International Journal of Scientific and Technical Advancements
ISSN: 2454-1532

Synthesis of Elephant Calls Using HNM and Effect of Voice Part on Call Quality

Arun Jasrotia[1], Randhir Singh[1], Parveen Lehana[2#], Saleem Khan[2]
[1]Sri SAI College of Engineering & Technology, Punjab, India
[2]Dept. of Physics & Electronics, University of Jammu, Jammu
[#]Email address:pklehanajournals@gmail.com

Abstract— Elephants have a well-developed system of communication. Acoustic communication is most important among other forms of communication used by elephant as they can generate both very high and very low frequency signals. In this research paper, elephant calls are synthesized using harmonic plus noise model (HNM) technique. Five calls of Asian elephant are recorded with high quality recording system which is used as source signal for synthesis process. Quality of the synthesized elephant calls has been tested using ITU-T standard perceptual evaluation of speech quality (PESQ). Also effect of voice part on the quality of synthesized elephant calls has been investigated. Various levels of the voice part ranging from 1% to 10% and constant noise part 10% is taken. Results shows that the PESQ score of the synthesized calls increases till the voice and noise part are equal in magnitude.

Keywords— HNM; PESQ; Voice; Nois.

I. INTRODUCTION

Elephants are highly social animals, which have well developed communication that uses all of their senses: hearing, smell, vision and touch [1]. Elephants produce a broad range of sounds from very low frequency to higher frequency calls. Elephant calls can be divided into three categories: laryngeal, trunk, and imitated calls.

Laryngeal Call: Sounds originating in the larynx are referred to as laryngeal calls such as rumbles, revs, roars, cries, grunts, barks and husky-cries. Among these calls rumbles are the most frequently heard call type across both sexes and all ages of elephants. This call may be easily distinguished from other call types by their very low frequencies and clear harmonic structure. Rev is a short tonal harmonic vocalization, less than a second in duration, and almost always followed immediately by a rumble. Its fundamental frequency is between 50 and 90 Hz which is significantly higher than any known rumble [2-4].

Trunk calls: sound originating from the long muscular trunk falls in this category like trumpet, Nasal-Trumpet, and, snort. Trumpets are produced by expulsion of air through trunk. These are tional sounds with harmonics overlaid with noise. Nasal trumpets sounds are noisier than trumpets having low frequency. Intense social excitement may elicit nasal trumpets. Another form of trunk call is snort. These are short, noisy, broadband sounds produced by blowing air through trunk. Elephants snort when they are surprised by something.

Imitated calls: elephants have the capacity of vocal production learning or imitation. These calls include croaking, squelching, purring and truck-like sounds. Croaking is pulsating sound lasting between 1 and 10 seconds. These sounds are produced by sucking of water or odors into the mouth and usually occurred in a series of two or three croaks. Squelching is produced by forcing air through a “scrunched-up” trunk. These sounds are most often heard when elephants are relaxed [5-7].

In this research paper elephant calls are synthesized using HNM technique. Analyses of the synthesized calls are carried out using PESQ score obtained with respect to the originally recorded elephant calls. Also investigations are carried out evaluate the effect of level of voice part on the synthesized calls.

HNM decomposes signal into harmonic component and a noise component. HNM based analysis and synthesis is performed in a frame-by-frame basis [8]. If a frame is unvoiced, the signal is only modeled by the noise part as an AR process. When the frame is voiced, then speech is modeled as the sum of two components:

$$s(n) = h(n) + u(n) \tag{1}$$

where h(n) is the harmonic part and u(n) is the noise part. Harmonic part is described by a sum of harmonically related sinusoids:

$$h(n) = \sum_{k=-L}^{L} a_k e^{j2\pi k(f_0/f_s)n} \tag{2}$$

where L denotes the number of harmonics, f_0 denotes the fundamental frequency, a_k are the complex amplitude of the k^{th} harmonic and f_s is the sampling frequency [9-10].

The estimation of the unknown complex amplitudes, a_k, is obtained by minimizing a weighted time domain least-squares criterion with respect to a_k,

$$\varepsilon_\alpha = \sum_{n=-T}^{n=T} w^2(n)(s(n) - h(n))^2 \tag{3}$$

where s(n) denotes the original speech signal, h(n) denotes the harmonic signal, w(n) denotes the weighted window.

II. METHODOLOGY

Research work is carried out to synthesize and analyses elephant calls using HNM technique. Also the effect of the level of sound part on synthesised elephant call is investigated.

Arun Jasrotia, Randhir Singh, Parveen Lehana and Saleem Khan, “Synthesis of elephant calls using HNM and effect of voice part on call quality,” *Journal of Scientific and Technical Advancements*, Volume 1, Issue 2, pp. 9-,11 2015.

Perceptual Evaluation of Speech Quality (PESQ) test is used for the comparison of the original and synthesized elephant calls. The block diagram of the methodology is shown in Fig. 1. A high quality sound recording system is used to record the sound produced by Asian elephant. Four sound signals are taken are source to be synthesized the calls using HNM techniques. Also a range of voice level from 1 to 10 was taken with constant noise level to evaluate the effect on the synthesized calls.

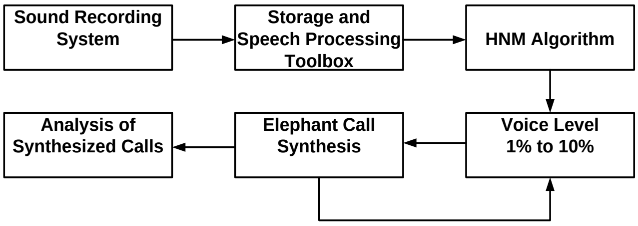

Fig. 1. Schematic of methodology.

III. Results

Figure 2 (a-e) shows the normalized recorded elephant calls and their respective spectrogram. These calls were used as source for synthesizing calls using HNM technique. Figure 3 shows the PESQ score of the synthesized elephant calls at 10% of voice and noise part. After synthesizing the calls effect of voice part on the call generation is investigated. The computed PESQ score of five synthesized elephant calls with various level of voice part with constant noise part are given in Table I. Figure 4 shows the pictorial view of the PESQ score of synthesized elephant calls with different voice level in synthesis process. In the graph x-axis represents the voice to noise level in HNM ranging from 1% to 10% and y-axis is the magnitude of the PESQ score. The inference drawn from the results is that at lower voice level, the PESQ score obtained is below 2 which represent very poor quality of synthesized call. As the voice level is increased the PESQ score reaches a magnitude of approximately 2.5 at unity ratio of voice and noise level, which is acceptable range of PESQ. This provides evidence that HNM model works well with elephant call generation. The quality of the elephants calls obtained shows a gradual increase until the ratio of voice and noise part approaches unity.

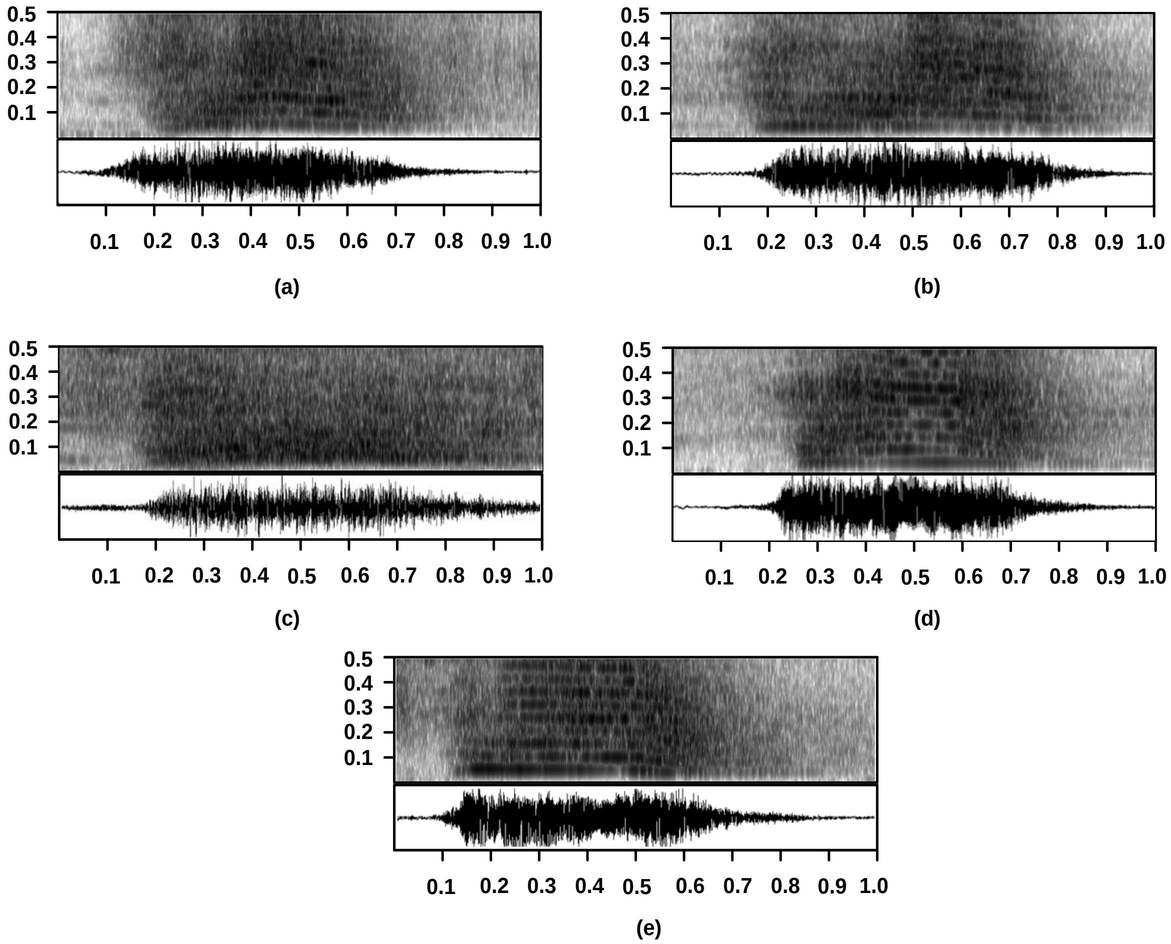

Fig. 2(a-e). Normalized signal and spectrogram of five different calls of Asian elephant.

Arun Jasrotia, Randhir Singh, Parveen Lehana and Saleem Khan, “Synthesis of elephant calls using HNM and effect of voice part on call quality,” *Journal of Scientific and Technical Advancements*, Volume 1, Issue 2, pp. 9-,11 2015.

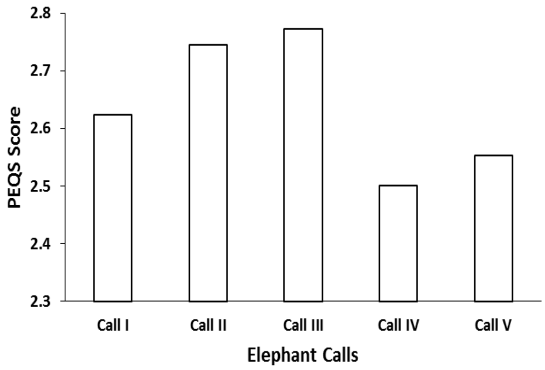

Fig. 3. PESQ score of five different synthesized calls w.r.t original call.

Table I. Computed PESQ score for different calls

Voice: Noise Part Ratio	PESQ Score				
	Call I	Call II	Call III	Call IV	Call V
v1n10	2.1668	2.2086	1.7700	1.9551	1.7423
v2n10	2.1983	2.5095	2.3925	2.1052	1.9951
v3n10	2.2017	2.6430	2.5919	2.063	2.2573
v4n10	2.3297	2.7304	2.5786	2.0566	2.3005
v5n10	2.3053	2.7493	2.6193	2.1800	2.2952
v6n10	2.4021	2.7473	2.6178	2.1538	2.2712
v7n10	2.3595	2.7249	2.6548	2.1518	2.3877
v8n10	2.3409	2.7756	2.6664	2.2307	2.4170
v9n10	2.4620	2.8010	2.6594	2.3410	2.5886
v10n10	2.4238	2.7445	2.7722	2.5010	2.5530

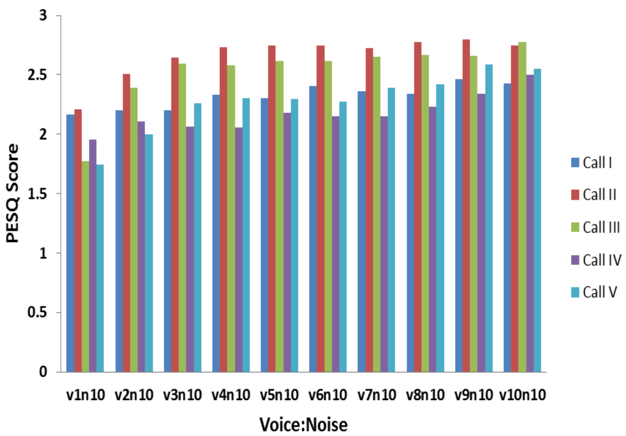

Fig. 4. PESQ score obtained at different level of voice and constant noise level.

IV. Conclusion

A preliminary investigation of Asian elephant calls synthesis using HNM technique has been presented. The effect of the proportion of voice part on the synthesized elephant call quality has been discussed. PESQ is used as the evaluation method for the quality elephant calls. The quality of the synthesized call obtained shows a gradual increase until the value of voice part is 10% and noise part is 10%. From the results it is quite apparent that HNM model is an efficient model for synthesizing elephant calls. Future studies are required to gain insight of acoustic signals generated by Asian elephants and different techniques to synthesis the calls.

References

[1] T. Reuter, S. Nummela, and S. Hemila, "Elephant hearing," *J. Acoust. Soc. Am.*, vol. 104, pp.1122-1123, 1998.

[2] C. O'Connell-Rodwell, J. D. Wood, C. Kinzley, T. C. Rodwell, J. Poole, and S. Puria, "Wild African elephants (Loxodonta africana) discriminate between familiar and unfamiliar conspecific seismic alarm calls," *J. Acoust. Soc. Am.*, vol. 122, pp. 823-830, 2007.

[3] C. O'Connell-Rodwell, B. T. Arnason, and L. A. Hart, "Seismic properties of Asian elephant (Elephas maximus) vocalizations and locomotion," *Acoust. Soc. Am.*, vol. 108, pp. 3066-3072, 2000.

[4] C. F. Eyring, "Jungle acoustics," *J. Acoust. Soc. Am.*, vol. 18, pp.257-270, 1946.

[5] U. Ingard, "A review of the influence of meteorological conditions on sound propagation," *J. Acoust. Soc. Am.*, vol. 25, pp. 405-411, 1953.

[6] W. R. Jr. Langbauer, K. B. Payne, R. A. Charif, L. Rapaport, and F. Osborn, "African elephants respond to distant playbacks of low-frequency conspecific calls," *Journal Experimental Biology*, 157, 35-46, 1991.

[7] W. R. Langbauer Jr., "Elephant communication," Zoo Biology, vol. 19, pp. 425-445, 2000.

[8] Y. Pantazis and Y. Stylianou, "Improving the modeling of the noise part in the harmonic plus noise model of speech," *in Proc. IEEE Acoustics, Speech and Signal Processing*, 2008.

[9] Y. Stylianou J. Laroche, and E. Moulines, "High quality speech modification based on a harmonic + noise model," EUROSPEECH, 1995.

[10] Y. Stylianou and C. Olivier, "A system for voice conversion based on probabilistic classification and a harmonic plus noise model," *in Proc. IEEE Acoustics, Speech and Signal Processing*, vol. 1, 1998.

Arun Jasrotia, Randhir Singh, Parveen Lehana and Saleem Khan, "Synthesis of elephant calls using HNM and effect of voice part on call quality," *Journal of Scientific and Technical Advancements*, Volume 1, Issue 2, pp. 9-,11 2015.