Vietnamese Journal of Primatology (2013) vol. 2(2), 75-82

Vocalization of red- and grey-shanked douc langurs (*Pygathrix nemaeus* and *P. cinerea*)

Isidoro Riondato, Marco Giuntini, Marco Gamba and Cristina Giacoma

University of Torino, Department of Life Sciences and Systems Biology, Via Accademia Albertina 13, 10123 Torino, Italy.

Isidoro Riondato and Marco Giuntini contributed equally to this work.
Corresponding author: Isidoro Riondato <isidoro.riondato@unito.it>

Key words: Vocalization, red-shanked douc langur, grey-shanked douc langur, *Pygathrix nemaeus*, *Pygathrix cinerea*

Summary

The vocal repertoires of red- and grey-shanked douc langurs (*Pygathrix nemaeus* and *P. cinerea*) have never been described before. Information derived from the recordings of the captive groups of the Endangered Primate Rescue Center, Cuc Phuong National Park, Vietnam, was used for a first quantitative description of these species' calls based on the calculation of pairwise dissimilarity indexes and cluster analysis. Distinctive features were found across the 8 vocal types identified for each species. Some of the clusters are feasibly representing call variants belonging to an acoustic continuum, but others clearly represent distinct types. The potential changes occurring due to sex- and age-dependent characteristics were not investigated in the present study.

Nghiên cứu về âm thanh học của hai loài chà vá chân nâu và chà vá chân xám (*Pygathrix nemaeus* và *P. cinerea*)

Tóm tắt

Nghiên cứu về âm thanh học của loài chà vá chân nâu và chà vá chân xám chưa từng được thực hiện trước đây. Những dữ liệu về tiếng kêu thu thập từ những đàn chà vá nuôi nhốt tại Trung tâm cứu hộ thú linh trưởng, Vườn quốc gia Cúc Phương, Việt Nam đã được phân tích một cách định lượng căn cứ vào các chỉ số không đồng nhất từng cặp và theo nhóm. Những đặc điểm khác biệt đã được tìm thấy ở 8 loại tiếng kêu khác nhau ở mỗi loài. Một số nhóm âm thanh thể hiện sự biến thiên đa dạng của tiếng kêu trong cùng một dải âm vực, tuy nhiên một số khác lại thể hiện sự khác biệt rõ ràng. Những thay đổi về âm thanh do sự khác nhau về giới tính và độ tuổi chưa được tìm hiểu kỹ trong nghiên cứu này.

Introduction

The study of primate vocal behaviour can offer important insights into the communication processes within a species, especially when seen in the light of the ecological and social information available (Maretti et al., 2010). Knowledge of the vocal repertoire can also be crucial to understand the biology of a species and set the appropriate policies for its conservation (Laiolo, 2010).

Previous studies have shown a strong influence of genetics on primate vocal repertoires (Hammerschmidt & Fischer< 2008) frequently leading to species-specific vocal types (Gamba & Giacoma, 2005) or showing shared types differing across species by more subtle acoustic properties basically due to morpho-anatomic variability (Gamba et al., 2012a). Sex-, age- and individual-specific features of primate calls may also derive from morpho-anatomical or physiological differences (Gamba et al., 2012b).

In this study we investigated the vocal repertoires of the red- and grey-shanked douc langur (*Pygathrix nemaeus* and *P. cinerea*). We identified the vocal types emitted by captive individuals using Affinity Propagation (hereafter AP) clustering. The resulting clusters were then confirmed as discrete vocal categories using multivariate analysis.

The information obtained is feasibly not sufficient to describe the complete vocal repertoire but largely increase knowledge of the species vocal communication. The description of the vocal types can be also valuable to recognize these monkeys during acoustic surveys in their natural habitats.

Material and Methods

The animals recorded in this study were housed at the Endangered Primate Rescue Center, Cuc Phuong National Park, Vietnam. This institution is part of the 'Vietnam Primate Conservation Programme' of the Frankfurt Zoological Society. For each species we had a different number of study groups and individuals. We recorded 10 individuals of *Pygathrix cinerea* hosted in 4 different enclosures, 15 *P. nemaeus* kept in 3 different enclosures (Table 1).

Table. 1. Composition and sex ratio of the study groups.

Species	N_{males}	$N_{females}$	$N_{youngsters}$	Total
Pygathrix cinerea	3	4	3	10
Pygathrix nemaeus	4	6	5	15

Recordings

We recorded the animals between March 4th and March 23rd, 2010. Recordings were collected using solid-state audio recorders Sound Devices 702 and Marantz PMD 671, equipped with Sennheiser ME66 (with windshield Sennheiser MZW66) and G.R.A.S. 40 BE microphones. During all sessions we used the *ad libitum* sampling method (Altmann, 1974). Vocal signals were recorded at a sampling rate of 96000 Hz (16 bit depth). All vocalizations were spontaneously emitted and nothing was done to elicit the emission of vocal signals, no vocal or visual playbacks were used. All subjects were treated according to the guidelines for animal behaviour studies (Anonymous, 2002).

Acoustic analyses

The recording sessions were split into shorter files using SoundForge 7.0 (Madison Media Software Inc., Sony Corporation of America, U.S.A.) and Adobe Soundtrack 2 (Apple Inc., U.S.A.). Once a vocalization was isolated from the recording session file, it has been saved in a new file and a silence of 0.5 sec was inserted at the beginning and the end. All sounds were normalized using the function "scale to peak" in the software Praat (Boersma & Weenink, 2006). The files prepared in this way were then treated in two different ways: (a) they were submitted to a similarity measuring procedure or (b) they were analysed using a semi-automatic extraction of the acoustic properties.

The generation of a pairwise similarity index (a) across the species-specific vocalizations was performed using an implementation of the dynamic time warping. The dynamic time warping is a procedure of spectrogram alignment increasingly used in bioacoustics to compare sounds belonging to large datasets. At the basis of this procedure, there is a method commonly used in speech science relying on the calculation of cepstrum coefficients (Davis & Mermelstein, 1980). These coefficients provide a representation of the energy distributed at the various frequencies in the sound spectrum and, even if the computation of cepstral coefficients is usually performed to match the sensitivity of the human ear, they have been shown to be useful in the study of animal calls (Trawicki et al., 2005; Clemins & Johnson, 2006; Ranjard & Ross, 2008; Tao et al., 2008; Brown & Smaragdis, 2009; Ranjard et al., 2010). We used a method currently implemented in a free package called DTWAVE (University of Auckland). A sequence of cepstrum coefficients was computed for each call by means of a mel filterbank (Ranjard et al., 2010) using the Hidden Markov Model Toolkit (Young, 1994). When they were submitted to the Hidden Markov Model Toolkit we used a target rate of 50000 ns and a window size of 100000 ns. Once all the cepstral coefficients were aligned and rescaled, the software constructed an average vector sequence. Then the dynamic time warping calculated the pairwise distances between all the calls in the dataset until only the sequence representing an average of all call sequences remained (see Ranjard & Ross, 2008).

The measurement of the acoustic properties of each vocalization was performed in Praat in the perspective of the source filter theory (Fant, 1960; Gamba & Giacoma, 2007; Taylor & Reby, 2010). To detect source features (MeF0, MiF0, MaF0, enstF0, sumvar, rangeF0, F0stdev, F0start, F0end), Fast Fourier transforms were generated for each call (frequency range: 0-12000 Hz; maximum: 100dB/Hz; dynamic range: 60 dB; pre-emphasis: 6.0 dB/Oct; dynamic compression: 0.0). The fundamental frequency and its variation were measured by using the autocorrelation method [“Sound: To pitch (ac)…”]. We adjusted the analysis parameters according to the range of variation in each vocalization (Gamba & Giacoma, 2005; Gamba et al., 2012a; 2012b).

Slightly different settings were used for those calls showing very low pitch. The files containing these calls were resampled to 22050 Hz. For these calls we have also used a ‘stop Hann band’ filter in Praat for frequency between 0 and 100 Hz (smoothing factor: 100 Hz) and were then analysed with specific settings [(advanced pitch settings) max number of candidates: 35; voice threshold: 0.15; octave-jump cost: 0.01; voiced/unvoiced cost: 0.15]. All measurements were taken during a step-by-step monitored process, where the operator could interrupt the analysis and modify the analysis parameters. A Praat script was used to automate file opening and editing and file saving of the measurements.

Quantitative and statistical analyses

The matrix of dissimilarity indexes generated from DTWAVE was submitted to a clustering procedure called AP (Frey & Dueck, 2007) using the *apcluster* package in R (Hornik, 2013). The algorithm used a squared negative Euclidean distance to measure dissimilarity and identify clusters. The number of clusters and the number of vocalizations used as exemplars were not user-defined. The result of the application of AP to measure vocal signal similarity were diagonally symmetric matrices, one for each species, in which each matrix element quantified the pairwise agreement between two of the calls. We extracted two parameters from this analysis: the number of the clusters and the filename of the calls assigned to each cluster. Then, the information concerning each vocalization and the cluster to which it was assigned was transferred into a table and used as ‘grouping variable’ in IBM SPSS Statistics 20 for Mac (IBM Corporation, Armonk, New

York, USA) for a stepwise Discriminant analysis (hereafter DFA). It is to be noticed that neither the parameters used during the dynamic time warping procedure nor the information derived from the clustering process were directly related with the acoustic measurements used in the Discriminant process. We have reported the average percentage of correctly classified instances (CCI) to the cluster and the cross-validated percentage of correctly classified instances (c.-v.) for each DFA.

Results

Description of the vocal repertoire

High quality calls entered the procedure for the calculation of the pairwise distances using DTWAVE. We set a procedure for each species.

The vocal emissions of *Pygathrix cinerea*.

We aimed to understand whether groups of the recorded calls could be identified using the dissimilarity matrix generated in DTWAVE. The AP analysis yielded 8 clusters (number of samples = 179; number of iterations = 191; input preference = -3.811533; sum of similarities = -32.84342; sum of preferences = -34.3038; net similarity = -67.14722). The cluster membership was then used as grouping variable in the DFA. For *P. cinerea* we found a significant model and high rate of correct classification [N = 179, Wilks' Lambda = 0.027, $F_{14,340}$ = 139.785, p < 0.001, CCI = 91.1% (c.-v. 89.4%)]. The most important parameters in this Discriminant model were duration (0.992) and F0start (0.078).

The spectrograms of the vocal types identified during the clustering process and then validated using the DFA as discrete categories (Fig. 1, Table 2).

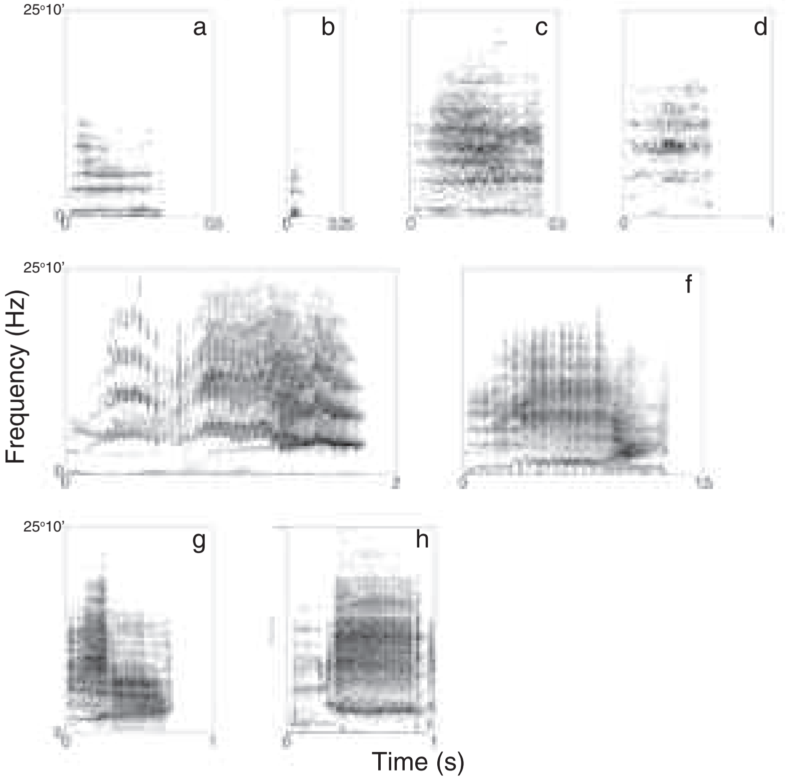

Fig.1. The spectrograms of the vocal types we recorded for *Pygathrix cinerea*: (a) Wroo, (b) Up, (c) Tzii, (d) Trill, (e) Friii, (f) Trill-wroo, (g) Phre, (h) Uo-traa. All spectrograms were generated in Praat with the following parameters: window length: 0.025 sec, time range as shown (0.25-2.00 s); frequency range: 0-24000 Hz; maximum: 50 dB/Hz; dynamic range: 35-55 dB; pre-emphasis: 6.0 dB/Oct; dynamic compression: 0.0.

Table. 2. Description of the *Pygathrix cinerea* vocal types. The spectrogram of a call resulting near to the vocal type centroid during the multivariate analysis is shown in Fig. 1.

Wroo	Medium-range low-pitched guttural call. The sonograms show a clear formants structure (Fig. 1a).
Up	Low-pitched short harsh vocalization (Fig. 1b).
Tzii	High-pitched vocalization showing frequency modulation. Compared to the Trill the harmonic structure is less visible due to the remarkable chaotic pattern (Fig. 1c).
Trill	Short high-pitched vocalization with frequency modulation (Fig. 1d).
Friii	Long and harsh high-pitched vocalization with remarkable frequency modulation (Fig. 1e).
Trill-wroo	High-pitched vocalization showing with frequency modulation followed by a short low-frequency emission (with a wroo-like structure, Fig. 1f).
Phre	Similar to *trill-wroo* but shorter and with different fundamental frequency values in the final portion (Fig. 1g).
Uo-traa	Vocalization with a high-pitched modulated part preceded and followed by short low-pitched units (with visible harmonics, Fig. 1h).

The vocal emissions of *Pygathrix nemaeus*.

The AP analysis of the pairwise distances generated for the calls of *Pygathrix nemaeus* yielded 8 clusters (number of samples = 264; number of iterations = 143; input preference = -4.804802;

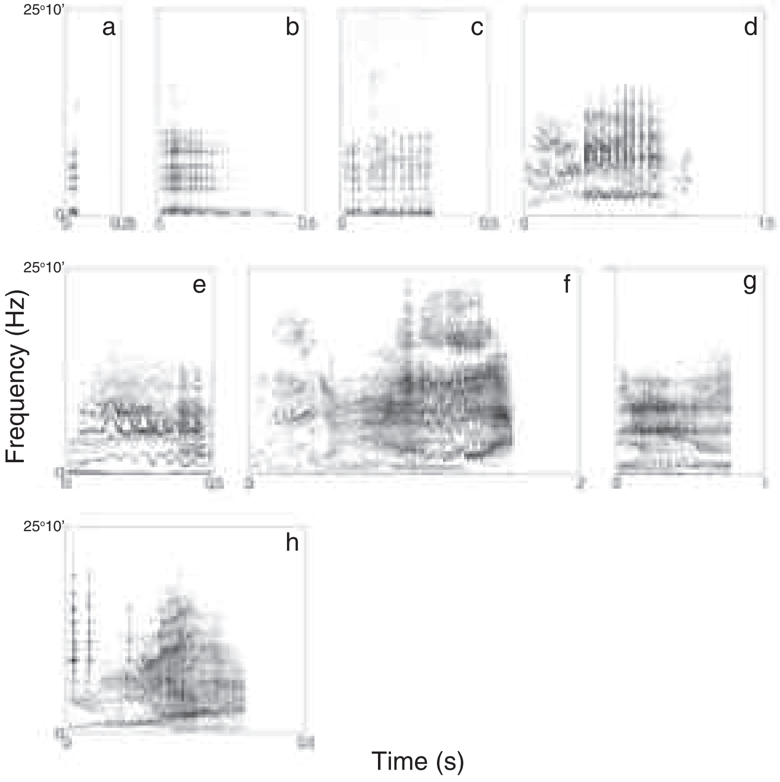

Fig.2. The spectrograms of the vocal types we recorded for *Pygathrix nemaeus*: (a) Up, (b) Wroo, (c) Slow Wroo, (d) Friii, (e) Trill, (f) Ruae-Ruae, (g) Uo-traa, (h) Trill-wroo. All spectrograms were generated in Praat with the following parameters: window length: 0.025 sec, time range as shown (0.25-2.00 s); frequency range: 0-24000 Hz; maximum: 50 dB/Hz; dynamic range: 35-55 dB; pre-emphasis: 6.0 dB/Oct; dynamic compression: 0.0.

sum of similarities = -65.32299; sum of preferences = -43.24321; net similarity = -108.56628), which were then submitted to DFA. The classification process showed a consistent part of the set correctly assigned to the previously identified clusters (CCI = 84.8%, c.-v. 83.0%) and a significant Discriminant model (N = 264, Wilks' Lambda = 0.036, $F_{14,510}$ = 172.351, $p < 0.001$). The six acoustic measures participating primarily to the Discriminant model were duration (0.937), enstF0 (0.118), sumvar (0.062), rangeF0 (0.149), F0stdev (0.143), and F0start (0.018). The spectrograms of the call types identified using AP and classified in the DFA (Fig. 2, Table 3).

Table. 3. Description of the *Pygathrix nemaeus* vocal types. The spectrogram of a call resulting near to the vocal type centroid during the multivariate analysis is shown in Fig. 2.

Up	Low-pitched and short harsh vocalization (Fig. 2a).
Wroo	Low-pitched guttural calls with clear formants structure (Fig. 2b).
Slow Wroo	Low-pitched guttural calls with clear formants structure. They are usually given at low intensity (Fig. 2c).
Friii	Long and harsh high-pitched vocalization with remarkable frequency modulation (Fig. 2d).
Trill	Short high-pitched vocalization with frequency modulation (Fig. 2e).
Ruae-ruae	Long and harsh high-pitched vocalization with frequency modulation. Units with low pitch occur during the emission (Fig. 2f).
Uo-traa	Long and harsh high-pitched vocalization, with frequency modulation preceded and followed by a low-pitched unit (Fig. 2g).
Trill-wroo	Long duration high-pitched vocalization with frequency modulation follow by short low-frequency emission (with a wroo-like structure, Fig. 2h).

Discussion

The recorded calls of *P. cinerea* and *P. nemaeus* showed an acoustic variability that allowed identifying 8 different vocal types for each species, all of which are distinguishable by visual inspection of the spectrograms and by ear. As we mentioned before, taking into account that the dataset has not been checked for sex-, age- or individual-specific characteristics, the vocal types identified in the present study can potentially represent individually given variants of the same call, although in most cases spectrographic inspection seems actually to support the discreteness of most vocal types.

Comparison between *Pygathrix cinerea* and *P. nemaeus*

We can notice that calls with similar acoustic structure are present in the repertoire of both species. Six clusters for each species are showing calls with similar acoustic structure: the *up*, the *wroo*, the *frii*, the *trill*, the *trill-wroo*, and the *uo-traa*. The *ups* are very short low-pitched vocalizations with visible formants. They sound like a burp. The *wroos* are low-pitched calls showing longer duration when compared to the *ups*.

A common acoustic unit across many of the call types is that they include a *trill* unit. The *trills* are high-pitched calls showing remarkable frequency modulation. In addition to the *trills*, we found other three vocal types showing a *trill*-like unit. The *friiis* are similar to *trills*, but compared to these, they are longer in duration and with stronger frequency modulation. The other two vocal types, the *trill-wroo* and the *uo-traa* sound harsh. They are both high-pitched and showed remarkable frequency modulation. Moreover is important to highlight that all the vocal types showing a *trill*-like structure often occur with subharmonics and other nonlinear phenomena.

Comparison with other species

Qualitatively speaking, we screened the published spectrograms of taxa phylogenetically close to the genus *Pygathrix*, as the species belonging to the genera: *Nasalis* (Messeri & Trombi, 2000; Srivathsan & Meier, 2011), *Presbytis* (Wich et al., 2008; Meyer et al., 2012), *Simias* (Schneider et al., 2007) and *Trachypithecus* (Krishnamurthy, 1991).

Messeri & Trombi (2000) recognized six vocal types in *Nasalis larvatus* using acoustic measurement partially overlapping those used in this study. Thus the repertoire size of *N. larvatus* seems reasonably similar to those found for *Pygathrix cinerea* and *P. nemaeus* in the present study.

Previous studies often concentrated on specific aspects of communication signals but only provided information on context-specific utterances In as contact calls (Rawson, 2009) or male loud calls of other species (Wich & Nunn, 2002). From these investigations we should expect various loud call types showing their main energies concentrated in the spectral window with the lowest background noise (e.g. low-pitched calls) and having a whistle-like structure with scarce frequency modulation (Schneider et al., 2007). This is not what we have found in the present study, where most types showed remarkable frequency modulation and a relatively high pitch. It is to be noticed that very high-pitched calls were recently found in many primate species and our findings support and even larger presence of these vocalizations in the repertoire of colobines. In fact, Riondato (*unpublished data*) also found high-pitched frequency modulated vocalizations in *Rhinopithecus brelichi*. The adaptive significance of vocalizations with such a high fundamental frequency is still unclear but their acoustic structure could actually mediate communication between conspecifics in a noisy and densely forested environment (Snowdon & Hodun, 1981).

It would be of great relevance to perform further recordings and analyses and more fieldwork on the study species, including eco-ethological observations, could provide a valuable framework for generalisation and functional interpretation of the results shown in this paper.

Acknowledgements

We are grateful to Truong Quang Bich, Director of Cuc Phuong National Park and Mr. Do Van Lap, Chief of the Department of Scientific and International Collaborations for granting permission to conduct research at the Endangered Primate Rescue Center. We are especially grateful to Tilo Nadler, Director of the EPRC and Nguyen Thi Thu Hien, Manager of the EPRC for their continued support of our research.

References

Altmann J (1974): Observational study of behavior: sampling methods. Behaviour 49, 227–265.

Boersma P & **Weenink D** (2006): Praat: Doing phonetics by computer (Version 4.3.20). Retrieved from: http://www.praat.org.

Brown J & **Smaragdis P** (2009): Hidden Markov and Gaussian mixture models for automatic call classification. J. Acoustical Society of America Express Letters, DOI: 10.1121/1.3124659

Clemins PJ & **Johnson MT** (2006): Generalized perceptual linear prediction features for animal vocalization analysis. J. Acoustical Society of America 120, 527–534.

Davis SB & **Mermelstein P** (1980): Comparison of parametric represen- tations for monosyllabic word recognition in continuously spoken sentences. IEEE transactions on acoustics, speech, and signal processing 28(4), 357–365.

Fant G (1960): Acoustic Theory of Speech Production. Mouton, The Hague.

Frey BJ & **Dueck D** (2007): Clustering by passing messages between data points. Science 315, 972–976.

Gamba M & **Giacoma C** (2005): Key issues in the study of primate acoustic signals. J. Anthropol. Sciences 83, 61-87.

Gamba M & **Giacoma C** (2007): Quantitative acoustic analysis of the vocal repertoire of the crowned lemur. Ethology Ecology & Evolution 19, 323-343.

Gamba M, **Colombo C** & **Giacoma C** (2012a): Acoustic cues to caller identity in lemurs: a case study. J. Ethology 30(1), 191-196.

Gamba M, **Friard O** & **Giacoma C** (2012b): Vocal tract morphology determines species-specific features in lemur vocal signals (Eulemur). Int. J. Primatol. 33(6), 1453-1466.

Hammerschmidt K & **Fischer J** (1998): The vocal repertoire of Barbary macaques: a quantitative analysis of a graded signal system. Ethology 104, 203–216.

Hornik K (2013): The R FAQ, http://CRAN.R-project.org/doc/FAQ/R-FAQ.html

Krishnamurthy RS (1991): Vocalization patterns of captive francois' langurs (*Presbytis francoisi*). Ph.D. Dissertation, Oregon State University.

Laiolo P (2010): The emerging significance of bioacoustics in animal species conservation. Biological Conservation 143(7), 1635–1645.

Maretti G, **Sorrentino V**, **Finomana A** & **Giacoma C** (2010): Not just a pretty song: an overview of the vocal repertoire of Indri indri. J. Anthropol. Sciences 88, 151–165.

Messeri P & **Trombi M** (2000): Vocal repertoire of proboscis monkey (*Nasalis larvatus, L.*) in Sarawak. Folia Primatol. 71 (4), 268–87.

Meyer D, **Hodges JK**, **Rinaldi D**, **Wijaya A**, **Roos C** & **Hammerschmidt K** (2012): Acoustic structure of male loud-calls support molecular phylogeny of Sumatran and Javanese leaf monkeys (genus *Presbytis*). Evolutionary Biology 12(16) 1-11.

Ranjard L, **Anderson MG**, **Rayner MJ**, **Payne RB**, **McLean I**, **Briskie JV**, **Ross HA**, **Brunton DH**, **Woolley SMN** & **Hauber ME** (2010): Bioacoustic distances between the begging calls of brood parasites and their host species: a comparison of metrics and techniques. Behavioral Ecology and Sociobiology, 64(11), 1915–1926.

Ranjard L & **Ross HA** (2008): Unsupervised bird song syllabe classification using evolving neural networks. J. Acoustical Society of America 123(6), 4358-4368.

Rawson BM (2009): The socio-ecology of the black-shanked douc (*Pygathrix nigripes*) in Mondulkiri Province, Cambodia. PhD thesis, Australian National University, Canberra.

Schneider C, **Hodges K**, **Fisher J** & **Hammerschmidt K** (2007): Acoustic Niches of Siberut Primates. Int. J. of Primatol. 29, 601–613.

Snowdon CT & **Hodun A** (1981): Acoustic adaptions in pygmy marmoset contact calls: Locational clues vary with distances between conspecifics. Behav Ecol. Sociobiol 9, 295-300.

Srivathsan A & **Meier R** (2011): Proboscis monkeys (*Nasalis larvatus* Wurmbier, 1787) have unusually high-pitched vocalizations. The Raffles Bull. of Zool. 59(2), 319–323.

Tao J, **Johnson MT** & **Osiejuk TS** (2008): Acoustic model adaptation for ortolan bunting (*Emberiza hortulana L.*) song-type classification. J. Acoustical Society of America 123 (3), 1582–1590.

Taylor AM & **Reby D** (2010): The contribution of source-filter theory to the study of mammal vocal communication (Review). J. Zool. 280(3), 221-236.

Trawicki MB, **Johnson MT** & **Osiejuk TS** (2005): Automatic song-type classification and speaker identification of Norwegian ortolan bunting (*Emberiza hortulana*) vocalizations. IEEE Workshop on Machine Learning for Signal Processing, Mystic, CT, 277–282.

Young SJ (1994): The HTK Hidden Markov Model Toolkit: Design and Philosophy. Cambridge University, Engineering Department.

Wich S, **Schel AM** & **De Vries H** (2008): Geographic Variation in Thomas Langur (*Presbytis thomasi*) Loud Calls. Am. J. Primatol. 70, 566–574.

Wich SA, **Nunn CL** (2002). Do Male "Long-distance Calls" Function in Mate Defense? A comparative Study of Long-distance Calls in Primates. Behavioral Ecology and Sociobiology 52, 474-484.