Supplementary Materials

Materials and Methods

Study site

We carried out this study in the Banli area of the Guangxi Chongzuo White-headed Langur National Nature Reserve (N22°24-22°46, E107°22-107°42) in southwestern Guangxi, China. The study area lies in karst monsoon seasonal forest at an elevation of 400–600 m. The topography consists of flatlands, cliffs, and karst hilltops (Lu et al., 2021). Most flatlands have been cultivated by local people for cereal crops, resulting in severe habitat fragmentation (Huang et al., 2002, Huang et al., 2008). Mean annual rainfall is 1 200 mm, with a dry season from October through March and a wet season from April through September. Mean annual temperature is 22.8 ℃, with a minimum monthly mean temperature (in January) of 15.0 ℃ and maximum monthly mean temperature (in July) of 28.7 ℃ (Jin et al., 2019; Yin et al., 2013).

Study group

We studied two well-habituated groups of white-headed langurs from July 2020 to September 2021. During the study period, the Zhiwuyuan group contained 28 individuals, including one adult male, 10 adult females, 12 subadults or juveniles, and five infants. From July 2020 to January 2021, the Dushan group contained 39 individuals, including one adult male, 16 adult females, 18 subadult or juveniles, and four infants. In January 2021, the Dushan resident male was replaced by a new adult male, resulting in the disappearance of four infants and dispersal of 15 subadults or juveniles. From February 2021 to September 2021, the Dushan group consisted of 20 individuals, including one adult male, 16 adult females, and three subadults or juveniles. We recorded vocalizations from the two study groups under favorable weather conditions (i.e., not raining or windy). As individuals could not always be identified, data were collected at the age/sex-class level. The following age/sex classes were distinguished: adult male, adult female, subadult or juvenile, and infant.

Ethics statement

Data collection was performed non-invasively and no animal was injured. All research protocols complied with the laws and regulations of the People's Republic of China and were approved by the Committee of Animal Welfare and Ethics of the Chinese Academy of Sciences, Institute of Zoology, and Guangxi Normal University. All authors declare no conflicts of interest.

Statistical analysis

We used principal component analysis (PCA) to summarize the 25 acoustic variables extracted from the high-quality call samples. We performed varimax rotation to generate linear combinations of variables (principal component, PCs) with a loading factor >0.6, which reduced the effects of related variables.

We used the Kolmogorov-Smirnov test to assess variable normality. We classified and compared call types using forward stepwise linear discriminant function analysis (DFA). The grouping variable was defined as the pre-classified call type based on human ear and spectrogram assessment of each call. As unbalanced data can adversely affect DFA, we weighted the analyses by observed group size. To improve cross-validation power, we used the leave-one-out classification method to assign call types, which uses all cases as a training set except the one assigned. All data analyses were performed in R v4.1.2 (R Development Core Team, 2021).

Results

Classification of call types

Based on 1 884 h of direct observations over 284 days, we recorded 28 697 call samples (mean±SD=101±62, range=41–235), yielding 409 high-quality calls for acoustic parameter extraction. Based on aural perception and spectrogram assessment, we identified 17 call types (Figure 1). We excluded four call types in statistical analysis due to their small sample size. Thus, 12 call types (387 call samples) were statistically analyzed.

The PCA results converted 25 variables into 10 PCs, which explained 93.4% of total variance in the dataset. The 10 PCs were most strongly associated with MaxF_Mean, HNR_Centre, Bandwidth_End, HNR_End, HNR_End, Bandwidth_Start, HNR_Mean, HNR_Start, MinF_End, and duration, respectively (Supplementary Table S4). Based on the 10 PCs, DFA correctly assigned 80.9% of the call samples to the pre-classification catalogue. The correct rate of cross-validation was 71.1%, better than the 8.3% excepted by random chance (binomial test, $P<0.001$). The DFA results generated 10 linear discriminant functions (LD), the first three of which cumulatively explained 93.8% of variance among call types (LD1: 75.7%; LD2:11.8%; LD3: 6.3%). LD1 was mainly associated with PC1, PC2, and PC4, LD2 was primarily associated with PC2 and PC10, and LD3 was primarily associated with PC4 and PC10 (Supplementary Table S5).

The PCA and DFA results supported our observed vocal repertoire classification. Cross-validation accuracy of each call type was as follows: snort 80.0%, roar 88.9%, wahoo 94.5%, purr 50.0%, squeal 50.0%, long squeal 30.8%, two-syllabled chuck 58.9%, three-syllabled chuck 68.8%, modulated tonal scream 100%, squeak 84.6%, compound squeak 100%, and infant scream 92.3%. Long squeals (30.8%) had the lowest validation and were assigned as squeaks because they had similar spectral-based acoustic features. The mean spectral-based parameters of purrs and squeals were similar, accounting for their relatively low validation among call types (Supplementary Table S3).

Description of call types

We identified 17 call types using auditory perception, spectrogram inspection, and statistical analysis of acoustic parameters. The vocal repertoire of white-headed langurs was influenced by both social and environmental factors. Some call types, such as snorts, roars, wahoos, purrs, soft snorts, punks, and squeaks, were emitted in multiple social-ecological contexts. Other call types, such as squeals, modulated tonal screams, grunts, and infant coos, were emitted only in a single context. Some of the call types in the vocal repertoire were age- and sex-specific. Adult males produced seven call types, i.e., snorts, roars, wahoos, purrs, soft snorts, punks, and leap grunts. Adult females produced six call types, i.e., punks, squeals, long squeals, two-syllabled chucks, three-syllabled chucks, and leap grunts; of these, punks and leap grunts were used by both sexes. Five call types were specific to adult males, two to adult females, three to subadults and juveniles, and three to infants (Supplementary Table S2).

Snorts (Figure 1A-a)

Snorts are monosyllabic, explosive, and harsh calls emitted exclusively by adult males. They are of short duration (0.178±0.050 s, mean±SD), with a median peak frequency of 1 193±615 Hz and bandwidth of 4 139±1 340 Hz. They are used during vigilance (91%), in response to natural disturbance (4%) and human disturbance (1%), and in other unknown contexts (4%). In these four contexts, adult males frequently produced snorts at irregular intervals, up to 165 times over 40 min. The caller is usually vigilant, staring intently, and may jump suddenly from branch to branch or rock to rock.

Roars (Figure 1A-b)

Roars are impressive calls with an obvious exhalation and inhalation sound. They are atonal and harsh, with a relatively low mean peak frequency of 1 195±73 Hz. They are used by adult males within their own home range. The call is uttered in a phrase, which generally includes 5–10 roar calls of 2.850±0.547 s duration each. Roars are given in a variety of social-ecological contexts, including vigilance (71%), natural disturbance (2%), human disturbance (1%), and unknown factors (4%). Individuals also display jumping behavior from branch to branch or rock to rock when roaring.

Wahoos (Figure 1A-c)

Wahoo calls are atonal, characterized by median duration (0.394±0.105 s), second lowest mean peak frequency (314±27 Hz), relatively high mean harmonics-to-noise ratio HNR (31.22±8.41), and lowest mean entropy. They are only produced by adult males and are emitted in bouts with regular intervals (range=2–28) of up to 15 s between calls. Most are uttered during leaping displays upon detection of bachelor males (84%) and thus likely serve to warn group members of spatial encroachments by intruders. Sequences consisting of snorts, roars, and wahoos are routinely emitted by adult males. Sequences are uttered at any time of the day, but most frequently in the early morning (07:00–10:00 h). Before calling, males typically monitor their surroundings from a high point on a rock or from a tree on a hilltop. The sequences are very loud and likely function as long-distance signals used for intergroup communication. Adult males living in adjacent one-male-multi-female groups may respond to each other's call sequences in rotation, suggesting that they are notifying one another of territorial boundaries.

Purrs (Figure 1A-d)

Purrs are used by adult males. They are non-tonal and contain several trains of pulsed units of relatively short duration (0.290±0.034 s), low mean peak frequency (1067±456 Hz), and wide mean frequency bandwidth (6 242±1 673 Hz). These vocalizations sound like tongue-clicks to the human ear. These calls occur most frequently during vigilance (51%) but are also produced while moving (25%) and feeding (24%).

Soft snorts (Figure 1A-e)

Soft snorts are emitted by adult males and are comprised of a series of pulsed units of relatively short duration (0.187±0.035 s), very low frequency (93±33 Hz), and narrow frequency bandwidth (1 204±467 Hz). They are heard in two contexts. Most soft snorts (99%) are vocal responses to roars or wahoos of an adjacent adult male. In this context, soft snorts can grade into snorts as the calls of the extra-group male become more intense (louder or more frequent). Soft snorts also accompany male-male agonistic interactions (1%) and are usually uttered by the retreating male.

Punks (Figure 1A-f)

Punk calls exhibit some harmonic structures and are uttered by adult males, adult females, subadults, and juveniles. They are of short duration (0.169±0.027 s) and low frequency (260±254 Hz) and occur often in all age-sex classes, except infants, during group traveling (52%), foraging (32%), resting (15%), and vigilance (1%). As individuals often exchange punks during group movement, foraging, and resting, these calls are likely close-range communication signals that maintain group cohesion. These vocalizations are only moderately loud, which may prevent eavesdropping by extra-group individuals.

Squeals (Figure 1A-g) *and long squeals* (Figure 1A-h)

Squeals and long squeals are two separate call types emitted exclusively by adult females in agonistic within-group interactions. Both calls are harsh and noisy but are distinct. For example, squeals are explosive syllables of moderate duration (0.496±0.281 s) and relatively low mean peak frequency (114±36 Hz), while long squeals have a longer duration (0.512±0.261 s), higher mean peak frequency (989±1 155 Hz), wider bandwidth (7 139±1 515 Hz), higher mean entropy (0.53±0.06), and lower mean HNR (11.80±9.43). Generally, dominant females utter squeals during mild competitive interactions with other females over food or position (100%). If subordinate females avoid conflict, no other behaviors or vocalizations are observed. In contrast, long squeals are emitted by dominant females during physical attacks when conflict escalates. Long squeals are produced during intense female-female conflicts over food or position (90%) and some unidentified contexts (10%) and serve as a threat and dominance signal.

Two-syllabled chucks (Figure 1A-i) *and three-syllabled chucks* (Figure 1A-j)

Two kinds of chucks, i.e., two-syllabled and three-syllabled chucks, were recorded. They are harsh non-tonal calls produced by adult females, subadults, and juveniles, with durations of 0.204±0.064 s and 0.226±0.024 s, respectively. Two-syllabled chucks contain two harsh syllables with a mean peak frequency of 1808±636 Hz. This vocalization is directed towards strangers who are in a stationary state. Three-syllabled chucks have a higher mean peak frequency (1 866±1 080 Hz) and are directed towards unknown moving animals (we could not confirm species due to visual obstruction) (86%) and approaching strangers (14%). These calls usually prompt the rapid movement of group members away from the potential risks and are accompanied by chuck calls.

Modulated tonal screams (Figure 1A-k)*, grunts* (Figure 1A-l)*, and squeaks* (Figure 1A-m)

Modulated tonal screams, grunts, and squeaks are produced by subadults and juveniles. Modulated tonal screams have rich harmonics and frequency modulations and are of long duration (1.575±0.221 s) and high mean peak frequency (6 849±239 Hz). These calls rapidly decrease in frequency, then increase before ending with a slight drop. On occasion, several modulated tonal screams are emitted in succession, followed by grunts of median duration (0.617±0.111 s), low mean peak frequency (970±350 Hz), and high mean HNR (37.06±9.22). These vocalizations appear to be contact calls, as they are emitted when contact with the social group is lost. Generally, upon hearing this call, the adult male (unit leader) responds by stopping his activity, while adult females look in the call direction and respond with punks.

Squeaks are harsh noisy calls of long duration (1.231±0.351) and wide mean frequency bandwidth (10 883±2 863 Hz). These calls are heard predominantly when subadults or juveniles stray from the group (84%), although 16% of contexts were undetermined. When heard, adult females sometimes moved to assist the subadults and juveniles.

Infant coos (Figure 1A-n)*, compound squeaks* (Figure 1A-o)*, and infant screams* (Figure 1A-p)

Infants produce three kinds of call types. Infant coo calls are uttered by infants less than one month of age. They are typically heard when an infant is carried by its mother. This is the first call mastered by this species, which sounds like the vowel “a” to the human ear. Infant coos have a medium duration (0.404±0.043 s), relatively low mean peak frequency (1 790±315 Hz), and high mean HNR (29.91±3.15).

Compound squeaks are explosive calls mixed with some harmonics, frequency modulations, and harsh noisy elements. They are different from squeaks in their relatively short duration (1.003±0.310 s), high mean peak frequency (4 969±1 083 Hz), and wide bandwidth (12 838±1 558 Hz). Compound squeaks occur in the context of infants being held by their mother or other individuals during moving (64%), followed by infants moving independently near their mother (19%), with 17% of contexts undetermined. This call likely indicates fear.

Infant screams have a harmonic-like structure with some noisy elements, characterized by relatively long duration (0.796±0.064 s), high mean peak frequency (4 953±3 004 Hz), and wide frequency bandwidth (8 891±3 072 Hz). Infant screams were recorded from a two-month-old infant, who uttered approximately 300 screams during a 2.5 h period when isolated from the group. Surprisingly, these screams did not elicit a response from the other group members.

Leap grunts

Leap grunts are uttered when adult individuals chase each other with rapid leaping in trees or running on rocks. Unfortunately, only low-quality recordings of these calls were possible due to their low loudness (dB) and excessive background noise from shaking branches. Therefore, no spectrogram parameters are reported here.

Supplementary Table S1 Definition of each acoustic parameter measured from vocalizations of white-headed langurs

Location	Abbreviation	Definition
Entire call	Duration	Time span of the entire call
	PF_Mean	Peak frequency derived from the averaged spectrum of the entire call
	MinF_Mean	Minimum frequency derived from the averaged spectrum of the entire call
	MaxF_Mean	Maximum frequency derived from the averaged spectrum of the entire call
	Bandwidth_Mean	Frequency range calculated by subtracting MinF_Mean from MaxF_Mean
	Entropy_Mean	Spectral flatness derived from the averaged spectrum of the entire call
	HNR_Mean	Ratio of harmonic to nonharmonic energy derived from the averaged spectrum of the entire call
Start	PF_Start	Peak frequency at the beginning of the call
	MinF_Start	Minimum frequency at the beginning of the call
	MaxF_Start	Maximum frequency at the beginning of the call
	Bandwidth_Start	Frequency range calculated by subtracting MinF_Start from MaxF_Start
	Entropy_Start	Spectral flatness at the beginning of the call
	HNR_Start	Ratio of harmonic to nonharmonic energy at the beginning of the call
Centre	PF_Centre	Peak frequency at the middle of the call
	MinF_Centre	Minimum frequency at the middle of the call
	MaxF_Centre	Maximum frequency at the middle of the call
	Bandwidth_Centre	Frequency range calculated by subtracting MinF_Centre from MaxF_Centre
	Entropy_Centre	Spectral flatness at the middle of the call
	HNR_Centre	Ratio of harmonic to nonharmonic energy at the middle of the call
End	PF_End	Peak frequency at the ending of the call
	MinF_End	Minimum frequency at the ending of the call
	MaxF_End	Maximum frequency at the ending of the call
	Bandwidth_End	Frequency range calculated by subtracting MinF_End from MaxF_End
	Entropy_End	Spectral flatness at the ending of the call
	HNR_End	Ratio of harmonic to nonharmonic energy at the ending of the call

Supplementary Table S2 Social-ecological contexts of call types and occurrences (percentage) of all call types in different contexts

Call type	Caller	Context category	Context description	Number (%) of vocalizations observed
Snort	AM	Vigilance	Sitting, monitoring another bi-sexual group, signal bachelor males, or all-male group. Responding to snort, roar or wahoo uttered by resident male in adjacent group	6382 (91%)
		Natural disturbance	Thunder, loud calls from red-bellied tree squirrel (*Callosciurus erythraeus*) or red junglefowl (*Gallus gallus*); boulder rolling downslope.	249 (4%)
		Human disturbance	Agricultural vehicle passing by; fireworks	96 (1%)
		Unknown factor		279 (4%)
Roar	AM	Vigilance	Sitting and monitoring another bi-sexual group, signal bachelor males, or all-male group; Responding to snort, roar or wahoo uttered by resident male or adjacent group	3339 (71%)
		Natural disturbance	Thunder, loud calls from red-bellied tree squirrel (*Callosciurus erythraeus*) or red junglefowl (*Gallus	85 (2%)

			gallus). Large boulder rolling downslope.	
		Human disturbance	Agricultural vehicle passing by; fireworks	35 (1%)
		Unknown factor		1236 (26%)
Wahoo	AM	Vigilance	Sitting and monitoring another bi-sexual group, signal bachelor males, or all-male group; Responding to snort, roar or wahoo uttered by resident male or adjacent group	981 (84%)
		Unknown factor		180 (16%)
Purr	AM	Vigilance	Sitting and monitoring another monkey group	146 (51%)
		Traveling	Moving directionally	71 (25%)
		Foraging	Moving and foraging	68 (24%)
Soft snort	AM	Vigilance	Responding to roar or wahoo uttered by males of adjacent groups	677 (99%)
		Agonistic interactions	Male-male conflict with no physical contact	10 (1%)
Punk	AM, AF	Traveling	Moving directionally	3815 (52%)
		Foraging	Moving and foraging	2326 (32%)
		Resting	Stationary	1080 (15%)
		Vigilance	Sitting and monitoring another bi-sexual group, bachelor, or all-male group	111 (1%)

Squeal	AF	Agonistic interactions	Mild female-female competition for food or position	85 (100%)
Long Squeal	AF	Agonistic interactions	Intense female-female competition for food or position	111 (90%)
		Unknown factors		13 (10%)
Two syllabled chuck	AF, S, J	Human disturbance	Proximity to strangers who are in a stationary state	109 (100%)
Three syllabled chuck	AF, S, J	Natural disturbance	Other species of animals in the vicinity of the langurs	725 (86%)
		Human disturbance	Strangers who are approaching	119 (14%)
Modulated tonal scream	S, J	Traveling	Losing contact during group movement	1745 (100%)
Grunt	S, J	Traveling	Losing contact during group movement	36 (100%)
Squeak	S, J	Traveling	Moving alone on cliffs	612 (70%)
		Natural disturbance	Responding to roar emitted by adult male	118 (14%)
		Unknown factors		139 (16%)
Infant coo	I	Traveling	Being carried by mother	72 (100%)
Compound squeak	I	Traveling	Being carried by mother and other group members	2119 (64%)
		Traveling	Moving independently but in close proximity to mother	620 (19%)
		Unknown factors		546 (17%)
Infant scream	I	Traveling	Losing contact during group	300 (100%)

			movement	
Leap grunt	AM, AF	Playing	Chasing each other in excited state	62 (100%)

AM: adult male, AF: adult female, S: subadult, J: juvenile, I: infant

Supplementary Table S3 Acoustic parameters (mean±SD) of each call type for white-headed langurs

Call type	Serial number	N	Correct rate	Duration	PF_Mean	MinF_Mean	MaxF_Mean	Bandwidth_Mean	Entropy_Mean	HNR_Mean
Snort	1	95	80.0%	0.178±0.050	1193±615	46±21	4185±1336	4139±1340	0.41±0.04	10.67±8.35
Roar	2	90	88.9%	0.386±0.054	1195±73	148 ± 80	2744±371	2596±406	0.31±0.04	24.29±11.00
Wahoo	3	55	94.5%	0.394± 0.105	314±27	88±51	727±323	639±345	0.17±0.02	31.22±8.41
Purr	4	20	50.0%	0.290±0.034	1067±456	40±0	6282±1673	6242±1673	0.45±0.03	15.88±8.07
Soft snort*	-	7	-	0.187±0.035	180±262	40±0	1244±467	1204±467	0.23±0.06	26.84±16.27
Punk*	-	6	-	0.169±0.027	260±254	40±0	2227±1011	2187±1011	0.30±0.04	30.08±10.52
Squeal	5	16	50.0%	0.155±0.038	729±601	40±0	5376±2932	5336±2932	0.43±0.10	17.16±12.44
Long squeal	6	13	30.8%	0.512±0.261	989±1155	40±0	7179±1515	7139±1515	0.53±0.06	11.80±9.43
Two syllabled chuck	7	17	58.9%	0.204±0.064	1808±636	48±31	5512±1664	5465±1669	0.43±0.05	12.01±6.39
Three syllabled chuck	8	16	68.8%	0.226±0.024	1866±1080	71±70	5761±1252	5689±1212	0.44±0.05	14.70±7.41
Modulated tonal scream	9	15	100%	1.575±0.221	6849±239	4305±1991	13665±1722	9360±2995	0.43±0.08	23.27±5.57
Grunt*	-	6	-	0.617±0.111	970±350	40±0	2155±218	2115±218	0.29±0.01	37.06±9.22
Squeak	10	13	84.6%	1.231±0.351	2585±2299	40±0	10923±2863	10883±2863	0.63±0.07	6.43±2.15
Infant coo*	-	3	-	0.404±0.043	1790±315	1047±464	3857±224	2810±297	0.32±0.03	29.91±3.15
Compound squeak	11	24	100%	1.003±0.310	4969±1083	2061±567	14899±1748	12838±1558	0.66±0.05	24.13±3.71
Infant scream	12	13	92.3%	0.796±0.064	4953±3004	1929±653	10820±3016	8891±3072	0.56±0.06	25.38±3.08

Continued Table S3.

Call type	PF_Start	MinF_Start	MaxF_Start	Bandwidth_Start	Entropy_Start	HNR_Start	PF_Centre	MinF_Centre	MaxF_Centre
Snort	1836±1013	246±420	5873±4204	5627±4144	0.39±0.08	25.98±13.63	1004±607	53±28	4279±2111
Roar	1254±183	68±43	3396±824	3329±827	0.33±0.03	26.29±6.61	1186±83	255±160	2685±411
Wahoo	269±66	56±38	1786±837	1730±835	0.23±0.04	38.25±10.33	293±56	86±55	1067±623
Purr	690±1262	229±575	7326±3344	7097±3299	0.37±0.12	35.97±18.72	1141±1375	65±57	6557±3042
Soft snort*	40±0	40±0	567±486	527±486	0.17±0.06	43.70±19.05	377±398	40±0	1403±193
Punk*	132±95	40±0	2550±2714	2510±2714	0.24±0.09	37.68±5.39	523±421	53±19	2745±1390
Squeal	536±608	156±306	5354±2880	5199±2903	0.33±0.06	41.74±18.58	1018±466	61±47	6322±4463
Long squeal	1538±2133	192±340	8671±3983	8479±3827	0.47±0.14	24.82±14.03	2377±1440	112±159	7058±1571
Two syllabled chuck	1156±890	389±573	10099±3486	9710±3705	0.48±0.08	29.11±15.83	1932±820	472±660	6924±2894
Three syllabled chuck	893±808	43±10	7259±1833	7217±1835	0.44±0.05	21.67±7.19	2112±408	906±658	6242±1593
Modulated tonal scream	3348±2618	2114±561	13847±5658	11733±5405	0.54±0.10	38.82±3.82	7272±1596	6453±972	13464±1867
Grunt*	712±453	205±369	2163±891	1958±780	0.25±0.06	47.12±10.23	913±177	90±112	1855±176
Squeak	832±1043	116±264	5716±3693	5600±3782	0.39±0.11	20.66±12.72	3056±2223	434±1341	11295±2979
Infant coo*	2290±816	1517±148	3840±145	2323±69	0.20±0.04	60.04±10.28	1347±476	1143±561	3240±448
Compound squeak	1985±874	1425±461	6535±4821	5110±4812	0.33±0.11	48.13±14.25	5528±1914	2805±558	15070±1857
Infant scream	2308±228	1760±315	19676±5035	17916±5221	0.63±0.07	43.21±5.92	8502±624	5851±2781	11861±3739

Continued Table S3.

Call type	Bandwidth_Centre	Entropy_Centre	HNR_Centre	PF_End	MinF_End	MaxF_End	Bandwidth_End	Entropy_End	HNR_End
Snort	4226±2114	0.36±0.05	18.79±8.45	787±854	347±460	4609±3939	4262±3845	0.33±0.09	33.27±15.28
Roar	2431±488	0.26±0.04	25.47±6.57	1266±532	57±32	3612±1073	3555±1074	0.35±0.04	24.79±6.79
Wahoo	981±641	0.18±0.04	37.87±8.40	288±158	53±37	1994±2591	1941±2596	0.22±0.04	37.90±9.71
Purr	6492±3061	0.41±0.09	22.65±9.97	597±942	264±469	6001±4746	5738±4514	0.29±0.13	51.18±14.03
Soft snort*	1363±193	0.23±0.03	33.22±12.96	107±149	101±135	1834±3323	1733±3349	0.20±0.13	40.67±17.03
Punk*	2692±1384	0.27±0.03	39.82±16.06	297±448	190±335	2458±1492	2268±1407	0.28±0.09	34.73±12.63
Squeal	6261±4484	0.41±0.14	19.62±8.95	118±239	40±0	3233±2957	3193±2957	0.29±0.11	39.44±14.62
Long squeal	6946±1620	0.45±0.06	20.09±8.13	1628±1683	225±455	9383±4785	9158±4532	0.46±0.13	29.72±25.22
Two syllabled chuck	6452±3159	0.42±0.08	16.69±3.41	731±1204	488±977	6539±4969	6051±4280	0.38±0.15	33.74±14.64
Three syllabled chuck	5336±2103	0.35±0.07	19.52±5.89	183±176	40±0	4029±2237	3989±2237	0.37±0.07	23.08±11.73
Modulated tonal scream	7011±2057	0.19±0.08	39.90±3.33	6491±344	3996±1151	15163±5410	11167±5811	0.52±0.10	38.33±6.17
Grunt*	1765±281	0.25±0.02	45.56±11.20	527±529	363±457	6835±7231	6472±6792	0.28±0.09	55.21±13.29
Squeak	10861±3009	0.53±0.08	17.31±4.28	535±941	272±791	6710±4777	6438±4238	0.36±0.08	18.32±6.59
Infant coo*	2097±930	0.21±0.05	44.79±6.85	3123±1366	2520±984	8810±7109	6290±6129	0.27±0.08	55.10±10.66
Compound squeak	12265±1795	0.51±0.10	36.14±4.52	5064±1039	3923±1176	7922±2882	3999±3055	0.35±0.06	38.10±8.56
Infant scream	6010±5548	0.36±0.08	42.22±8.60	6977±245	4196±1872	20952±2152	16756±3225	0.63 ± 0.09	42.28 ± 13.40

*We excluded these call types for statistics analysis due to small sample size.

Supplementary Table S4 Rotated factor loadings of measured acoustic variables on 10 PCs with eigenvalues >0.6

Acoustic variables	Components									
	1	2	3	4	5	6	7	8	9	10
Duration	0.201	0.164	0.162	0.227	0.127	0.158			0.222	0.609
PF_Mean	0.244	0.108		0.183		0.106			0.265	-0.164
MinF_Mean	0.219	0.215		0.19	-0.171				-0.312	0.234
MaxF_Mean	0.259		0.209						0.224	
Bandwidth_Mean	0.237	-0.147	0.238						0.352	-0.103
Entropy_Mean	0.212	-0.266	0.2						0.245	-0.239
HNR_Mean		0.379		-0.107	0.282		-0.546	-0.477		-0.209
PF_Start	0.144	-0.132	0.149	-0.564	-0.474	-0.461	-0.153		0.182	0.142
MinF_Start	0.234	0.161			-0.265	-0.209		0.148		-0.106
MaxF_Start	0.213		-0.293	-0.2	-0.157	0.339	-0.129	0.106		
Bandwidth_Start	0.194	-0.123	-0.315	-0.215	-0.131	0.39	-0.141			
Entropy_Start	0.188	-0.194	-0.327		-0.163	0.259				
HNR_Start		0.269	0.146	-0.472		-0.143	-0.151	0.615	0.19	
PF_End	0.244	0.177				-0.12			-0.225	-0.278
MinF_End	0.237	0.164	0.107			-0.102	0.19	0.12	-0.383	-0.197
MaxF_End	0.232		-0.253	-0.11	0.29	-0.276	0.119			0.147
Bandwidth_End	0.192		-0.329	-0.135	0.352	-0.288			0.122	0.235
Entropy_End	0.196	-0.108	-0.308		0.31	-0.245				-0.153
HNR_End		0.16		-0.613	-0.234		0.425	-0.514		
PF_Centre	0.258									-0.198
MinF_Centre	0.236	0.204		0.155		0.133	0.198		0.145	-0.101
MaxF_Centre	0.25	-0.104	0.224						-0.249	0.179
Bandwidth_Centre	0.189	-0.221	0.302	-0.11			-0.185		-0.369	0.263
Entropy_Centre	0.12	-0.366	0.258	-0.205			-0.169		-0.125	-0.128
HNR_Centre		0.415		-0.132	0.153	0.213	-0.247			
Eigenvalue	3.462	1.919	1.599	1.219	0.98	0.887	0.8	0.696	0.642	0.605
% Variance explained	47.9	14.7	10.2	5.9	3.8	3.1	2.6	1.9	1.6	1.5
Cumulative%	47.9	62.7	72.9	78.8	82.7	85.8	88.4	90.3	92.0	93.4

Supplementary Table S5 DFA structure matrix showing pooled within-group correlations among discriminating variables and first 10 standardized canonical discriminant functions with eigenvalues >0.6

PC scores	Discriminant functions									
	1	2	3	4	5	6	7	8	9	10
PC1	1.362	0.180	-0.067	-0.046	0.033	0.022	-0.031	-0.009	-0.024	-0.022
PC2	0.776	-0.943	-0.076	0.293	0.040	-0.068	-0.030	0.037	0.096	-0.068
PC3	0.460	0.400	0.723	0.521	-0.117	-0.145	0.074	0.021	0.066	0.080
PC4	0.733	-0.453	0.239	-0.423	-0.787	0.130	0.185	-0.009	0.059	0.184
PC5	0.007	-0.267	0.446	0.112	0.267	0.963	-0.312	0.007	-0.002	0.375
PC6	0.374	-0.163	0.226	-0.138	0.482	0.333	1.001	0.126	0.016	-0.140
PC7	0.401	0.039	-0.253	-0.332	0.424	-0.483	0.061	0.203	0.601	0.868
PC8	-0.026	0.051	-0.171	0.136	0.057	0.088	0.156	-1.342	0.448	0.016
PC9	-0.249	0.343	0.655	-0.556	0.029	0.225	-0.315	0.292	1.117	-0.774
PC10	0.412	-0.999	1.744	-1.075	0.627	-0.600	-0.271	-0.338	-0.478	-0.054
Eigenvalue	3.462	1.919	1.599	1.219	0.98	0.887	0.8	0.696	0.642	0.605
% of variance	75.74	11.79	6.34	2.98	1.56	1.25	0.22	0.08	0.04	0
Cumulative%	75.74	87.53	93.87	96.85	98.41	99.66	99.88	99.96	100	100

REFERENCES

Huang CM, Li Y, Zhou Q, Feng Y, Chen Z, Yu H, Wu Z. 2008. Karst habitat fragmentation and the conservation of the white-headed langur (*Trachypithecus leucocephalus*) in China. *Primate Conservation*, **23**: 133-139.

Huang CM, Wei FW, Li M, Quan G, Li H. 2002. Current status and conservation of white-headed langur (*Trachypithecus leucocephalus*) in China. *Biological Conservation*, **104**(2): 221-225.

Jin T, Wang DZ, Zhao Q, Yin L, Qin D, Ran W, Pan W. 2009. Social organization of white-headed langurs (*Trachypithecus leucocephalus*) in the Nongguan karst hills, Guangxi, China. *American Journal of Primatology*, **71**(3): 206-213.

Lu S, Chen T, Huang Z, Li Y, Lu C. 2021. Interannual variation in food choice of white-headed langur inhabiting limestone forests in Fusui, southwest Guangxi, China. *Ecology and Evolution*, **11**: 9349-9360.

R Core Team. 2021. R: A language and environment for statistical computing. R Foundation for Statistical Computing.

Yin L, Jin T, Watanabe K, Qin D, Wang D, Pan W. 2013. Male attacks on infants and infant death during male takeovers in wild white-headed langurs (*Trachypithecus leucocephalus*). *Integrative Zoology*, **8**(4): 365-377.