Acoustic Detection of Elephant Presence in Noisy Environments

Matthias Zeppelzauer
Vienna University of Technology
Interactive Media Systems Group
Vienna, Austria
mzz@ims.tuwien.ac.at

Angela S. Stöger
University of Vienna
Department of Cognitive Biology
Vienna, Austria
angela.stoeger-horwath@univie.ac.at

Christian Breiteneder
Vienna University of Technology
Interactive Media Systems Group
Vienna, Austria
cb@ifs.tuwien.ac.at

ABSTRACT

The automated acoustic detection of elephants is an important factor in alleviating the human-elephant conflict in Asia and Africa. In this paper, we present a method for the automated detection of elephant presence and evaluate it on a large dataset of wildlife recordings. We introduce a novel technique for signal enhancement to improve the robustness of the detector in noisy situations. Experiments show that the proposed detector outperforms existing methods and that signal enhancement strongly improves the robustness to noise sources from the environment. The proposed method is a first step towards an automated detection system for elephant presence.

Categories and Subject Descriptors

[**H. Information Systems**]: H3. Information storage and retrieval—*H3.3 Information Search and Retrieval*

General Terms

Algorithms, Experimentation

Keywords

Audio retrieval, sound detection, feature extraction, sound enhancement

1. INTRODUCTION

The human-elephant conflict is a serious conservation problem in Africa and Asia. Due to the rising number of elephants and the increasing human population, the habitat of elephants becomes increasingly narrow. Due to the lack of habitat, elephants enter new territory, which often coincides with agricultural areas or human villages. The consequence is the involuntary confrontation of people and elephants, which claims the lives of many animals and humans every year [10]. Different efforts have been undertaken to alleviate this conflict, such as the establishment of electric fences, which is, however, not practicable to cover larger areas. Early warning systems are required that monitor travel routes of elephants and alert humans to avoid involuntary confrontations.

Elephants communicate with each other by low-frequency sounds, which travel distances of several kilometers. The most common elephant call is the *rumble*, which extends into the infrasound band. The rumble is a harmonic sound with a fundamental frequency in the range of 15-35Hz and a duration between 0.5 and 5s [12]. Figure 1 shows a typical rumble with a high signal-to-noise ratio (SNR). The acoustic detection of elephants by their calls is currently the most promising approach towards an early warning system that is able to detect the presence of elephants over large distances.

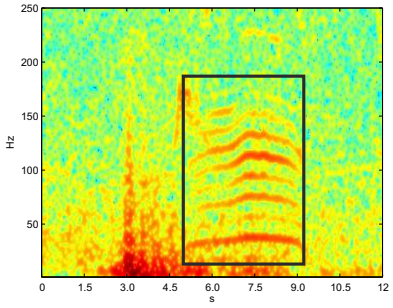


Figure 1: A typical elephant rumble.

Attempts towards acoustic detection (and localization) of elephants exist in literature [11, 4]. However, the large variety of noise sources present in the wild impede automated analysis methods. As a result, no system exists so far that is ready to operate in the field. So far research on acoustic analysis of elephant calls has addressed highly selective tasks, such as the identification of elephants by their calls [3] and the analysis of particular call types, e.g. rumble types [15]. The automated *detection* of elephant calls, which is the basis for the above mentioned tasks, has rarely been investigated.

There are two major challenges in the detection of elephants in wildlife recordings. The first challenge is the large variety of uncontrollable noise sources. Noise originates for example, from wind, rain, cars, and airplanes, which particularly pollute the low-frequency channel where elephant

MAED 2013, Barcelona Spain

http://dx.doi.org/10.1145/2509896.2509900.

calls reside. Additionally, human speech and sounds from other animals disturb the automated detection. The second challenge is the sparsity and irregularity of elephant calls, which makes it difficult to predict the occurrence of a call.

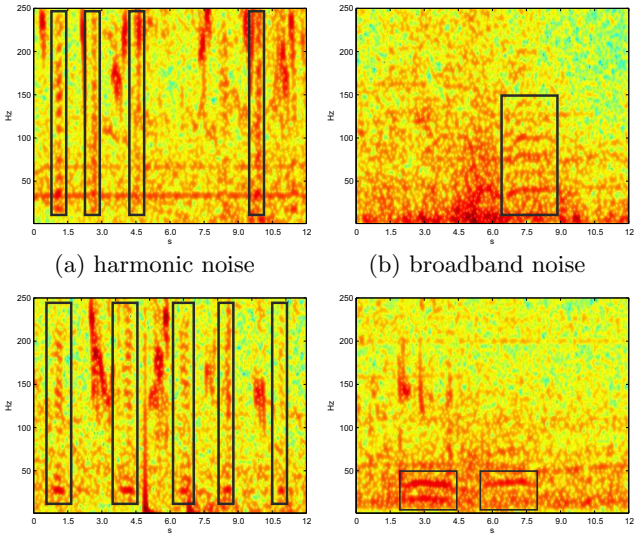
The contribution of this paper is a robust method for the detection of elephant presence. For this purpose, we employ an audio representation that models psychoacoustic properties of the elephant's hearing system. We introduce a novel method for the enhancement of signal quality to improve the noise robustness of the representation. The detector is evaluated on a large dataset of wildlife recordings to simulate a real-life scenario.

The paper is structured as follows. In Section 2 we review related approaches on elephant detection. Section 3 describes the acoustic detector and the proposed method for sound enhancement. The experimental setup and the evaluation of the proposed approach are presented in Section 4. Finally, we conclude our work in Section 5.

2. RELATED WORK

Sound detection has a long history [9]. There are two general approaches to sound detection: *template-based* methods and *feature-based* methods. Template-based methods successively match a given sound example (template) to a (longer) sound recording, in order to find occurrences of the template in the recording. A straight-forward approach is the matched filter method where two spectrograms are directly matched to each other. The method is optimal to find occurrences of the template itself in the recording, but suboptimal if similar signals to the template should be found or complex noise sources are present [13]. [9] propose the spectrogram correlation technique, which employs more abstract templates to make the matched filter approach more robust. The templates represent the coarse spectro-temporal energy distribution of the searched-for sound and improve the tolerance of the matching process. The spectrogram correlation technique has been applied to elephant call detection in [13]. However, results are reported to be suboptimal. One reason is that elephant calls vary significantly in duration and spectrogram correlation is not able to model variances in duration. Figure 2 shows the large variation in duration of rumbles (from 0.5s to 2.5s). A more promising template-based method for call detection has been recently introduced in [7]. The authors perform semi-supervised learning to select *the* sound snippet (template) that best discriminates between the positive and negative sound samples in a provided training set. For sound detection, the spectrograms of the template and of the recording are compared to each other using a distance measure that builds upon MPEG compression [2] and which allows for a certain tolerance in time and frequency. The approach has not been applied to elephant calls so far. We evaluate the approach on elephant call detection and compare it with the proposed approach in Section 4.

The second class of approaches for sound detection are feature-based techniques. The advantage of feature-based techniques over template-based techniques is the additional layer of abstraction introduced by the features, which provide a higher-level representation of the sounds. For the detection of elephant calls, feature-based techniques have been developed that build upon certain acoustic characteristics of elephant rumbles. [13] exploit the harmonic structure of rumbles and perform pitch detection using a sub-band



(a) harmonic noise (b) broadband noise

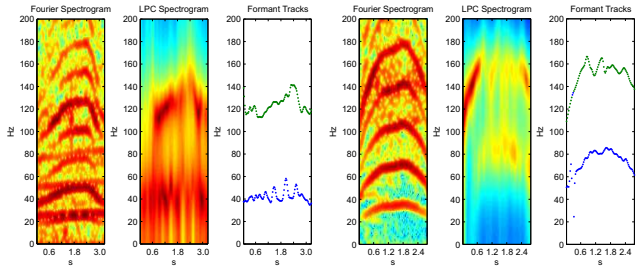
(c) short rumbles (d) missing harmonics

Figure 2: Rumbles with different interfering noise sources.

pitch analysis. The authors report good performance as long as the harmonic structure of the rumbles is not buried in background noise and at least three harmonics can be clearly distinguished. In practice, we observe that the harmonic structure of rumbles is often covered by noise, which is introduced by wind and other low-frequency disturbers like cars and airplanes. [13] report that engine noises lead to false positive detections if they have stronger harmonics than the rumbles. Figure 2(a) shows a rumble in the presence of narrow-band noise introduced by a car engine. The engine sound has a harmonic at 70Hz, which is particularly misleading for detectors that rely on harmonic structure. There are, however, additional factors that corrupt the harmonic structure. Figure 2(b) shows a rumble superimposed by broadband noise where the harmonic structure is hardly visible. The harmonic structure for short rumbles (see Figures 2(a) and 2(c) is less salient than for rumbles with a longer duration (see Figure 1). Additionally, the number of harmonics decreases with the distance of the caller to the microphone. Figure 2(d) shows two distant rumbles where the higher harmonics are completely missing, which impedes pitch detection as reported by [13].

Based on these observations, [14] proposes formant analysis for the detection of elephant rumbles. The formants are derived from the peaks of the transfer function of the all pole filter obtained by linear predictive coding (LPC). The basic assumption of the approach is that the first and the second formant are stationary during a rumble. This assumption does not hold in general as illustrated in Figure 3. Figure 3(a) shows the formant tracks of a rumble, which is superimposed by narrow-band harmonic noise. The resulting formant tracks show a high variation over time of approximately 20Hz. The second example in Figure 3(b) shows a clean rumble with a strong temporal modulation. The formant tracks reflect this modulation, which results in a variation of more than 30Hz.

From the existing investigations we conclude that a more holistic representation of the frequency distribution is re-



(a) modulation (b) noise

Figure 3: Formant tracks of elephant rumbles.

quired that does not rely on specific (and partially difficult to detect) sound attributes such as pitch, harmonics, and formants. A promising signal representation based on MFCC features has been proposed in [3]. The authors of [3] extend MFCCs by replacing the Mel-scaled filter bank with a more general Greenwood-scaled filter bank [6]. The Greenwood scale is a logarithmic scale that models the critical bands and that can be adapted to all mammals for which the corresponding hearing range is known (which is the case for elephants). Exchanging the Mel scale by the Greenwood scale in MFCC computation results in Greenwood Function Cepstral Coefficients (GFCCs), which provide a well-suited representation of elephant calls and have been successfully applied for call-specific speaker identification in [3]. The sounds used in [3] were recorded by microphones that are directly mounted on a collar worn by the elephant. Due to the low distance between sound source and microphone the sounds are expected to have a high signal-to-noise ratio.

In this paper, we investigate the *detection* of elephant calls in a real-life scenario. In this scenario a broad range of noise sources exist that decrease the signal-to-noise ratio. Additionally, the investigated calls originate from near and far distant elephants and from many different individuals. This results in a more complex setting for automated analysis than in [3]. We employ GFCCs as a basis for signal representation. Our experiments, however, show that additional processing steps are necessary to improve the noise robustness of the representation.

3. PROPOSED METHOD

The proposed approach comprises the following processing steps: First, the input signal is framed and transformed into frequency domain (see Section 3.1). Next, we perform signal enhancement by applying a spectro-temporal structure analysis to reduce the influence of noise (see Section 3.2). In a next step, we apply the species-specific Greenwood filter bank and transform the filter energies into cepstral coefficients (see Section 3.3). In the following, we aggregate the cepstral features of subsequent frames to obtain a more expressive and robust representation (see Section 3.4). Finally, a classifier is trained on a randomly chosen training set of a few positive and negative examples (see Section 3.5). Elephant call detection is performed by applying the trained classifier to unseen test data.

3.1 Preprocessing

In preprocessing we split the input signal into short audio frames and transform each framed signal into the Fourier domain by FFT. Since the energy of elephant rumbles is mostly concentrated below 500Hz we limit the analyzed frequency range to 0-500Hz. The analysis window is set to 300ms to capture the infrasound components with an adequate frequency resolution similarly to [3]. Temporal smoothness is obtained by a small step size between successive frames of 30ms.

3.2 Signal enhancement

Environmental sounds, such as wind and rain generate broadband noise which reduces the signal-to-noise ratio (see Figures 1 and 2). The background noise masks the fine harmonic structures of the rumbles and makes them hard to detect. Signal enhancement tries to emphasize spectro-temporal structures to facilitate their automated detection. Sounds like the rumbles generate spectral structures which extend in frequency as well as in temporal dimension. A pure intra-frame analysis as performed by most spectral features is not sufficient for signal enhancement since it is not able to exploit the temporal structure and relations of the sound of interest.

In a first step of signal enhancement, we group temporally adjacent spectral vectors and form a spectrogram. Important components that make up the spectro-temporal structure of a sound are *frequency contours* and *spectral peaks*. The detection of contours and peaks in a spectrogram is similar to the detection of edges and corners in images. A powerful method for the detection of such structures is the *structure tensor* which describes the image gradients and is frequently used for edge- and corner detection [8]. We apply the structure tensor to the spectrogram to enhance spectro-temporal structures and to increase the signal-to-noise ratio.

The structure tensor has been applied to spectral data in [1] for the detection of local feature points. In contrast to [1] we employ the structure tensor to generate a weighting filter that is applied to the entire spectrogram. Note, that in contrast to [1] this operation does not introduce additional detection thresholds.

The structure tensor is derived from the gradients of an image. In our case the input image is a logarithmized spectrogram S with elements $S(t, f)$ along time t and frequency f. For each element $S(t, f)$ in S we compute the gradients $\nabla_t(t, f)$ and $\nabla_f(t, f)$ from the partial derivatives along time and frequency as follows:

$$\nabla_t(t, f) = \frac{dS(t, f)}{dt} = S(t, f) - S(t + 1, f) \ ,$$

$$\nabla_f(t, f) = \frac{dS(t, f)}{df} = S(t, f) - S(t, f + 1) \ .$$

The tensor T at position (t, f) is constructed from the gradients and is defined as:

$$T(t, f) = \begin{pmatrix} \nabla_t(t, f)^2 & \nabla_{tf}(t, f) \\ \nabla_{tf}(t, f) & \nabla_f(t, f)^2 \end{pmatrix} ,$$

where $\nabla_{tf}(t, f) = \frac{dS(t,f)}{dtdf} = \nabla_t(t, f) \cdot \nabla_f(t, f)$. The tensor represents the local gradient structure for a particular position (t, f). Since the computation of the tensor depends only on neighboring elements from S, the tensor is prone to noise. To make the tensor more robust, the gradients are first smoothed along the time and frequency axis by a two-dimensional Gaussian filter of bandwidth b and duration d. The standard deviation of the filter is $\sqrt{bd}/4$. A

tensor that results from the smoothed gradients of a larger neighborhood represents larger and more salient structures.

The eigenvalues λ_1 and λ_2 of the tensor are well-suited indicators for the description of the local gradient structure. Since T is a symmetric matrix, the eigenvalues can be computed as follows:

$$\lambda_{1,2} = \frac{1}{2}\left(\left({\nabla_t}^2 + {\nabla_f}^2\right) \pm \sqrt{\left({\nabla_t}^2 - {\nabla_f}^2\right)^2 + 4{\nabla_{tf}}^2}\right).$$

The eigenvalues provide information about the local structure at a given position (t, f). If $\lambda_1 > \lambda_2$, then λ_1 represents the amount of variation along the gradient and λ_2 represents the amount of variance orthogonal to the gradient. If a perfect edge is found, $\lambda_2 = 0$ and $\lambda_1 > \lambda_2$. If both eigenvalues are equal, $\lambda_1 = \lambda_2$, the underlying structure is rotational symmetric. If both eigenvalues are zero the underlying structure is homogeneous.

From the eigenvalues we compute the *coherence* c which is a combined measure that provides the amount and type of structure at a given position. The coherence at a position (t, f) is defined as:

$$c(t, f) = \frac{\lambda_1(t, f) - \lambda_2(t, f)}{\lambda_1(t, f) + \lambda_2(t, f))} \ .$$

The coherence is 0 for completely isotropic structures, 1 for perfect edges, and undefined for homogeneous structures. Note, that the last case does usually not occur since spectrograms show hardly completely homogeneous areas in practice. Since the coherence quantifies structure, we employ the coherence as a weighting filter for the spectrogram. The enhanced spectrogram $\widehat{S}(t, f)$ is computed as: $\widehat{S}(t, f) = S(t, f) \cdot \kappa \cdot (c(t, f) + 1)$, where κ controls the strength of the weighting (structure amplification). The default value for κ is 1. In this case the largest possible weight is 2.

The effect of tensor filtering is shown in Figures 4 and 5. The figures show the input spectrograms in row 1, the corresponding coherence values in row 2 and the enhanced spectrogram in row 3. The figures show that the coherence gives larger weights to edge-like structures and lower weights to nearly homogeneous and isotropic structures. Figure 4 shows a rumble at 35s with background noise in its surrounding. The coherence is significantly higher in the area of the rumble due to the edge-like spectral contours. As a result the rumble is emphasized in the enhanced spectrogram. For the broadband noise at 4s (label A) the confidence is nearly zero. Consequently, the broadband noise is attenuated in the enhanced spectrogram. Other noise sources, such as the low-frequency spike at 30s (label B) are attenuated as well.

The example in Figure 5 (best viewed in color) shows a series of short rumbles from seconds 1.5 to 12. Again, the coherence yields the highest values for the rumbles while most noise sources receive lower confidence. The background noise level is reduced over the entire spectrogram and the structure of the rumbles is preserved well. Strong noise components, such as the one labeled C at 5s are attenuated. The particularly sharp noise contour at 10s (label D) remains due to its edge-like shape. The spectrogram additionally shows the sound of a car engine starting at 16s (label E) at a frequency of approximately 40Hz. Along the respective frequency contour higher coherence values are observed. However, compared to the input spectrogram, where the rumbles and the contour of the engine sound have similar energy, the difference in the respective energies increases strongly after signal enhancement.

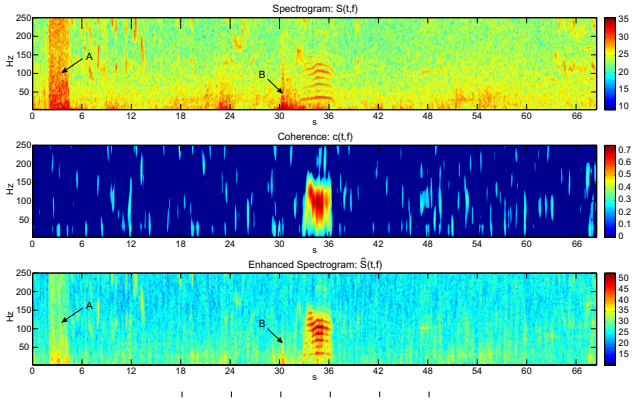


Figure 4: Spectrogram enhancement. The SNR in the recording is significantly enhanced.

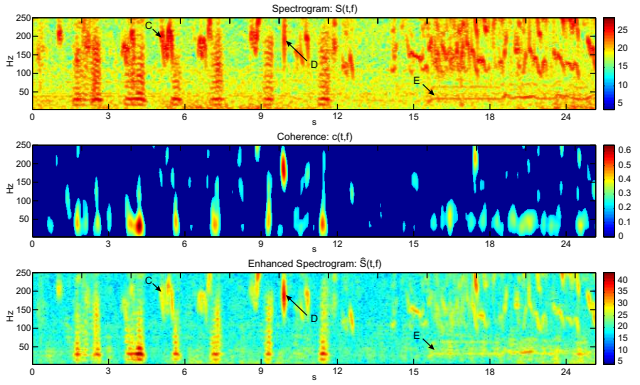


Figure 5: Spectrogram enhancement. Regions with rich structure, such as rumbles are enhanced.

3.3 Cepstral Feature Extraction

We apply a Greenwood-scaled filter bank to the enhanced spectrogram to account for the critical bands similarly to [3]. The Greenwood scale requires the definition of three species-specific parameters: the hearing range of the animal species (f_{min} and f_{max}) and a value k, which is assumed to be $k = 0.88$ for elephants [3]. The hearing range is set to $f_{min} =$ 10Hz and $f_{max} =$ 10000Hz according to [3]. The Greenwood filter bank consists of 30 bandpass filters and is scaled to the frequency range of 10Hz to 500Hz to account for the frequency range relevant for rumbles.

After the application of the filter bank, the filter energies are logarithmically scaled to compress their value range. Finally, a DCT is performed in a frame-wise manner to obtain cepstral coefficients for each audio frame. We select the first 18 cepstral coefficients as features to represent the coarse spectral shape of the audio frame.

3.4 Feature Aggregation

We aggregate the short-time cepstral coefficients to obtain a more robust and expressive representation. For this purpose successive audio frames (8 frames at each given position) with an overlap of 50% (4 frames) are grouped to-

gether. For each set of aggregated frames we compute the mean and variance of each cepstral coefficient over time.

3.5 Training and Detection

The detector is trained on the aggregated features. The preferred classifier is a linear support vector machine (SVM). The linear SVM is chosen, since it provides good results even for small training sets, has only a few parameters to specify, and exhibits a strong generalization ability due to its low-complex decision boundary. Additionally, the linear SVM has outperformed other classifiers (allowing more complex decision boundaries) in preliminary experiments. For detection, we apply the trained classifier on the test dataset, which has not been employed during training. The detection is performed for window sizes which equal that of the aggregated features.

4. EXPERIMENTS AND RESULTS

4.1 Dataset

The investigated data have been recorded at *Adventures with Elephants*[1] in Bela Bela, South Africa in 2011. The collected dataset comprises in total six hours of continuous wildlife recordings and contains numerous calls of African elephants (*Loxodonta africana*). To our knowledge, this is the most comprehensive dataset evaluated so far for elephant detection. [14] employ a dataset represented by only 2800 audio frames (which corresponds to approximately 420 seconds at a window size of 300ms and 50% overlap). [13] employ a dataset comprising four hours of recordings but with only 28 rumbles (with can be derived from Figure 4 in [13]). In contrast to this our dataset contains 635 rumbles which have different durations, fundamental frequencies, harmonic structure, and signal-to-noise ratios. All rumbles have been maually annotated by domain experts.

In practice, recording elephant rumbles for the training of the detector is a time-consuming and expensive task. The proposed method should be applicable in different sites without lengthy setup and training times. This means that the method has to learn from a few recorded samples only to adapt to a new site. We simulate this real-life requirement in the experiments by employing training sets with only a few positive examples.

We partition the dataset into three sets: a positive and negative training set and a test set. The positive training set contains 10 randomly selected rumbles. The negative training set is generated from the background sound between annotated rumbles randomly. The remaining dataset, which makes up 95% of the data, is used for testing the detector. We compute four such partitions of the data set randomly and independently from each other in order to reduce the dependence on the training set. We perform all experiments on these four partitions and average the results.

4.2 Evaluation

For evaluation of the method we compute three different performance measures: The *detection rate* D is the percentage of detected rumbles in all rumbles. A rumble is declared as detected if at least one automated detection intersects with a segment annotated as rumble. A segment that is hit several times is counted only once. The *false positive rate* FP is the percentage of falsely detected rumbles. Each automated detection that does not overlap with an annotated rumble is counted as one false positive detection. Finally, the *percentage of falsely classified frames* FCF is computed by dividing the number of false positives by the total number of audio frames that are input to detection. This measure provides the estimate of the total amount of background sound that is falsely classified.

[1]http://www.adventureswithelephants.co.za

4.3 Results

We compare the proposed approach with two baseline approaches: (i) the semi-supervised template-based approach by [7], short "B1" and (ii) call detection based on the GFCC features as proposed by [3], short "B2". Table 1 provides the performance measures of the baseline methods and the proposed approach (short "P").

The baseline method "B1" selects meaningful templates automatically from the training set in a semi-supervised way. Experiments show that the selected templates clearly capture parts of rumbles, which confirms that the template selection works well. The detection of calls by the templates, however, yields suboptimal results. While the detection rate is 71.5% the false positive rate of 62.1% is considerably high. About 3.5% of the background data is falsely classified. Rumbles have highly varying spectro-temporal characteristics, such as duration and fundamental frequency. The template-based matching is not able to take these variations into account. To be fair, it must be said that the method does not take any characteristics of the elephant species (except for the considered frequency range) into account.

Table 1: Average performance of the compared methods over all partitions of the dataset.

Abbr.	Method	D	FP	FCF
B1	Hao et al. [7]	71.5%	62.1%	3.5%
B2	Clemins et al. [3]	88.1%	43.4%	3.8%
B2a	"B2" with 0-500Hz	91.9%	45.0%	4.9%
P	Proposed method	91.0%	26.6%	1.8%
Pa	"P" with high-dim.	92.0%	25.8%	1.3%
Pb	"Pa" with feature sel.	91.8%	25.2%	1.2%

The second baseline ("B2") clearly outperforms the approach of [7], both in detection rate and in false positive rate. The percentage of falsely classified frames is similar. The method employs the GFCC features introduced by [3], which use a frequency range from 0 to 150Hz. We observe that rumbles frequently exceed this frequency range and thus we extend the upper frequency limit to 500Hz. This additionally enables an objective comparison with the proposed method.

The baseline method with the extended frequency range ("B2a") yields an increase of detection rate by 3.8% and a slight increase of false positive rate by 1.6%. For both variants of the second baseline, however, nearly every second detection represents a false detection.

The proposed approach ("P") yields a detection rate that is slightly lower than that of "B2a" (-0.9%). However, at the same time the false positive rate strongly decreases to only 26.6% (-18.4%) and the percentage of falsely classified frames drops to 1.8%. The spectro-temporal enhancement of the signal has a strong beneficial effect on the false positive rate and makes the method more robust to noise. The

overall performance (considering the tradeoff between detection rate and false positive rate) of the proposed method is strongly improved compared to the baseline approaches.

The Greenwood filter bank employed so far consists of 30 logarithmically spaced frequency bands. We observe that the filter bank considerably reduces the resolution of the spectrum. As a result many fine details (e.g. harmonics), which may be beneficial for the detection of rumbles, are lost. Additionally, the cepstral compression further removes fine details. In the following, we investigate the impact of this repeated data reduction in the GFCC features. For this purpose, we increase the number of bands to 50 and employ all cepstral coefficients as feature components. As a result the cepstral transform only decorrelates the filter energies but does not remove any information. The resulting variant "Pa" of the proposed method improves the overall performance as shown in Table 1. This demonstrates that fine spectral details are valuable for the detection of rumbles.

Especially when the training set is small, the higher-dimensional variant "Pa" may be prone to the curse-of-dimensionality. To reduce the dimensionality of the feature vectors, we apply feature selection prior to classification. For this purpose, we compute the Fisher criterion for each feature component on the training set [5]. Features with a high value for the criterion function separate the underlying classes well. We sort the feature components by their respective values and remove the third of the feature components with the lowest values. The results for the respective method "Pb" show that the feature dimension can be reduced without a loss in performance. We conclude that dimension reduction by feature selection is more efficient than the dimension reduction performed in the computation of GFCC features. The reason for this is that feature selection incorporates the training set and thus enables the proposed method to adaptively reject irrelevant or less relevant information. The dimension reduction in GFCC features, however, is blind to the training set.

We further investigate the false detections of the proposed approach. The investigation shows that aside from rumbles also other elephant calls are detected. Elephant trumpets, for example, exhibit a spectral structure similar to that of rumbles and generate "false detections". Since the presented study focuses solely on rumbles (because they are the most common call type), we have to count such detections as false detections to keep the evaluation objective. The integration of further call types into the detector remains future work. Other false detections are introduced by airplanes whose engines have a fundamental frequency similar to that of rumbles and which exhibit numerous strong harmonics with a similar spacing as the harmonics of rumbles.

5. CONCLUSIONS

We have presented a novel method for the detection of elephant presence in wildlife recordings. The presented evaluation is the most comprehensive study so far in this domain with regard to the amount of data and number of elephant calls. The novel spectro-temporal method for signal enhancement based on the structure tensor strongly improves the robustness of the detector in noisy situations. For the application of the proposed method as an early warning system in situ, the false positive rate is still too high. We currently investigate hierarchical classification and additional features to further reduce the amount of false detections.

6. ACKNOWLEDGMENTS

We thank "Adventures With Elephants" for their support. This work received financial support from the Austrian Science Fund (FWF) under grant number P23099.

7. REFERENCES

[1] R. Bardeli. Similarity search in animal sound databases. *IEEE Trans. on MM*, 11(1):68–76, 2009.

[2] B. Campana and E. Keogh. A compression-based distance measure for texture. *Statistical Analysis and Data Mining*, 3(6):381–398, 2010.

[3] P. J. Clemins, M. B. Trawicki, K. Adi, J. Tao, and M. T. Johnson. Generalized perceptual features for vocalization analysis across multiple species. In *Proc. of the IEEE Int. Conf. on Acoustics, Speech and Signal Proc.*, volume 1, pages 253–256, 2006.

[4] C. M. Dissanayake, R. Kotagiri, M. N. Halgamuge, B. Moran, and P. Farrell. Propagation constraints in elephant localization using an acoustic sensor network. In *6th IEEE Int. Conf. on Information and Automation for Sustainability*, pages 101–105, 2012.

[5] R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern classification*. Wiley, 2nd edition, 2001.

[6] D. Greenwood. Critical bandwidth and the frequency coordinates of the basilar membrane. *The Journal of the Acoustical Society of America*, 33:1344–1356, 1961.

[7] Y. Hao, B. Campana, and E. Keogh. Monitoring and mining animal sounds in visual space. *Journal of Insect Behavior*, 26(4):466–493, 2012.

[8] C. Harris and M. Stephens. A combined corner and edge detector. In *4th Alvey vision conference*, pages 147–151. Manchester, UK, 1988.

[9] D. K. Mellinger and C. W. Clark. Recognizing transient low-frequency whale sounds by spectrogram correlation. *The Journal of the Acoustical Society of America*, 107(6):3518–3529, 2000.

[10] C. Santiapillai, S. Wijeyamohan, G. Bandara, R. Athurupana, N. Dissanayake, and B. Read. An assessment of the human-elephant conflict in Sri Lanka. *Ceylon Journal of Science*, 39(1):21–33, 2010.

[11] L. Seneviratne, G. Rossel, H. L. Gunasekera, Y. Madanayake, and G. Doluweera. Elephant infrasound calls as a method for electronic elephant detection. In *Proc. of the Symp. on Human- Elephant Relationships and Conflicts*, pages 1–7, 2004.

[12] A. Stöger, G. Heilmann, M. Zeppelzauer, A. Ganswindt, S. Hensman, and B. Charlton. Visualizing sound emission of elephant vocalizations: Evidence for two rumble production types. *PloS one*, 7(11):e48907, 2012.

[13] P. J. Venter and J. J. Hanekom. Automatic detection of african elephant (loxodonta africana) infrasonic vocalisations from recordings. *Biosystems engineering*, 106(3):286–294, 2010.

[14] J. Wijayakulasooriya. Automatic recognition of elephant infrasound calls using formant analysis and hidden markov model. In *6th IEEE Int. Conf. on Industrial and Information Sys.*, pages 244–248, 2011.

[15] J. D. Wood, B. McCowan, W. Langbauer, J. Viljoen, and L. Hart. Classification of african elephant loxodonta africana rumbles using acoustic parameters and cluster analysis. *Bioacoustics*, 15(2):143–161, 2005.