APRIL 18 2017

Automated detection of low-frequency rumbles of forest elephants: A critical tool for their conservation

Sara C. Keen; Yu Shiu; Peter H. Wrege; Elizabeth D. Rowland

J. Acoust. Soc. Am. 141, 2715–2726 (2017)
https://doi.org/10.1121/1.4979476

Automated detection of low-frequency rumbles of forest elephants: A critical tool for their conservation

Sara C. Keen,[a)] Yu Shiu, Peter H. Wrege, and Elizabeth D. Rowland
Bioacoustics Research Program, Cornell University, 159 Sapsucker Woods Road, Ithaca, New York 14850, USA

(Received 16 September 2016; revised 3 March 2017; accepted 16 March 2017; published online 18 April 2017)

African forest elephants (*Loxodonta cyclotis*) occupy large ranges in dense tropical forests and often use far-reaching vocal signals to coordinate social behavior. Elephant populations in Central Africa are in crisis, having declined by more than 60% in the last decade. Methods currently used to monitor these populations are expensive and time-intensive, though acoustic monitoring technology may offer an effective alternative if signals of interest can be efficiently extracted from the sound stream. This paper proposes an automated elephant call detection algorithm that was tested on nearly 4000 h of field recordings collected from five forest clearings in Central Africa, including sites both inside protected areas and in logging concessions. Recordings were obtained in different seasons, years, and under diverse weather conditions. The detector achieved an 83.2% true positive rate when the false positive rate is 5.5% (approximately 20 false positives per hour). These results suggest that this algorithm can enable analysis of long-term recording datasets or facilitate near-real-time monitoring of elephants in a wide range of settings and conditions.
 [http://dx.doi.org/10.1121/1.4979476]

[AMS] Pages: 2715–2726

I. INTRODUCTION

For species that occupy large ranges or live in habitats that prevent receiving information through visual cues, acoustic signaling is often a primary means of communication, e.g., whales coordinating movements across oceans (Payne and Webb, 1971; Tyack and Clark, 2000), birds singing to attract mates in thick forest habitat (Slabbekoorn *et al.*, 2002; Catchpole and Slater, 2003), and amphibians using long-ranging vocalizations to establish territories and attract mates (Littlejohn, 1977; Wells, 1988). African savannah elephants (*Loxodonta africana*), which live in highly fluid fission-fusion societies (Poole *et al.*, 1988; Poole 1994; McComb *et al.*, 2000; Wittemyer *et al.*, 2005; Archie *et al.*, 2006), have been shown to use infrasonic vocal signals, hereafter referred to as rumbles, to communicate and facilitate social interactions across large distances (Poole *et al.*, 1988; Langbauer *et al.*, 1991; McComb *et al.*, 2003). These low frequency vocal signals can carry farther than calls in higher frequency bands, such as trumpets or roars, and therefore may be more useful in communicating over large distances or through thick vegetation that could more easily dampen high frequency signals. Similar vocal signals are frequently used by African forest elephants (*Loxodonta cyclotis*), which live in the dense rain forests of Central Africa and occupy ranges of up to 2000 km^2 (Blake, 2003; Kolowski *et al.*, 2010). Forest elephants often exhibit coordinated social behaviors, such as synchronized gatherings near resources in forest clearings (Turkalo and Fay, 1995, 2001), which may be made possible by these infrasonic rumbles (Thompson *et al.*, 2010). Although vocal signals of savannah elephants have been studied extensively (e.g., McComb *et al.*, 2000; McComb *et al.*, 2003; Poole *et al.*, 2005), relatively little is known about the vocal behavior of forest elephants.

Forest elephant populations have declined dramatically in recent years, largely due to illegal poaching for ivory (Maisels *et al.*, 2013; Wittemyer *et al.*, 2014). There is a critical need to closely monitor key forest elephant populations and better understand their distribution and habitat use. However, forest elephants largely inhabit closed-canopy forest where direct observation is difficult and aerial census methods not applicable. Traditional methods for estimating population sizes of large mammal species, in particular line-transect methods, are highly cost and labor intensive, limiting their frequency of use and rendering them unsuitable for monitoring seasonal shifts in habitat use and behavior. Furthermore, forest elephants are known to alter behavior and spatial distribution in response to anthropogenic presence and activity (Blake *et al.*, 2008; Wrege *et al.*, 2010; Rabanal *et al.*, 2010; Wrege *et al.*, 2012), meaning that traditional survey methods could be disruptive and therefore are poorly suited for an endangered population. Passive acoustic monitoring is a highly suitable method of monitoring forest elephants, as it is non-invasive and less labor intensive and as such could facilitate more frequent tracking of population trends and behavior patterns than traditional methods (Thompson, 2009; Wrege *et al.*, 2012). This approach is also closely aligned with the principles guiding the burgeoning field of ecoacoustics (Sueur and Farina, 2015), a framework for investigating ecological processes through the analysis of environmental recordings.

Since 2000, the Elephant Listening Project (ELP) at Cornell University has been collecting sound data in Gabon,

[a)]Electronic mail: sck74@cornell.edu

Congo-Brazzaville, Cameroon, and Central African Republic. These data constitute the largest collection of wild forest elephant recordings (and indeed a wide spectrum of all forest sounds) that the authors know to currently exist, including over 700 000 h of recordings from 14 study sites spanning 7 years. While these data represent an invaluable ecological record of Central African rainforest habitats, manual analysis of recordings to find elephant vocalizations or other signals of interest is extremely time-intensive and costly, limiting the amount of sound that can be reviewed, the scope of questions that can be answered with the data, and increasing the time-frame from data collection to providing critical information to protected-area managers and conservation organizations tasked with preserving forest elephant populations.

As using acoustic surveys to conduct population monitoring becomes more feasible with advanced recording technology, numerous techniques have been developed to enable automated detection of animal vocalizations within a sound stream. Among the most commonly used methods is feature-based classification of individual frames of a recording to determine whether each time window is likely to contain a signal of interest (Fristrup and Watkins, 1992). In early bioacoustics studies, researchers often developed classifiers using spectro-temporal features that were initially developed for automatic speech recognition, such as mel-frequency cepstral coefficients (Davis and Mermelstein, 1980) and linear predictor coefficients (Hermansky, 1990), though more recently several feature sets have been developed for the purpose of detection and classification of animal vocalizations, such as the R packages seewave (Sueur *et al.*, 2008) and monitoR (Hafner and Katz, 2015), Acoustat sound measurements (Fristrup and Watkins, 1992), and Sound Analysis Pro song features (Tchernichovski *et al.*, 2000). These measurements are typically fed into a statistical classifier that has been trained with features extracted from exemplar signals, often using previously developed classifiers that are tuned to their dataset, e.g., artificial neural networks trained to detect northern Atlantic right whale vocalizations (Pourhomayoun *et al.*, 2013) or to classify the songs of passerine bird species (Fox *et al.*, 2008), using Gaussian mixture models to classify calls of toothed whales (Roch *et al.*, 2007), or employing support vector machines to classify bat echolocation calls (Redgwell *et al.*, 2009).

Fundamental frequency estimation of speech and music signals has been studied extensively; techniques based on autocorrelation (de Cheveigne and Kawahara, 2002; Boersma, 1993) and harmonicity (Klapuri, 2003) have yielded highly accurate results. Additionally, fundamental frequency estimation has been successfully applied in the analysis of animal sounds, e.g., to accurately trace the time-frequency contours of dolphin whistles in the presence of noise (Mallawaarachchi *et al.*, 2008) and to correctly identify individual grey wolves based on the fundamental frequency and amplitude of their howls (Root-Gutteridge *et al.*, 2014). Passive acoustic monitoring of forest elephants is necessarily different from previous analytical work on the problem of fundamental frequency estimation and harmonic characterization described above in that most of the time a fundamental frequency contour is expected and analysis is performed on short sound streams. Here, the technical challenge is acoustic object detection in the context of information retrieval, where target sound objects occur occasionally but not frequently, and therefore an approach based on feature extraction and classification is better suited for this problem.

Although other species have received more attention from bioacoustics researchers, recent significant progress has been made towards the goal of automated detection of elephant rumbles. Venter and Hanekom (2010) developed a novel method for detecting rumbles by estimating the fundamental frequency of adjacent windows in the sound stream and determining whether this value was constant over time. Another promising method proposed by Wijayakulasooriya (2011) uses linear predictive coding to measure formant frequencies in the sounds stream and then classifies each sound segment using a hidden Markov model. However, although the aforementioned methods are technically rigorous they have not been sufficiently tested on large datasets in which rumbles are rare. A third detection-classification method was put forth by Zeppelzauer *et al.* (2015), which uses several stages of spectrogram pre-processing before extracting cepstral features from the sound stream and then classifying each time frame using a support vector machine. This method appears to perform well in the presence of wind and rain noise and was tested more extensively than the former techniques, but was designed to detect rumbles from savannah elephants in a single reserve in South Africa. The algorithm presented here was designed to consistently detect forest elephant rumbles in dense rain forests under widely varying habitat conditions, across seasons, and over long time spans. It was tested extensively on recordings collected from numerous locations and under diverse conditions.

This paper describes a novel automated method to locate forest elephant rumbles within recordings. A central design goal was to create a generalizable detection algorithm that could be applied to a diverse array of tonal harmonic signals, while also offering increased processing efficiency and better performance than previous detectors. The detector described here can produce a greater than 300-fold increase in efficiency as compared to manually browsing sound data, exhibits comparable or better performance than previously published techniques, and has been demonstrated to perform consistently well in a variety of habitats and locations. The methods presented here are highly generalizable and could be applied to the detection of Asian or savannah elephant rumbles, as well as most low frequency, tonal, harmonic signals. This method may have the potential to radically change the manner in which elephant population surveys are conducted, and could also become the basis for a real-time monitoring system that could track elephant distribution and population size, and serve as an early warning system for altered behavioral patterns resulting from anthropogenic activity or other threats.

II. METHODS

A. Data collection

Over 90% of the recordings used in this study were collected from five sites in Gabon and Central African Republic between 2007 and 2012. Recording sites were each situated in forest clearings (also called "bais"), and included four "typical" bais varying in size from approximately 25 to 50 ha (Djidji, CEB1, CEB4, Dzanga), and one site near a large river (Kessala). Three of these sites (Djidji, CEB1, CEB4) were within logging concessions. A map depicting study site locations is shown in the supplementary material.[1] Audio files were recorded by autonomous recording units (ARUs) (Calupca *et al.,* 2000; Bioacoustics Research Program, 2014) which recorded audio signals at 16-bit resolution and sampling rates of 2000 or 4000 Hz. ARUs can detect elephant rumbles produced within a radius of approximately 0.8 km. Recordings were saved as time-stamped aif or wav files.

B. Acoustic characteristics of elephant rumbles

This study focuses on elephant rumbles, which are frequency-modulated vocal signals that lie in the frequency range of 8–180 Hz and usually last from 2 to 8 s. Rumbles have a distinct harmonic structure, with the fundamental frequency occurring in the infrasonic frequency band, typically between 8 and 34 Hz, median 15 Hz (Pool, 2011). Rumbles may be attenuated in recordings (i.e., have only a single formant visible in the spectrogram) either because they were produced far from the ARU and degraded while traveling over a large distance or because they were produced with low source amplitude (Thompson, 2009). For this reason, it is not possible to determine proximity of the elephant based on attenuation. When rumbles are not attenuated, typically three to eight harmonics are visible in the spectrogram [Fig. 1(a)].

C. Training and testing datasets

Two datasets were created for the purposes of training and testing the proposed rumble detection-classification system: A 522-h training dataset which was collected from eight study sites and contains 10 721 elephant rumbles, and a 3983-h testing dataset that consists of recordings from five sites and includes 110 042 rumbles. The testing and training datasets did not contain any of the same sound files, though both included recordings from three of the same study sites (CEB1, CEB4, and Djidji). The number of recording hours and study sites used for training and testing are summarized in Tables I and II. In both datasets all elephant rumbles with durations of at least 2 s were manually tagged by members of the ELP team, providing a beginning and ending time for each truth event. Spectrograms were generated with 1024-sample Hanning windows with a 200-point advance, as these parameters allow for favorable time-frequency resolution of rumbles. With information on the bounds of each truth event,

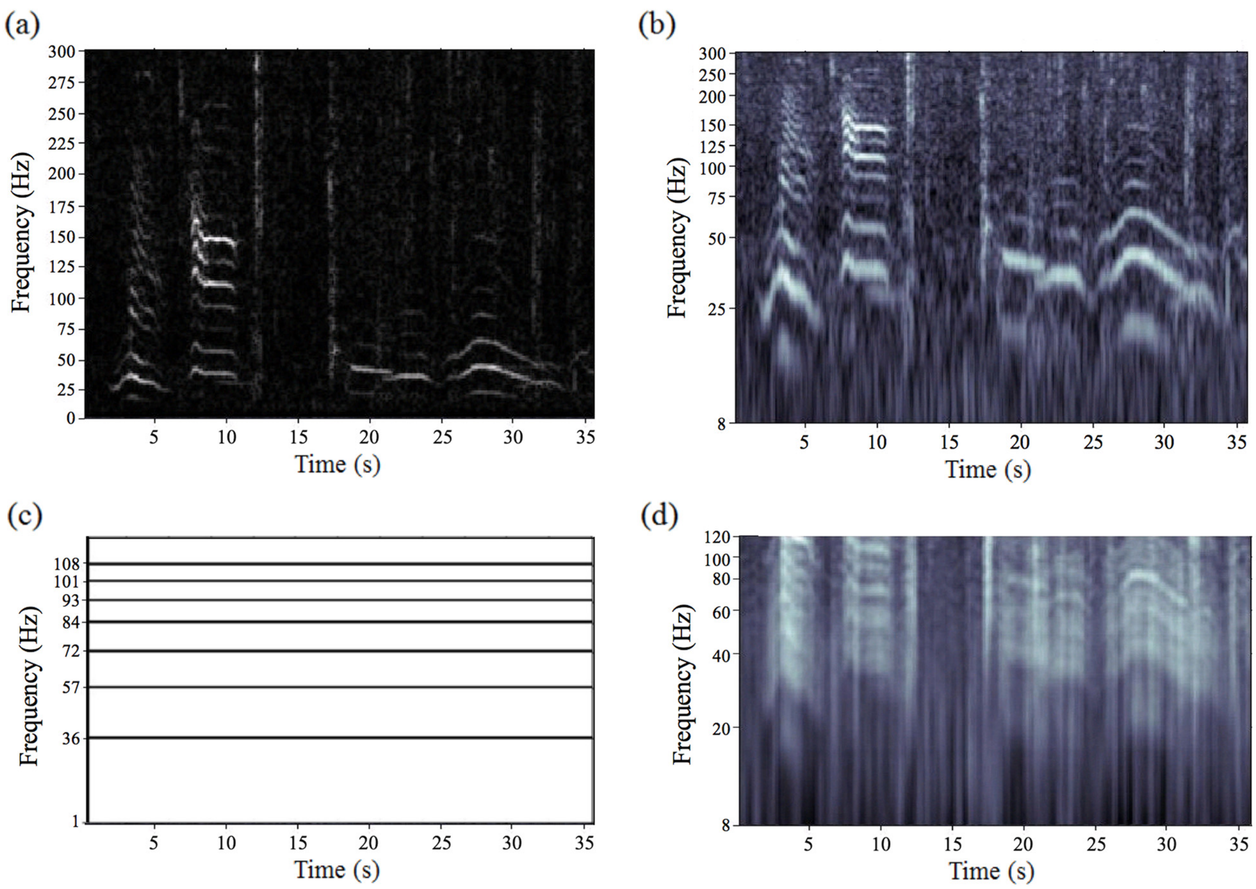

FIG. 1. (Color online) Elephant rumbles depicted in a standard spectrogram and the variants of spectrograms used during feature extraction. (a) Standard spectrogram containing elephant rumbles which illustrate the tonal, harmonic structure of rumbles, (b) log-frequency spectrogram of the same audio samples, (c) the harmonic template used to measure power at harmonic intervals, and (d) harmonic map calculated from the convolution of the log-frequency spectrogram with the harmonic template.

TABLE I. Summary of training dataset. Recordings from eight sites were selected in order to represent a diverse array of habitats and acoustic environments. The number of truth events that are attenuated to the point that only a single frequency band is visible in the spectrogram is also included.

Location	Days used	Hours used	Hours analyzed per day (mean ± SD)	Total rumbles in recordings	Number of attenuated rumbles
CEB1	9	180	20 ± 3	5348	2040
CEB2	4	90	22.5 ± 3	909	482
CEB3	1	24	24 ± 0	80	17
CEB4	1	18	18 ± 0	254	44
Djidji	2	48	24 ± 0	678	188
Djou	1	24	24 ± 0	291	39
Jobo	3	7	24 ± 0	2592	943
Langoue	2	48	24 ± 0	569	381
Total	24	522	21.75 ± 2.97	10 721	4134

each 100-ms time frame (see below) in the training and testing recordings to be labeled as either a positive or negative truth event (i.e., containing or not containing an elephant rumble). Time frames classified as negative truth events could include any sounds that were not rumbles, such as vocal signals from other species, vehicles, thunder, rain, etc.

D. Design of automated detection algorithm

An algorithm was created to automatically detect elephant rumbles in recordings by dividing the sound stream into 100-ms time frames, assigning each frame a score indicating the likelihood that it contains an elephant rumble. This process includes three main steps: pre-processing, feature extraction, and classification of frames, each of which is described below. All computation for was completed using MATLAB version 7.10 (The Mathworks, Inc., Natick, MA).

1. Pre-processing of sounds

For historical reasons the detector was optimized for sound files recorded at a sampling rate of 2 kHz. Recordings at higher sampling rates were down-sampled to 2 kHz. The short-time Fourier transform (STFT) was then calculated using 1024-sample Hanning windows with a 200-point overlap, resulting in a spectrogram with frequency resolution of 1.9531 Hz and a time resolution of 100 ms. The spectrogram representations of the audio windows were then split into 20-s blocks in order to reduce the amount of local memory used to calculate and store spectrograms and to increase processing efficiency. The spectrogram representation of each 20-s block of sound data was then converted to a logarithmic frequency scale to facilitate feature extraction; this representation is advantageous because the harmonics produced by different fundamental frequencies are at constant locations on a logarithmic scale. These 20-s long log-frequency spectrograms were calculated for $\Omega = 6$ octaves, spanning frequencies from $F_{\min} = 8$ Hz to an $F_{\min} \cdot 2^{\Omega} = 512$ Hz, and had $\beta = 36$ bins per octave. Note that in the lowest octaves adjacent frequency bins in the resulting log-frequency spectrogram will contain identical values as they are extracted from the same FFT cell, as the frequency resolution in the original spectrogram is 2 Hz. The conversion between the frequency f Hz in the linear frequency scale and the bin number b in the log-frequency scale was achieved using the following equations:

$$f = F_{\min} \cdot 2^{b/\beta},$$
$$\mathrm{b} = \beta \cdot \log_2(f/F_{\min})\,. \tag{1}$$

Although elephant rumbles exhibit high levels of variation, all rumbles share the common trait of having a quasi-stationary distribution of spectral energy, i.e., the distribution of power among different frequencies is stable for short periods of time throughout a rumble. For this reason, it was determined that a frame-based classifier that was based on the pattern of spectrum over a short period of time was best suited for this application, rather than a template or image-based classifier that examines temporal and spectral variation throughout an entire rumble. Therefore, a vector of spectro-temporal features was calculated for each 100-ms time frame and a classifier then estimated the probability of an elephant rumble being present during this window. Although adjacent time frames have high overlap, as specified by the spectrogram parameters, this technique enables detection in cases where fundamental frequency is stationary

TABLE II. Summary of testing dataset. These recordings were selected from the subset of sounds that were previously tagged by ELP researchers such that a variety of habitats and seasons would be represented. The mean ± SD of recording hours used from each study site was 818.2 ± 535.9. Due to the large size of the testing dataset it was not possible to count the number of rumbles that were attenuated.

Location	Days used	Hours used	Hours analyzed per day (mean ± SD)	Total rumbles in recordings
CEB1	96	1560	16.42 ± 7.54	20 162
CEB4	20	180	9 ± 5.3	4245
Djidji	44	1056	24 ± 0	8360
Dzanga	60	221	3.68 ± 2.17	41 602
Kessala	70	1074	15.34 ± 7.62	40 610
Total	290	3983	13.41 ± 6.75	110 255

for only a fraction of a signal, as is typical of the majority of elephant rumbles.

2. Feature selection

Two novel sets of feature measurements were developed for this application: (1) "harmonic features," which quantify the harmonic structure of the audio data within the time frame, and (2) "horizontal features," which measure the presence of energy distributed horizontally on the spectrogram, i.e., the amount of power present at specific frequencies. These features account for different ways in which information is encoded within forest elephant rumbles and also serve complimentary purposes: harmonic features search for characteristics of rumbles that are not attenuated whereas the horizontal features are better suited for finding attenuated rumbles. Together, the horizontal and harmonic feature measurements resulted in a feature vector of 46 elements for each time frame. The process of developing these measurements and extracting them from the sounds stream is described below.

a. Harmonic features. These feature measurements take advantage of the harmonic structure of forest elephant rumbles. Harmonic signals are characterized by high energy at the fundamental frequency F_0 as well as in harmonics: $2 \cdot F_0$, $3 \cdot F_0$, $4 \cdot F_0$, and so on. After replacing the linear-frequency scale with the log-frequency scale during preprocessing of the spectrogram, the harmonics become, by Eq. (1), $\beta \cdot (\log_2 F_0 - \log_2 F_{\min}$, $\log_2 F_0 + \log_2 2 - \log_2 F_{\min}$, $\log_2 F_0 + \log_2 3 - \log_2 F_{\min}, \ldots)$. Therefore, the frequency difference between F_0 and the harmonics, which was originally $(F_0, 2 \cdot F_0, 3 \cdot F_0, 4 \cdot F_0, 5 \cdot F_0, \ldots)$, will become $\beta \cdot (\log_2 2$, $\log_2 3$, $\log_2 4$, $\log_2 5, \ldots)$. An example log-frequency spectrogram is shown in Fig. 1(b).

Although the log-frequency axis is useful for display, converting to this representation also facilitates the measurement of harmonic features as follows. First, a time-frequency domain "harmonic map," $H_{\mathrm{map}}(b,t)$ with bin number b and frame number t, is generated by correlating the target log-frequency spectrogram with a binary mask (hereafter referred to as a "harmonic template") which has a value of 1 in bin numbers $\beta \cdot (\log_2 2; \log_2 3; \log_2 4; \ldots)$ and a value of 0 in all other bins. The harmonic template used for this study is shown in Fig. 1(c), and a harmonic map generated by correlating the above log-frequency spectrogram with this harmonic template is shown in Fig. 1(d). To determine the amount of power present across all harmonic intervals for a given time frame during this calculation, the correlation values are summed along the time axis. In addition, to reduce differences in amplitude among signals within the spectrogram, the log-frequency spectrogram $S(b,t)$ is raised to the power of $\tau = 0.25$ before being correlated with the binary mask. In short,

$$H_{\mathrm{map}}(b,t) = \sum_{d=1}^{D-1} M(d) \cdot S(b+d,\, t), \tag{2}$$

where M is the binary mask encoding the harmonic structure onto the fixed bin numbers $\beta \cdot (\log_2 2; \log_2 3; \log_2 4; \ldots)$ and D is the length of M's bin numbers.

For many elephant rumbles used in this study, relatively low energy was observed near F_0, as shown in Fig. 1(a), and typically signal energy was most prominent around 20–24 Hz, or $2 \cdot F_0$. This is likely due to a decrease in microphone sensitivity in low frequencies and may also result from forest elephants shifting energy out of the fundamental frequency of rumbles (Thompson, 2009). Using the harmonic map it was possible to estimate the fundamental frequency F_0 even with little or no recorded energy at the fundamental frequency. Five harmonic features were then extracted from each 100-ms time frame in the harmonic map:

(1) The value of the spectrogram bin containing the peak harmonic $H\mathrm{map}(b_{\mathrm{peak}}, t_i)$, where t_i is the current frame, and its corresponding bin number b_{peak}, which is the estimated bin number b_0 of the fundamental frequency F_0.
(2) The bin number, b_{peak}, which contains the peak harmonic in the current spectrogram frame
(3) Signal-to-noise ratio (SNR) of the estimated fundamental frequency F_0. The noise floor of the log-frequency spectrogram for each bin b was found by calculating the first percentile of power measured within the log-frequency spectrogram time frame, and the signal was defined as the power of $S(b_0, t_i)$.
(4) SNR of estimated first harmonic $2 \cdot F_0$, where the signal is the power of $S(b_0 + \beta \cdot \log_2 2, t_i)$.
(5) The number of harmonics, which was calculated as the sampling rate divided by b_{peak}.

b. Horizontal features. The "horizontal" features measure the degree to which each time frame in the log-frequency spectrogram contains power that is present at a constant frequency over time. This is calculated by convolving the log-frequency spectrogram with a set of image filters, each of which corresponds to a different pattern, scale, and direction. This is a variant of Leung-Malik (LM) filter bank (Leung and Malik, 2001), which has been used widely in object and texture recognition in computer vision (e.g., Koch and Brilakis, 2011; Ngan *et al.*, 2011).

An LM filter bank comprising eight image filters was applied to each 20-s log-frequency spectrogram. These filters include first- and second-order derivatives at one orientation on four scales, and are shown in Fig. 2. Four filters used on each spectrogram frame were designed to match horizontal "bars" in the spectrogram. The width of these bars was constant on the vertical axis because the spectrograms being filtered were on the log-frequency scale and therefore the bandwidth of elephant rumbles did not vary significantly. Four different lengths of horizontal bars were used in order to detect power in the focal frequency band that may only be present for a short time. Four image filters were also designed which search for signals of a particular duration by clipping the ends of the horizontal bars. These filters reward rumbles of the same duration as one of the four filters and penalize all others. For example, noise from the storage hard drive within the ARU typically appeared in spectrograms as a long horizontal line; such signals were penalized by these four filters

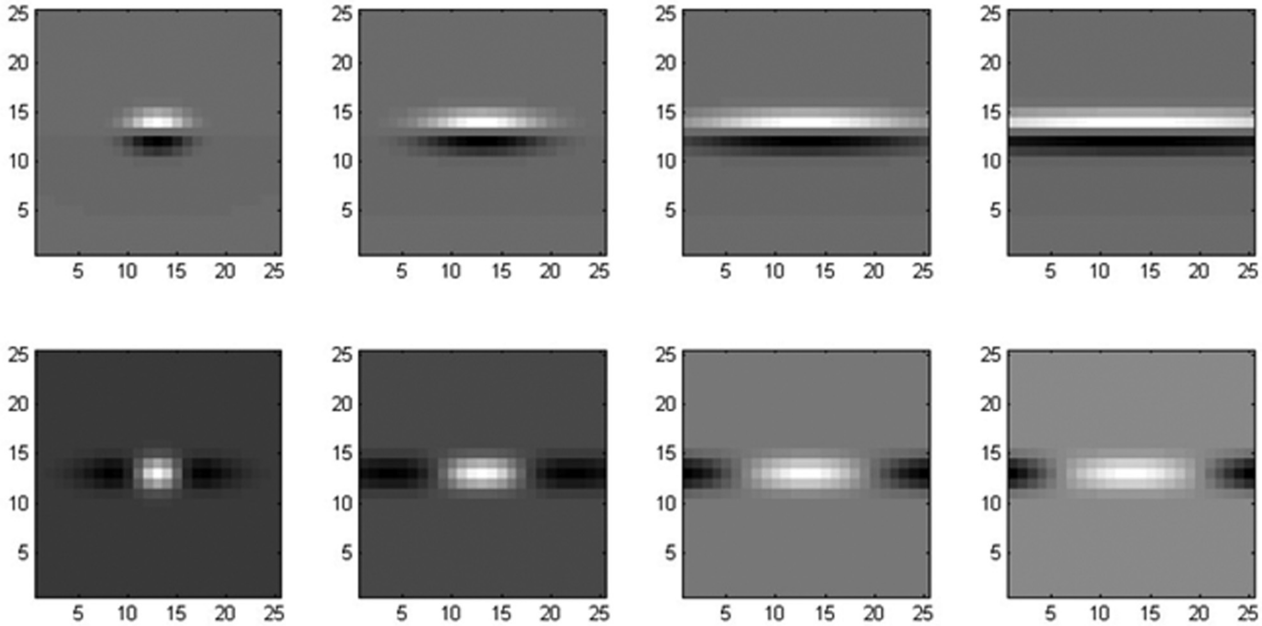

FIG. 2. Image filter bank used for the feature extraction. These filters are a subset of the commonly used Leung-Malik filter bank, which includes 48 filters in total. X and Y axis units are pixels; each 25 by 25 pixel image is convolved with a 2.5 s segment of the log-frequency spectrogram of the acoustic recording being analyzed. Lighter pixels represent areas where the log-frequency spectrogram was amplified and darker pixels represent where it was attenuated.

because the white area in the image filters amplifies values in the log-frequency spectrogram, and the black areas attenuate values. All filters are 25 by 25 pixel images, which corresponds to 2.5 s in the log-frequency spectrogram. Adjacent spectrogram frames that are classified as elephant rumbles are joined together at a later step, ensuring that rumbles longer than 2.5 s can be detected. Only a subset of the 48 filters used in the original LM filter bank design were used in order to reduce computation time and data size and to limit filter matches to signals that have power at fairly constant frequencies.

During this step the log-frequency spectrogram was limited to frequencies between 8−64 Hz, as either the fundamental frequency or a formant frequency is visible within this range in all rumbles regardless of degree of attenuation. This upper bound also avoided noise generated by some recorders due to data storage on a hard drive. Convolving the log-frequency spectrogram with the filter bank yielded 109 horizontal feature measurements ($=36 \times 3 + 1$) for every time frame. To reduce the number of feature measurements and generalize the feature space, an overlapping window was applied to group adjacent values in the feature vector, similar to the triangular overlapping windows used in mel-frequency cepstral coefficients (MFCCs). A maximum function was used to replace the triangular function. This resulted in a vector of 41 horizontal features derived for each 100-ms frame. Combined with the harmonic features, this yields a feature vector of 46 elements for every frame.

c. Feature set comparison. To assess the performance of the selected feature sets relative to other commonly used acoustic features, comparative tests were performed using several different suites of measurements. In addition to the horizontal and harmonic feature sets, linear predictor coefficients (hereafter LPCs) and MFCCs were also evaluated, as these could be considered a baseline to test against. Feature comparison was accomplished by creating several classification models using the Adaboost machine learning algorithm, a technique which combines the weighted output of several weak classifiers in order to create a more accurate prediction rule (Freund and Schapire, 1996). The models used here were created using the Matlab implementation by Kroon (2010) with default parameter settings, as this could easily provide a fair comparison between feature sets; selection of a classifier model for the final design occurred afterwards. Classification models were each trained with a different subset of features and the relative performance of each model was evaluated.

This classifier assigned each time frame to either class C_1 (i.e., a time frame that was thought to contain an elephant rumble), or class C_0 (i.e., a time frame that was thought not to contain an elephant rumble). To do this, the algorithm first calculated a score for every frame which represented the relative strength between classes C_1 and C_0 such as $P(x|C=1)/P(x|C=0)$, effectively indicating the likelihood that a given time frame contained or did not contain a rumble. The sequence of scores corresponding to all time frames in the sound data was then smoothed using overlapping Gaussian windows, as this is computationally efficient and is an effective means of edge detection in noisy data streams (Canny, 1986). A user-selected threshold between 0 and 1 was then applied. Sections of the smoothed series of time frame scores which were above the threshold and which included more than 2 s (i.e., more than twenty 100-ms time frames) above the threshold were returned by the classifier as "events." The mean score of the smoothed time frames and the start and stop time of the event in the sound data, (t_1, t_2), were reported to the user as the event score and time boundaries.

To compare feature sets, classifier models were both trained and tested using the training dataset as follows: a single day was first removed from this dataset and then a model was trained using features extracted from the remaining 23 days. The model was then tested on the excluded day by calculating feature vectors for every time frame within each day of recording and then comparing the time bounds of test events and truth events. This was repeated 24 times, each time leaving out a different day, providing an estimate of model performance across the entire 24 days. This was done for each of the four feature sets (harmonic features, horizontal features, MFCCs, and LPCs), as well as the combined harmonic and horizontal features.

d. Performance evaluation of feature sets. All classifiers were tested by comparing the time bounds (t_1, t_2), of "test events," i.e., selections marked as elephant rumbles by each classifier, to the time bounds of "truth events," i.e., those classified as elephant rumbles by human experts. Each truth event was classified as either a true positive (TP_{truth}) or false negative (FN). If at least 10% of a truth event was overlapped by a test event it was classified as TP_{truth}, otherwise is was marked as FN. Test events that overlapped with truth events by at least 10% were marked as $TP_{test,}$ and all others were marked as false positives (FP). This 10% threshold ensured that human experts visually inspecting test events returned by the algorithm could easily confirm whether detections indeed contained elephant rumbles. The following metrics were then calculated in order to evaluate classifier performance:

True positive rate (Recall): $TPR = TP_{truth}/(TP_{truth} + FN)$,
False positives per hour: $FP/h = FP/(\text{number of hours in the recording})$,
False positive rate: $FPR = FP/(FP + TN)$,
$\text{Precision} = TP_{test}/(TP_{test} + FP)$,
$\text{F1 score} = 2 \cdot (\text{precision} \cdot \text{recall})/(\text{precision} + \text{recall})$.

In order to calculate FPR it was necessary to determine the number of test events that were considered true negatives (TN). Because the classifiers were operating on continuous sound streams, rather than discrete events, TN was calculated as follows. The duration of all TP_{truth} and TP_{test} events was first calculated, ensuring that the duration of overlapping events was counted only once. The total duration of all FPs was also calculated. These totals were then subtracted from the total duration of the recording(s) on which the classifier was run, which produced the amount of time remaining in the recording which could be considered TN test events. To determine the number of TNs contained within this remaining time in the recording, this time was divided by the average duration of TP truth events that were found at the current classifier threshold. Using this approach, it was possible to plot TPR vs FPR, producing a standard region of convergence (ROC) curve that would be comparable to other published techniques, as well as a modified ROC curve, that compared TPR to false positive per hour (FP/h). When measuring these performance characteristics, a minimum detector threshold of 0.2 was used as values below this would yield an unmanageable number of false positive detections. To better understand how the number of total test events varies with the proportion of truth events detected, precision vs recall was also plotted for each threshold value. Additionally, the F1 score was calculated at each threshold, as it is a measure of detector accuracy that is based on both precision and recall, and therefore could indicate the optimal operating point for the detector.

The performance of each feature set is shown in Fig. 3. The feature set composed of both the horizontal and harmonic measurements performed the best among the models

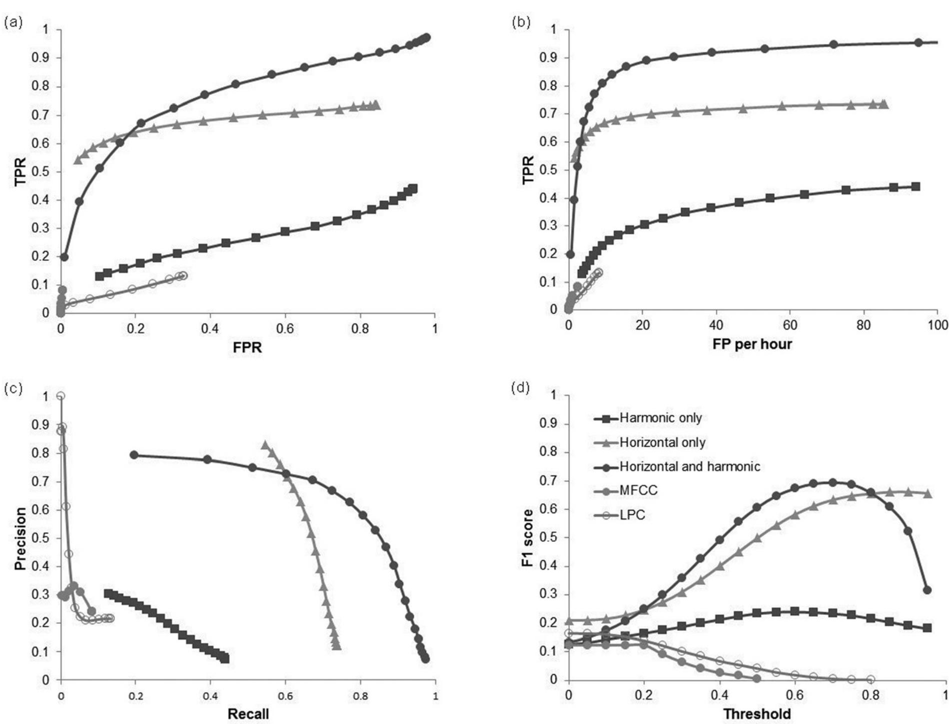

FIG. 3. Performance comparison of feature sets. (a) Traditional ROC curves for all features sets, (b) modified ROC curves which use FP/h, (c) precision vs recall, (d) F1 score vs threshold.

tested, achieving a significantly higher TPR than other feature sets and a maximum F1 score of 0.695, higher than any other suite of features. Although LPCs and MFCCs are typically considered to be highly effective in the detection of tonal signals such as human speech, the noisy state of the forest recordings and the high attenuation of rumble harmonics likely contributed to the poor performance of these feature sets, confirming that they were not appropriate for this study.

3. Binary classification

Three classifiers were considered in the design of this algorithm: linear support vector machines (SVM; Cortes and Vapnik, 1995), random forest (Breiman, 2001), and Adaboost. These methods were selected as the best candidates for this design as they are each able to be trained using a known dataset and applied to novel test data. However, these approaches are quite different from one another. Briefly, SVMs are discriminative classifiers well-suited to separating two classes of data, random forests are decision trees composed of a series of weak classifiers, and Adaboost is an adaptive boosting method that is similar to random forest, but with weighting applied to the most discriminatory features. To train a model for each of the candidate classifiers, feature vectors and their corresponding class labels were input into an SVM, random forest, and Adaboost classifier. Additional details about classifier training can be found in the appendix.[1]

Performance of all candidate classifiers was highly similar, with maximum F1 scores of the SVM, random forest, and Adaboost classifiers being 0.70, 0.75, and 0.73, respectively. Thus, these methods were viewed as interchangeable for this application. Ultimately, the Adaboost classifier was selected for feature vector classification because of its fast training and testing speed relative to the SVM and random forest.

4. Training and testing of detection-classification system

A final Adaboost classifier model was trained using the training dataset and default parameters. Performance was then evaluated by testing the model on the 3983-h testing dataset and calculating ROC curves, precision-recall curves, and F1 scores with classification thresholds of 0.2–0.95, as described above. Performance was evaluated individually for each of the five study sites included in the testing dataset. The proposed detector was also tested on the publicly available dataset presented in Zeppelzauer *et al.* (2015) in order to provide a relative comparison with the novel method proposed here.

III. RESULTS

The algorithm was tested using approximately 4000 h of recordings from five ELP study sites collected across six years. At a classification threshold of 0.4 (giving a reasonable limit to the number of FP/h—see below), the algorithm obtained a TPR between 79.6% and 89.35% across sites and a FPR between 0.75% and 8.66% [Fig. 4(a)]. At all study sites the detector reached a TPR of at least 90% when run at the lowest classification threshold of 0.2, and had FPRs of 2.1%–44.4%. Although the test dataset was composed in all cases of "new" recordings, algorithm performance was best at the sites that had been included in the training dataset. A more relevant performance metric in terms of analysis efficiency is the relationship between TPR and FP/h [Fig. 4(b)] because this models the number of output events that must be reviewed and the tradeoff between signals of interest and classification errors.

As would be expected, there was variation in precision (the number of returned test events that are elephant rumbles) and recall (the proportion of rumbles that are found by the algorithm) at different sites but overall performance was quite high [Fig. 4(c)]. At a very acceptable level of 80% recall, at least 70% of events returned by the algorithm were elephant rumbles and at some sites the precision was 90%. F1 score, a weighted average of precision and recall, is highest when a threshold of 0.6 or 0.65 is used, suggesting that these are optimal operating points. However, if increasing TPR is a priority, higher rates can be achieved with this detector-classifier at the expense of lower precision.

To calculate average processing time, the proposed algorithm was run on a Dell laptop with 8 cores and 16 G RAM as well as a high performance computer with 64 cores and 192 G RAM. On the laptop the detection algorithm processed 24 h of recordings in approximately nine minutes. Using the high performance computer, the algorithm processed 180 days of recordings in 2.5 h.

The proposed detector was also tested on the publicly available dataset presented in Zeppelzauer *et al.* (2015). The proposed algorithm yielded a recall of approximately 85% and precision of 88% [Figs. 4(e) and 4(f)].

IV. DISCUSSION

The automated detection algorithm described here for locating forest elephant vocalizations in field recordings represents a tool that enables efficient large scale monitoring of wild elephant populations. Although improvements are planned, the proposed algorithm nonetheless performed well when tested on an extremely large dataset (3983 h) that included recordings from diverse habitats, including two that were not part of the training data. With recall rates above 70%, automated detection is a reliable tool for monitoring elephant activity and represents a greater than 300-fold increase in analysis efficiency over locating elephant vocalizations by hand. This increase in efficiency and performance, particularly in diverse habitats, fulfills the goals laid out before creating the algorithm.

Elephant rumble vocalizations present a particular challenge for automatic detection because the salient features are strongly modified as distance to the recording device increases and as the source amplitude of the call varies. This means that a given rumble will show strong harmonic features at sufficiently close distances, but is reduced to a single relatively unmodulated tone farther away (termed "attenuated" below). Therefore, for realistic application of acoustic

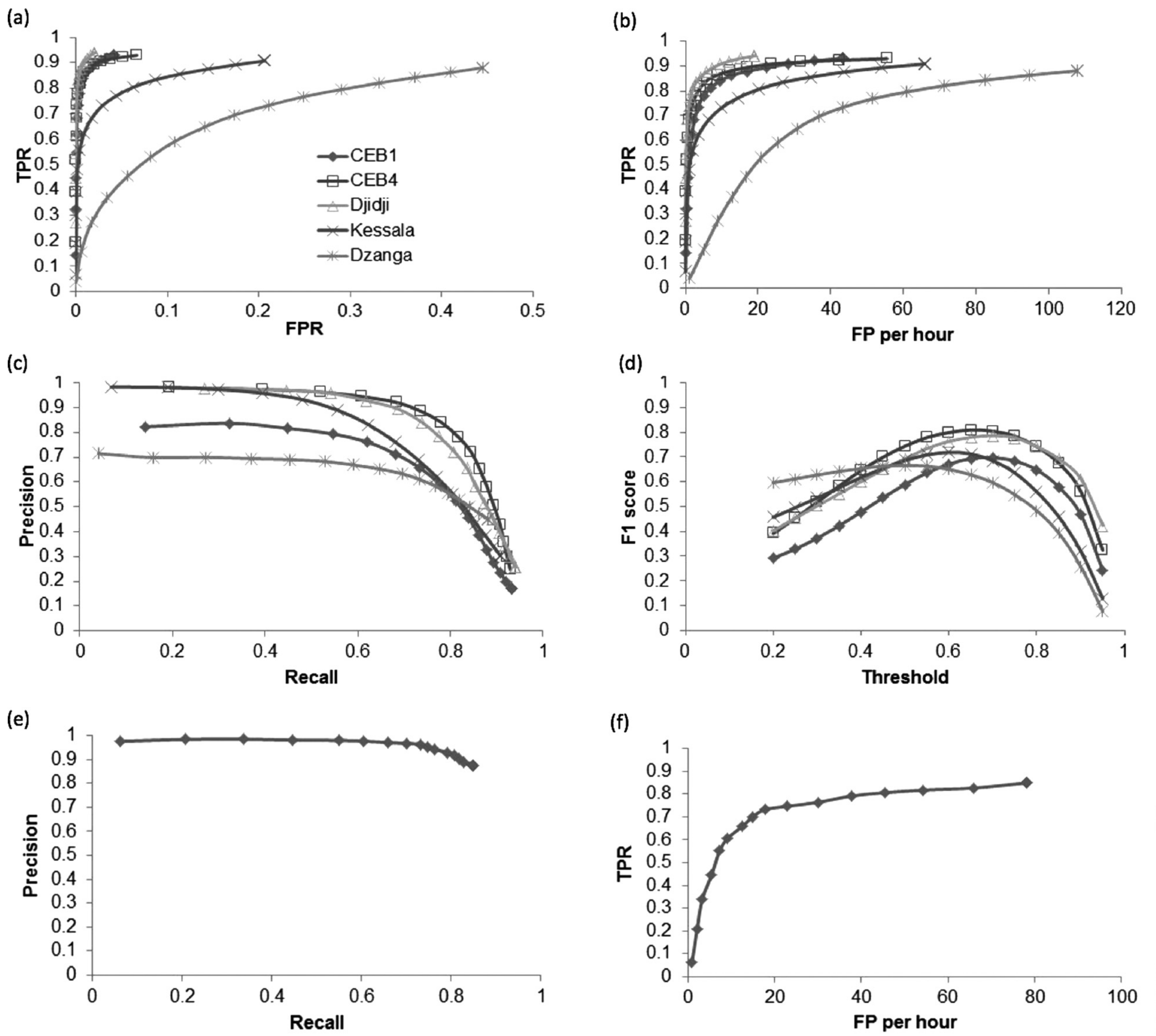

FIG. 4. Performance plots at classification thresholds 0.2–0.95. (a) Individual ROC curves for all study sites, (b) individual modified ROC curves which use FP/h for all study sites, (c) precision vs recall curves for all study sites, (d) F1 score vs threshold for all study sites, (e) precision-recall curve of proposed method when using the publicly available dataset provided by Zeppelzauer *et al*. (2015), (f) modified ROC curve using FP/h when using the Zeppelzauer *et al*. (2015) dataset.

monitoring where huge areas of habitat are under study, both the harmonic and horizontal classes of features must be used (see Fig. 3).

Even at lowest classification thresholds, the proposed algorithm did not detect all elephant rumbles that were tagged by human review. While this is nearly universal for automated detection, the characteristics of these false negatives are informative for improving the algorithm. In the test dataset false negatives were nearly all short rumbles (<3 s) represented in sound files only by the first harmonic (i.e., heavily attenuated). In addition, this persistent frequency band was often highly frequency modulated, rising or falling rapidly within the call. Although the algorithm training set included "attenuated" rumbles (approximately 40% of the dataset), it had not been optimized to classify rumbles with varying fundamental frequency.

The most common causes of false positive detections were low frequency, harmonic signals in which frequency varied little over time, i.e., signals that closely resembled elephant rumbles. Such signals were often caused by distant trucks or planes, thunder, and insects flying close to ARUs. Human experts could readily discriminate between engine noise and elephant rumbles because these signals typically continue for several minutes and can be easily noticed when viewing long time segments in the spectrograms. The algorithm could be modified by adding a post-processing step to reject stable, low frequency harmonic signals with durations much longer than typical elephant rumbles. Thunder produces broadband signals that may fool the detector because they contain power in the frequency bands tested by the harmonic template. Again, a post-processing step could be created to address these detections because there should be relatively lower energy *between* the frequency bands of the harmonic template.

The proposed detector performed remarkably well across the study sites tested but nonetheless performance

TABLE III. Comparison of the proposed detection algorithm and previously published methods. A threshold of 0.3 was used to test the proposed method. The FPR in this table (as reported in Zeppelzauer *et al.*, 2015) is calculated as FP/(FP + TP), or the proportion of all events returned that are incorrectly classified. The TPR and FPR values reported for the other published methods were obtained from Zeppelzauer *et al.* (2015). A fourth method, proposed by Wijayakulasooriya (2011) could not be include because this was presented as a proof of concept and was tested on only a few minutes of sound.

Detection method	TPR (%)	FPR (%)	Test dataset length (h)	Rumbles in test dataset
Proposed method	87.2	8.9	3983	110 042
Venter and Hanekom (2010)	85.7	14.2	4	28
Zeppelzauer *et al.* (2015)	88.2	13.7	6	635
Proposed method using Zeppelzauer (2015) dataset	87.3	9.2	6	635

was significantly lower at the two sites that were completely novel in the test dataset (i.e., no examples from Dzanga or Kessala were in the training data; see Fig. 3). These differences might simply reflect different forest environments where novel background sounds are introduced. The sites used for training were all forest clearings in central Gabon, sharing a broadly similar landscape, while Dzanga and Kessala are clearings surrounded by quite different forest. Dzanga in particular is extremely busy, with 60 or more elephants often present simultaneously (Turkalo *et al.*, 2013). In this more behaviorally charged atmosphere, rumbles may more often have features that were rare in the training data (e.g., high emotive states are reflected in less stable and higher fundamental frequencies; Soltis *et al.*, 2009). Regardless of cause, this site to site variation in performance points out the critical need to evaluate any detector not only in each new study location, but also across seasons–and such evaluation should be repeated at sites where monitoring continues for long time periods.

The proposed detection-classification algorithm performed well when tested on the only other publicly available dataset of elephant vocalizations (Zeppelzauer *et al.*, 2015). This is particularly notable because these calls are of savannah elephants recorded in a vastly different acoustic environment and of course not used in the training data. Table III shows a comparison of three proposed elephant rumble detectors, all trained and tested on different sets of examples, as well as testing the proposed algorithm on the dataset of Zeppelzauer *et al.* (2015). Performance of the proposed algorithm on the alternate dataset may in fact be better than presented in Table III; the algorithm found several "attenuated" rumbles in this alternate dataset which were not marked as truth events and so contributed to the FP count.

Table III also illustrates the disparity in the size of test datasets used for each method; having access to an extremely large dataset collected in diverse locations and conditions likely helped to create this extremely robust and accurate detection algorithm. In order to contribute to the research community and assist others in developing advanced elephant detection algorithms, the authors will make the test dataset available to any interested party. This publicly available dataset will include all test recordings and tables containing the time and frequency bounds of elephant rumbles identified by human experts.

A central goal in the design of this algorithm is to enable practical application in the monitoring of African forest elephant populations. The proposed algorithm is currently in use to monitor seasonal and temporal elephant activity in large tracts of forest and at attractive resources like forest clearings. By regularly retrieving sound files from recording units and applying the proposed algorithm, researchers and park managers can accurately track landscape use and relative elephant density over short time intervals. The short processing time and high accuracy of the detection algorithm could also allow for its use in a real time detection system that would alert farmers and local landowners of nearby elephant activity, potentially reducing human-elephant conflict.

V. CONCLUSIONS

Automated elephant call detection offers the possibility of analyzing large acoustic datasets in order to conduct population monitoring with lower cost and time investment than any method currently in use. Acoustic monitoring can give researchers insights into spatial and seasonal use of landscapes as well as monitor the effects of human activity on elephant behavior and distribution. The proposed method is a valuable contribution to all researchers interested in the detection of tonal low frequency signals in diverse environments and conditions, and the automated detection system described here could be easily implemented in the field or in a laboratory setting. The authors hope that this method may radically change the way in which forest elephant conservation is conducted and will help ensure the survival of elephant populations.

ACKNOWLEDGMENTS

The authors thank A. Turkalo and M. Thompson for their expertise and insight into acoustic monitoring of forest elephants in Central Africa, ELP volunteers for their help in tagging events in training and testing datasets, M. Zeppelzauer and A. Stoeger for sharing their acoustic data, and the Bioacoustics Research Program, and the Cornell lab of Ornithology for enabling this work. For permission to work in the Dzanga-Ndoki Park the authors thank the Dzanga Sangha Project, the Central African Republic Ministry of Water and Forest Resources. Permission to work in Gabon was generously granted by the Republic of Gabon Centre National de la Recherché Scientifique et Technologique. The Wildlife Conservation Society provided essential logistics support in the field. This study was supported by a generous donation from Scott and Karen Harder, and grants to P.H.W. from the U.S. Fish and Wildlife Service (96200-9-G231, 98210-7-G205, 96200-0-G301), the International Fund for Animal

Welfare, The Robert G. and Jane V. Engle Foundation, and private contributors.

[1]See supplementary material at http://dx.doi.org/10.1121/1.4979476 for Matlab implementation of proposed algorithm.

Archie, E. A., Morrison, T. A., Foley, C. A., Moss, C. J., and Alberts, S. C. (**2006**). "Dominance rank relationships among wild female African elephants, *Loxodonta africana*," Anim. Behav. **71**, 117–127.
Blake, S. (**2003**). "The ecology of forest elephant distribution and its implications for conservation," Doctoral dissertation, University of Edinburgh.
Blake, S., Deem, S., Strindberg, S., Maisels, F., Moment, L., Isia, I. B., Douglas-Hamilton, I., Karesh, W. B., and Kock, M. D. (**2008**). "Roadless wilderness area determines forest elephant movements in the Congo Basin," PLoS One **3**, e3546.
Bioacoustics Research Program. (**2014**). Raven Pro: Interactive sound analysis software (Version 1.5), Ithaca, NY: The Cornell Lab of Ornithology. Available from http://www.birds.cornell.edu/raven.
Boersma, P. (**1993**). "Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound," Proc. Inst. Phonet. Sci. **17**, 1993.
Breiman, L. (**2001**). "Random forests," Mach. Learn. **45**, 5–32.
Catchpole, C. K., and Slater, P. J. (**2003**). *Bird Song: Biological Themes and Variations* (Cambridge University Press, Cambridge, UK), pp. 1–248.
Calupca, T. A., Fristrup, K. M., and Clark, C. W. (**2000**). "A compact digital recording system for autonomous bioacoustic monitoring," J. Acoust. Soc. Am. **108**, 2582.
Canny, J. (**1986**). "A computational approach to edge detection," IEEE Trans. Pattern Anal. **PAMI-8**, 679–698.
Cortes, C., and Vapnik, V. (**1995**). "Support-vector networks," Mach. Learn. **20**, 273–297.
Davis, S. B., and Mermelstein, P. (**1980**). "Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences," IEEE Trans. Acoust. Speech **28**, 357–366.
de Cheveigné A., and Kawahara, H. (**2002**). "YIN, a fundamental frequency estimator for speech and music," J. Acoust. Soc. Am. **111**, 1917–1930.
Fox, E. J., Roberts, J. D., and Bennamoun, M. (**2008**). "Call-independent individual identification in birds," Bioacoustics **18**, 51–67.
Freund, Y., and Schapire, R. E. (**1996**). "Experiments with a new boosting algorithm," in *Proceedings of the Thirteenth International Conference of Machine Learning*, Vol. 96, pp. 148–156.
Fristrup, K. M., and Watkins, W. A. (**1992**). "Characterizing acoustic features of marine animal sounds," Woods Hole Oceanographic Institution Technical Report WHOI-92-04.
Hafner, S., and Katz, J. (**2015**). "monitoR: Acoustic template detection in R," R package version 1.0.4, http://www.uvm.edu/rsenr/vtcfwru/R/?Page=monitoR/monitoR.htm.
Hermansky, H. (**1990**). "Perceptual linear predictive (PLP) analysis of speech," J. Acoust. Soc. Am. **87**, 1738–1752.
Klapuri, A. P. (**2003**). "Multiple fundamental frequency estimation based on harmonicity and spectral smoothness," IEEE Trans. Acoust. Speech. **11**, 804–816.
Koch, C., and Brilakis, I. (**2011**). "Pothole detection in asphalt pavement images," Adv. Eng. Inform. **25**, 507–515.
Kolowski, J. M., Blake, S., Kock, M. D., Lee, M. E., Henderson, A., Honorez, A., and Alonso, A. (**2010**). "Movements of four forest elephants in an oil concession in Gabon, Central Africa," Afr. J. Ecol. **48**, 1134–1138.
Kroon, D. (**2010**). "Classic adaboost classifier," Department of Electrical Engineering Mathematics and Computer Science (EEMCS), University of Twente, Netherlands.
Langbauer, W. R., Payne, K. B., Charif, R. A., Rapaport, L., and Osborn, F. (**1991**). "African elephants respond to distant playbacks of low-frequency conspecific calls," J. Exp. Biol. **157**, 35–46.
Leung, T., and Malik, J. (**2001**). "Representing and recognizing the visual appearance of materials using three-dimensional textons," Int. J. Comp. Vision. **43**, 29–44.
Littlejohn, M. J. (**1977**). "Long-range acoustic communication in anurans: An integrated and evolutionary approach," in *The Reproductive Biology of Amphibians*, edited by D. H. Taylor and S. I. Guttman (Springer, New York), pp. 263–294.
Maisels, F., Strindberg, S., Blake, S.,Wittemyer, G., Hart, J., Williamson, E. A., Aba'a, R., Abitsi, G., Ambahe, R. D., Amsini, F., Bakabana, P. C., Hicks, T. C., Bayogo, R. E., Bechem, M., Beyers, R. L., Bezangoye, A. N., Boundja, P., Bout, N., Akou, M. E., Bene Bene, L., Fosso, B., Greengrass, E., Grossmann, F., Ikamba-Nkulu, C., Ilambu, O., Inogwabini, B.-I., Iyenguet, F., Kiminou, F., Kokangoye, M., Kujirakwinja, D., Latour, S., Liengola, I., Mackaya, Q., Madidi, J., Madzoke, B., Makoumbou, C., Malanda, G.-A., Malonga, R., Mbani, O., Mbendzo, V. A., Ambassa, E., Ekinde, A., Mihindou, Y., Morgan, B. J., Motsaba, P., Moukala, G., Mounguengui, A., Mowawa, B. S., Ndzai, C., Nixon, S., Nkumu, P., Nzolani, F., Pintea, L., Plumptre, A., Rainey, H., Bokoto de Semboli, B., Serckx, A., Stokes, E., Turkalo, A., Vanleeuwe, H., Vosper, A., and Warren, Y. (**2013**). "Devastating decline of forest elephants in Central Africa," PLoS One **8**, e59469.
Mallawaarachchi, A., Ong, S. H., Chitre, M., and Taylor, E. (**2008**). "Spectrogram denoising and automated extraction of the fundamental frequency variation of dolphin whistles," J. Acoust. Soc. Am. **124**, 1159–1170.
McComb, K., Moss, C., Sayialel, S., and Baker, L. (**2000**). "Unusually extensive networks of vocal recognition in African elephants," Anim. Behav. **59**, 1103–1109.
McComb, K., Reby, D., Baker, L., Moss, C., and Sayialel, S. (**2003**). "Long-distance communication of acoustic cues to social identity in African elephants," Anim. Behav. **65**, 317–329.
Ngan, H. Y., Pang, G. K., and Yung, N. H. (**2011**). "Automated fabric defect detection—A review," Image Vision Comput. **29**, 442–458.
Payne, R., and Webb, D. (**1971**). "Orientation by means of long range acoustic signaling in baleen whales," Ann. N.Y. Acad. Sci. **188**, 110–141.
Poole, J. H. (**1994**). "Sex differences in the behaviour of African elephants," in *The Differences between the Sexes*, edited by R. V. Short and E. Balaban (Cambridge University Press, Cambridge, UK), pp. 331–346.
Poole, J. H. (**2011**). "Behavioral contexts of elephant acoustic communication," in *The Amboseli Elephants: A Long-Term Perspective on a Long-Lived Animal*, edited by C. J. Moss, H. Croze, and P. C. Lee (University of Chicago Press, Chicago), pp. 125–161.
Poole, J. H., Payne, K., Langbauer, W. R., and Moss, C. J. (**1988**). "The social contexts of some very low-frequency calls of African elephants," Behav. Ecol. Sociobiol. **22**, 385–392.
Poole, J. H., Tyack, P. L., Stoeger-Horwath, A. S., and Watwood, S. (**2005**). "Animal behaviour: Elephants are capable of vocal learning," Nature **434**, 455–456.
Pourhomayoun, M., Dugan, P., Popescu, M., and Clark, C. (**2013**). "Bioacoustic signal classification based on continuous region processing, grid masking and artificial neural network," arXiv:1305.3635.
Rabanal, L. I., Kuehl, H. S., Mundry, R., Robbins, M. M., and Boesch, C. (**2010**). "Oil prospecting and its impact on large rainforest mammals in Loango National Park, Gabon," Biol. Cons. **143**, 1017–1024.
Redgwell, R. D., Szewczak, J. M., Jones, G., and Parsons, S. (**2009**). "Classification of echolocation calls from 14 species of bat by support vector machines and ensembles of neural networks," Algorithms **2**, 907–924.
Roch, M. A., Soldevilla, M. S., Burtenshaw, J. C., Henderson, E. 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.
Root-Gutteridge, H., Bencsik, M., Chebli, M., Gentle, L. K., Terrell-Nield, C., Bourit, A., and Yarnell, R. W. (**2014**). "Identifying individual wild Eastern grey wolves (*Canis lupus lycaon*) using fundamental frequency and amplitude of howls," Bioacoustics **23**, 55–66.
Slabbekoorn, H., Ellers, J., and Smith, T. B. (**2002**). "Birdsong and sound transmission: The benefits of reverberations," Condor **104**, 564–573.
Soltis, J., Leighty, K. A., Wesolek, C. M., and Savage, A. (**2009**). "The expression of affect in African elephant (*Loxodonta africana*) rumble vocalizations," J. Comp. Psychol. **123**, 222–225.
Sueur, J., Aubin, T., and Simonis, C. (**2008**). "Seewave: A free modular tool for sound analysis and synthesis," Bioacoustics **18**, 213–226.
Sueur, J., and Farina, A. (**2015**). "Ecoacoustics: The ecological investigation and interpretation of environmental sound," Biosemiotics **8**, 493–502.
Tchernichovski, O., Nottebohm, F., Ho, C. E., Bijan, P., and Mitra, P. P. (**2000**). "A procedure for an automated measurement of song similarity," Anim. Behav. **59**, 1167–1176.
Thompson, M. (**2009**). "African forest elephant (*Loxodonta africana cyclotis*) vocal behavior and its use in conservation," Ph.D. dissertation, Cornell University, Ithaca, New York.
Thompson, M. E., Schwager, S. J., Payne, K. B., and Turkalo, A. K. (**2010**). "Acoustic estimation of wildlife abundance: Methodology for vocal mammals in forested habitats," Afr. J. Ecol. **48**(3), 654–661.

Turkalo, A. K., and Fay, J. M. (**1995**). "Studying forest elephants by direct observation: Preliminary results from the Dzanga clearing, Central African Republic," Pachyderm **20**, 45–54.
Turkalo, A. K., and Fay, J. M. (**2001**). "Forest elephant behavior and ecology: Observations from the Dzanga saline," in *African Rain Forest Ecology and Conservation*, edited by W. Webber, L. J. T. White, A. Vedder, and L. Naughton-Treves (Yale University Press, New Haven, CT), pp. 207–213.
Turkalo, A. K., Wrege, P. H., and Wittemyer, G. (**2013**). "Long-term monitoring of Dzanga Bai forest elephants: Forest clearing use patterns," PLoS One **8**, e85154.
Tyack, P. L., and Clark, C. W. (**2000**). "Communication and acoustic behavior of dolphins and whales," in *Hearing by Whales and Dolphins*, edited by W. W. L. Au, A. N. Popper, and R. R. Fay (Springer, New York), pp. 156–224.
Venter, P. J., and Hanekom. J. J. (**2010**). "Automatic detection of African elephant (*Loxodonta africana*) infrasonic vocalisations from recordings," Biosyst. Eng. **106**, 286–294.
Wells, K. D. (**1988**). "The effect of social interactions on anuran vocal behavior," in *The Evolution of the Amphibian Auditory System*, edited by B. Fritzsch, M. J. Ryan, W. Wilczynski, T. E. Hetherington, and W. Walkawiak (Wiley, New York), pp. 433–454.
Wijayakulasooriya, J. V. (**2011**). "Automatic recognition of elephant infrasound calls using formant analysis and hidden Markov model," in *Proceedings of the International Conference on Industrial and Information Systems*, pp. 244–248.
Wittemyer, G., Douglas-Hamilton, I., and Getz, W. M. (**2005**). "The socioecology of elephants: Analysis of the processes creating multitiered social structures," Anim. Behav. **69**, 1357–1371.
Wittemyer, G., Northrup, J. M., Blanc, J., Douglas-Hamilton, I., Omondi, P., and Burnham, K. P. (**2014**). "Illegal killing for ivory drives global decline in African elephants," Proc. Natl. Acad. Sci. **111**, 13117–13121.
Wrege, P. H., Rowland, E. D., Bout, N., and Doukaga, M. (**2012**). "Opening a larger window onto forest elephant ecology," Afr. J. Ecol. **50**, 176–183.
Wrege, P. H., Rowland, E. D., Thompson, B. G., and Batruch, N. (**2010**). "Use of acoustic tools to reveal otherwise cryptic responses of forest elephants to oil exploration," Cons. Biol. **24**, 1578–1585.
Zeppelzauer, M., Hensman, S., and Stoeger, A. S. (**2015**). "Towards an automated acoustic detection system for free-ranging elephants," Bioacoustics **24**, 13–29.