Vocal classification of vocalizations of a pair of Asian Small-Clawed otters to determine stress

Peter M. Scheifele, Michael T. Johnson, Michelle Fry, Benjamin Hamel, and Kathryn Laclede

Citation: The Journal of the Acoustical Society of America **138**, EL105 (2015); doi: 10.1121/1.4922768
View online: http://dx.doi.org/10.1121/1.4922768
View Table of Contents: http://asa.scitation.org/toc/jas/138/1
Published by the Acoustical Society of America

Vocal classification of vocalizations of a pair of Asian Small-Clawed otters to determine stress

Peter M. Scheifele,[1,a)] Michael T. Johnson,[2] Michelle Fry,[3] Benjamin Hamel,[2] and Kathryn Laclede[1]

[1]*FETCHLAB, Department of Audiology, University of Cincinnati, 3202 Eden Avenue, Cincinnati, Ohio 45267, USA*

[2]*Electrical and Computer Engineering Department, Marquette University, 1515 West Wisconsin Avenue, Milwaukee, Wisconsin 53233, USA*

[3]*Newport Aquarium, 1 Aquarium Way, Newport, Kentucky 41071, USA*

scheifpr@ucmail.uc.edu, mike.johnson@mu.edu, mfry@newportaquarium.com, benjamin.hamel@mu.edu, lacledkm@mail.uc.edu

Abstract: Asian Small-Clawed Otters (*Aonyx cinerea*) are a small, protected but threatened species living in freshwater. They are gregarious and live in monogamous pairs for their lifetimes, communicating via scent and acoustic vocalizations. This study utilized a hidden Markov model (HMM) to classify stress versus non-stress calls from a sibling pair under professional care. Vocalizations were expertly annotated by keepers into seven contextual categories. Four of these—aggression, separation anxiety, pain, and prefeeding—were identified as stressful contexts, and three of them—feeding, training, and play—were identified as non-stressful contexts. The vocalizations were segmented, manually categorized into broad vocal type call types, and analyzed to determine signal to noise ratios. From this information, vocalizations from the most common contextual categories were used to implement HMM-based automatic classification experiments, which included individual identification, stress vs non-stress, and individual context classification. Results indicate that both individual identity and stress vs non-stress were distinguishable, with accuracies above 90%, but that individual contexts within the stress category were not easily separable.

[WA]

Date Received: January 23, 2015 **Date Accepted:** June 3, 2015

1. Introduction

Asian Small-Clawed Otters (*Aonyx cinerea*) are a small species that inhabit freshwater coastal areas throughout China, Southeast Asia, and Indonesia. They are currently a protected species threatened by hunting and pollution (Hussain and de Silva, 2008). They are gregarious and live in monogamous pairs for their lifetimes (Wozencraft, 2005). Interspecies communication is accomplished primarily via scent and acoustical means (Perdue *et al.*, 2013; Lemasson *et al.*, 2014). Classification of their vocalizations will give conservationists a better idea of exactly how they live and perhaps enable us to assist in preserving the species. The only prior study of this species' vocalizations is from Lemasson *et al.* (2013) although they are a popular species kept under professional care at various aquaria and zoos throughout the world. As such, monitoring the health and well-being of these otters by staff biologists is paramount.

Monitoring for care requires the ability to observe when an animal is under stress, and vocalizations are a significant indicator of stress. This study focused on the analysis of stress and non-stress calls from a sibling pair in captivity. Vocalization classification experiments were implemented to examine differences between the two individuals and differences between specific stress and non-stress vocalizations. Automated monitoring of the backup area where these otters sleep, eat, and where husbandry takes place would be ideal for their daily care. In addition, since the pair have to be separated when one or the other is taken to the veterinary clinic for routine health preventative care such monitoring would allow the keepers to better manage the lone animal during that time.

Classification was implemented using a hidden Markov model (HMM) approach. HMMs (Rabiner, 1989) are statistical state machines, commonly used for

[a)]Author to whom correspondence should be addressed.

human speech recognition, and which have shown significant promise in classification of bioacoustics vocal patterns in a number of species (Adi *et al.*, 2010; Clemins, 2005; Clemins *et al.*, 2004; Clemins and Johnson, 2006; Ren *et al.*, 2009; Trawicki and Johnson, 2005).

2. Methods

Vocalizations were recorded from a sibling pair, one male (Gyan) and one female (Malena) from July of 2014 through March of 2015. This pair were monitored and recorded in the backup area of an aquarium (from 3 m) using a Zoom H1 recorder (Zoom Corp., Tokyo, Japan) (sample rate 48 000 Hz, resolution 32 bits, recording duration 30 min) and Schur SM86 microphone (50 Hz to 20 kHz frequency response with an open circuit sensitivity of −50 dB V/Pa). The data had a primary sampling rate of 44.1 kHz and an average duration of sound files was 21.65 s ($\sigma = 33.64$ s).

Initial annotation of behavioral context was made onsite by the primary keeper and trainer based on behavior during this study. The training set was made by removing the female from the backup area and recording the male who exhibited separation anxiety while she was away for veterinary services. In addition, audio and video recordings were made of this pair before, during and after feeding, while at play, and during husbandry training sessions. The videos established a behavioral pattern within each of these except for the pain category. These behaviors were then matched with the reported vocalizations as the "expert classification" of each vocalization. This was done by the otter trainer, keeper and acoustician collectively from the recordings. Stressful contexts included four categories: aggression (S1), separation anxiety (S2), pain (S3), and prefeeding (S4). Non-stressful contexts included three categories: feeding (NS1), training (NS2), and play (NS3). The seven behaviors were further segregated by individual identify (or equivalently by gender), yielding 14 total categories.

In addition, vocalizations were manually segmented and labeled according to general vocal type, including the four call types chirps, squeals, barks, screams, as well as an unknown/other type, by visual inspection of the spectrograms and the use of auditory factors (such as length of segment, tone, and volume). Since the vocal repertoire of the species is not clearly established (Lemasson *et al.*, 2013) and this experiment represents a small number of individuals in a captive environment, this separation into vocal type is not intended to represent a comprehensive repertoire analysis but simply provides a broad acoustic categorization for further analysis. Extraneous sounds (such as doors closing or background noise) were also removed from sound files and labeled separately. Most of the calls fell into the chirp or squeal category, and the remaining analysis focuses on only those vocal types.

The spectrograms shown in Fig. 1 represent the general patterns observed for the four basic call types. The average duration of a vocalization segment was 1.02 s ($\sigma = 0.61$ s). Squeal segments averaged 1.16 s in duration ($\sigma = 0.60$ s), chirp segments averaged 0.33 s in duration ($\sigma = 0.12$ s), scream segments averaged 0.90 s in duration ($\sigma = 0.19$ s), and bark segments averaged 0.42 s in duration ($\sigma = 0.04$ s).

Recordings that included either multiple individuals or multiple behaviors in a single file were discarded. Table 1 shows the total number of recordings for each of the 14 possible label types, 73 in total. In order to ensure sufficient data, behavioral categories with 0 or 1 recordings were eliminated from the study, reducing the total number of categories from the original 14 possible categories to just the following subset: GS2, GS4, MS2, MS4, and MNS3. The biggest limitation of this is that there were insufficient non-stress vocalizations for the male otter which limits the study of stress/non-stress vocal differences to within a single individual.

After segmentation, the vocalized and background portions of the recordings were analyzed to estimate the signal-to-noise ratio (SNR) for each vocalization segment. Vocalizations with SNR below 0 dB (noise background of greater magnitude than the signal) were eliminated from the study. Once all vocalizations had been segmented, categorized by basic vocalization type, and quantified by SNR, a data analysis of the total number of calls was done. Table 2 shows the number of vocalizations as a function of SNR and contextual category for the remaining 343 vocalizations with SNR above zero.

3. Experimental design and results

An analysis of SNR as a function of vocalization type and behavioral context indicated that a study of relatively "clean" vocalizations, defined by SNRs greater than 15 dB, would be limited to MS2 vs MS4, i.e., separation anxiety versus prefeeding vocalizations within one female subject. The other three categories—GS2, GS4, and MNS3—contain mostly noisy calls in the 0 dB to 5 dB SNR range. In addition, most

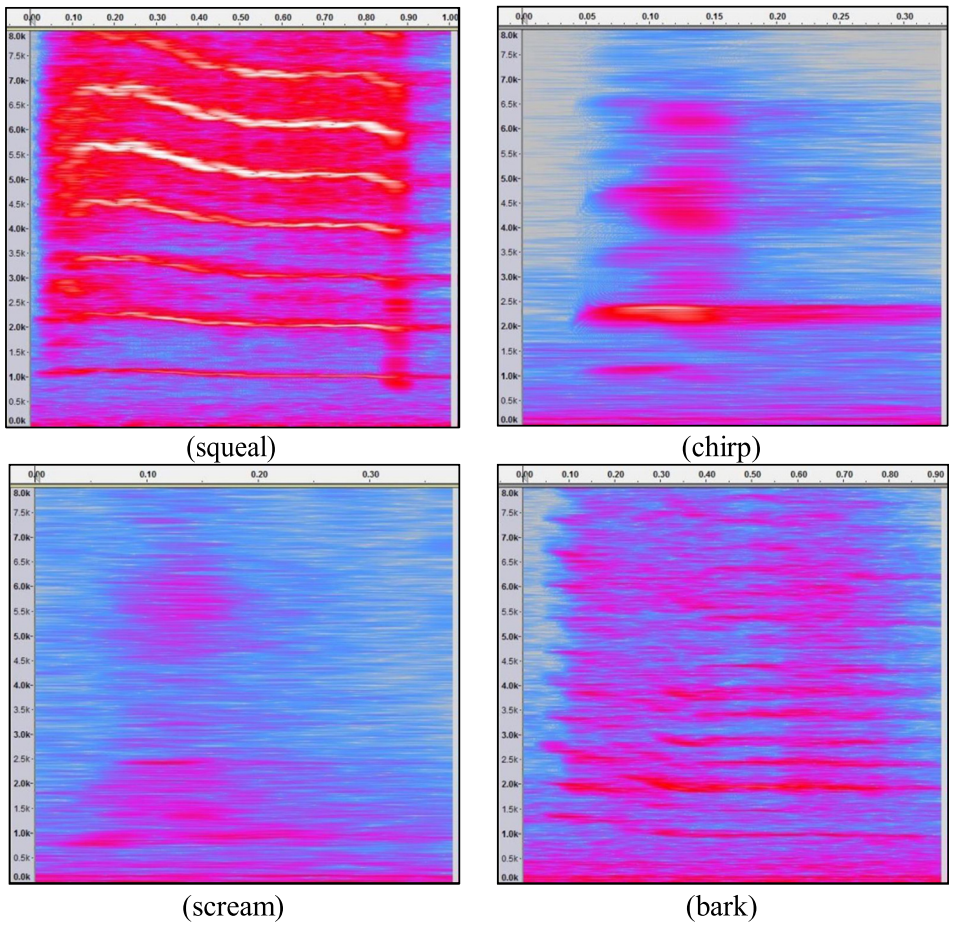

Fig. 1. (Color online) Illustrative spectrograms for the four call categories: (clockwise from top left) squeal; chirp; scream; bark. Horizontal axis indicates time in seconds, and vertical axis indicates frequency in kHz.

of the GS2, GS4, and MNS3 calls are chirps, while the MS2 and MS4 calls are primarily squeals with a smaller subset of chirps.

However, since the MS2 and MS4 are by far the largest groups, there are still enough chirps available to enable comparison all five behavioral categories within the chirp vocalization subtype of a single individual, which makes it possible to compare vocalization differences related to behavior with the smallest number of confounding variables.

On the basis of the data availability and focus of the study, several classification experiments were designed, including the following:

- Individual identify—GS4 vs MS4, among all chirps (SNR > 0 dB);
- Stress vs non-stress—MNS3 vs (MS2 + MS4), among chirps with SNR > 0 dB;
- Separation anxiety vs prefeeding—MS2 vs MS4, low noise (squeals with SNR > 15 dB) and general (SNR > 0 dB) sub-cases.

In order to maximize the use of data, classification experiments were implemented using a tenfold cross-validation protocol. This protocol consisted of splitting

Table 1. Distribution of behavioral labels across recording sessions. Total 73 files. G = Gyan (male), M = Malena (female), S = stress context, NS = non-stress context.

Label	Quantity	Label	Quantity
GS1	0	MS1	0
GS2	2	MS2	22
GS3	0	MS3	0
GS4	6	MS4	35
GNS1	1	MNS1	0
GNS2	0	MNS2	0
GNS3	1	MNS3	6

Table 2. Total number of vocalizations (all call types) as a function of minimum signal-to-noise ratio (SNR). The first column shows the distribution of all 343 calls, while the remaining columns show the distribution remaining when restricted to a specific minimum SNR.

Voc\SNR	>0	>5	>10	>15
GS2	3	0	0	0
GS4	25	5	2	1
MNS3	26	4	3	2
MS2	188	115	78	44
MS4	343	271	164	132

the dataset into ten random label-balanced subsets, and implementing ten experimental runs such that in each run 90% of the data was used for model training and the remaining 10% was held out for testing.

In each experiment, a three-state single Gaussian HMM was trained for each classification category. Input features were mel-frequency cepstral coefficients and normalized energy, along with the first and second derivatives (Young, 1997), calculated on 25 ms long windows with a step size of 10 ms. Feature computation and HMMs were implemented using the free Hidden Markov Model Toolkit (HTK) (Young, 1997) in combination with a freely available MATLAB-based graphic user interface recognition toolkit (Johnson *et al.*, 2009). Model training in this toolset is accomplished via the standard Baum-Welch expectation maximization algorithm, and recognition via the standard Viterbi algorithm (Young, 1997). Confusion matrices from results of the classification experiments are shown in Table 3 below. Overall, the results indicate the following:

- Individual identity—91% accuracy separating Gyan and Malena vocalizations, within calls representing a single vocal type and behavior pattern;
- Stress vs non-stress—93% accuracy separating NS3 play vs S2 separation anxiety and S4 prefeeding, (classification within single individual);
- Separation anxiety vs prefeeding—no significant separation of these stress subcategories (65%–70% accuracy, not significantly higher than chance).

Table 3. Classification accuracy confusion matrix results: (a) individual identity GS4 vs MS4 (b) stress vs non-stress MNS3 vs MS2 and MS4 combined; (c) stress sub-category MS2 vs MS4 (low-noise squeals); (d) stress sub-category MS2 vs MS4 (all squeals). In each confusion matrix, rows represent actual categories and columns represent classified vocalizations, so that entries on the diagonal represent correct classifications and entries off-diagonal represent errors.

(a) Individual identity (91% accuracy) GS4 vs MS4, chirps > 0 dB

	GS4 (classified)	MS4 (classified)
GS4 (actual)	14	6
MS4 (actual)	7	118

(b) Stress vs non-stress (93.7% accuracy) MNS3 vs MS2/4, chirps > 0 dB

	MNS3 (classified)	MS2/4 (classified)
MNS3 (actual)	21	0
MS2/4 (actual)	12	157

(c) Stress sub-category separation anxiety vs prefeeding (74.6% accuracy) MS2 vs MS4, squeals > 15 dB

	MS2 (classified)	MS4 (classified)
MS2 (actual)	23	21
MS4 (actual)	22	103

(d) Stress sub-category separation anxiety vs prefeeding (69.4% accuracy) MS2 vs MS4, squeals > 0 dB

	MNS3 (classified)	MS2/4 (classified)
MNS3 (actual)	102	42
MS2/4 (actual)	86	188

These results indicate that individual otter vocalizations can be differentiated and that, within individuals, the presence of stress may be differentiated by vocalizations as well, as seen by the 93% accuracy separating NS3 play vs S2 separation anxiety and S4 prefeeding. We were able to classify stress calls by individual (male versus female) and identify stress due to separation anxiety and hunger. Future work may include a more in-depth classification using additional data in situations such as temperature changes, crowding, reaction to unfamiliar objects or altercations, as well as the comparison of vocalizations between wild and captive animals.

4. Conclusions

This work has presented an initial study of the potential for automatic classification of otter vocalizations, with a focus on recognition of behavioral contexts. Knowledge of these vocal patterns as well as tools for automatic differentiation have the potential to offer professional caregivers better mechanisms to acoustically monitor the otters and enhance caregiving especially in situations where a family group of more than two otters is being kept. Potential benefits of vocalization monitoring include identification of aggression within the group, improvement of husbandry behaviors to reduce stress, and better management of environmental noise. Results indicate that both individual identify and stress vs non-stress are vocally distinguishable with accuracies above 90%, suggesting that vocal differentiation of stress is feasible and supporting the need for further continued work in this direction.

Acknowledgments

The authors thank the Newport Aquarium biology and caregiving staff for logistical and biological support. We particularly thank Chris Pierson, Operations Director, and Mark Dvornak, General Curator, for technical and animal health guidance.

References and links

Adi, K., Johnson, M. T., and Osiejuk, T. S. (**2010**). "Acoustic censusing using automaticvocalization classification and identity recognition," J. Acoust. Soc. Am. **127**(2), 874–883.

Clemins, P. (**2005**). "Automatic classification of animal vocalizations," Ph.D. thesis, Marquette University, Milwaukee, WI.

Clemins, P., and Johnson, M. T. (**2006**). "Generalized perceptual linear prediction (gPLP) features for animal vocalization analysis," J. Acoust. Soc. Am. **120**(1), 527–534.

Clemins, P. J., Johnson, M. T., Leong, K. M., and Savage, A. (**2004**). "Automatic classification and speaker identification of African Elephant (*Loxodonta Africana*) vocalizations," J. Acoust. Soc. Am. **117**(2), 956–963.

Hussain, S. A., and De Silva, P. K. (**2008**). "Aonyx cinerea," in *IUCN 2008: IUCN Red List of Threatened Species*, International Union for Conservation of Nature and Natural Resources report.

Johnson, M. T., Conley, D., Delfrate, D., and Bost, C. (**2009**). "Recognition toolkit for MATLAB," http://speechlab.eece.mu.edu/dolittle/proj_reu.html (Last viewed January 2015).

Lemasson, A., Mikus, M. A., Blois-Heulin, C., and Lodé, T. (**2013**). "Social partner discrimination based on sounds and scents in Asian Small-Clawed Otters (*Aonyx cinereus*)," Commun. Natrurwissenschaften **100**, 275–279.

Lemasson, A., Mikus, M. A., Blois-Heulin, C., and Lodé, T. (**2014**). "Vocal repertoire, individual distinctiveness, and social networks in a group of captive Asian Small-Clawed Otters (*Aonyx cinerea*)," J. Mammal. **95**(1), 128–139.

Perdue, B. M., Snyder, R. J., and Maple, T. L. (**2013**). "Cognitive research in Asian Small-Clawed Otters," Int. J. Comp. Psychol. **26**(1), 105–113.

Rabiner, L. R. (**1989**). "Tutorial on hidden Markov Models and selected applications in speech recognition," Proc. IEEE **77**, 257–286.

Ren, Y., Johnson, M. T., Clemins, P. J., Darre, M., Glaeser, S. S., Osiejuk, T. S., and Out-Nyarko, E. (**2009**). "A framework for bioacoustic vocalization analysis using hidden Markov models," Algorithms **2**(4), 1410–1428.

Trawicki, M., and Johnson, M. T. (**2005**). "Automatic song-type classification and speaker identification of Norwegian Ortolan Bunting (Eberiza Hortulana)," in *IEEE International Conference on Machine Learning in Signal Processing*, June 1992.

Wozencraft, W. C. (**2005**). "Order Carnivora," in *Mammal Species of the World*, 3rd ed., edited by D. E. Wilson, and D. M. Reeder (Johns Hopkins University Press, Baltimore, MD), pp. 532–628.

Young, S., Odell, J., Ollason, D., Valtchev, V., and Woodland, P. (**1997**). *The HTK Book* (Microsoft Corp.), pp. 995–1999.