Recognition of Vowels for Sri Lankan Traditional *Pirith* Chanting Using Formant Variation

M.A.C.P. Gunawardana*, S.S.N. Gamage

**osheenchamee1989@gmail.com*
[1]University of Sri Jayewardenepura, Gangodawila, Nugegoda, 10250, Sri Lanka

Abstract
Pirith is believed to be a doctrine preached by Lord Buddha and regarded to obtain protection from evil, promote health and well-being. Voice source produces a harmonic series, consisting of the fundamental frequency, F_0 and harmonic frequencies called as formants, F_n. Vowels can be mapped using the relationship between lip opening width to the first formant frequency, F_1 and tongue constriction width to second formant frequency, F_2. This work is dedicated to developing phonetic picture on Pirith chants and analyze acoustic properties using computer-aided tools. As reported by several other studies, characteristic vowels and high frequency formants are identified in chanting in contrast to normal speaking. The motive of this study is to investigate vowel distribution of Pirith chants with the aim of special pattern recognition. Samples of Ratana, Karaneeya and Angulimala Suttas recited by male monk chanters were recorded using high precision microphone array and 15 samples of each Sutta were analyzed. Recorded samples were then subjected to splitting of smaller voiced segments of frame length 10 ms using sampling 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. The vowel distribution reveals that a higher number of open vowels are articulated in Pirith chanting with totally disuse of closed vowels.

Keywords: *Formant frequencies, Vowels*

1. Introduction

Human speech is a complex signal that carries many kinds of information. Since the dawn of humankind people follow a recitation of prayers believing that those chants can protect them from super natural powers. *Pirith* refers to Buddhist scriptures, believed to be recited by Lord Buddha. It is a style of intoned recitation based on phonological properties of the *Pali* language. Voice source produces a harmonic series, consisting of the fundamental frequency and harmonic frequencies called as formants. There are several formants in vocal tracks, each at different frequencies, roughly one in each 1000 Hz band. The formant band width shows the sharpness of the resonance peak and in many chanting styles the formants are quite sharply tuned compared to speech.

Vowels are determined by the wide mouth opening and tongue position at the instant of pronunciation. Subsequently, vowels can be mapped using the relationship between lip opening width to the first formant frequency, and tongue constriction width to second formant frequency. This was first demonstrated by Essner (1947) and Joos (1948) and since then, the F_1-F_2 plane has become one of the standard ways of comparing vowel quality in a whole range of studies in linguistic phonetics (Ladefoged, 1971) and in many other fields.

This work is dedicated to developing phonetic picture on *Pirith* chants using computer-aided tools. As reported by several other studies, characteristic vowels and high frequency formants are identified in chanting in contrast to normal speaking and singing. 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 like what is obtained under a meditative trance. The motive of this study is to investigate vowel distribution of *Pirith* chants with the aim of special pattern recognition.

2. Experimental Section

Samples of *Ratana Sutta*, *Karaniya Metta Sutta* and *Angulimala Sutta* recited by male monk chanters were recorded under high precision conditions and 15 samples of each *Sutta* (Individual chanting) were subjected to analysis. Voiced recording was 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.

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 coefficients. 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,

$$\frac{S(z)}{U(z)} = \frac{G}{1 + \sum_{k=1}^{p} a_k z^{-k}} \qquad (1)$$

where, $S(z)$: Speech output, $U(z)$: Excitation, G : Gain and a_k : Predicted Coefficient

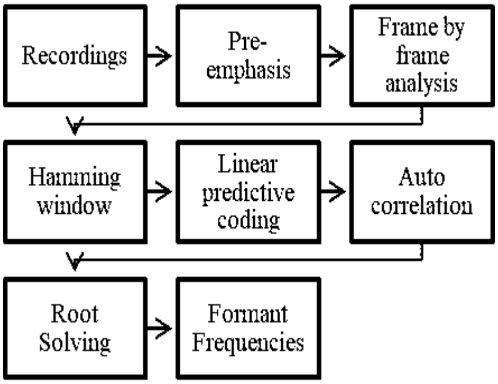
$$S(n) = \sum_{k=1}^{p} \alpha_k S(n-k) + GU(n) \qquad (2)$$

where, $S(n)$: Estimated current value, α_k : 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,

$$S(n) = \alpha_1 S(n-1) + \alpha_2 S(n-2) + \alpha\_3\, S(n-3) \ldots + \alpha\_p\, S(n-p) \qquad (3)$$

By minimizing the square of the error $\{e(n)^2\}$ filter coefficients can be generated. After finding the location 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.



Scheme1 - Block diagram of the experiment setup of formant extraction

2.1. Vowel Distribution

Vowels can be mapped using the relationship between lip opening width to the first formant frequency, F_1 and tongue constriction width to second formant frequency, F_2. Cardinal vowels, which are not of any language but a measuring system in describing sounds of languages, are used as a set of reference vowels in this work. These vowel sounds demonstrate if the tongue is in an extreme position, either front or back, high or low with combining the shape of lips. The cardinal vowel system was modified by Daniel Jones (Jones, 1953) based on the original idea proposed by earlier phoneticians, notably Ellis and Bell (Bell, 1867).

3. Results and Discussion

3.1. Vowel Distribution

Figure 1, figure 2 and figuer 3 show the vowel distribution for individual *Pirith* chants of *Ratana Sutta*, *Karaneeya Sutta* and *Angulimala Sutta* respectively. Corresponding vowel distributions for figure 1a, figure 2a and figure 3a are shown as percentages constitutional to each predefined territory are shown in inset figure b, while a further analysis of denser areas is indicated by inset figures c.

In this analysis, vowel distribution recognized a common area from all three types of solo *Pirith* chants with the identified percentage values of 63.4%, 65.6% and 71.6% of vowels concentrate around the frequency range of F_1, 0-1500 Hz and F_2, 0-1500 Hz for all individual *Pirith* chants of *Ratana Sutta*, *Karaneeya Sutta* and *Angulimala Sutta* respectively.

Figures 4-6 offer comparisons of vowel distribution of *Pirith* chants with primary cardinal vowels. In *Ratana Sutta*, majority of the vowels are pronounced when tongue is positioned high, besides, a considerably high number of vowels are pronounced when tongue positioned front low and lip shaped unrounded and when reciting *Karaneeya Sutta*, tongue has moved at higher and lower levels omitting most of the intermediate levels while lip shape remained unrounded when pronouncing vowels. While in *Angulimala Sutta*, significantly high number of vowels produced at low positioned tongue moving front and back while lip shape vary from rounded to unrounded.

As a summery, for all chants, irrespective of type of chanting, the densest vowel areas are overlapped with cardinal vowel [a] and [ɛ]. It conveys the idea of tendency towards the front positioned tongue while lips positioned unrounded and tongue height changed from the lowest to lower-mid level when chanting *Pirith Suttas*. Conversely, vowel distributions show zero percentage on cardinal vowel [i] and [e] inferring tongue has never been in high or upper limits while pronouncing *Pirith* chants.

Most of *Pirith Suttas* identify a common region of F_1, 0-1000 Hz and F_2, 1000-1500 Hz in vowel map. This result can be interpreted by using the cardinal vowel map with center vowels. It illustrates most of the vowels gathered towards cardinal center vowel [a] and [ɛ]. These two vowels generate from center positioned tongue and unrounded lips while tongue height changes upper lower to lower mid-level, showing a clear trend to chant center unrounded vowels in all *Pirith Suttas* subjected to analysis.

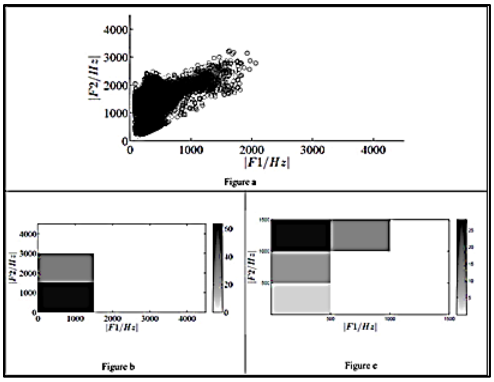


Figure 1: Vowel distribution for *Ratana Sutta*. (a) Vowel distribution in F_1-F_2 plane. (b) Percentage vowel distribution in the range of F_1 upto 4500 Hz, F_2 upto 4500 Hz. (c) Further analysis of the densest area, F_1 upto 1500 Hz and F_2 upto 1500 Hz. Area corresponding to the range of F_1, 0-500 Hz and F_2, 1000-1500 Hz highlightes as the densest area with a vowel distribution precentage of 28.15%.

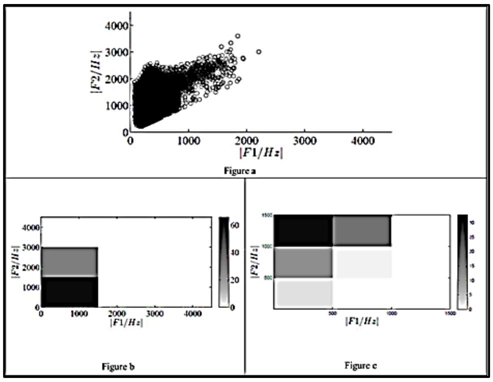


Figure 2: Vowel distribution for *Karaneeya Sutta*. Figure a shows vowel distribution in F_1-F_2 plane. (b) Percentage vowel distribution in the range of F_1 up to 4500 Hz, F_2 up to 4500 Hz. (c) Further analysis of the densest area, F_1 up to 1500 Hz and F_2 up to 1500 Hz. Area corresponding to the range of F_1, 0-500 Hz and F_2, 1000-1500 Hz highlights as the densest area with a vowel distribution percentage of 32.63%

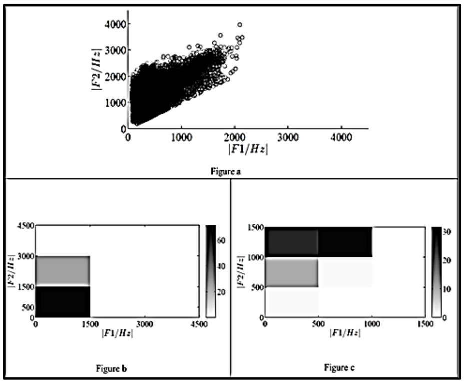


Figure 3: Vowel distribution for *Angulimala Sutta*. (a) Vowel distribution in F_1-F_2 plane. (b) Percentage vowel distribution in the range of F_1 upto 4500 Hz, F_2 upto 4500 Hz. (c) Further analysis of the densest area, F_1 upto 1500 Hz and F_2 upto 1500 Hz. Area corresponding to the range of F_1,500-1000 Hz and F_2, 1000-1500 Hz highlightes as the densest area with a vowel distribution precentage of 31.50%.

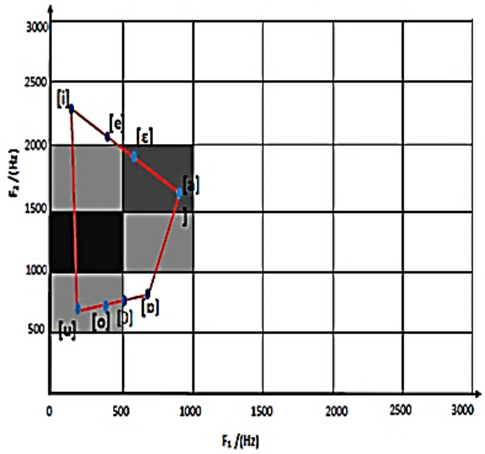


Figure 4: The comparison between vowel distribution of *Ratana Sutta* and primary cardinal vowels. The vowel distribution is denser at vowel areas of [a] and [u] while showing zero density at vowel areas [i] and [ɑ].

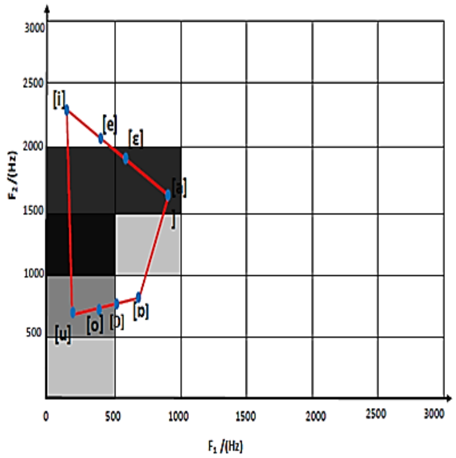


Figure 5: The comparison between vowel distribution of *Karaneeya Sutta* and primary cardinal vowels. The vowel distribution is denser at vowel area of [a] and [ɛ] while showing zero density at vowel areas [i] and[ɑ]

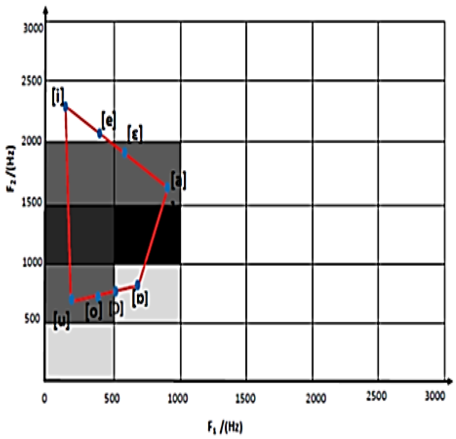


Figure 6: The comparison between vowel distribution of Angulimala Sutta and primary cardinal. The vowel distribution is denser at vowel area of [a] and [ɛ] while showing zero density at vowel areas [i]

4. Conclusions

When analysing vowel distribution for individual chanting of *Pirith Suttas*, it is observed that about 65% of vowels generate with unrounded lips with front or centre positioned tongue at lower levels. Further, about 15% of vowels are developed with back positioned tongue and rounded lips.

Pronunciation of vowels from high positioned tongue is very unlikely to occur when chanting *Pirith Suttas*. Besides, vowels generated with lower positioned tongue are called open vowels while high positioned ones are closed vowels. Therefore, it can be concluded that the larger part of open vowels articulated in *Pirith* chanting with obsolescence of closed vowels.

Acknowledgements

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

References

Bell, A. (1867). *Visible speech: the science of universal alphabetics*. [London]: Simpkin, Marshall & Co.

Bruce, H. (2018). Linguistics 103 - Bruce Hayes - UCLA. [Online] Linguistics.ucla.edu. Available at: http://linguistic.ucla.edu/people/hayes/103/index.htm [Accessed 26 Dec. 2017].

Gunawardana, C. and Gamage, S. (2017). Analysis of vocal parameters: Angulimala Sutta using voiced to unvoiced ratio and formant variations. Proceedings of the 4th International

Conference on Multidisciplinary Approaches. Nugegoda: Faculty of Graduate Studies, University of Sri Jayewardenepura., p.19.

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

Hardcastle, W., Hardcastle, W., Laver, J. and Gibbon, F. (2012). *The handbook of phonetic sciences*. Malden, MA: Wiley-Blackwell.

Jayaratne, C. (2010). 7th UN Day of Vesak Conference "Global Recovery – The Buddhist Perspective." In Buddhist paritta (Pirit) chants for mental and physical wellbeing - a scientific approach

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

Ladefoged, P., (1971). *Preliminaries to Linguistic Phonetics.* Chicago: University of ChicagoPress.