Published in final edited form as:

. 2014 March 5; 23(3): 231–246. doi:10.1080/09524622.2014.888375.

Age group estimation in free-ranging African elephants based on acoustic cues of low-frequency rumbles

Angela S. Stoeger^{1,*}, Matthias Zeppelzauer^{2,*}, and Anton Baotic¹

¹Department of Cognitive Biology, University of Vienna, Vienna, Austria

²Institute for Software Technology and Interactive Systems, Vienna University of Technology, Vienna, Austria

Abstract

Animal vocal signals are increasingly used to monitor wildlife populations and to obtain estimates of species occurrence and abundance. In the future, acoustic monitoring should function not only to detect animals, but also to extract detailed information about populations by discriminating sexes, age groups, social or kin groups, and potentially individuals. Here we show that it is possible to estimate age groups of African elephants (*Loxodonta africana*) based on acoustic parameters extracted from rumbles recorded under field conditions in a National Park in South Africa. Statistical models reached up to 70 % correct classification to four age groups (infants, calves, juveniles, adults) and 95 % correct classification when categorising into two groups (infants/calves lumped into one group versus adults). The models revealed that parameters representing absolute frequency values have the most discriminative power. Comparable classification results were obtained by fully automated classification of rumbles by high-dimensional features that represent the entire spectral envelope, such as MFCC (75 % correct classification) and GFCC (74 % correct classification). The reported results and methods provide the scientific foundation for a future system that could potentially automatically estimate the demography of an acoustically monitored elephant group or population.

Keywords

Loxodonta africana; acoustic cues; age groups; acoustic monitoring

Introduction

Elephants (*Elephas maximus* and *Loxodonta* sp.) live in various habitats, from the desert to savannahs and dense forests. The need to control and monitor elephant populations is evident everywhere (most importantly in remote areas) because poaching (Blake et al. 2007), human disturbances, habitat loss and the resulting human-elephant conflict (Santiapillai et al. 2010; Lee and Graham 2006, Dublin and Hoare 2004; Sitati and Walpole 2006, Hedges and Gunaryadi 2010, O'Connell-Rodwell et al. 2000) pose serious threats to elephant populations worldwide.

*Corresponding authors: angela.stoeger-horwath@univie.ac.at, zeppelzauer@ims.tuwien.ac.at.

Acoustic recordings are an efficient way (apart from cost-intensive and invasive GPS and satellite tracking) to sample populations and to obtain reliable estimates of species occurrence and, potentially, abundance (Blumstein et al. 2011; Wrege et al. 2010). Today, technological advances including the development of autonomous and wireless recording devices increase flexibility and enable acoustic recordings in multiple locations over time (Blumstein et al. 2011). In the future, acoustic monitoring systems should not only automatically detect and discriminate between species in order to investigate biodiversity at different habitats and locations, but extract more detailed information about populations of particular species. This includes discriminating sexes, age groups, social or kin groups and, potentially, individuals (Blumenstein et al. 2011).

Elephants produce powerful sounds with fundamental frequencies in the infrasonic range (termed “rumbles”), which travel distances of up to several kilometres (Garstang 2004). Accordingly, elephants are ideally suited for acoustic monitoring even in dense forests. Payne et al. (2003) and Thomson et al. (2009a, b) showed that acoustic monitoring is a valuable tool for estimating African forest elephant (*Loxodonta cyclotis*) abundance. Here we provide evidence that it is further possible to discriminate age groups based on acoustic parameters extracted from elephant (*Loxodonta africana*) rumbles recorded under field conditions in a National Park in South Africa.

Vocalizations are physiologically constrained and therefore have characteristics directly related to intrinsic properties of the caller (Maynard-Smith and Harper 2003). There is strong evidence that elephants produce the low-frequency rumble via flow-induced self-sustained vocal fold vibration (Herbst et al. 2012), following the same physical principles of voice production as most mammals, including humans (Titze 1994). Frequencies produced via this myoelastic-aerodynamic mode are tightly limited by the physical size of the oscillators, such as the vocal folds (Titze 1994). In mammals, there is a direct interspecific relationship between body mass, vocal fold size, and the fundamental frequency. Within a given species it has been shown that body size is related to fundamental frequency across age categories and among adult females (Morton 1977; Titze 1994; August and Anderson 1987; Pfefferle and Fischer 2006; Fischer et al. 2002; Collins and Missing 2003). In fact, the fundamental frequency of infant elephant rumbles is not in the infrasonic range, and thus considerably higher than the average fundamental frequency of adult rumbles (Stoeger-Horwath et al. 2007; Wesolek et al. 2009; Poole 2011). We therefore expect that fundamental frequency and/or the frequency of higher harmonics of elephant rumbles can be used to discriminate between age groups.

This paper identifies acoustic parameters for the discrimination of age groups and investigates the automated classification of calls into predefined age categories. We describe the variability of temporal and spectral features within age groups, the overlap of parameters between age groups, and the overall age group classification success of rumbles recorded of a free-ranging African elephant population. A future elephant monitoring system requires identifying the most characteristic parameters in order to develop reliable and robust automatic feature extraction methods. We investigate the parameters identified from the statistical analysis in the context of automated acoustic classification. Furthermore, we

evaluate the suitability of different content-based audio features for the prediction of age groups.

Materials and method

Study population

Recordings were collected at the Addo Elephant National Park, (Addo: 33°30'S, 25°45'E), Eastern Cap Province, South Africa, at the 'Main Camp' section and the 'Colchester section', which were conjoined in 2010. In 2008, the elephant population numbered 481 individuals with an annual rate of increase of 5.81 % (between 1976-2002, Whitehouse et al. 2008), split into 7 family groups, the A, B (subdivided into two subgroups Ba and Bb, H, L, M, P and the R family. Within each group (except for the L-group), one adult female wears a GPS collar, sending a GPS signal twice a day.

Data collection and annotation

Data were collected during June and July 2011 and July and August 2012, which resulted in 101.4 h of recordings. We used a vehicle with two observers to find and follow the elephants and to approach the groups, yielding recording distances from about 10 to 60 m. The recording equipment was fixed on a custom-built tripod on the car. We performed stereo-recordings with a directional AKG microphone (AKG 480 B CK 69, frequency response 8 Hz - 20 Hz $\pm$ 0.9 dB) and an omni-directional Neumann microphone (KM 183) modified for recording frequencies below 20 Hz (flat recording down to 5 Hz) connected to a 722 Sound Device HDD recorder. Concomitant video recordings were done using a Sony DHC-SD909 camcorder in HD quality. This helped to verify field notes later in the lab during data annotation.

Recording sessions were conducted throughout the day until about 9 p.m. During recordings, the location, the elephant group, the overall group activity (feeding/browsing, locomotion, drinking, ...), the context in which the vocalization was uttered (however, very difficult to assess in many instances), the approximate distance from the microphone to the vocalizing elephant, the gender and the estimated age group (discriminating infant, calf, juvenile and subadult/adult) of the vocalizing individual were noted (the name/ID of the caller was sometimes determined in adult females; in the data set we have 33 rumbles of 24 known females (and one identified male); from 6 female individuals we do have two rumbles, from one female we recorded 3 rumbles, respectively); ID was rarely defined in calves or infants). The family groups have been defined based on individual females known to belong to a particular group (in most instances the collared female) that could be clearly identified based on ear marks (the pattern of notches and holes).

Caller identification works best during calm contexts such as browsing, where groups divide up and decentralize, and the elephants are wide spread within a particular area (or having for example just one smaller group at the waterhole during drinking). In browsing situations, elephants will also regularly rumble in order to keep vocal contact, infants and calves often suckle (a situation where young and mothers are likely to vocalize), and calves and juveniles engage in social play. In order to allocate vocalizations to individual elephants, we focussed

on particular individuals of the group. We chose our focus elephants based on accessibility (identifying callers works best when the focus elephants are close; observer < 20m from the elephant), available age groups and also particular promising situations (e.g. suckling calves, two young elephants playing, an adult bull approaching females). We observed the focus elephants (without ear phones, since wearing ear phones inhibits sound localization), and after perceiving a vocalization we used optical cues such as an open mouth, lifted or spread ears, and general body postures and changes of posture to identify the vocalizing individual. In addition we did use video recordings to verify vocalizing individuals during data annotation.

Since elephants grow throughout their lifetime, the age of free-ranging individuals can be roughly estimated based on their relative size and developmental stage (Whitehouse et al. 2008, see Figure 1 and Table 1, Varma et al. 2012).

Data annotation was performed using a customized annotation tool from S_Tools Stx (Acoustic Institute, Austrian Academy of Science). Each rumble was identified based on field notes and by examining the spectrogram. The start and end cue of each rumble was tagged and the corresponding annotation was added.

Data analysis

Acoustic analysis—Rumbles are highly harmonic and modulated sounds. In order to identify and extract basic signal parameters, we developed a semi-automatic analysis tool in Matlab (Figure 2). The tool takes the segmented rumbles as input, computes a Fourier spectrogram for the frequency range of 0 to 400Hz using a frame size of 300ms and a step size of 40ms, and shows it to the user. The tool enables the user to interactively trace frequency tracks (contours) in the spectrogram. Each contour can be annotated using predefined labels (e.g. ‘fundamental frequency’, ‘2nd harmonic’, etc). The harmonics of a periodic waveform are partials with frequencies at integer multiples of the fundamental frequency. If the fundamental frequency has f Hz, the second harmonic has frequency $2*f$ Hz, the third harmonic $3*f$ Hz, and so on. Due to the strong relation between the fundamental frequency and harmonics, they share similar properties and progression.

We analysed 526 rumbles recorded from 7 elephant groups, varying genders (in infants, calves and juveniles) and of varying contexts. We annotated the frequency contour of the second harmonic (H2), because it was more consistent than the fundamental frequency, which was often masked with low-frequency environmental noise. From the contours of the second harmonic, a number of features (four feature groups with a total of 104 dimensions) were extracted automatically (see Table 2). The features comprised a set of basic contour parameters (BASIC) according to Wood et al. (2005), enhanced with additional features. This group contains frequency-related parameters of the contour (e.g. finish frequency, start frequency, median frequency), relative parameters (e.g. frequency range), shape parameters (e.g. jitter factor and frequency variability), and temporal parameters (e.g. duration, peak frequency location). The resulting BASIC set of contour parameters has 24 dimensions in total. Additionally, we extracted two further feature groups: firstly, a Fourier descriptor which primarily represents the shape (frequency modulation) of a contour (Kunttu et al. 2003); secondly, a descriptor which directly contains the sampled frequencies from the raw

contours (*raw samples*) and thus represents purely frequency-related information. All features are exported into comma-separated files, which form the input of statistical analysis.

Statistical analysis—Statistical tests were conducted using IBM SPSS 20. In order to increase normality and bring in outliers, all variables were $\log_{10}$ -transformed. In order to detect the most discriminative variables, we first ran an ANOVA to test whether the mean values for each parameter of the BASIC feature set differed significantly between age groups in general. A Post-Hoc Bonferroni Test was further applied to identify significant differences in pairwise comparisons of age group categories.

A principle components analysis (PCA) was made to reduce the 24, partly correlated, log-transformed variables to four factors explaining 78 % of the variation (factors with eigenvalues above 1.0 were retained and varimax rotated, Table 3). Another PCA was performed applying only parameters that were significantly distinct in the ANOVA (excluding five variables; COFM, frequency variation, peak by mean frequency, start slope and middle slope; the remaining were reduced to three factors explaining 78.4% of variation). The factor scores were saved as variables using the regression method and used for further analysis.

In order to test whether acoustic parameters can be used to predict the age group of the caller, we conducted discriminant analyses (DA), entering age group as categorical variable. We tested age group classification with resubstitution. The results are expressed as percentage of correct classification, and normalized against expected rates in terms of relative error reduction (Bachorowski and Owren 1999). This last metric takes into account the chance error rate, producing an unbiased measure of classification accuracy.

First, we performed a DA, entering the four PCA factors, which were retained from all features of the BASIC set, as discriminant variables. In addition, we compared the classification success of those features that differed significantly between age groups in the previous ANOVA (entering the three factors retained of this reduced BASIC data set). We further entered the Fourier descriptor parameters as discriminative variables to test whether parameters related with the shape of the frequency contour would be useful to discriminate between age groups.

Automated classification—We developed an automated system to classify age groups. The system takes the raw signals of the segmented rumbles as inputs and extracts acoustic features from the input signals. The features are extracted for short audio frames and are then aggregated over time. The temporally aggregated features form the input to a classifier. The classifier is trained on a randomly selected subset of the input samples (by cross-validation). The trained classifier is then applied to the remaining samples (test data) and automatically predicts the most likely age group. Figure 3 illustrates the architecture of the automated classification system.

We evaluated different audio features for automated age group classification. Initially, we skipped acoustic feature extraction and employed the set of the semi-automatically estimated

contour parameters (BASIC) for automated age group classification. The feature set contains parameters directly derived from the manually annotated contours of the second harmonic. Thus, they are expected to be accurate and not influenced by noise present in the recordings. The classification results based on these features provide an upper limit of performance that an automated classification system is able to achieve.

In a real-world scenario, an automatic classification system needs to be independent of any human intervention. Hence, the features in the BASIC feature set would not be applicable because they require a manual annotation of frequency contours in the spectrogram. To evaluate how well a fully automated approach is able to separate age groups, we replaced the contour-based parameters by features that were fully automatically computed from the audio signals. We extracted features that estimate the frequency track of the fundamental frequency. The fundamental frequency is strongly correlated with the second harmonic and such features thus represent good candidates for the classification of age groups. We applied three alternative algorithms to estimate fundamental frequency: zero crossing rate (ZCR), fundamental frequency based on the subharmonic-to-harmonic ratio (SHTHR), and fundamental frequency estimated by autocorrelation as specified in the MPEG-7 standard, AFF (ISO/IEC 2002; Mitrovic et al. 2010). The features were extracted directly from the raw audio signals without first considering the manually annotated contours.

Additionally to fundamental frequency estimation, we extracted higher-dimensional features that represent the entire spectral envelope of the investigated sounds: Mel-frequency cepstral coefficients (MFCC; Davis and Mermelstein 1980) and Greenwood Function cepstral coefficients (GFCC; Clemins et al. 2006). For a given audio signal, we first computed the Fourier spectrogram and then applied different psychoacoustic filterbanks (critical band filterbanks) to the spectrogram. For MFCCs a Mel-scaled filterbank was used that models the critical bands of the human ear. We shifted the Mel filterbank into the frequency range of 0-500Hz, which is the relevant frequency range for the analysis of elephant rumbles. For GFCCs we applied a Greenwood-scaled filterbank, which is a generalization of the Mel scale to mammals (Greenwood 1961). For both features the filterbank outputs (filter energies) were logarithmized and input to a discrete Cosine transform (cepstral transform) that decorrelates the feature components. The first 18 low-frequency components (Cosine coefficients) were selected as features. The low-frequency coefficients robustly represent the coarse envelope of the spectral energy distribution.

All described acoustic features were extracted for audio frames of 300ms with 100ms step size. The resulting feature vectors for each audio frame of a sound sample were aggregated over time by computing their mean and variance. The temporal mean and variance of all components form the input to classification.

The employed classification technique is a sensitive parameter that strongly influences the overall performance of an automated classification system. In order to estimate the performance of the different features independent from the employed classifier, we evaluated the performance of automatic age group classification with different classifiers and averaged the results. This enables the evaluation of the features' performance independently from a particular classification technique. For classification we employed

powerful and well-established techniques such as linear Support Vector Machines, (SVMs; Cortes and Vapnik 1995), Nearest Neighbor (NN), K-Nearest Neighbor with K=5 (KNN; Cover and Hart 1967) as well as Linear Discriminant Analysis, (LDA; Duda et al. 2001). We applied 10-fold cross-validation in all experiments to assure independence from the training data. 10-fold cross validation generates 10 different partitions of the dataset into disjoint test and training sets. Each of the training sets contains 474 randomly chosen samples while the test set is formed by the remaining 52 samples. The distribution of the age classes among the training and test set is kept equal for all partitions. Each training sets contains 58 infants, 92 calves, 24 juveniles, and 300 adults. Each test set contains 6 infants, 10 calves, 3 juveniles, and 33 adults.

Results

Statistical analysis

In the ANOVA, most frequency and temporal parameters (BASIC feature set) varied between age groups. Most importantly, features representing absolute frequency values of the second harmonic differed significantly in the post hoc test for each pairwise comparison. The frequency in rumbles decreased consistently with age (Figure 4 and 5, Table 1). Duration tended to increase with age, but pairwise comparison of infants to calves, and calves to juveniles, did not yield significant results. Subadult/adult rumbles differed significantly from the other age groups in almost all tested parameters.

In order to verify whether age groups can indeed be classified based on acoustic parameters (and to further verify the most descriptive features), we performed several DAs. Entering all 24 features of the BASIC data set into the analysis resulted in 68 % correct classification (57 % error reduction). Applying only parameters that were significantly distinct in the ANOVA yielded 70 % correct classification (60 % error reduction). The confusion matrices revealed that mainly infant and calf, calf and juvenile, and juvenile and adult rumbles mixed up in the analysis (Table 4a). Despite this, infants and calves were reliably discriminated from subadults/adults (Table 4b). Applying DA discriminating infants/calves (lumped into one category) and subadults/adults (excluding juveniles) yielded a classification result of 95 % (92.8 % error reduction). The parameters representing absolute frequency values (e.g. minimum, start, maximum frequency, medianFreq_log) contributed most to the classification result (they had strong loadings on the first factor in the PCA, which again loaded strongest in the structure matrix of the DA). Call duration and features expressing the frequency modulation (e.g. jitter), relative relations (e.g. frequency range) and temporal features were of less value. Similarly, applying DA on the Fourier descriptors (representing mainly the frequency contour) resulted in only 56.3% of cases correctly classified (41.7% error reduction).

Automated classification

The results of automated age group classification are summarized in Table 4. In a first run, we evaluated the performance of the system based on the BASIC feature set. The classification system yielded an accuracy of 75% on average for all employed classifiers. This result is comparable to the performance obtained in statistical analysis, which

demonstrates that the classification system works properly. In a next step, we removed all features from the BASIC set, which were not significant in the statistical analysis (BASIC (Sig) in Table 2), obtaining a 2% increase in classification performance (77%).

The features related to the fundamental frequency (ZCR, SHTHR, MPEG7-AFF) yielded strongly varying classification accuracies between 46% and 63%. The feature based on subharmonic-to-harmonic ratio clearly outperformed ZCR as well as the MPEG7-AFF descriptor. Nonetheless, the performance is still considerably lower than that of the BASIC feature set. The high-dimensional features (MFCC and GFCC) clearly outperformed the features related to the fundamental frequency. Both features yielded similar accuracy on average over all classifiers (75% and 74%, respectively).

Discussion

We showed that frequency parameters of the second harmonic of African elephant rumbles can be used to classify elephants into age groups, because frequency in rumbles generally decreases with age. Discriminating adults from infants and calves (lumped together in one group) resulted in 95 % correct classification. This highlights that it is possible to detect the presence of infants and calves within an elephant group based on vocalizations with great accuracy. Statistical models revealed that parameters representing absolute frequency values have the most discriminative power. Applying parameters that represent the contour and modulations of the second harmonic yielded lower classification results, indicating that the frequency contour does not convey information about age in elephants, but that absolute-pitch-related parameters seem to be more reliable age cues.

Frequency can be modulated by varying the tension of the vocal folds (Hast 1966) and in elephants has been shown to differ according to specific contexts and/or motivational states in adult females (Soltis 2010; Soltis et al. 2009; King et al. 2010; Poole et al. 1988; Poole 2011; Stoeger et al. 2012) and infants (Stoeger-Horwath et al. 2007; Wesolek et al. 2009). We also obtained considerable overlap in parameters between adjacent groups, particularly infant and calve, and juvenile and adult vocalizations mixed up in the DA classification. This result is not unexpected when investigating elephant rumbles of a population recorded in various contexts and situations.

Other acoustic parameters that correlate with overall size in vertebrates are formant frequencies (or vocal tract resonances). Formant frequency values are generally determined by the length and shape of the vocal tract, which is constrained by skeletal structures and therefore closely tied to body size (Fitch 1997). Vocal tract lengths can be calculated based on formant frequency spacing (Fitch 1997; Reby and McComb 2003). It has been suggested that formants are a more reliable cue to body size in mammals than the fundamental frequency (Fitch 1997; Riede and Fitch 1999). This also holds true because vocal folds are highly sensitive to changes in testosterone (Beckford et al. 1985; Dabbs and Mallinger 1999) and are thus a poor indicator to discriminate body size between adult males (Fitch 2000; Reby and McComb 2003; Fischer et al. 2004).

Stoeger et al. (2012), however, recently demonstrated that African elephants significantly vary their formant structures by switching from nasal to oral sound production. In elephants, the nasal vocal tract (up to 2.5 m long in adult females) is strongly elongated in relation to the oral path (0.7 to 1 m in adult females). Accordingly, an individual elephant potentially lowers its formants ca. threefold by using the nasal path, compared to rumbles uttered orally (Stoeger et al. 2012). Estimations of vocal tract lengths for infant and calves yield an oral vocal tract length of 30 cm and a nasal vocal tract length of about 90 cm (Sikes 1971; Laws et al. 1975; Soltis 2010). Therefore, formant values from oral rumbles by juveniles and adults might overlap with formants of nasal rumbles uttered by infants and calves. In addition, formant frequency values can vary significantly with arousal and context (King et al. 2010), and potentially with social rank (Soltis et al. 2009). Therefore, formant frequency analysis of elephant vocalizations is highly complex and intricate, since formant values are even more variable than fundamental frequency.

Pfefferle and Fischer (2006), for example, found that the fundamental frequency in hamadryas baboons (*Papio hamadryas*) was more closely related to most physical measurements across age classes and among adult females than formant dispersion. Depending on the call type, and possibly the sex of the caller, the fundamental frequency can thus serve as a reliable indicator of the physical characteristics of individuals (Pfefferle and Fischer 2006).

The best results in automatic classification experiments were obtained by the subset of the BASIC features that contains only the significant feature components, BASIC (Sig). The performance of these features is high because they were derived from manually annotated contours and thus were not corrupted by noise. This result shows that the parameters extracted from the annotated contours represent a solid basis for age group classification. A fully automated classification system, however, cannot rely on manually annotated contours. Furthermore, the automated extraction of frequency contours (e.g. of fundamental frequency) from an audio signal is a challenging task. Noise often masks the frequency contours, making the contours difficult to detect and track automatically. This was also observed for all three features employed to estimate the fundamental frequency (ZCR, SHTHR, and MPEG-7 AFF). Especially in noisy recordings the features were unable to detect the fundamental frequency robustly and generated false estimates. This weakness is directly reflected in the classification performance, which is markedly below that of the BASIC features.

The high-dimensional features (MFCC and GFCC) clearly outperform the features related to the fundamental frequency. The high performance shows that information about the spectral envelope is beneficial for age group classification. An advantage of these features is that they do not rely on extracting particular sound attributes such as fundamental frequency, but instead capture the entire spectral envelope. As a result the features are less prone to noise and provide a robust signal representation. The small performance difference between both features (74% vs. 75%) indicates that the different psychoacoustic filterbanks only minimally influence performance. Remarkably, the overall performance for both features is comparable to that of the BASIC features even though they were extracted fully automatically without human intervention. From the experiments we conclude that MFCC

and GFCC are promising signal representations for a future monitoring system that autonomously estimates elephant age based on vocalizations.

Generally, certain call types themselves (such as elephant barks and roars) are characteristic for age groups (Stoeger-Horwath et al. 2007; Poole 2011). Infants and calves are much more likely to produce barks and roars (in particular tonal and mixed roars, Stoeger-Horwath et al. 2007; Stoeger et al. 2011) than juveniles and adult females. These vocalizations are often applied by infants and calves to indicate suckle intention (barks) or as protest vocalizations (e.g. roars during suckle protest). Juvenile or adult elephants hardly bark and only roar when highly distressed (Poole 2011). This results in a very chaotic call structure, in contrast to protest roars uttered by infants and calves, which do possess harmonic (or subharmonic) call sections (Stoeger et al. 2011). These particular age-dependent call types provide additional clues to detect the presence of elephants of a particular age group.

Our results reveal that it is possible to estimate the age of an elephant based on its vocalizations, and that it is further possible to acoustically classify age groups automatically. This provides the scientific foundation for a future system that could estimate the demography of an acoustically monitored elephant group or population.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We thank John Adendorff, the rangers from the Addo Elephant National Park, and Simon Stoeger from the Vienna Zoo for their support during data collection. We further thank W. Tecumseh Fitch, Department of Cognitive Biology, University of Vienna, and Christian Breiteneder, Institute for Software Technology and Interactive Systems, Vienna University of Technology, for strongly supporting our elephant research at their institutions. In addition we acknowledge Michael Stachowitsch for editing the manuscript. This work was funded by the FWF, the Austrian Science Fund [P23099].

References

- August PV, Anderson JGT. Mammal sounds and motivation-structural rules: a test of the hypothesis. *Journal of Mammology*. 1987; 68:1–9.
- Bachorowski JA, Owren MJ. Acoustic correlates of talker sex and individual talker identity are present in a short vowel segment produced in running speech. *Journal of the Acoustical Society of America*. 1999; 106:1054–1063. [PubMed: 10462810]
- Beckford NS, Rood SR, Schaid D, Schanbacher B. Androgen stimulation and laryngeal development. *The Annals of Otology, Rhinology and Laryngology*. 1985; 94:634–640.
- Blake S, Strindberg S, Boudjan P, Makombo C, Bila-Isia I, Ilambu O, Grossmann F, Bene-Bene L, de Semboli B, Mbenzo V, S'hwa D, Bayogo R, Williamson L, Fay M, Hart J, Maisels F. Forest elephant crisis in the Congo Basin. *PLoS Biology*. 2007; 5:e111. doi:10.1371/journal.pbio.0050111. [PubMed: 17407383]
- Blumstein DT, Mennill DJ, Clemins P, Girod L, Yao K, Patricelli G, Deppe JL, Krakauer AH, Clark C, Cortopassi KA, Hanser SF, McCowan B, Ali AM, Kirschel ANG. Acoustic monitoring in terrestrial environments using microphone arrays: applications, technological considerations and prospectus. *Journal of Applied Ecology*. 2011; 48:758–767.
- Clemins PJ, Trawicki MB, Adi K, Tao J, Johnson MT. Generalized perceptual features for vocalization analysis across multiple species. *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*. 2006:253–256.

- Collins SA, Missing C. Vocal and visual attractiveness are related in women. *Animal Behaviour*. 2003; 65:997–1004.
- Cortes C, Vapnik V. Support-vector networks. *Machine Learning*. 1995; 20:273–297.
- Cover T, Hart P. Nearest neighbor pattern classification. *IEEE Transactions on Information Theory*. 1967; 13:21–27.
- Dabbs JM, Mallinger A. High testosterone levels predict low voice pitch among men. *Personality and Individual Differences*. 1999; 27:801–804.
- Davis S, Mermelstein P. Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences. *IEEE Transactions on Acoustic and Speech Signal Processing*. 1980; 28:357–366.
- Dublin HT, Hoare RE. Searching for Solutions: The Evolution of an Integrated Approach to Understanding and Mitigating Human-Elephant Conflict in Africa. *Human Dimensions of Wildlife: An International Journal*. 2004; 9:271–278.
- Duda, R.; Hart, P.; Stork, D. *Pattern Classification*. 2nd edition. Wiley; 2001.
- Fischer J, Kitchen DM, Seyfarth RM, Cheney DL. Baboon loud calls advertise male quality: acoustic features and their relation to rank, age, and exhaustion. *Behavioral Ecology and Sociobiology*. 2004; 56:140–148.
- Fischer J, Hammerschmidt K, Cheney DL, Seyfarth RM. Acoustic features of male baboon loud calls: Influences of context, age, and individuality. *Journal of the Acoustical Society of America*. 2002; 111:1465–1474. [PubMed: 11931324]
- Fitch WT. Vocal tract length and formant frequency dispersion correlate with body size in rhesus macaques. *Journal of the Acoustical Society of America*. 1997; 102:1213–1222. [PubMed: 9265764]
- Fitch WT. The evolution of speech: a comparative review. *Trends in Cognitive Science*. 2000; 4:258–267.
- Garstang M. Long-distance, low-frequency elephant communication. *Journal of Comparative Physiology A*. 2004; 190:791–805.
- Greenwood D. Critical bandwidth and the frequency coordinates of the basilar membrane. *The Journal of the Acoustical Society of America*. 1961; 33:1344–1356.
- Hast MH. Physiological Mechanisms of Phonation: Tension of the Vocal Fold Muscle. *Acta Otolaryngologica*. 1966; 62:309–318.
- Hauser MD. The evolution of nonhuman primate vocalizations: effects of phylogeny, body weight and social context. *American Naturalist*. 1993; 142:528–542.
- Hedges S, Gunaryadi D. Reducing human-elephant conflict: do chillies help deter elephants from entering crop fields? *Oryx*. 2010; 44:139–146.
- Herbst CT, Stoeger AS, Frey R, Lohscheller J, Titze IR, Gumpenberger M, Fitch WT. How low can you go - physical production mechanism of elephant infrasonic vocalization. *Science*. 2012; 337:595–599. [PubMed: 22859490]
- ISO/IEC. Information Technology – Multimedia Content Description Interface. First Edition. 2002. 15938
- King LE, Soltis J, Douglas-Hamilton I, Savage A, Vollrath F. Bee threat elicits alarm call in African elephants. *PloS One*. 2010; 5:e10346. doi:10.1371/journal.pone.0010346. [PubMed: 20436682]
- Kunttu I, Lepistö L, Rauhamaa J, Visa A. Multiscale Fourier Descriptor for Shape Classification. *Proceedings of the International Conference on Image Analysis and Processing*. 2003; 1:536–541.
- Laws, RM.; Parker, ISC.; Johnstone, RCB. *Elephants and their habitats: the ecology of elephants in North Bunyoro, Uganda*. Oxford University Press; London: 1975.
- Lee PC, Graham MD. African elephants *Loxodonta africana* and human-elephant interactions: implications for conservation. *International Zoo Yearbook*. 2006; 40:9–19.
- Maynard, JS.; Harper, D. *Animal Signals*. Oxford University Press; New York: 2003.
- Mitrovic, D.; Zeppelzauer, M.; Breiteneder, C. Features for Content-Based Audio Retrieval. In: Zelkowitz, MV., editor. *Advances of Computers* vol. 78. Academic Press; Burlington: 2010. p. 71-150.

- Morton ES. On the occurrence and significance of motivation-structural rules in some bird and mammal sounds. *American Naturalist*. 1977; 111:855–869.
- O'Connell-Rodwell CE, Rodwell T, Rice M, Hart LA. Living with the modern conservation paradigm: can agricultural communities co-exist with elephants? A five year case study in East Caprivi, Namibia. *Biological Conservation*. 2000; 93:381–391.
- Payne KB, Thompson M, Kramer L. Elephant calling patterns as indicators of group size and composition: the basis for an acoustic monitoring system. *African Journal of Ecology*. 2003; 41:99–107.
- Pfefferle D, Fischer J. Sound and size: identification of acoustic variables that reflect body size in hamadryas baboons, *Papio hamadryas*. *Animal Behaviour*. 2006; 72:43–51.
- Poole JH, Payne K, Langbauer WR Jr, Moss C. The social contexts of some very low frequency calls of African elephants. *Behavioral Ecology and Sociobiology*. 1988; 22:385–392.
- Poole, JH. Behavioral contexts of elephant acoustic communication. In: Moss, CJ.; Croze, H.; Lee, PC., editors. *The Amboseli elephants: a long-term perspective on a long-lived mammal*. The University of Chicago Press; Chicago: 2011. p. 125–161.
- Reby D, McComb KE. Anatomical constraints generate honesty: acoustic cues to age and weight in the roars of red deer stags. *Animal Behaviour*. 2003; 65:519–530.
- Riede T, Fitch WT. Vocal tract lengths and acoustics of vocalization in the domestic dog (*Canis familiaris*). *The Journal of Experimental Biology*. 1999; 202:2859–2867. [PubMed: 10504322]
- Santiapillai C, Wijeyamohan S, Bandara G, Athurupana R, Dissanayake N, Read B. An assessment of the human-elephant conflict in Sri Lanka. *Ceylon Journal of Science (Biological Sciences)*. 2010; 39:21–33.
- Sikes, SK. *The natural history of the African elephant*. Weidenfeld and Nicolson; London: 1971.
- Sitati NW, Walpole MJ. Assessing farm-based measures for mitigating human-elephant conflict in Transmara District, Kenya. *Oryx*. 2006; 40:279–286.
- Soltis P, Leighty KA, Wesolek CM, Savage A. The expression of affect in African elephants (*Loxodonta africana*) rumble vocalizations. *Journal of Comparative Psychology*. 2009; 123:222–225. [PubMed: 19450029]
- Soltis J. Vocal communication in African elephants (*Loxodonta africana*). *Zoo Biology*. 2010; 29:192–209. [PubMed: 19434672]
- Stoeger-Horwath AS, Stoeger S, Schwammer HM, Kratochvil H. Vocal repertoire of infant African elephants – First insights into the early vocal ontogeny. *Journal of the Acoustical Society of America*. 2007; 121:3922–3931. [PubMed: 17552738]
- Stoeger AS, Charlton BD, Kratochvil H, Fitch WT. Vocal cues indicate level of arousal in infant African elephant roars. *Journal of the Acoustical Society of America*. 2011; 130:1700–1711. [PubMed: 21895107]
- Stoeger AS, Heilmann G, Zeppelzauer M, Ganswindt A, Hensman S, Charlton BD. Visualizing Sound Emission of Elephant Vocalizations: Evidence for Two Rumble Production Types. *PLoS ONE*. 2012; 7:e48907. doi:10.1371/journal.pone.0048907. [PubMed: 23155427]
- Thompson ME, Schwager SJ, Payne KB, Turkalo AK. Acoustic estimation of wildlife abundance: methodology for vocal mammals in forested habitats. *African Journal of Ecology*. 2009a; 48:654–661.
- Thompson ME, Schwager SJ, Payne KB. Heard but not seen: an acoustic survey of the African forest elephant population at Kakum Conservation Area, Ghana. *African Journal of Ecology*. 2009b; 48:224–231.
- Titze, IR. *Principles of Voice Production*. Prentice Hall; Englewood Cliffs, New Jersey: 1994.
- Varma, S.; Baskaran, N.; Sukumar, R. Resource material for synchronized elephant population count using block count, line transect dung count method and waterhole count. Asian Nature Conservation Foundation, Innovation Centre, Indian Institute of Science; Bangalore - 560 012, Karnataka: 2012. Field Key for Elephant Population Estimation and Age and Sex Classification. and Centre for Ecological Sciences, Indian Institute of Science, Bangalore - 560 012, Karnataka
- Whitehouse, A.; Irwin, P.; Gough, K. *A field guide to the Addo elephants*. second edition. Print Grahamstown; Duplin: 2008.

- Wesolek CM, Soltis J, Leighty KA, Savage A. Infant African elephant rumble vocalizations vary according to social interactions with adult females. *Bioacoustics*. 2009; 18:227–239.
- Wood JD, McCowan B, Langbauer W Jr, Viljoen J, Hart L. Classification of African elephant *Loxodonta africana* rumbles using acoustic parameters and cluster analysis. *Bioacoustics*. 2005; 15:143–161.
- Wrege PH, Rowland ED, Thompson BG, Batruch N. Use of Acoustic Tools to Reveal Otherwise Cryptic Responses of Forest Elephants to Oil Exploration. *Conservation Biology*. 2010; 24:1578–1585. [PubMed: 20666800]

Figure 1.
Photograph of an elephant group at the Addo Elephant National Park showing the different age-group categories, infant, calf, juvenile and adult.

Figure 2.

The semi-automatic contour annotation tool. Contours are labelled and drawn with the mouse. From the contours the tool automatically extracts different features. Drawing a contour into the spectrogram. The fundamental frequency is entirely masked with noise.

Figure 3.

Architecture of the automated age group classification system. The input signal (a rumble) is first framed. For each frame, content-based features are extracted (e.g. fundamental frequency or GFCCs). The frame-based features are aggregated over time and input to a classifier which has previously been trained on a number of training samples. The result of classification is a label which specifies the most likely age group of the individual that elicited the rumble in the input signal.

Figure 4.

Labelled contours in a rumble of an adult female (a) and an infant (b). The fundamental frequency is visible in both examples. The fundamental frequency (labelled in black) and the 2nd harmonic (labelled in blue) and the spacing of the harmonics are lower in the adult (one higher harmonic has been additionally labelled, respectively). Although the frequency contour is different in these examples, we found that, over the entire data, the contour has relatively less power in discriminating age compared to absolute frequency values.

Figure 5.

Box-plot presentation showing the general decrease in frequency (mean frequency of the second harmonic) with age in African elephant rumbles.

Table 1

Age class categories, the age in years, and the corresponding description, the number of calls in the analyses (*N calls*; *male*, *fem=female*), the elephant groups that individuals have been recorded of, mean rumble duration ($\text{Dur} \pm \text{SD}$) in seconds (s) and mean frequency in Hertz (Hz) of the second harmonic $\pm$ standard deviation ($\text{H2} \pm \text{SD}$) are listed.

Age class	Age (yrs)	Description	N calls	Elephant groups (N calls per group and per gender (in juv and adults))	Dur $\pm$ SD (s)	H2 $\pm$ SD (Hz)
Infant	<1	Infants can walk underneath mother, regularly observed to suckle	66	A* (8), B** (Ba8; Bb11), H(8), M (5), L (2), R(19), In 5 rumbles, the group was not defined	1.48 $\pm$ 0.73	59.8 $\pm$ 7.7
Calf	1-4	Calf still suckling and still the youngest calf of the mother. Tusks visible of calves from approximately 2.5 years.	98	A (20), Ba(7); Bb(2), 12 not allocated to Ba or Bb), H (4), M (17), P (3), R (9) In 24 rumbles, the group was not defined	1.79 $\pm$ 0.79	54.2 $\pm$ 7.3
Juvenile	5-12	Weaned, and often have a younger sibling. By ~10-years, elephants reach ~ 3/4 the size of an adult female	27 (17males, 10fem)	A (3, all males), Ba (1 male; not allocated to Ba or Bb: 5 fem), H (1, male), M (1, fem), R (3, fem/5, males)	2.42 $\pm$ 0.89	48.1 $\pm$ 8.7
Subadults and adults	13<	In females, breasts start to develop with the pregnancy of their first calf, males tend to become independent, Fully mature adult females (20<) usually have more than one calf	333 (males: 5-not allocated to groups), fem: 328)	A* (fem: 33), Ba (fem: 39), Bb (fem: 21) B (fem: 45 - not allocated to Ba or Bb, H (fem: 26), M (fem: 31), L* (fem: 14), P (fem: 34), R (fem: 45), In 73 rumbles, the group was not defined	3.34 $\pm$ 1.44	38.0 $\pm$ 5.8

* The A-group was only recorded in 2011. The L-group was only recorded in 2012. Otherwise all groups were recorded in both years of data collection.

** The B group (the biggest group) is subdivided into two subgroups (Ba led by Bubbles, Bb led by Catharina), which appear sometimes separated, but do also conjoin regularly.

Table 2

The different features (grouped into sets), which are extracted from the contour of the second harmonics together with their dimension (Dim).

Feature Set	Description	Dim.
Raw samples	A resampling of a contour by a fixed number of points. 60 uniformly sampled frequencies along the contour are extracted	60
Fourier descriptor	The logarithmized low-frequency Fourier coefficients of a contour based on the descriptor of Kunttu et al. (2003)	20
Basic	A set of basic shape descriptors based on Wood et al. (2005) including: coefficient of frequency modulation, jitter factor, frequency variability, inflection factor, start frequency, mid frequency, finish frequency, minimum and peak frequency, time between minimum and peak frequency, mean frequency, mean frequencies of first, second, and third third of the sound segment, median frequency, frequency range, max/mean frequency, mean/min frequency, peak frequency location, minimum frequency location, duration, start slope, middle slope, final slope	24

Table 3

Results of the PCA with varimax rotation performed on the BASIC data set on 526 vocalizations.

	Factor1	Factor2	Factor3	Factor4
Eigenvalue	10.518	4.899	2.260	1.260
% Variance	43.8	20.2	9.4	5.0

Variables	Principle Component			
	Factor1	Factor2	Factor3	Factor4
COFM_log	.120	.761	.075	.016
JitterFactor_log	-.298	.588	-.176	-.183
FreqVariability_log	.023	.915	-.130	.061
InflectionFactor_log	-.280	-.408	-.368	.097
FinishFreq_log	.954	-.135	.157	-.061
MinFrq_log	.970	-.191	.034	.014
MaxFreq_log	.967	.239	-.004	.017
MeanFreq_log	.993	.094	.003	.000
FreqRange_log	.368	.900	-.071	.011
PeakByMeanFreq_log	-.049	.870	-.024	.125
MeanByMinFreq_log	-.111	.885	-.080	-.020
PeakFreqLoc_log	.115	.004	.698	-.280
MinFreqLoc_log	-.007	.032	-.418	.723
Duration_log	-.580	.333	.244	.234
StartSlope_log	.281	.490	-.134	-.560
MiddleSlope_log	-.071	-.160	.782	-.081
FinalSlope_log	-.412	-.269	.458	.144
StartFreq_log	.953	-.098	-.124	.184
MidFreq_log	.970	.184	.003	-.041
TimeMinMax_log	.136	.098	-.108	.749
mean1stThird_log	.970	.090	-.145	.091
mean2ndThird_log	.977	.175	.005	-.037
mean3rdThird_log	.977	.005	.155	-.055
medianFreq_log	.989	.112	.001	-.005

Table 4

Classification results in percent (and absolute numbers) of the DA entering all BASIC features as discriminative variables and (1) all four age groups as categorical variables, and (2) testing only two categorical variables (infants and calves lumped into one category, versus subadults/adults, excluding juveniles).

(a) Age group	Predicted group membership			
	Infants	Calves	Juveniles	Subadults/Adults
Infants	64% (41)	28% (18)	8% (5)	0% (0)
Calves	28% (28)	38% (39)	33% (34)	1% (1)
Juveniles	18.5% (5)	7.4% (2)	41% (11)	33% (9)
Subadults/Adults	0.6% (2)	2% (5)	16% (54)	82% (272)

(b) Age group	Predicted group membership	
	Infants/Calves	Subadults/Adults
Infants/Calves	93% (155)	7% (11)
Subadults/Adults	3% (16)	95% (317)

Table 5

Overall classification rate for automated age group classification with 10-fold cross-validation. All features are evaluated with four different classifiers (SVM, NN, 5-NN, and LDA). The last column shows the mean and standard deviation (SD) of classification accuracy over all classifiers.

Feature	SVM	NN	5-NN	LDA	Mean (SD.)
BASIC	78%	71%	75%	75%	75% ($\pm 2.5\%$)
BASIC (Sig)	78%	73%	79%	76%	77% ($\pm 2.3\%$)
ZCR	63%	48%	60%	35%	52% ($\pm 11.1\%$)
SHTHR	64%	59%	66%	62%	63% ($\pm 2.6\%$)
MPEG7-AFF	63%	55%	55%	12%	46% ($\pm 20.0\%$)
MFCC	76%	72%	75%	75%	75% ($\pm 1.6\%$)
GFCC	74%	75%	75%	70%	74% ($\pm 2.1\%$)