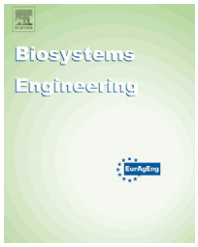
BIOSYSTEMS ENGINEERING XXX (2010) 1–9

Available at www.sciencedirect.com

journal homepage: www.elsevier.com/locate/issn/15375110



Research Paper

Automatic detection of African elephant (Loxodonta africana) infrasonic vocalisations from recordings

Pieter J. Venter [a], *Johan J. Hanekom* [a,*]

[a] *Department of Electrical, Electronic and Computer Engineering, University of Pretoria, Pretoria, 0002, South Africa*

ARTICLE INFO

Article history:
Received 5 August 2008
Received in revised form
23 March 2010
Accepted 1 April 2010
Published online xxx

Recordings of elephant vocalisations can be used to determine the size and composition of the herd, the sexual state, as well as the emotional condition of an elephant. Manual analysis of recordings (by listening to these and by visual inspection of spectrograms) to locate vocalisations is tedious. The automatic detection of vocalisations in recordings is explored. Important signal characteristics of elephant vocalisations were identified from spectrograms and a technique, based on the principles of existing voice activity detection algorithms, was developed to exploit these. Results obtained suggest that the algorithm can reliably detect elephant vocalisations from noisy recordings as long as the harmonic structure of vocalisations is not buried in background noise.

1. Introduction

The best known elephant vocalisations are low frequency rumbles or higher frequency trumpets (McComb, Reby, Baker, Moss, & Sayialel, 2003), but researchers agree that elephants can produce at least 10 different sound types (Clemins & Johnson, 2003; Leong, Ortolani, Burks, Mellen, & Savage, 2002; Soltis, Leong, & Savage, 2005b). Most vocalisations are produced in the form of rumbles with infrasonic pitch that is too low to be easily perceived by humans.

The importance of elephant vocalisations to researchers in the field of elephant behaviour is largely due to the abundance of information that can be retrieved from it (Garstang, 2004; Langbauer, 2000; Langbauer, Payne, Charif, & Thomas, 1989; McComb et al., 2003; O'Connell-Rodwell, Arnason, & Hart, 2000; Poole, Tyack, Stoeger-Horwath, & Watwood, 2005; Wood, McCowan, Langbauer, Viljoen, & Hart, 2005). The number of rumbles observed within a certain time can be used to determine the size of an unseen group of elephants as well as the number of males, females and calves in such a group (Payne, Thompson, & Kramer, 2003). Each elephant has specific voice characteristics which means that individuals may be recognised by their vocalisations (Clemins & Johnson, 2003; Clemins, Johnson, Leong, & Savage, 2005; Soltis et al., 2005b). Information about the sexual state of individual elephants can also be determined by analysing their rumbles (Leong, Burks, Rizkalla, & Savage, 2005; Poole, 1999; Soltis, Leong, & Savage, 2005a). As is the case with humans, some parameters of rumbles can be used to determine the emotional state of an elephant (Clemins et al., 2005; Soltis et al., 2005b).

These infrasonic rumbles have a fundamental frequency of between 15 and 25 Hz and harmonics ranging several hundred Hz (Langbauer, 2000). The harmonic component at approximately 125 Hz has been shown to be the most important frequency needed for an elephant in the group to correctly establish the identity of the caller (Langbauer, 2000). Fig. 1a shows a spectrogram of a typical elephant rumble. The harmonic nature of a typical elephant rumble is clear from the spectrogram. Harmonics, appearing in the figure from around 2.3 s–5.3 s, are indicated by horizontal lines at around 4 s in

* Corresponding author. Tel.: +27 12 4202461; fax: +27 12 3611629.
E-mail addresses: pjventer@tuks.co.za (P.J. Venter), johan.hanekom@up.ac.za (J.J. Hanekom).
1537-5110/$ – see front matter © 2010 IAgrE. Published by Elsevier Ltd. All rights reserved.
doi:10.1016/j.biosystemseng.2010.04.001

Nomenclature

SNR	Signal to Noise Ratio (dB)
ERB	Equivalent Rectangular Bandwidth (Hz)
A	normalised autocorrelation
c	channel number
r	filter output
N	number of samples used for the calculation of the autocorrelation
τ	number of autocorrelation lag steps
j	centre position of current analysis window
F_s	sampling frequency (Hz)
s	number of lag steps to first peak of autocorrelation
p	pitch (Hz)
P	array containing pitch estimates for all samples
C	array containing pitch estimates of rumbles only
L	primary pitch difference limit in pitch tracking algorithm (Hz)
LT	secondary pitch difference limit in pitch tracking algorithm (Hz)
S	primary sample number limit in pitch tracking algorithm
ST	secondary sample number limit in pitch tracking algorithm

the spectrogram. Observations from available data (around 40 h of data) are that the spectral energy of a rumble is largely located below 250 Hz.

Elephant rumbles in recordings are usually isolated by experts who manually analyse spectrograms of recordings and listen to them played back at higher rates, so that the elephant sounds appear in the frequency range audible to humans. This procedure is labour intensive and time consuming, especially for studies where large numbers of vocalisations are needed. In view of these limitations, the possibility of adapting speech processing techniques that are usually applied to human speech have been considered for the

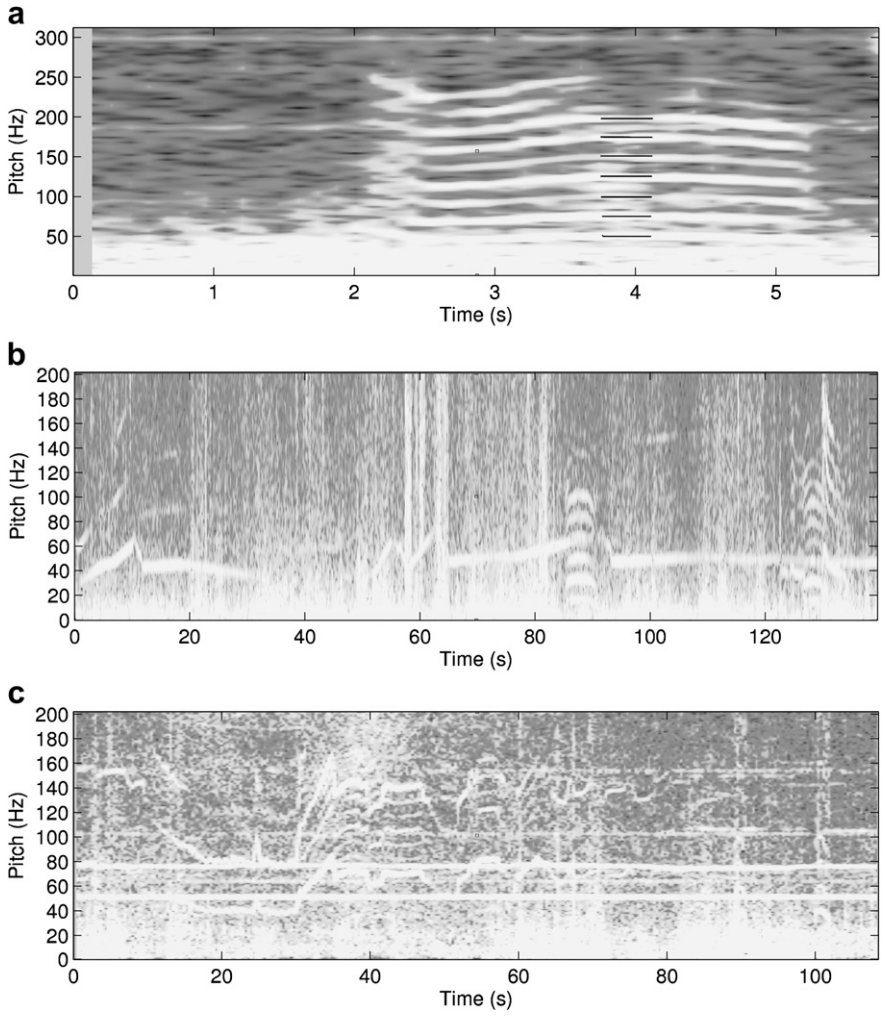


Fig. 1 – Spectrograms of elephant rumbles. (a). Spectrogram containing a typical elephant rumble. Eight harmonics of the rumble (occuring from 2.3 s to 5.3 s) are indicated with lines between 3.8 s and 4.1 s (b). Spectrogram of a segment of sound containing narrow band and broadband noise as well as elephant rumbles. (c). Spectrogram of a recording with loud unwanted periodic sounds.

Please cite this article in press as: Pieter J. Venter, Johan J. Hanekom, Automatic detection of African elephant (Loxodonta africana) infrasonic vocalisations from recordings, Biosystems Engineering (2010), doi:10.1016/j.biosystemseng.2010.04.001

extraction of elephant rumbles from recordings. Infrasonic elephant rumbles are produced by vocal cords (Garstang, 2004; McComb et al., 2003; Soltis et al., 2005b) so that it may be expected that the resulting sounds would have characteristics similar to voiced human speech (i.e., the harmonic nature seen in Fig. 1a). In fact, studies have shown that most mammalian vocal production and reception systems are very similar (Bradbury & Vehrencamp, 1998; Titze, 1994).

Therefore, existing techniques used for voice activity detection of human speech may also be suitable for elephant rumble detection. Because of the harmonic nature of voiced speech, one way of detecting the presence of speech during a particular interval may be to apply a pitch detection algorithm. The application of pitch detection algorithms used in voice activity detectors described in the literature is often limited to noiseless speech in a telecommunications context. However, elephant rumble recordings usually contain unwanted sounds, including sounds produced by other animals, wind, elephants walking through the bush and other sounds that occur in the wild. In addition, low frequency sounds travel several kilometres, resulting in distant sounds caused by motor vehicles of tourists occuring commonly in wild elephant recordings. Therefore, a pitch detection algorithm that is robust against noise is required to detect elephant rumbles effectively in typical recordings.

The simplest time-domain pitch detection algorithm counts the number of zero-crossings of the recorded signal. However, this method is inaccurate when the signal contains noise or, in the case of a harmonic signal, when the fundamental frequency is less energetic than any of the higher harmonics. Autocorrelation analysis (Takagi, Seiyama, & Miyasaka, 2000) works well with lower frequencies and when harmonic structure is present. The addition of noise, however, degrades the definition of the peaks of the autocorrelation function. Similarly, added noise in cepstrum based techniques (Kim & Chung, 2004; Noll, 1967) diminishes the peak indicating the fundamental. Although frequency domain pitch detection algorithms (Zhang, Zhang, Lin, & Quan, 2006) are computationally inexpensive, they are also not robust against noise.

More robust are time-frequency domain algorithms (Wu, Wang, & Brown, 2003; Zhao & Ogunfunmi, 1999). They firstly filter the original signal into sub-bands and then perform time-domain analysis on the band-filtered signals, as is believed to be done by the cochlea. Although this method is computationally expensive, due to autocorrelations being computed for every sub-band (as explained in more detail later), its greater robustness against noise makes a sub-band pitch estimation algorithm attractive for the present application.

2. Methods

A number of recordings, with a total duration of around 40 h made by the research group of Dr. William Langbauer Jr. in the southern parts of the Kruger National Park, South Africa during 2003, were made available for the present study. These recordings were made in the mornings using a handheld sound recorder. Some were of good quality, whilst the signal to noise ratio (SNR) in others was poor.

Properties identified from these recordings (Langbauer, 2000) were that rumbles usually contained a number of harmonics, had a fundamental of between 15 and 25 Hz and duration of 0.5–5 s, and usually commenced and ended at the same pitch with a rise in pitch in the middle of the rumble.

These characteristics were used to assist in the development of an algorithm to detect rumbles in recordings. The proposed algorithm is based on the work of (Wu et al., 2003).

The first step in the proposed algorithm was to obtain pitch estimates. Recorded sound was filtered into bands using a bank of fourth-order gammatone filters (Lee & Ellis, 2006) evenly distributed on the equivalent rectangular bandwidth (ERB) scale. These filters were adapted from human gammatone filters, the latter being based on the particular sensitivity of the human cochlea to pitch (Johannesma, 1972). The elephant cochlea is somewhat larger than the human cochlea and is sensitive in a frequency range of 17–10.5 kHz (Reuter, Nummela, & Hemila, 1998). Evidence exists that the elephant cochlea analyses pitch information, and indeed analyses frequency information on a logarithmic scale similar to the human ERB scale (Heffner & Heffner, 1982). Based on the differences between the pitch ranges of the human voice and elephant rumbles, the human ERB scale was shifted to lower frequencies by a factor of 10 to obtain an adapted elephant ERB scale. As may have been expected, pilot tests showed that the rumble detection algorithm performed better using this adapted ERB scale than when using the human ERB scale.

A bank of 32 gammatone filters with centre frequencies evenly distributed on the shifted ERB scale between 12 Hz and 30 Hz were used. Gammatone filters are defined in Katsiamis, Drakakis, and Lyon (2007), but for the present algorithm the implementation of the filters was done using the appropriate function from the Signal Processing Toolbox of Matlab R2007b (The MathWorks, Natick, MA, USA).

After filtering, the normalised autocorrelation, A, of each of the bands or channels, c, was calculated using Eq. (1) (Wu et al., 2003). The output of a specific filter is denoted as $r_c(t)$ in this equation, while N is the number of samples used for the calculation of the autocorrelation. The choice of N determines the resolution of the processed data. A window size of 40 ms was used, corresponding to $N = 128$ samples in a signal sampled at 3 kHz. The number of lag steps, τ, used within the autocorrelation determines the lowest frequency that may be detected (as explained in the next paragraph) and needs to be at least 300 samples to detect a frequency component of 10 Hz. The position in the centre of the presently processed window is denoted as j, and index n steps through this window of N samples.

$$A_c(j,\ \tau) = \frac{\sum_{n=-N/2}^{N/2} r_c(j+n) r_c(j+n+\tau)}{\sqrt{\sum_{n=-N/2}^{N/2} r_c(j+n)}\ \sqrt{\sum_{n=-N/2}^{N/2} r_c(j+n+\tau)}} \quad (1)$$

The numerator is the usual autocorrelation function, while the denominator performs normalisation.

The autocorrelation calculation was performed on each of the channels originating from the filter bank. The amplitude of the first positive peak in the correlogram of each channel gives an indication of the fundamental frequency component present in that channel and the maximum amplitude of the positive peaks in the correlogram gives an indication of the

Please cite this article in press as: Pieter J. Venter, Johan J. Hanekom, Automatic detection of African elephant (Loxodonta africana) infrasonic vocalisations from recordings, Biosystems Engineering (2010), doi:10.1016/j.biosystemseng.2010.04.001

amount of noise present in the channel for the analysed segment. Only channels with a maximum normalised amplitude of more than 0.945 were selected; a threshold established empirically in Wu et al. (2003). This method of channel selection enabled the algorithm to detect faint rumbles in recordings with severe broadband disturbances, since only channels with energy at specific pure tone frequencies were selected regardless of the strength of the pure tone component. A summed correlogram was calculated by adding all the autocorrelations of the selected channels together, forming large peaks where peaks in individual channels coincided. The largest positive peak in the summed correlogram was selected, and the corresponding pitch was calculated using Eq. (2). The number of lag steps at which the first positive peak is located can be used to determine the pitch of the input signal,

$$p = \frac{F_s}{s}, \tag{2}$$

with F_s the sampling frequency of the input signal and s the number of lag steps before the first positive peak occurs.

An elephant rumble occurs from around 2.3 s to around 5.3 s in the spectrogram shown in Fig. 1a. After performing the signal processing steps on this sound segment, the pitch estimates shown in Fig. 2a were calculated from the maximum peak in each of the summed correlograms. The pitch estimate was noisy when no rumble was present, but smoother where the rumble occurred. There were, however, some data points (e.g. around 4 s) where the algorithm detected pitch incorrectly.

Irrespective of the presence or absence of elephant rumbles, pitch estimates were obtained for each channel for all the recorded data. The second step of the algorithm was to (i) detect valid rumbles and extract only these sections containing elephant rumbles while discarding pitch estimates that were judged not to originate from rumbles, and (ii) to construct continuous pitch tracks in these sections. Comparing Figs. 1a and 2a, it should be noted that the sections of the estimated pitch where rumbles were present were smooth, but they also contained some discontinuities. This was typical of the available recordings.

A number of techniques were considered as candidates to track rumbles automatically from the pitch estimates. These include computationally expensive techniques involving the use of hidden Markov models (Paris & Jauffret, 2003; Xie & Evans, 1993), neural networks (Adams & Evans, 1994) and Kalman filters (Mustafa & Bruce, 2006). However, we noted that sections containing valid pitches were easy to locate by inspection from the pitch estimates of step 1. Mimicking this process of detecting valid pitches by inspection, a simple algorithm with low computational demands was developed. In brief, the algorithm scans through the pitch estimates, looking for smooth parts that indicates the presence of a rumble and automatically discards discontinuities within an identified smooth section if they are of short duration.

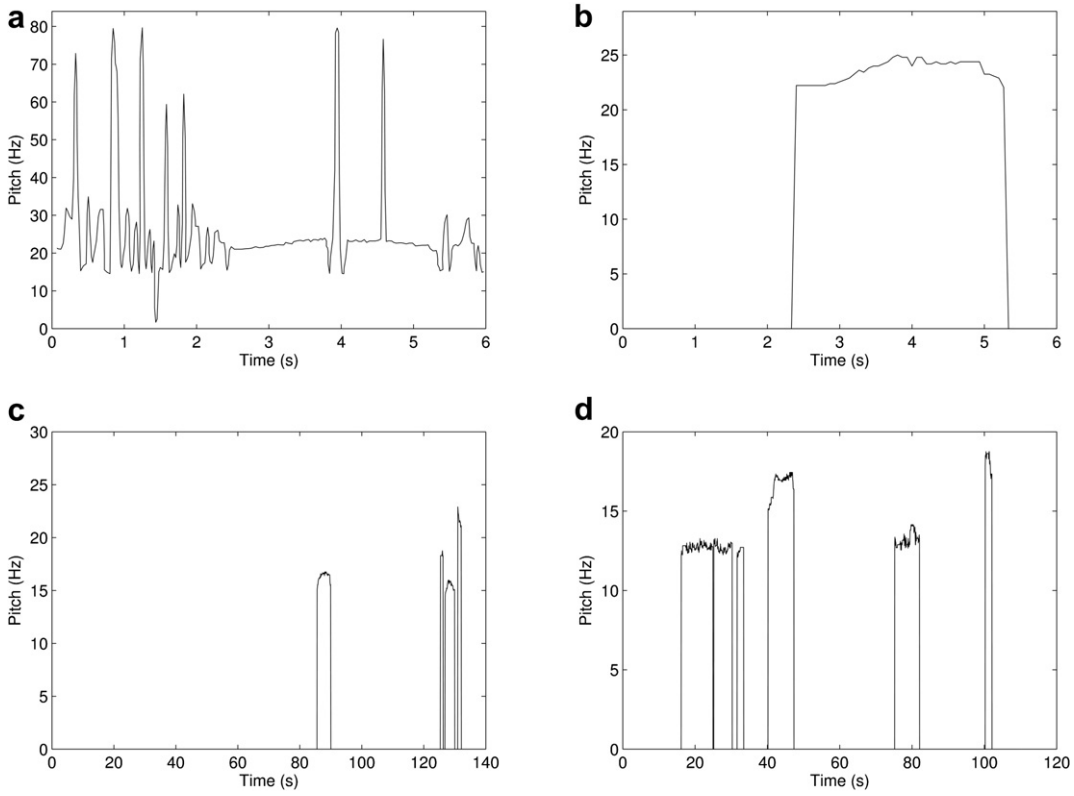


Fig. 2 – Sub-band pitch detector outputs. (a). Output of step 1 – pitch estimates as calculated for the sound segment shown in Fig. 1a before pitch tracking algorithm. (b). Output of step 2 for the sound segment of Fig. 1a. (c). Output of the pitch detector when the spectrogram of Fig. 1b was analysed. (d). Output of the pitch detector when the spectrogram of Fig. 1c was analysed.

Please cite this article in press as: Pieter J. Venter, Johan J. Hanekom, Automatic detection of African elephant (Loxodonta africana) infrasonic vocalisations from recordings, Biosystems Engineering (2010), doi:10.1016/j.biosystemseng.2010.04.001

More formally, this second step of the algorithm (pitch tracking) performs the following steps. Pitch estimates are stored in an array P, with an estimate being available for each sample of the original signal. The objective is to create a new array C containing only the pitch estimates of rumbles and zeros elsewhere. If the difference in pitch value of S consecutive samples in P is less than L Hz, those samples are considered part of a rumble and are transferred to C, Eq. (3).

$$C(k) = P(k) \quad for\ i \le k \le i + S, \tag{3}$$

where i is the index of the first sample of the rumble and k is a sample counter. As soon as two successive samples differ by more than L Hz, the next ST samples are examined to determine whether it is within LT Hz of the last valid pitch estimate. If so, all samples between these two valid pitch samples are equated to the last valid pitch estimate (Eq.(4)) to form a continuous pitch track,

$$C(k) = P(i + n) \quad for\ i + n \le k \le i + n + ST, \tag{4}$$

where the last valid sample was $i + n$ (n samples after the first sample of the rumble). Otherwise, the end of the rumble has been reached and all other samples set to zero. Parameter values were determined empirically to match the available data. They were $S = 14$ samples, $ST = 10$ samples, $L = 1$ Hz and $LT = 1.3$ Hz.

Together, the two steps of the algorithm are referred to as sub-band pitch detection in the rest of this article. The algorithm (step 1, sub-band pitch estimation and step 2, pitch tracking) was implemented in Matlab on a Windows PC with a 2.8 GHz processor. It took approximately 1 min of processing time for every minute of a recording to be analysed, although the processing time depended on the choice of a number of parameters. Halving the number of channels significantly reduced the analysis time, while reducing the accuracy of the algorithm only marginally. The autocorrelations that are performed on each channel were certainly the most computationally expensive part of the algorithm. Using fewer points in the autocorrelation function reduced processing time at the cost of having a higher minimum detectable pitch. Based on these considerations, the number of frequency channels arrived at was 32 and the correlations were 300 samples long. The sampling frequency of a recording did not have an impact on the processing time as the input was always re-sampled to 3 kHz.

To establish a performance baseline, the proposed sub-band pitch detector was compared to four other algorithms. Each of these methods requires some tuning of parameters for best results. In the energy threshold method, implemented using Ishmael software (Mellinger, 2001), the energy in the selected band of frequencies is determined and it is assumed that a vocalisation is present in the recording when the energy is greater than a given threshold. Typically, the threshold may be varied until an optimal value is found. This optimal value will depend on the intensity of background noise in a recording and on the distance from the elephants when the recording was made. For the present tests, following the suggestions in the help files of Ishmael, this threshold was set to just above the average energy of the sound. This is probably not an optimal setting for this particular problem.

Matched filtering is a second method that may be used to locate vocalisations. Here, the recording is cross-correlated with a recording of a standard vocalisation. Matched filtering is optimal for detecting a known signal in white Gaussian noise (e.g. Van Trees, 1968), but may be expected to be less successful when the signal to be detected differs somewhat from the template as would be the case for elephant vocalisations. A wide variety of vocalisations may be tested to see which would give the best results, but since elephant vocalisations are not identical, no single vocalisation will always provide good results. For the present application, one particular vocalisation was selected as the template. Both matched filtering and spectrogram correlation were tested in Ishmael 1.0 (free software, National Oceanic and Atmospheric Administration (NOAA), Washington DC, USA).

Spectrogram correlation is related to matched filtering. In this method, a spectrogram of the input recording is cross-correlated with a synthetic time-frequency template that mimics the pitch contour of the vocalisation. This technique has been used successfully for the detection of whale sounds in recordings (Mellinger & Clark, 2000). For the present work, a template was created to resemble typical elephant rumble data. Elephant rumbles typically rise in pitch initially, have constant pitch in the middle of the rumble and then drop in pitch towards the end of the rumble. Typical rumble durations vary between 0.5 s and 5 s. However, not all rumbles follow this pattern. The 3 s template for spectrogram correlation had a linear rise in frequency from 22.5 to 25 Hz between 0 and 0.5 s, a steady frequency of 25 Hz until 2.5 s, and a linear decline in frequency to 22.5 Hz over the last part of the template rumble.

Finally, a pitch tracking method, applicable to human speech (Boersma, 1993) and also based on autocorrelation but without the use of sub-bands, was tested using Praat 5.1.11 software (free software, Phonetic Sciences, University of Amsterdam, The Netherlands; Boersma & Weenink, 2009). A number of parameters must be specified; although applicable to humans, the recommended values as suggested in Boersma (1993) were used, with the exception of one parameter that was adjusted for elephants. Specifically, a pitch floor of 15 Hz and a pitch ceiling of 40 Hz were chosen; the (typical pitch range of elephant rumbles.

3. Results

Fig. 3 characterises the proposed sub-band pitch detector and compares it to the other methods explained earlier. As explained earlier, each method was defined as a two-step procedure. In the first step, pitch estimates were formed, while pitch tracking was done in the second step. For all the methods presented in this figure, the pitch track was obtained in exactly the same way as explained before (Eqs. (3) and (4)). Thus, the methods differed only in the pitch estimate step, providing a way to compare them directly. In each case, the input signal was an artificial harmonic complex with 25 Hz fundamental and with a varying number of equally strong harmonics band limited to 1500 Hz, and was presented in zero mean additive white Gaussian noise.

Matched filter A contained an exact (but noiseless replica) of the input 5-partial harmonic complex. Matched filter B

Please cite this article in press as: Pieter J. Venter, Johan J. Hanekom, Automatic detection of African elephant (Loxodonta africana) infrasonic vocalisations from recordings, Biosystems Engineering (2010), doi:10.1016/j.biosystemseng.2010.04.001

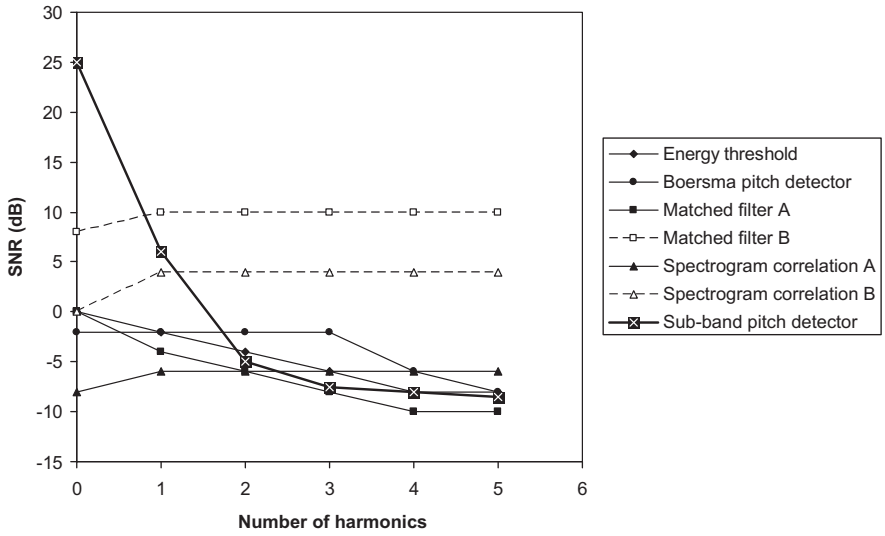


Fig. 3 – Performance of the various pitch detectors in noise as a function of the number of harmonics available when using an artificial harmonic complex as test signal. The SNR value indicated was that at which the pitch tracking first failed.

showed the result if the matched filter is offset slightly – the fundamental here was 24.8 Hz. Similarly, spectrogram correlations A and B considered the effect when the duration of the template and the rumble to be detected were either the same (A) or where the actual rumble was shorter than the template (B, 75% of template duration).

The SNR value, indicated on the ordinate, was that at which the pitch tracking first failed. E.g., when 3 harmonics were present, matched filter B fared worst (failed already at 10 dB SNR), while matched filter A performed best (pitch track was maintained down to −8 dB SNR). Note also that under the conditions of matched filter B, performance was best when no harmonics were present. This is because the frequency errors of the harmonics were greater than that of the fundamental.

The sub-band pitch detector performed almost as well as the matched filter, but (as will be explained later) is better suited to the conditions of the particular application. Fig. 3 shows that the pitch of a signal containing more than one harmonic can be tracked in noisy conditions with the proposed method. It may also be noted that the sub-band pitch detector fares worst of all the considered methods when no harmonics are present. It uses the sum of the autocorrelations of different channels to form a peak that is used for pitch detection. When fewer than two harmonics are present, the magnitude of this peak is not large enough to detect pitch in noisy recordings. This observation is useful for ignoring non-harmonic sounds.

The performance of the proposed algorithm on actual recordings may be predicted by its characteristics shown in Fig. 3, which indicates conditions under which a sound with harmonic structure will be detected. The pitch tracking output of step 2 when the sound segment of Fig. 1a was processed by the sub-band pitch detection algorithm is shown in Fig. 2b. This rumble possessed at least seven harmonics (Fig. 1a) and background noise approximating white noise was present. As expected from the characterisation curve in Fig. 3, the rumble was accurately detected. In addition, the algorithm determined the beginning of the rumble to be at 2.4 s, the rumble end at 5.2 s, the lowest pitch to be 22.3 Hz and the highest pitch 24.8 Hz. As a comment, calculation of pitch cannot easily be done by visual inspection of a spectrogram and may be valuable, since the pitch of a rumble can communicate information about the emotional state of an elephant (Soltis et al., 2005b).

From Fig. 3, it can be seen that narrow band noises without harmonic structure will allow detection of elephant rumbles. Fig. 1b shows a spectrogram of a recording with narrow band background noise present. Energetic components visible on the spectrogram as lighter areas include horizontal white lines visible between 20 and 60 Hz in the first 30 s as well as from 50 s to 150 s, these being unwanted narrow band noises produced by a car engine. The vertical lines visible throughout the spectrogram at different intensities are broadband noises typically caused by wind, breaking of branches or disturbance of the microphone itself. Four elephant rumbles with clearly defined harmonic structures can be seen on the spectrogram. The first rumble is located at approximately 85 s, while the other three rumbles are located in short succession starting at approximately 125 s Fig. 2c shows the output obtained after processing the recording with the sub-band pitch detector. Unwanted noises have been rejected, while the harmonic structures of the elephant rumbles have all been identified and their pitches extracted.

With some restrictions, given below, the elephant rumble detection algorithm reliably detected infrasonic elephant rumbles in noisy conditions from the available recordings. Unwanted low frequency sounds that did not have higher harmonics, such as distant motor vehicles, were also rejected.

However, it is clear from Fig. 3 that rumbles with weak harmonic structure will not be detected reliably. These rumbles are also hard to detect manually and are usually rumbles that have been recorded from a long distance. Fig. 1c shows the spectrogram of a recording with energetic undesired frequency components at 50 Hz and 75 Hz, along with the engine noises of a motor vehicle. Two elephant rumbles

Please cite this article in press as: Pieter J. Venter, Johan J. Hanekom, Automatic detection of African elephant (Loxodonta africana) infrasonic vocalisations from recordings, Biosystems Engineering (2010), doi:10.1016/j.biosystemseng.2010.04.001

occurred at 90 s and 100 s, but the energy of the unwanted components appears stronger than those of the rumbles on the spectrogram. The algorithm falsely detected (Fig. 2d) the engine noises as rumbles (because of the presence of the stronger harmonics) and at certain instants the pitch of the unwanted frequency components as well. The last of the two elephant rumbles was less faint and was correctly identified.

The quality of the recording of Fig. 1c was low. The sources of the unwanted frequency components appear to have been close to the microphone, while the elephant rumbles were distant. A recording as poor as this is unlikely to occur if an elephant call recording collar were to be used. Under these conditions, it appears that unwanted sounds with strong harmonics in the infrasonic band may cause false rumble detections, while distant elephant rumbles, which have lost their upper harmonics, may not be detected.

Finally, Fig. 4 shows the results from detections of actual recording of rumbles. The rumbles used here had to be analysed manually by inspecting spectrograms and by listening to accelerated recordings of the data. In total, 4 h of analysed data are reflected here, with the rumble detection task ranging from easy to difficult in these recordings. The results from the five methods are ordered from best to worst (i.e., most correct detections to least correct detections) in Fig. 4. Note that as the number of correct detections decreased, the number of false detections and missed rumbles increased.

4. Discussion

When comparing to previous research, possibly the only other published work that applied speech processing techniques to elephant vocalisations were studies by Clemins and co-workers (Clemins & Johnson, 2003; Clemins et al., 2005). Theirs was a different application than the present: they performed automatic classification of rumble type and speaker identification on a collection of vocalisations using techniques typically used for automatic speech recognition of human speech. Five elephant vocalisations were classified using Mel-frequency cepstral coefficients as features, and speaker recognition was achieved using hidden Markov models.

Similarly, techniques applicable to speech processing of human speech may be applied to the present task of detecting elephant rumbles from noisy recordings. Elephant rumbles are not unique, in the sense that they do not differ greatly from human speech in spectral structure. Both have a harmonic structure that may be exploited to detect vocalisations in noise. For example, in human speech, the harmonic structure is important in auditory stream segregation (Bregman, 1999) and this is taken advantage of in segregating simultaneous voices is speech enhancement applications (Hu & Wang, 2004). However, techniques that perform well under other circumstances do not necessarily perform well for the present task of detecting elephant rumbles in noisy recordings.

Considering then the efficacy of the proposed sub-band pitch detector in comparison to other previously applied methods, Fig. 3 characterised the proposed method and compared it to other methods. The figure shows that when a signal contains a few harmonics, with the exception of matched filter A, the sub-band pitch detector performs best. For signal in detection in noise, a matched filter is optimal. Thus, if the expected signal was known precisely, and the task was simply to detect the presence of this signal in noise, the matched filter would be expected to perform best. However, elephant rumbles vary in fundamental, harmonics and duration. When the expected signal (or the matched filter template) is slightly mistuned to the actual signal arriving in noise, this method fails, as is clearly shown in Fig. 3. Thus, matched filtering will work well when searching for a specific rumble in noise, but will not work if the rumbles in the recording differ, even slightly, from the template.

Spectrogram cross-correlation performs well when the signal to be detected varies little in duration (A), but fails if the template duration differs from the duration of the signal to be detected (B). Although this method had previously been applied successfully to whale sounds, the duration of a whale

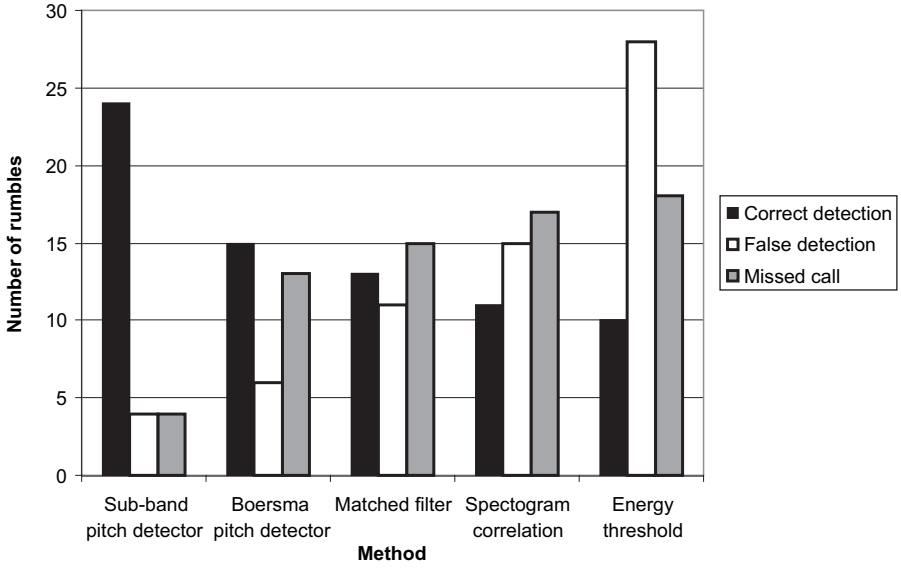


Fig. 4 – Performance of the various pitch detectors in noise when applied to recorded elephant rumbles.

Please cite this article in press as: Pieter J. Venter, Johan J. Hanekom, Automatic detection of African elephant (Loxodonta africana) infrasonic vocalisations from recordings, Biosystems Engineering (2010), doi:10.1016/j.biosystemseng.2010.04.001

endnote segment varies with less than 0.139 s (Mellinger & Clark, 2000). In contrast, elephant rumble duration in the present recordings varied over a range of 0.5–5 s, which makes spectrogram cross-correlation not ideal for this task.

The Boersma pitch detector (Boersma, 1993) was designed for human voice, but has been slightly adapted for elephant rumbles. This method works well for elephant rumbles with different durations and pitches. However, periodic sounds other than elephant rumbles (such as distant engine noises, or other sounds without harmonics) are falsely detected as valid pitches, as pitch is detected in noise even when no harmonics are present.

The energy thresholding method is simple and performs well as long as no other energetic sounds appear in a recording. However, this method cannot distinguish between harmonic complexes (elephant rumbles) and other loud sounds in recordings (e.g. elephants crashing through the bush, wind, or other animal sounds). Also, faint rumbles are not detected.

Fig. 3 showed that the sub-band pitch detection algorithm performed best when elephant rumbles in a recording have three or more clear harmonics. The algorithm could detect the pitch of a sound with harmonic structure in an SNR as low as −8 dB, which means that the algorithm should have good rejection of broadband noise like that caused by wind. This result also implies that the weakness of the algorithm lies in its inability to reject noises with a harmonic structure.

Thus, of the methods compared here, the sub-band pitch detector appears to be best suited for the task of detecting elephant rumbles. This is confirmed in Fig. 4, where actual rumbles from the recordings had to be detected. Although the sub-band pitch detector did miss some rumbles, faint rumbles in the presence of noise sources with harmonic content, it outperformed the other methods tested by some measure. Although this cannot be deduced from Fig. 4, none of the other techniques could detect the vocalisations that were missed by the sub-band pitch detector.

It should be noted that each of the methods requires some hand-tuning to perform best. Most of the tuning is required in step 2 (pitch tracking). Possibly, the pitch estimators compared to the sub-band pitch estimator may have performed better than shown in Fig. 4 if they were properly tuned. However, (i) Fig. 3 compares the methods directly (step 1 that does the pitch estimation) without the need for the tuning in step 2, and as such gives a good indication of which methods provide the best pitch estimates for step 2, and (ii) the shortcomings of each of the methods discussed in the previous paragraphs cannot be overcome with tuning.

The pitch tracking algorithm of step 2 (Eq. (1) and Eq. (2)) requires some further comments. Four parameters had to be obtained empirically from the available data. In essence, these parameters characterise the dynamics of pitch changes during a typical rumble. Although some tuning was necessary to obtain optimal parameters for the present data set, this is not unlike any other parametric pitch tracking algorithm. For example, if the pitch tracking were to be performed using a Kalman filter, it would have been necessary to infer the dynamics of typical rumble pitch tracks from available data in order to define the Kalman filter's internal model.

Finally, when two rumbles overlap in time, only the more energetic rumble is identified. The algorithm may be adapted slightly to allow tracking of two pitch tracks simultaneously. In the single-track algorithm, the sample in the summed correlogram with the maximum value is used to calculate each data point of the pitch estimate array. In the dual-track version, the two peaks in each summed correlation with the highest values are used to generate two pitch estimate arrays. The estimates contained in these two arrays are then processed in the same way as was the case for the single pitch estimated array, but two pitch tracks are processed. This allows for detecting overlapping rumbles within a recording. Either the single or the dual pitch track algorithm may be selected, depending on the need of the user.

5. Conclusions

A collection of typical elephant rumble recordings were analysed, where "typical" was defined as recordings containing only natural noises, but with the inclusion of sounds from far-off motor vehicles occasionally occuring. However, it should be noted that both the manual detection of elephant rumbles and the definition of "typical" elephant rumble recordings are subjective to some extent.

Irrespective of this integral subjectivity, some generalisations may be made about the proposed sub-band pitch detection algorithm. In summary, sub-band pitch detection worked reliably for tracking of elephant rumbles in most conditions tested, but not in all. Weak rumbles were not detected when their harmonic structure was lost, and false alarms increased when the pitch of overlapping rumbles had to be detected using the dual-track option of the algorithm. In applications where all possible candidate rumbles need to be detected in a recording, the dual track algorithm may prove valuable despite the likely increase in false alarms.

The practical value of the proposed elephant rumble detection algorithm will only be established through extensive testing with recorded data and comparison with manual identification of rumbles. A next step that may have potential value to elephant call researchers would be to automatically detect rumbles from a large collection of unprocessed recordings (like those obtained from an elephant recording collar), and automatically isolate, label (e.g. with time stamps) and save detected calls to individual sound files. The algorithm may also provide a foundation for further signal processing steps, e.g. the automatic classification of rumble types and speaker identification, as has been carried out by Clemins and Johnson (2003) using previously identified rumbles.

REFERENCES

Adams, G. J., & Evans, R. J. (1994). Neural networks for frequency line tracking. *IEEE Transactions on Signal Processing, 42*, 936–941.

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

Boersma, P., & Weenink, D. (2009). Praat: doing phonetics by computer (version 5.1.11) [computer software]. Available from http://www.praat.org/.

Please cite this article in press as: Pieter J. Venter, Johan J. Hanekom, Automatic detection of African elephant (Loxodonta africana) infrasonic vocalisations from recordings, Biosystems Engineering (2010), doi:10.1016/j.biosystemseng.2010.04.001

Bradbury, J. W., & Vehrencamp, S. L. (1998). *Principles of animal communications*. Sinauer Associates.

Bregman, A. S. (1999). *Auditory scene analysis. The perceptual organization of sound* (2nd ed.). Cambridge, Massachusetts: MIT Press.

Clemins, P. J., & Johnson, M. T. (2003). Application of speech recognition to African elephant (Loxodonta Africana) vocalizations. In *Proceedings of ICASSP* (pp 484–487). Hong Kong: IEEE International Conference on Acoustics, Speech and Signal Processing.

Clemins, P. J., Johnson, M. T., Leong, K. M., & Savage, A. (2005). Automatic classification and speaker identification of African elephant (Loxodonta africana) vocalizations. *Journal of the Acoustical Society of America, 117*, 956–963.

Garstang, M. (2004). Long-distance, low-frequency elephant communication. *Journal of Comparative Physiology A: Neuroethology, Sensory, Neural, and Behavioral Physiology, 190*, 791–805.

Heffner, R. S., & Heffner, H. E. (1982). Hearing in the elephant (Elephas maximus): absolute sensitivity, frequency discrimination, and sound localization. *Journal of Comparative and Physiological Psychology, 96*, 926–944.

Hu, G., & Wang, D. L. (2004). Monaural speech segregation based on pitch tracking and amplitude modulation. *IEEE Transactions on Neural Networks, 15*, 1135–1150.

Johannesma, P. I. M. (1972). The pre-response stimulus ensemble of neurons in the cochlear nucleus. In *Symposium on hearing theory* (pp 58–69), Eindhoven, Holland.

Katsiamis, A. G., Drakakis, E. M., & Lyon, R. F. (2007). Practical gammatone-like filters for auditory processing. *Eurasip Journal on Audio, Speech, and Music Processing*. art. no. 63685.

Kim, Y. J., & Chung, J. H. (2004). Pitch synchronous cepstrum for robust speaker recognition over telephone channels. *Electronics Letters, 40*, 207–209.

Langbauer, J. (2000). Elephant communication. *Zoo Biology, 19*, 425–445.

Langbauer, W. R., Jr., Payne, K. B., Charif, R. A., & Thomas, E. M. (1989). Responses of captive African elephants to playback of low-frequency calls. *Canadian Journal of Zoology, 67*, 2604–2607.

Lee, K., & Ellis, D. P. W. (2006). Voice activity detection in personal audio recordings using autocorrelogram compensation. In *Interspeech 2006 and ninth international conference on spoken language processing* (pp 1970–1973), Pittsburgh, PA, USA.

Leong, K. M., Burks, K., Rizkalla, C. E., & Savage, A. (2005). Effects of reproductive and social context on vocal communication in captive female African elephants (Loxodonta africana). *Zoo Biology, 24*, 331–347.

Leong, K. M., Ortolani, A., Burks, K. D., Mellen, J. D., & Savage, A. (2002). Quantifying acoustic and temporal characteristics of vocalizations for a group of captive African elephants Loxodonta africana. *Bioacoustics, 13*, 213–231.

McComb, K., Reby, D., Baker, L., Moss, C., & Sayialel, S. (2003). Long-distance communication of acoustic cues to social identity in African elephants. *Animal Behaviour, 65*, 317–329.

Mellinger, D. K. (2001). Ishmael: integrated system for holistic multi-channel acoustic exploration and localization (version 1.0) [computer software]. Available from http://www.pmel.noaa.gov/vents/acoustics/whales/ishmael/index.html.

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

Mustafa, K., & Bruce, I. C. (2006). Robust formant tracking for continuous speech with speaker variability. *IEEE Transactions on Audio, Speech and Language Processing, 14*, 435–444.

Noll, A. M. (1967). Cepstrum pitch determination. *Journal of the Acoustical Society of America, 41*, 293–309.

O'Connell-Rodwell, C. E., Arnason, B. T., & Hart, L. A. (2000). Seismic properties of Asian elephant (Elephas maximus) vocalizations and locomotion. *Journal of the Acoustical Society of America, 108*, 3066–3072.

Paris, S., & Jauffret, C. (2003). Frequency line tracking using HMM-based schemes. *IEEE Transactions on Aerospace and Electronic Systems, 39*, 439–449.

Payne, K. B., Thompson, M., & Kramer, L. (2003). Elephant calling patterns as indicators of group size and composition: the basis for an acoustic monitoring system. *African Journal of Ecology, 41*, 99–107.

Poole, J. H. (1999). Signals and assessment in African elephants: evidence from playback experiments. *Animal Behaviour, 58*, 185–193.

Poole, J. H., Tyack, P. L., Stoeger-Horwath, A. S., & Watwood, S. (2005). Elephants are capable of vocal learning. *Nature, 434*, 455–456.

Reuter, T., Nummela, S., & Hemila, S. (1998). Elephant hearing. *Journal of the Acoustical Society of America, 104*, 1122–1123.

Soltis, J., Leong, K., & Savage, A. (2005a). African elephant vocal communication I: antiphonal calling behaviour among affiliated females. *Animal Behaviour, 70*, 579–587.

Soltis, J., Leong, K., & Savage, A. (2005b). African elephant vocal communication II: rumble variation reflects the individual identity and emotional state of callers. *Animal Behaviour, 70*, 589–599.

Takagi, T., Seiyama, N., & Miyasaka, E. (2000). Method for pitch extraction of speech signals using autocorrelation functions through multiple window lengths. *Electronics and Communications in Japan, Part III: Fundamental Electronic Science (English Translation of Denshi Tsushin Gakkai Ronbunshi), 83*, 67–79.

Titze, I. R. (1994). *Principles of voice communication*. Englewood Cliffs, NJ: Prentice-Hall.

Van Trees, H. L. (1968)*Detection, estimation and modulation theory, Vol. 1*. New York: Wiley.

Wood, J. D., McCowan, B., Langbauer, J., Viljoen, J. J., & Hart, L. A. (2005). Classification of African elephant Loxodonta africana rumbles using acoustic parameters and cluster analysis. *Bioacoustics, 15*, 143–161.

Wu, M., Wang, D. L., & Brown, G. J. (2003). A multipitch tracking algorithm for noisy speech. *IEEE Transactions on Speech and Audio Processing, 11*, 229–241.

Xie, X., & Evans, R. J. (1993). Multiple frequency line tracking with hidden Markov models – further results. *IEEE Transactions on Signal Processing, 41*, 334–343.

Zhang, T., Zhang, Z., Lin, X., & Quan, J. (2006). Power spectrum reprocessing algorithm for pitch detection of speech. *Jisuanji Gongcheng/Computer Engineering, 32*, 1–3.

Zhao, W. W., & Ogunfunmi, T. (1999). Formant and pitch detection using time-frequency distribution. *International Journal of Speech Technology, 3*, 35–49.

Please cite this article in press as: Pieter J. Venter, Johan J. Hanekom, Automatic detection of African elephant (Loxodonta africana) infrasonic vocalisations from recordings, Biosystems Engineering (2010), doi:10.1016/j.biosystemseng.2010.04.001