Discrimination of individual tigers (*Panthera tigris*) from long distance roars

An Ji and Michael T. Johnson[a)]
Department of Electrical and Computer Engineering, Marquette University, 1515 West Wisconsin Avenue, Milwaukee, Wisconsin 53233

Edward J. Walsh and JoAnn McGee
Developmental Auditory Physiology Laboratory, Boys Town National Research Hospital, 555 North 30th Street, Omaha, Nebraska 68132

Douglas L. Armstrong
Omaha's Henry Doorly Zoo, 3701 South 10th Street, Omaha, Nebraska 68107

(Received 8 December 2011; revised 11 January 2013; accepted 16 January 2013)

This paper investigates the extent of tiger (*Panthera tigris*) vocal individuality through both qualitative and quantitative approaches using long distance roars from six individual tigers at Omaha's Henry Doorly Zoo in Omaha, NE. The framework for comparison across individuals includes statistical and discriminant function analysis across whole vocalization measures and statistical pattern classification using a hidden Markov model (HMM) with frame-based spectral features comprised of Greenwood frequency cepstral coefficients. Individual discrimination accuracy is evaluated as a function of spectral model complexity, represented by the number of mixtures in the underlying Gaussian mixture model (GMM), and temporal model complexity, represented by the number of sequential states in the HMM. Results indicate that the temporal pattern of the vocalization is the most significant factor in accurate discrimination. Overall baseline discrimination accuracy for this data set is about 70% using high level features without complex spectral or temporal models. Accuracy increases to about 80% when more complex spectral models (multiple mixture GMMs) are incorporated, and increases to a final accuracy of 90% when more detailed temporal models (10-state HMMs) are used. Classification accuracy is stable across a relatively wide range of configurations in terms of spectral and temporal model resolution. © *2013 Acoustical Society of America*. [http://dx.doi.org/10.1121/1.4789936]

PACS number(s): 43.80.Ka [JJF] Pages: 1762–1769

I. INTRODUCTION

Unlike its smaller relatives, the tiger is known as a roaring cat, a distinguishing vocal attribute shared only with other species belonging to the genus *Panthera*. Although not universally accepted, the capacity to roar is generally taken to be the vocal attribute of a specialized hyoid apparatus in which the normally ossified and consequently rigid epihyoidium exhibited by most other representatives of Felidae is instead ligamentous and, therefore, elastic among representatives of the genus, *Panthera*. This anatomical specialization reputedly allows tigers and other species within *Panthera* to increase the length of the vocal tract during the act of roaring and, as a consequence, produce the intense low-frequency signature of the call (Weissengruber *et al*., 2002). Roaring, however, is but one of numerous calls in the tiger's vocal repertoire. Hissing, grunting, growling, snarling, gasping, and chuffing are also prominent utterances that are used to express attitudes and intentions in a variety of social settings (Powell, 1957; Schaller, 1967; Peters, 1978). Some calls, like the full-throated confrontational roar, are impressively loud, while others, like chuffing, are just audible within a few feet of the source. This wide dynamic range is largely a manifestation of the tiger's larynx; the flat and broad medial surface of its relatively massive vocal folds (Hast, 1989; Weissengruber *et al*., 2002; Titze *et al*., 2010) enables the big cat to produce surprisingly low phonation thresholds and extraordinary output (Titze *et al*., 2010; Klemuk *et al*., 2011).

Many studies have shown the presence of distinctive vocal features across a wide range of animal species (McGregor, 1993; Suthers, 1994). The degree of individuality, and the difficulty in extracting and using acoustic cues to identify individuals, differs among species (Eakle *et al*., 1989; Gibert *et al*., 1994; Puglisi and Adamo, 2004). The goal of this study was to determine the extent to which individual tigers can be identified on the basis of the acoustical properties of one specific, representative call, the long distance roar (LDR) that is sometimes referred to as a territorial roar, an estrus roar, an intense mew, or a moan (Peters, 1978; Walsh *et al*., 2010; Walsh *et al*., 2011b). The LDR appears to be one, if not the most common vocalization produced by tigers both in captivity and in the wild, often being repeated frequently for a period of 1 or 2 h. While not extensively studied in an ethological context, the call appears to operate in a variety of settings and is clearly intended to advertise an individual's presence. The call, a deep throated

a)Author to whom correspondence should be addressed. Electronic mail: mike.johnson@marquette.edu

 0001-4966/2013/133(3)/1762/8/$30.00

ahh-rooom typically lasting between 1 and 2 s, has all of the acoustic properties necessary to propagate through the environment for long distances, and field biologists and naturalists refer to the call as "one of the most thrilling noises one can hear in the jungle" (Powell, 1957).

There is a wide variety of approaches commonly used to evaluate a species' vocal individuality. Multivariate statistical approaches are often employed to establish quantitative measures, including methods such as discriminant function analysis (DFA), multivariate analysis of variance (MANOVA) and principal components analysis (Fristrup and Watkins, 1992; Leong *et al.*, 2002; Riede and Zuberbuhler, 2003). Such approaches have been used to establish the presence of vocal individuality in studies of many different species, including avian (Bauer and Nagl, 1992; Peake *et al.*, 1998), canine (Durbin, 1998; Darden *et al.*, 2003), and primate (Jorgensen and French, 1998) species. Beyond the basic statistical approach, a number of more complex pattern recognition approaches have also been used in the context of individual identification. Neural networks have successfully identified individuals from their vocalizations in tungara frogs (Phelps and Ryan, 1998; Phelps and Ryan, 2000), fallow deer (Reby *et al.*, 1997; Reby *et al.*, 1998), Gunnison's prairie dogs (Placer and Slobodchikoff, 2000), and killer whales (Deecke *et al.*, 1999). Hidden Markov models (HMM) have been used to demonstrate the presence of vocal individuality in a few species including elephants (Clemins *et al.*, 2005) and song birds (Wilde and Menon, 2003; Trawicki *et al.*, 2005). Advantages of such pattern recognition approaches include the ability to incorporate more complex models of both spectral and temporal vocalization characteristics.

This paper investigates the extent of tiger vocal individuality through both qualitative and quantitative approaches, using the LDR taken from tigers at Omaha's Henry Doorly Zoo in Omaha, NE. Section II gives an overview of the data set used and describes the qualitative and quantitative measures, classification techniques, and the experimental design procedure. Section III presents the results, followed by a discussion in Sec. IV and final conclusions in Sec. V.

II. METHODS

A. Data set

LDR vocalizations were acquired from six tigers between November 2009 and March 2010 at Omaha's Henry Doorly Zoo. Representatives of the Amur, Bengal, and Malayan subspecies, *Panthera tigris altaica, Panthera tigris tigris,* and *Panthera tigris jacksoni*, respectively, were included in the study. Animals were housed individually within a single large indoor/outdoor complex, and animals were frequently rotated between indoor and outdoor exhibits. Each animal was recorded in multiple sessions in acoustically similar outdoor enclosures. Sessions generally lasted between 1 to 4 h and occurred between 7 a.m. and 2 a.m.

Vocalizations were recorded using an Earthworks QTC50 small-diaphragm omnidirectional condenser microphone (Earthworks Precision Audio, Milford, NH) that was spectrally flat between 3 Hz and 50 kHz (+/−1.5 dB). The microphone was fitted with an Earthworks OMW1 teardrop windscreen and interfaced to a Fusion high resolution digital audio recorder (Zaxcom, Inc., Pompton Plains, NJ) and data were acquired using 24 bits per sample at a sampling rate of 44.1 kHz. Prior to analysis, files were converted to 16 bits and parsed into segments that contained calls from a single individual, including contiguous segments of the background acoustical environment before and after each call for use in preprocessing. Calls containing artifact (e.g., noise from zoo visitors or vehicles) that overlapped with the tiger call of interest were excluded from analyses. Sound files were further parsed into single calls produced by a clearly identified individual, and then analyzed using both whole-vocalization measures and frame-based measures, as described in detail in Sec. II B 4. The ambient background noise varied within and across recording sessions, with an average signal-to-noise ratio (SNR) of 12.6 dB. SNR was directly calculated from the mean-squared energy of the signal segments, using neighboring silence regions to determine noise power as the mean-squared signal energy. Table I shows a profile of the data set, including a total of 306 calls included for analysis and classification, representing a total useable waveform time of 247 s and 16 476 individual acoustic frames.

B. Vocalization analysis and features

1. Preprocessing and signal enhancement

In order to improve signal quality and limit the impact of ambient background noise on the classification experiments, the recorded vocalizations were first processed using an Ephraim-Malah filter (Ephraim and Malah, 1985). The Ephraim-Malah filter is a well-established speech enhancement method commonly used for human speech. The method works in the frequency domain to estimate the maximum likelihood clean signal magnitude in each frequency bin,

TABLE I. Profile of vocalization data set.

Tiger ID no.	Sex	Subspecies	Number of calls used for analysis	Waveform time (s)	Number of analysis frames	Mean SNR (dB)
1	Male	Malayan	46	48	3196	17.5
2	Male	Amur	90	57	3796	9.3
3	Female	Bengal	16	24	1596	18.0
4	Female	Amur	49	37	2496	17.7
5	Female	Amur	14	21	1396	13.9
6	Female	Amur	91	60	3996	9.6
Total			306	247	16 476	

given the original signal and an estimate of the background noise from a neighboring silence region. Application of the Ephraim-Malah filter reduced and equalized background noise around the calls. Following the application of the preprocessing filter, the individual vocalizations were further segmented to remove silence regions prior to HMM classification.

2. Qualitative analysis

Vocalizations were qualitatively inspected and compared across individuals using traditional spectrogram analysis (Freeman, 2000; Hartwig, 2005), with emphasis on the fundamental frequency contour which is often a dominating feature for discriminating individuals across a single call type. The power spectrum of the central stationary portion of each vocalization was also calculated and plotted for qualitative comparison in the frequency domain. Beyond these basic approaches, statistical box plots and histograms were plotted for several of the whole-vocalization and frame-based features described in Sec. II B 3, as a method for qualitative interpretation of the differences across individuals.

3. Whole-vocalization measures

Four whole-vocalization measures were used to represent individual characteristics for this study: average duration, maximum f_0, minimum f_0, and average f_0. All of the measures were obtained using Praat (Boersma, 1993), a software application for acoustic signal analysis. Average duration was computed by directly averaging the durations of the manually-segmented calls for each individual. Fundamental frequency measures were extracted by applying pitch analysis using Praat software.

4. Frame-based spectral measures

Greenwood frequency cepstral coefficients (GFCCs) (Clemins and Johnson, 2006; Ren *et al.*, 2009) were used as the primary frame-based spectral features for the HMM-based classification experiments. GFCCs are a generalization of the Mel-frequency cepstral coefficient representation which is widely used in human speech processing and recognition, adapted to characterize a broader range of species. GFCCs are frequency-warped cepstral coefficients that represent the underlying spectral shape of each frame of data, calculated as the discrete cosine transform of log filter-bank energies taken from the signal's discrete Fourier transform, as illustrated in more detail in Fig. 1.

The GFCC feature representation has proven to be effective for many terrestrial and aquatic mammals' acoustic pattern classification and applications. The warping function is based on Greenwood's work in the mammalian auditory system (Greenwood, 1961), and requires species-specific parameters that can be determined from audiogram data if available or alternatively from approximate lower and upper extrema frequencies of the hearing range of the species under study, denoted $f_{\min}$ and $f_{\max}$. In this work, $f_{\min}$ and $f_{\max}$ are set to 50 and 5000 Hz, respectively, based on prior work in this species (Walsh *et al.*, 2008; Walsh *et al.*, 2011a). Because GFCCs have the capacity to incorporate a model of the species' perceptual auditory warping function, they are often an effective choice of spectral features for representing vocal characteristics.

The GFCC coefficients are computed and normalized by subtracting the mean value across the utterance. In addition to GFCCs, the normalized log energy in each frame is also used as a feature. This is a relative energy measure, taken as the log of the difference between the time-domain energy of each frame and that of the overall utterance. Following calculation of the GFCCs and log energy, the velocity (first derivative) and acceleration (second derivative) of the features are computed over a five-frame window and appended to create the final feature vector, as shown in Fig. 1.

In these experiments, features were extracted from vocalizations using a 25 ms moving Hamming window with 15 ms overlap. Twelve mean-normalized GFCC coefficients plus normalized log energy, along with velocity and acceleration, are computed for a total of 39 features per frame (Ren *et al.*, 2009). The programming toolkit used to implement feature extraction, as well as HMM training and testing, is the hidden Markov model toolkit from Cambridge University (Young *et al.*, 2006).

5. Classification and Voice Identification Methodology

a. Statistical and discriminative function analysis. Statistical analysis methods for this work include an analysis of variance (ANOVA) of the four whole-vocalization measures, as well as DFA. The F-test statistic and p-value from the ANOVA results were used to test for equality of means across the six individuals in the data set, and to identify which of the five measures were useful in discriminating across individuals. DFA was then performed, using all four measures taken together, as well as a subset of those measures indicating statistically significant differentiation with respect to individuals.

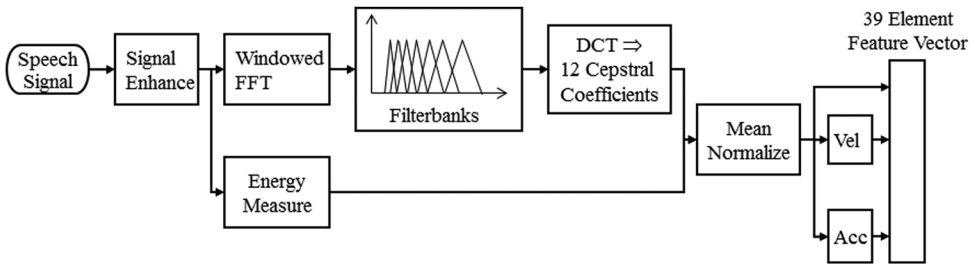

FIG. 1. Feature extraction in each frame, including front-end signal enhancement, cepstral coefficient and log energy calculation, mean normalization, and appended velocity and acceleration coefficients.

b. HMM. Temporal modeling and classification of the individual vocalizations was implemented using a HMM framework. The HMM is a statistical state machine model used in nearly all human speech processing and recognition studies (Juang, 1984). In recent years, the use of HMMs for animal vocalization classification in species such as elephants and dolphins (Roch *et al.*, 2007) has also achieved promising results. In essence, a HMM maps states in the model to a sequential pattern of acoustic observations, enabling calculation of a probabilistic match between the observation sequence and the underlying model. The individual statistical models within each state can be as simple as a single Gaussian distribution, represented by the mean and variance of frames that line up with that particular state. Typically more complex statistical models are used, such as Gaussian mixture models (GMMs), but the concept is the same. A simple illustration of this is shown in Fig. 2, using a sequence of three states where each state is modeled statistically as a Gaussian mixture model of the underlying features (Huang *et al.*, 2001).

Since each state corresponds to a single fixed statistical model, the number of states chosen for the HMM can be thought of as representing the number of different temporal segments in the vocalization type under study. The number of mixtures in the individual GMMs, then, can be thought of as representing the complexity of the spectral model used for each individual component. The primary limitation associated with increasing the number of states and number of mixtures used in a HMM is the amount of data available for estimation of model parameters.

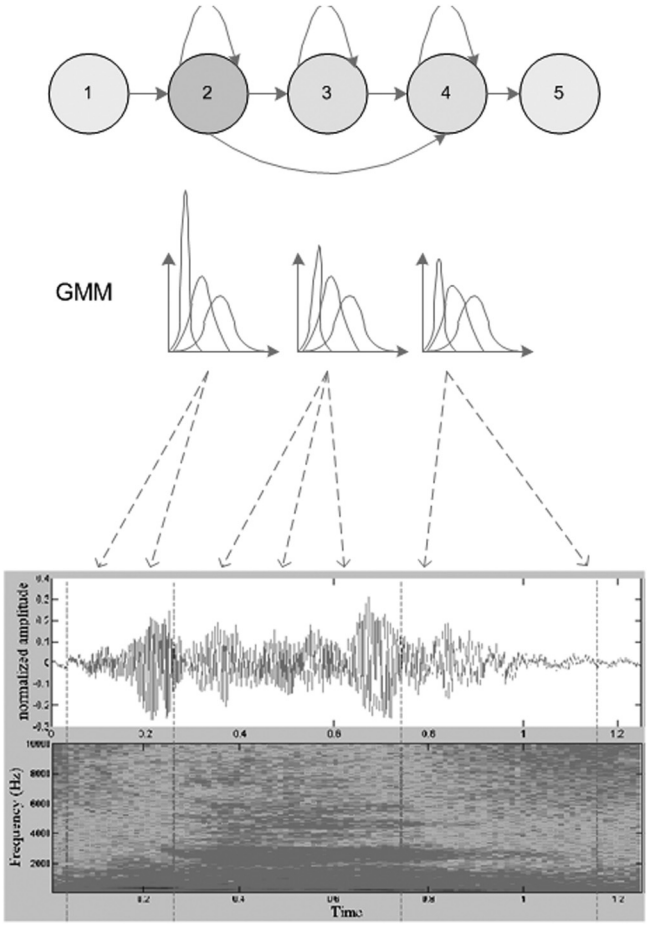

FIG. 2. Illustration of a HMM applied to vocalization modeling. HMM states align to the observed frame-based features using a maximum likelihood criterion, based on statistical transition and observation models.

c. Parameter variation and evaluation methodology. In addition to evaluating the presence and degree of vocal individuality within LDRs, one of the primary goals of this study was to investigate the extent to which that individuality was a function of differences in spectral characteristics versus differences in temporal characteristics. To accomplish this, the structure of the underlying HMM was systematically adjusted to vary the number of states, representative of temporal model complexity, and the number of mixtures in each state's GMM, representative of spectral model complexity. In the limit, a HMM with a single state is equivalent to a direct statistical classifier with a single GMM representing the overall average spectral characteristics without consideration of temporal pattern. Similarly, a HMM with multiple states but only a single Gaussian per state primarily focuses on the temporal pattern of the vocalization rather than fine details of the spectral characteristics. In the limiting case for both variables, a single-state single Gaussian HMM becomes a simple statistical classifier over average feature characteristics for the whole vocalization.

In the individual identification experiments implemented in this work, five-fold cross validation was used to split training and testing data. The full data set was split into five individual subsets, each containing one-fifth of the vocalizations from each individual, selected randomly. Classification was implemented five separate times, each time training on four subsets, or 80% of the data, and testing on the remaining subset, 20% of the data. This protocol ensures that all classification results were calculated from unseen data not used for training, while allowing for a larger training set size in each run.

The upper limit in terms of number of states or mixtures that can be reasonably considered is directly related to the amount of training data. In order to ensure that the model is sufficiently trained and results will be generalizable to new unseen data, a sufficient number of signal frames is required to estimate means and variances for each mixture in each state. If the number of parameters is increased beyond this point, the model will begin to overfit to the training data, and test set accuracy will begin to decrease. From Table I it can be seen that in this data set the minimum number of frames for an individual is about 1400, with approximately 80% used for training in any experimental run (depending on individual vocalization lengths, since cross-validation data splits were across files). Using the guideline that there should be at least ten examples for any parameter to be estimated, the total number of means to be estimated is equal to the number of states multiplied times the number of mixtures, and should not exceed approximately 100, e.g., 10 states with 10 mixtures each. In line with this, in these experiments the number of states and mixtures are each varied between 1 and 12, and the results shown in Sec. III confirm that the test set accuracy begins to drop once the number of parameters exceeds this level. Investigating the results separately as

spectral modeling and temporal modeling complexity increases is designed to lead to a better understanding of whether individual differences are due to overall spectral or temporal characteristics.

III. EXPERIMENTAL RESULTS

A. Qualitative visualization

1. Spectrograms and power spectra

The average fundamental frequency, f_0, of the LDR calls used in this study was approximately 150 Hz, and the bulk of broadcast energy was contained in a band of harmonically related frequency resonances ranging from about 100 to 800 Hz. Figure 3 illustrates examples of six individual tiger LDR waveforms with zoomed narrowband spectrograms. All of the calls exhibit harmonic structure and most energy is concentrated within lower frequencies. There are clear differences in temporal pattern across individuals, not only in terms of overall duration, but also in terms of energy contours. Spectrograms within the center portions of each call show clear differences in spectral structure.

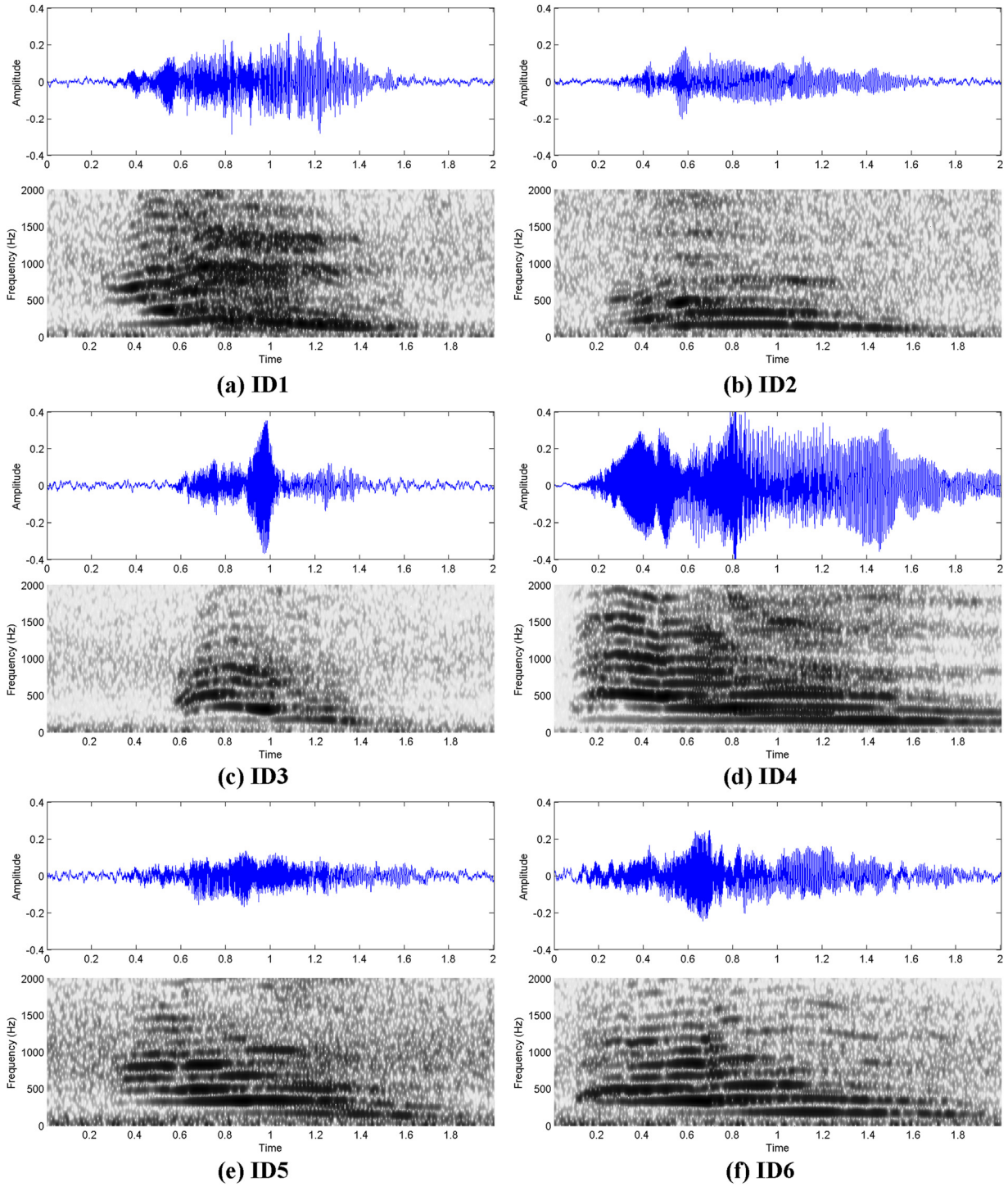

FIG. 3. (Color online) Comparative examples of LDR waveforms and zoomed narrow-band spectrograms for each individual in the study.

TABLE II. Whole vocalization measures across all six individuals.

Tiger ID	Average duration (s)	Maximum f_0 (Hz)	Minimum f_0 (Hz)	Average f_0 (Hz)
1	1.15	298	89	161
2	1.63	386	90	190
3	1.53	366	56	169
4	1.99	253	51	143
5	1.44	201	52	149
6	1.96	398	112	186

2. Whole-vocalization measures

Table II contains average whole-vocalization measures for the six individuals that were studied. The differences shown correspond well to the observations of the time-series and spectrograms in Fig. 3.

B. Statistical analysis

ANOVA analysis results are shown in Table III and Fig. 4. The boxplot results in Fig. 4 helps illustrate the degree of variation across individuals, leading to the corresponding F ratios and p values in Table III. Since high F ratios and small p values indicate highly significant differences across groups, this indicates that duration, min f_0, and average f_0, are all highly statistically significant, while max f_0 is not.

To combine these statistical measures, a MANOVA was implemented to consider duration, minimum f_0, and average f_0 as a set of variables. The MANOVA results for these three variables ($d=3$) were $p<0.0001$ for duration and min f_0, and $p=0.035$ for average f_0. These low p values indicate that there is strong statistical significance across vocalizers considering these three measures together.

To measure discriminative power, a DFA was implemented. For reference, the discriminative power of each individual measure taken separately ranges from a low of 23.0% (max f_0) to a high of 58.1% (duration). Applying DFA to all four parameters together results in a discrimination accuracy of 69.9%. Applying DFA to the three most significant parameters, duration, min f_0, and average f_0, the resulting accuracy drops slightly to 67.3%.

C. HMM classification

Results of the HMM classification system are shown in Table IV. In the upper left, the baseline accuracy of 71.1% for a single state and single mixture is only slightly higher than the simple discriminant analysis from Sec. III B. Focusing on the single-state case represented in the topmost row, it can be seen that increasing spectral modeling complexity without consideration of temporal pattern yields relatively little improvement in identification accuracy, increasing to a maximum of 75.5%. In contrast, the single-mixture case represented in the leftmost column illustrates that increasing the temporal resolution using a simple spectral model has more significant impact on identification accuracy, increasing to 80.5% as the number of states increases. This general pattern continues throughout the grid, improving slightly with increasing mixtures and more rapidly with increasing states, to a maximum of 90.5% at 11 (or 12) states and 10 mixtures.

TABLE III. ANOVA F statistics and p values showing discriminability for duration, maximum f_0, minimum f_0, and average f_0 measures.

ANOVA results	Duration	Maximum f_0 (Hz)	Minimum f_0 (Hz)	Average f_0 (Hz)
F ratio	58.3	0.46	13.2	15.2
p value	$p<0.0001$	0.81	$p<0.0001$	$p<0.0001$

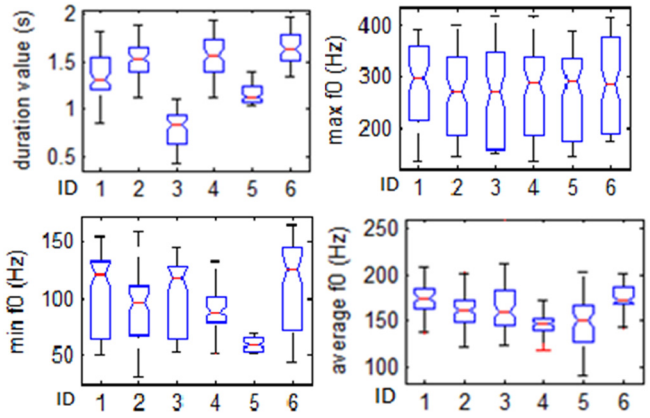

FIG. 4. (Color online) Boxplots for each of the four whole vocalization measures showing median, upper and lower quartile, and dynamic range.

To maximize generalizability to new test sets and ensure sufficient parameter training, the 10-state 10-mixture system was selected as the final classification system, even though it does not quite reach the highest level of accuracy. Past this point the model complexity and number of parameters has increased enough that there is a risk of some degree of overfitting, as discussed previously in Sec. II B 5 c. Overall, the accuracies are robust over a fairly wide range of parameters, exceeding 88% accuracy over the unseen test data in more than 25 different model cases in the bottom right of the chart, which suggests that the results are robust and likely generalizable in a broader context.

Examining the final system accuracy in more detail, Table V shows the corresponding confusion matrix for classification by a HMM with 10 states and 10 mixtures. In the confusion matrix, the row represents the actual ID of all test cases, while the column shows how these test cases were classified, with correct classifications on the diagonal. For visualization, the individuals are ordered to group them by error pattern, with individuals having mutually common identification errors highlighted in a block along the main diagonal. In this case it can be seen that confusions were highest between Tiger 2 and Tiger 6, representing 16 of the 36 total errors, and to a lesser extent between Tiger 4 and Tiger 5, representing 5 of the errors.

IV. DISCUSSION

Results of the ANOVA and MANOVA experiments indicated that duration, min f_0, and average f_0, were statistically significant factors in the vocal differentiation of

TABLE IV. Test set accuracy versus number of states and number of mixtures.

		Number of mixtures											
		1	2	3	4	5	6	7	8	9	10	11	12
Number of states	1	71.1	71.3	71.9	72.4	72.7	73.6	74.3	74.6	74.9	75.5	74.3	72.4
	2	73.0	73.3	75.3	76.4	76.1	77.0	77.6	77.3	79.0	79.6	77.1	75.3
	3	73.3	74.1	76.2	77.5	79.67	79.7	80.2	81.7	82.2	85.3	83.5	81.4
	4	74.3	74.4	77.2	78.7	77.3	79.8	79.5	81.0	82.4	85.7	84.1	83.3
	5	74.0	75.9	79.0	79.4	79.1	80.7	80.4	82.4	84.4	86.1	85.2	83.0
	6	74.7	76.3	77.2	79.5	80.7	80.3	82.5	85.8	86.1	88.0	85.7	85.3
	7	75.3	76.8	78.7	80.0	80.8	83.6	84.1	86.0	87.2	88.3	86.2	86.0
	8	77.7	79.1	79.3	80.0	81.0	81.3	85.0	86.7	88.4	89.4	88.8	86.3
	9	78.4	78.2	80.0	80.3	81.4	82.7	86.3	86.8	88.7	89.7	88.0	87.3
	10	78.5	79.1	80.3	80.6	81.2	81.7	85.4	86.8	89.3	90.2	89.7	89.3
	11	79.0	79.7	81.0	82.0	82.3	82.7	84.2	86.3	89.6	90.5	90.1	89.3
	12	80.5	81.7	81.3	82.1	84.6	85.3	88.7	89.1	90.4	90.5	90.3	89.7

individual tigers. The primary conclusion of the HMM studies is that the temporal patterns of the vocalizations are the single biggest factor in increasing individual discrimination accuracy. Incorporating both higher temporal resolution by increasing the number of states and more detailed spectral modeling by increasing the number of mixtures leads to a final accuracy of 90.2%. This corresponds to a relative error reduction of more than two-thirds, compared to the original DFA result of 69.9%. This conclusion also matches the findings associated with the whole-vocalization measures, in which the temporal characteristic of duration was the most discriminating of those measures. Several other studies have also found that that signal duration is a vocal parameter that cannot only be related to individuality, but may also encode contextual information such as emotion and stress (Janik *et al.*, 1994; Lengagne *et al.*, 1997). Examining the individual waveforms and spectrograms from Fig. 3, along with the final confusion matrix results of Table V, it is interesting to compare those individuals exhibiting the highest degree of confusion, Tiger 1 and Tiger 6, and note that the temporal similarities can be seen directly from the waveform examples.

TABLE V. Confusion matrix for the final system with 10 states and 10 mixtures. IDs are re-ordered to illustrate confusability between individuals. Confusions between Tigers 2 and 6 and Tigers 4 and 5 are highlighted, accounting for about 2/3 of all errors.

		Predicted ID					
		ID 2	ID 6	ID 4	ID 5	ID 3	ID 1
Actual ID	ID 2	83	6	0	0	0	1
	ID 6	7	83	0	0	0	1
	ID 4	1	0	45	3	0	0
	ID 5	0	1	2	11	0	0
	ID 3	0	2	0	1	12	1
	ID 1	1	2	0	1	0	42

V. CONCLUSION

In this paper, we have investigated the individuality of LDRs produced by tigers from both a qualitative and quantitative perspective and identified which vocal characteristics have the biggest impact in differentiating individuals within that context. Results from all the experiments clearly indicate the presence of vocal individuality for this call type, and suggest that the temporal pattern is the biggest factor in differentiation. The final identification accuracy of the system is 90.2% using a 10-mixture 10-state HMM with frame-based GFCC features.

ACKNOWLEDGMENTS

The authors would like to thank Adam Smith, Amy Larson, Ian Hoppe, Joseph Churilla, Alisa Theis, and staff of the Henry Doorly Zoo for their important contributions to the study. Work reported here was partially supported by the National Science Foundation Grant No. IOS 0823417.

Bauer, H. G., and Nagl, W. (**1992**). "Individual distinctiveness and possible function of song parts of short-toed treecreepers (*Certhia brachydactyla*). Evidence from multivariate song analysis," Ethology **91**, 108–121.

Boersma, P. (**1993**). "Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound," Proc. IFA **17**, 97–110.

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

Clemins, P. J., Johnson, M. T., Leong, K. M., and Savage, A. (**2005**). "Automatic classification and speaker identification of African elephant (*Loxodonta Africana*) vocalizations," J. Acoust. Soc. Am. **117**, 1–8.

Darden, S. K., Dablesteen, T., and Pedersen, S. B. (**2003**). "A potential tool for swift fox Vulpes velox Conservation: Individuality of Long-Range Barking Sequences," J. Mammal. **84**, 1417–1427.

Deecke, V. B., Ford, J. K. B., and Spong, P. (**1999**). "Quantifying complex patterns of bioacoustic variation: Use of a neural network to compare killer whale (*Orcinus orca*) dialects," J. Acoust. Soc. Am. **105**, 2499–2507.

Durbin, L. S. (**1998**). "Individuality in the whistle call of the Asiatic wild dog *Cuon Alpines*," Bioacoustics **9**, 197–206.

Eakle, W. L., Mannan, R. W., and Grubb, T. G. (**1989**). "Identification of individual breeding bald eagles by voice analysis," J. Wildl. Manage. **53**, 450–455.

Ephraim, Y., and Malah, D. (**1985**). "Speech enhancement using a minimum mean-square error log-spectral amplitude estimator," IEEE Trans. Acoust., Speech, Signal Process. **ASSP-33**, 443–445.
Freeman, P. L. (**2000**). "Identification of individual barred owls using spectrogram analysis and auditory cues," J. Raptor Res. **34**, 85–92.
Fristrup, K. M., and Watkins, W. A. (**1992**). Characterizing Acoustic Features of Marine Animal Sounds, Technical Report (Woods Hole Oceanographic Institution, Woods Hole, MA), pp. 84–126.
Gibert, G., McGregor, P. K., and Tyler, G. (**1994**). "Vocal individuality as a census tool, Practical Considerations Illustrated by a Study of two rare species," J. Field Ornithol. **65**, 335–348.
Greenwood, D. D. (**1961**). "Critical bandwidth and the frequency coordinates of the basilar membrane," J. Acoust. Soc. Am. **33**, 1340–1356.
Hartwig, S. (**2005**). "Individual acoustic identification as a non-invasive conservation tool: An approach to the conservation of the African wild dog *Lycaon Pictus*," Bioacoustics **15**(1), 35–50.
Hast, M. H. (**1989**). "The larynx of roaring and non-roaring cats," J. Anat. **163**, 117–121.
Huang, X., Acero, A. and Hon, H. (**2001**). "Hidden Markov models," in *Spoken Language Processing: A Guide to Theory, Algorithm, and System Development* (Prentice Hall, Upper Saddle River, NJ), pp. 377–414.
Janik, V. M., Dehnhardt, G., and Todt, D. (**1994**). "Signature whistle variations in a Bottlenose Dolphin, Tursiops Truncates." Behav. Ecol. Sociobiol. **35**, 243–248.
Jorgensen, D. D., and French, J. A. (**1998**). "Individuality but not stability in marmoset long calls," Ethology **104**, 729–742.
Juang, B. H. (**1984**). "On the hidden Markov model and dynamic time warping for speech recognition: A unified view," AT&T Tech. J. **63**, 1213–1243.
Klemuk, S. A., Riede, T., Walsh, E. J., and Titze, I. R. (**2011**). "Adapted to roar: Functional morphology of tiger and lion vocal folds," PLoS ONE **6**(11), 27–29.
Lengagne, T., Lauga, J., and Jouventin, P. (**1997**). "A method of independent time and frequency decomposition of bioacoustic signals: Inter-individual recognition in four species of penguins," C. R. Acad. Sci. III **320**, 885–891.
Leong, K. M., Ortolani, A., Burks, K. D., Mellen, J. D. M., and Savage, A. (**2002**). "Quantifying acoustic and temporal characteristics of vocalizations for a group of captive African elephants *Loxodonta Africana*," Bioacoustics **3**(13), 213–231.
McGregor, P. K. (**1993**). "Signaling in territorial systems: A context for individual identification, ranging and eavesdropping," Philos. Trans. R. Soc. London, Ser. B **340**, 237–244.
Peake, T. M., McGregor, P. K., Smith, K. W., Tyler, G., Gilbert, G., and Green, R. E. (**1998**). "Individuality in Corncrake Crex crex vocalizations," Int. J. Avian Sci. **140**(1), 120–127.
Peters, G. (**1978**). "Vergleichende Untersuchung zur lautgebung einiger Feliden (Mammalia, Felidae) [Comparative study of the vocalizations of some felids (Mammalia, Felidae)]," Spixiana Supplement **1**, 1–206.
Phelps, S. M., and Ryan, M. J. (**1998**). "Neural networks predict response biases of female tungara frogs," Proc. R. Soc. London, Ser. B **265**, 279–285.
Phelps, S. M., and Ryan, M. J. (**2000**). "History influences signal recognition: Neural network models of tungara frogs," Proc. R. Soc. London, Ser. B **267**, 1633–1639.
Placer, J., and Slobodchikoff, C. N. (**2000**). "A fuzzy-neural system for identification of species-specific alarm calls of Gunnison's prairie dogs," Behav. Processes **52**, 1–9.
Powell, A. N. W. (**1957**). *Call of the tiger, Technical Report*, Robert Hale Ltd., London, pp. 1–237.
Puglisi, L., and Adamo, C. (**2004**). "Discrimination of individual voices in male great bitterns (*Botaurus stellaris*) in Italy," Auk **121**, 541–547.
Reby, D., Joachim, J., Lauga, J., Lek, S., and Aulagnier, S. (**1998**). "Individuality in the groans of fallow deer (*Dama dama*) bucks," J. Zool. **245**, 78–84.
Reby, D., Lek, S., Dimopoulos, I., Joachim, J., Lauga, J., and Aulagnier, S. (**1997**). "Artificial neural networks as a classification method in the behavioural sciences." Behav. Processes **40**, 35–43.
Ren, Y., Johnson, M. T., Clemins, P. J., Darre, M., Stuart Glaeser, S., Osiejuk, T. S., and Out-Nyarko, E. (**2009**). "A framework for bioacoustic vocalization analysis using hidden Markov models," J. Algorithms **2**, 1410–1428.
Riede, T., and Zuberbuhler, K. (**2003**). "The relationship between acoustic structure and semantic information in Diana monkey alarm vocalization," J. Acoust. Soc. Am. **114**, 1132–1142.
Roch, M. A., Soldevilla, M. S., Burtenshaw, J. C., Henderson, E., and Hildebrand, J. A. (**2007**). "Gaussian mixture model classification of odontocetes in the Southern California Bight and the Gulf of California," J. Acoust. Soc. Am. **121**, 1737–1748.
Schaller, G. B. (**1967**). *The Deer and the Tiger* (University of Chicago Press, Chicago), pp. 1–384.
Suthers, R. A. (**1994**). "Variable asymmetry and resonance in the avian vocal tract: A structural basis for individually distinct vocalizations," J. Comp. Physiol., A **175**, 457–466.
Titze, I. R., Fitch, W. T., Hunter, E. J., Alipour, F., Montequin, D., Armstrong, D. L., McGee, J., and Walsh, E. J. (**2010**). "Vocal power and pressure-flow relations in excised tiger larynxes," J. Exp. Biol. **213**, 3866–3877.
Trawicki, M. B., Johnson, M. T., and Osiejuk, T. S. (**2005**). "Automatic song-type classification and speaker identification of the Norwegian Ortolan bunting,"in *IEEE International Conference on Machine Learning in Signal Processing (MLSP)*, Mystic, CT.
Walsh, E. J., Armstrong, D. L., and McGee, J. (**2011a**). "Comparative cat studies: Are tigers auditory specialists?" J. Acoust. Soc. Am. **129**(4), 2447.
Walsh, E. J., Armstrong, D. L., and McGee, J. (**2011b**). "Tiger bioacoustics: An overview of vocalization acoustics and hearing in *Panthera tigris*," in *3rd Symposium on Acoustic Communication by Animals*, Cornell University, Ithaca, NY.
Walsh, E. J., Armstrong, D. L., Napier, J., Simmons, L. G., Korte, M., and McGee, J. (**2008**). "Acoustic communication in *Panthera tigris*: A study of tiger vocalization and auditory receptivity revisited," J. Acoust. Soc. Am. **123**, 3507.
Walsh, E. J., Armstrong, D. L., Smith, A. B., and McGee, J. (**2010**). "The acoustic features of the long distance advertisement call produced by *Panthera tigris altaica*, the Amur (Siberian) tiger," J. Acoust. Soc. Am. **128**(4), 2485.
Weissengruber, G. E., Forstenpointner, G., Peter, G., Kubber-Heiss, A., and Fitch, W. T. (**2002**). "Hyoid apparatus and pharynx in the lion (*Panthera leo*), the jaguar (*Panthera onca*), the tiger (*Panthera tigris*), the cheetah (*Acinonyx jubatus*), and the domestic cat (*Felis silvestris f. catus*)," J. Anat. **201**, 195–209.
Wilde, M., and Menon, V. (**2003**). "Bird call recognition using hidden Markov models," Ph.D. dissertation, EECE Department, Tulane University, New Orleans, LA, pp. 35–59.
Young, S., Evermann, G., Gales, M., Hain, T., Kershaw, D., Liu, X., Moore, G., Odell, J., Ollason, D., Povey, D., Valtchev, V., and Woodland, P. (**2006**). *Hidden Markov Model Toolkit (HTK) Version 3.4 User's Guide* (Cambridge University Press, Cambridge, UK), pp. 1–384.