A COMPUTER-BASED ANALYSIS OF ACOUSTIC PROPERTIES OF SRI LANKAN PIRITH CHANTS USING VOICED TO UNVOICED RATIO AND PROBABILITY DISTRIBUTION FUNCTIONS

Chamali Gunawardana[1] and Shantha Gamage[2]
[1,2]Department of Physics, University of Sri Jayewardenepura, Sri Lanka
[2]*shanthagamage@sci.sjp.ac.lk*

Abstract - Pirith is a special type of chanting; believed to be a protective doctrine preached by the Load Buddha in Pali language. Generally, a voice signal consists of the fundamental frequency, F_0 and a series of harmonic frequencies called as formants, F_n. As reported by several other studies, characteristic formant frequency distributions are identified in chanting, contrast to normal speaking. This work is dedicated to developing a phonetic picture on Pirith chants investigating probability distribution of formants and quantify voiced to unvoiced ratio utilizing computer-aided tools. In this study, 25 samples of each Rathana, Karaneeya Metta and Angulimala Suttas recited by male monk chanters were recorded using high precision microphone array and then subjected to splitting of smaller voiced segments of frame length 10*ms* sampling at a rate of 44.1 *kHz*. In the computational speech model, a pre-emphasis filter is applied to the sampled time series of voiced segment to cancel out the effect of glottis. Then frame-by-frame analysis was used with hamming windows and liner predictive coding (LPC) and auto correlation to extract the formant values. Voiced to unvoiced ratio is assessed using zero crossing rate and energy content of the acoustic signal. Results of the Voiced to Unvoiced ratio over 75% of voiced frames in all types of Suttas despite number of monks involved in chanting. Having a high percentage of voiced frames interpret strong contribution of vibrating vocal folds involved in chanting of Pirith Suttas. Further, Probability Distribution Functions (PDFs) of each Sutta is generated and compared for first five formants. Angulimala Sutta and Ratana Sutta show similar patterns in terms of PDFs while Karaneeya Metta Sutta indicates a clear discrepancy demonstrating a unique set of characteristics.

Keywords - Formant frequencies, Voiced to unvoiced ratio, Probability distribution functions

I. INTRODUCTION

Speech production process begins at the point of converting an idea developed in the speakers' mind to a language code. With the aid of articulatory motion and vocal tract movement, the phonemes which are lined up in a set of sequences propagate outside as an acoustic waveform. *Pirith* means protection from all aspects and this protection is to be obtained by reciting or listening to *Pirith suttas*. The practice of reciting and listening to *Pirith suttas* began very early in the history of Buddhist culture.

Voiced to unvoiced ratio (V/UV ratio) indicates the involvement of speech production system with vibration of vocal codes. In this work, we combined the results of zero crossing rate (ZCR) and energy of short time segments of the signal to generate V/UV ratio. In voiced speech, the vibrating glottis generates periodic pulses which are resonate in the vocal tract. However, in the unvoiced speech, vocal chords held open and a continuous air beam flow through them. (Lee and Yoo, 2003). The zero crossing rate measures number of intersections a given signal makes with the time axis per unit time in an amplitude- time plot. Voiced speech shows a low zero-crossing rate due to the excitation of vocal tract by the periodic air flow, whereas the unvoiced speech shows high zero-crossing count as it is produced by the turbulent airflow flowed through the narrowed vocal tract (Bachu et al., 2010). Additionally, the voiced part of the speech has high energy content because of its periodicity.

According to the acoustic theory of speech production, vocal tract is modelled as a non-uniform tube closed at vocal folds and open at the lip end (Stevens and House 1955). Due to varying cross section along the vocal tract, different resonance frequencies (harmonics) are generated in response to varying vocal fold vibrations. Consequently, the complex output voice signal is composed of several harmonics called as formants (Fant,1973).

As reported by Jayaratne 2007, an experiment was performed at Kanduboda International Meditation Centre, Sri Lanka to understand the effect of *Pirith* on human beings. When a sample of human subjects could listen to *Pirith* chants, it is observed that within 10 minutes of the commencement of the chanting, their heart beat reduced, heart pulse amplitude halved and reached to an alpha state similar to what is obtained under a meditative trance. A famous Japanese researcher, Masaru Emoto had provided evidences that human thoughts can affect the molecular structure of water through words, ideas and human vibrational energy (healingsounds.com, Jan 2018).

In the literature, there are evidences a large majority of the studies carried out so far in this discipline, have used a qualitative approach to tackle the problem. Instead, we chose a rigorous but quantitative approach which exploits state-of-the-art computer aided tools such as high- performance computing facilities (supercomputer cluster) to evaluate acoustic properties of a vast range of *Pirith* chants.

II. METHODOLOGY AND EXPERIMENTAL DESIGN

In the analysing process, Samples of *Rathana, Karaneeya Metta* and *Angulimala Suttas* recited by male monk chanters were recorded using high precision microphone array and 25 samples of each *Sutta* were *analysed*. Recorded samples were then subjected to splitting of smaller voiced segments of frame length 10 ms using sampling rate of 44.1 kHz. This specific frame length was selected as vocal tract has fixed characteristics over a time interval of the order of 10 ms.

Fast Fourier Transformation (*FFT*) and Linear Predictive Coding (*LPC*) are techniques used in spectral analysis of the speech. Fast Fourier Transformation(*FFT*) develops a spectrum by decomposing a sound wave into sinusoidal components whereas Linear Predictive Coding (*LPC*) estimates formant frequencies associated with vocal tract.

The formant frequency estimation was performed using computational method as follows. A pre-emphasis filter is applied to the sampled time series of voiced segment to cancel out the effect of glottis. Then *frame-by-frame analysis* was used with *hamming windows* and *liner predictive coding* (*LPC*) and auto correlation to extract the formant values. Speech signal has been modelled as a combination of a source and a filter. Source-filter separation model is use as a fundamental method for formant frequency estimation. The modelled system and its frequency resonances are only considerable in this estimation and Liner Predictive coding (LPC) is used to find the best matching system. The LPC filter is a function with set of filter coefficients. Resonance of the filter is expressed by a pair of coefficient. As in every 10 ms vocal tract parameters are changed, creating new set of coefficients. When applying LPC, a speech sample approximated as a linear combination of past speech samples. Minimizing the sum of squared differences over a 10 ms frame between actual sample and linearly predicted sample, a set of predictor coefficients can be obtained.

According to the discrete-time model,

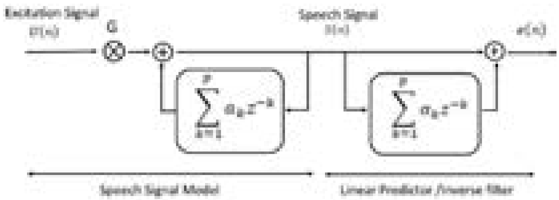

Figure 1- Linear Prediction model and

$$S(n) = \sum \alpha_k S(n-k) + GU(n)$$

k=1

where, $S(n)$: Estimated current value, αk : Predicted coefficients, $S(n-k)$: Past samples and $U(n)$: Excitation of vocal folds

Between two pitch pluses $U(n)$ is zero. Therefore, nth speech sample can be written as a linear combination as follows,

$$S(n) = \alpha_1 S(n-1) + \alpha_2 S(n-2)$$

$$+ \alpha_3 S(n-3) \ldots \ldots . . + \alpha_p S(n-p)$$

Figure1 demonstrate how the prediction error ($e(n)$) is defined. By minimizing the square of the error $\{e(n)2\}$ filter coefficients can be generated. After finding the locations of the resonance to extract the formant frequencies from the filter, the filter coefficients were treated as a polynomial and solved for the roots of the polynomial.

Voiced to unvoiced ratio is calculated by counting number of frames less than a reference zero-crossing rate and higher than a reference short time energy as voiced frames and others as unvoiced frames according to the algorithm shown in Figure 2. Zero-crossing rate is defined as the ratio between number of intersections given signal makes with time axis to number of sampled data points in the signal. The upper threshold zero crossing rate for qualifying a voiced frame is 0.1, greater than 0.1 but less than 0.3 for a silent frame and over 0.3 for an unvoiced frame. Probability Distribution Function (PDF) based quantitative analysis is carried out to identify the occurrence of formants. First five formant frequencies were extracted from the recorded voiced track with the aid of computer program scripted in MATLAB and corresponding PDFs are determined.

III. RESULTS AND DISCUSSION

MATLAB is used for scripting, calculations and analysis. In the frame by frame analysis, speech signals are divided into a non-overlapping frame of samples.

A. Voiced to unvoiced ratio

Voiced to unvoiced ratio is developed based on the algorithm shown in Fig 1 and corresponding results are summarized in Table 1. All types of chant samples demonstrated over 75% of voiced frames despite number of monks involved in chanting. Periodically vibrating vocal folds cause generation of voiced speech and hence, having a high percentage of voiced frames interpret strong contribution of vibrating vocal folds involved in chanting of *Pirith Suttas*. Usually, in speech, silent regions exist in between voiced and unvoiced regions yet without existence of silent regions the speech will not be

Table 1- Voiced to Unvoiced ratio and percentage of silent frames for Pirith Suttas

Name of the *Pirith Sutta*	Voiced to Unvoiced ratio	Percentage of silent frames
Rathana Sutta	87:1	4.68
Karaneeya Metta Sutta	88:1	2.7371
Angulimala Sutta	76:1	26.64

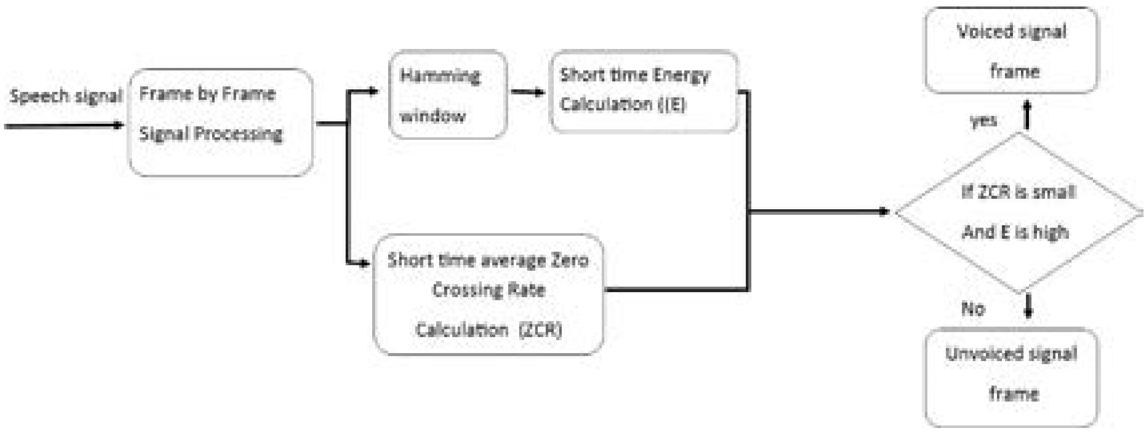

Figure 2 The algorithm to recognize Voiced to Unvoiced ratio

A. Probability Distribution Function (PDF)

In figures 3,4 and 5 the resulting variation of PDFs are shown. In these figures red, gray, black, blue and green solid lines show PDFs corresponding to formant frequency 1, 2, 3, 4 and 5 respectively. In Fig 6 the comparison for all suttas is shown. Plot dot shapes square, round and triangle indicate PDFs related to *Angulimala Sutta*, *Karaneeya Metta Sutta* and *Rathana Sutta* respectively.

Most common features of all Suttas are representing a considerable amount of probability for final three formants and probability of occurring second formant is comparatively lower.

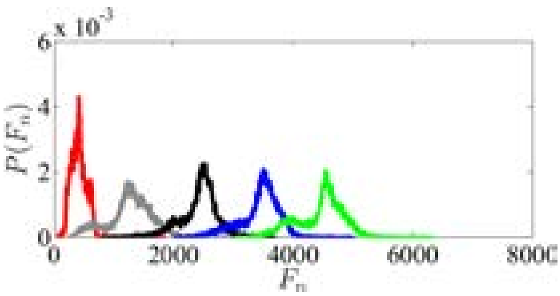

Figure 3. The figure represents probability of format frequencies of first five formants of Ratana Sutta. $f1$, $f2$, $f3$ $f4$ and $f5$ represent by red, ash, black, blue and green colour plots respectively. First formant ($f1$) shows highest probability of occurrence while third ($f3$) forth ($f4$) and fifth ($f5$) formants indicate second highest probability. Second formant ($f2$) shows a considerable low probability

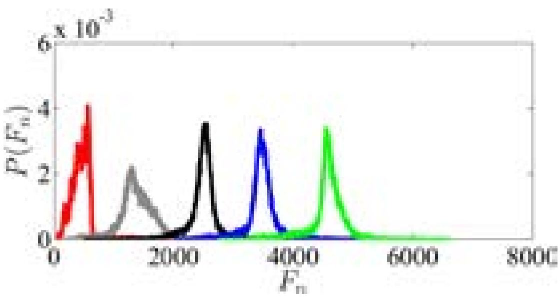

Figure 4. Probability distribution for Karaneeya Sutta as per contributed by first FIVE frequencies, $f1$, $f2$, $f3$ $f4$ and $f5$ represented by red, ash, black, blue and green colour plots respectively. In Karaneeya Sutta $f1$, $f3$ $f5$ and $f4$ shows probability in descending order while $f2$ remains at the least probability

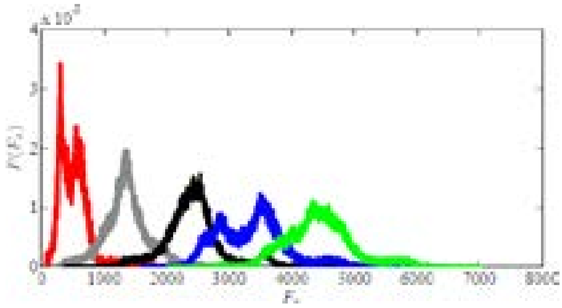

Figure 5. Probability distribution for Angulimala Sutta as per contributed by first FIVE frequencies, $f1$, $f2$, $f3$ $f4$ and $f5$ represented by red, ash, black, blue and green colour plots respectively. Occurring probabilities of formants, decrease continuously from $f1$ to f

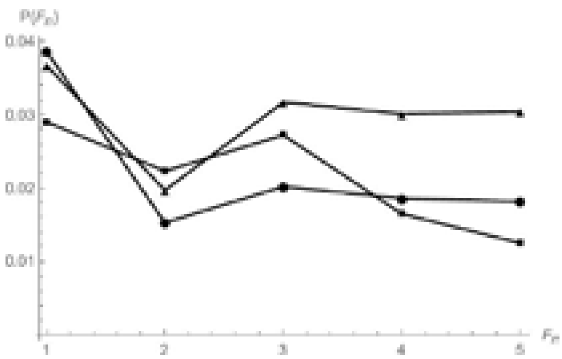

Figure 6. Mean Probability Distribution for all three kinds of Suttas. Plot dot shapes square, round and triangle indicate PDFs related to Angulimala Sutta, Karaneeya Metta Sutta and Rathana Sutta respectively

VI. CONCLUSION

As per deduced from the analysis of voiced to unvoiced ratio, demonstrating over 75% of voiced frames in all Pirith samples reveal a strong contribution of vocal folds in Pirith chanting with high vowel pronouncing probability.

The relative occurrence of each formant frequency varies according to one of the following two patterns shown below:

- First formant, $f1$ shows the highest probability of occurrence while $f3$, $f4$ and $f5$ contribute to probabilities in descending order while $f2$ remains at the least probability of occurrence
- Probability of occurrence of formants, decrease continuously from f to f.

In terms of third and fourth formant frequencies is more related to lip-spreading and to lip-protrusion. As reported previously and can be made much stronger in singing than in speaking. Trained singers can manipulate and by lowering the larynx and elevating the tongue blade to enhance this part of the spectrum and make it heard above an orchestral accompaniment.

As a future work, it would be challenging but rewarding to investigate on corresponding practical effects such as the effect of these properties on blood pressure, alpha-ray emittance of the brain in response to variation of these physical properties.

V. ACKNOWLEDGEMENT

Authors wish to acknowledge The University of Sri Jayewardenepura for providing financial support under the research grant No. ASP/01/RE/SCI/2016/14.

VI. REFERENCES

Bachu, R., Kopparthi, S. and Adapa, B. (2010). Voiced/Unvoiced Decision for Speech Signals Based on Zero-Crossing Rate and Energy. Advanced Techniques in Computing Sciences and Software Engineering. Springer, Dordrecht., pp.279-282.

Fant, G. (1973). Speech Sounds and Features. Cambridge, MA: MIT Press.

Gunawardana, M.A.C.P. and Gamage, S.S.N. (2017). Analysis of "Angulimala sutta" using voiced to unvoiced ratio and formant variations. In: 4th International Conference on multidisciplinary Approaches. Nugegoda, Sri Lanka: Faculty of Graduate Studies, University of Sri Jayewardenepura, Sri Lanka, p.19.

Gunawardana, C. and Gamage, S. (2018). Recognition of Vowel Distribution for Sri Lankan Traditional *Pirith* Chants Using Formant Variation. In: The 2nd International Research Symposium. Badulla, Sri Lanka: Uva Wellassa University, p.119.

Gunawardana, C. and Gamage, S. (2018). Analysis of voiced to unvoiced ratio and vowel distribution for Sri Lankan Traditional *Pirith* Chants. In: Vingnanam International Research Conference. Jaffna, Sri Lanka: Faculty of Science, University of Jaffna., p.09.

Jayaratne, D. C., 2007. Science behind the benefits of listening to Buddhist *paritta* (*Pirit*) Chants. Colombo, s.n.

Jones, D. (1953). An English pronouncing dictionary. London: J.M. Dent. New York.

Lee, J. and Yoo, C. (2003). Wavelet speech enhancement based on voiced/unvoiced decision. In: Korea Advanced Institute of Science and Technology The 32nd International Congress and Exposition on Noise Control Engineering. pp.4149-4156.

Stevens, K. and House, A. (1955). Development of a Quantitative