiers. Int J Primatol 20:567–583.

Oyakawa C, Koda H, Sugiura H, 2007. Acoustic features contributing to the individuality of wild agile gibbon Hylobates agilis agilis songs. Am J Primatol 69:777–790.

Rendall D, 2003. Acoustic correlates of caller identity and affect intensity in the vowel-like grunt vocalizations of baboons. J Acoust Soc Am 113:3390.

Robinson JG, 1979. An analysis of the organization of vocal communication in the titi monkey Callicebus moloch. Z Tierpsychol 49:381–405.

Robinson JG, 1981. Vocal regulation of inter- and intragroup spacing during boundary encounters in the titi monkey Callicebus moloch. Primates 22: 161–172.

Saar S, Mitra PP, 2008. A technique for characterizing the development of rhythms in bird song. PLoS ONE 3:e1461.

Sasahara K, Tchernichovski O, Takahasi M, Suzuki K, Okanoya K, 2015. A rhythm landscape approach to the developmental dynamics of birdsong. J R Soc Interface 12:20150802.

Seyfarth RM, Cheney DL, 2003. Signalers and receivers in animal communication. Annu Rev Psychol 54:145–173.

Shekelle M, 2008. Distribution of tarsier acoustic forms, North and Central Sulawesi: with notes on the primary taxonomy of Sulawesi's tarsiers. Primates Orient Night 1:35–50.

Shekelle M, Groves C, Merker S, Supriatna J, 2008. Tarsius tumpara: a new tarsier species from Siau Island, North Sulawesi. Primate Conserv 23: 55–64.

Shekelle M, Groves CP, Maryanto I, Mittermeier RA, 2017. Two new tarsier species (Tarsiidae, Primates) and the biogeography of Sulawesi, Indonesia. Primate Conserv 31:61–69.

Stivers T, Enfield NJ, Brown P, Englert C, Hayashi M et al., 2009. Universals and cultural variation in turn-taking in conversation. Proc Natl Acad Sci USA 106:10587–10592.

Suthers RA, 1994. Variable asymmetry and resonance in the avian vocal tract: a structural basis for individually distinct vocalizations. J Comp Physiol A 175:457–466.

Tecot SR, Singletary B, Eadie E, 2016. Why "monogamy" isn't good enough. *Am J Primatol* **78**:340–354.

Terleph TA, Malaivijitnond S, Reichard UH, 2015. Lar gibbon *Hylobates lar* great call reveals individual caller identity. *Am J Primatol* **821**: 811–821.

Terleph TA, Malaivijitnond S, Reichard UH, 2016. Age related decline in female lar gibbon great call performance suggests that call features correlate with physical condition. *BMC Evol Biol* **16**:4.

Terleph TA, Malaivijitnond S, Reichard UH, 2018. Male white-handed gibbons flexibly time duet contributions. *Behav Ecol Sociobiol* **72**:16.

Thorpe WH, Hall-Craggs J, Hooker B, Hutchison R, 1972. Duetting and antiphonal song in birds: its extent and significance. *Behaviour* **18**: 1–197.

Tincoff R, Hauser M, Tsao F, Spaepen G, Ramus F et al., 2005. The role of speech rhythm in language discrimination: further tests with a non-human primate. *Dev Sci* **8**:26–35.

Tobias ML, Viswanathan SS, Kelley DB, 1998. Rapping, a female receptive call, initiates male-female duets in the South African clawed frog. *Proc Natl Acad Sci USA* **95**:1870–1875.

Torti V, Bonadonna G, De Gregorio C, Valente D, Randrianarison RM et al., 2017. An intra-population analysis of the indris' song dissimilarity in the light of genetic distance. *Sci Rep* **7**:1–16.

Torti V, Valente D, De Gregorio C, Comazzi C, Miaretsoa L et al., 2018. Call and be counted! Can we reliably estimate the number of callers in the indri's (Indri indri) song? *PLoS ONE* **13**:e0201664.

Turesson HK, Ribeiro S, Pereira DR, Papa JP, De Albuquerque VHC, 2016. Machine learning algorithms for automatic classification of marmoset vocalizations. *PLoS ONE* **11**:e0163041.

Vehrencamp SL, Yantachka J, Hall ML, de Kort SR, 2013. Trill performance components vary with age, season, and motivation in the banded wren. *Behav Ecol Sociobiol* **67**:409–419.

Venables WN, Ripley BD, 2002. *Modern Applied Statistics with S. Fourth*. New York: Springer.

Volodin IA, Lapshina EN, Volodina EV, Frey R, Soldatova NV, 2011. Nasal and oral calls in juvenile goitred gazelles *Gazella subgutturosa* and their potential to encode sex and identity. *Ethology* **117**:294–308.

Zeileis A, Hothorn T, 2002. Diagnostic checking in regression relationships. *R News* **2**:7–10. Available from: https://CRAN.R-project.org/doc/Rnews/.