



# Evaluation of Acoustic Parameters for Angulimala Sutta using Voiced to Unvoiced Ratio and Vowel Distribution

Gunawardana MACP & Gamage Shantha SN  
Department of Physics, University of Sri Jayewardenepura, Sri Lanka

## ABSTRACT

*Pirith is believed as a protective doctrine preached by the Lord Buddha in Pali language. The aim of this study is to analyze acoustic properties of Pirith using computer-aided methods and identify special characteristics and patterns. In this study, two methods were used to identify special characteristics of Angulimala Sutta. First method calculates voiced to unvoiced ratio using zero crossing rate and energy content associated with the acoustic signal while second method recognizes vowel distribution using first and second formant frequencies. Results of the first method indicates approximately 96% of frames are voiced while the second method suggests approximately 72% of vowels concentrate in the square region of  $F_1, 0\sim 1500$  Hz and  $F_2, 0\sim 1500$  Hz demonstrating when chanting the Angulimala sutta most of the time the tongue height is low positioned in back levels while lips shaped unrounded.*

**KEYWORDS:** *Formant frequencies, Voiced to unvoiced ratio, Zero-Crossing rate, Vowel distribution*

## 1. 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) is an important parameter as it indicates the involvement of speech production system with vibration of vocal codes. In this work, we combined the results of V/UV ratio with zero crossing rate (ZCR) and energy of short time segments of the signal to strengthen the analysis.

In voiced speech, the vibrating glottis generates periodic pulses which resonate in the vocal tract. Therefore, when vowels are pronounced similar frequencies are generated. However, in the unvoiced speech, vocal chords held open and a continuous air beam flow through them. The air beam turns into a turbulent flow because of narrowed vocal tract and it creates, non-periodic, noise-like sounds (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 modeled as a non-uniform tube closed at vocal folds and open at the lip end (Stevens and House 1955). Cross sectional area of the vocal tract depends on the position of tongue, lips, jaw and velum. 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 which are clearly visible in spectrographic displays of voice segments. Normally, they occur on average at intervals of  $c/2L$ , where  $c$  is the speed of sound and  $L$  is the length of the vocal tract (Fant, 1973).

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 particular 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. 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).

## 2. LITERATURE SURVEY

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).

Several speech classification systems are proposed in literature. According to the Ayadi 2011, speech features can be grouped into four categories as continuous features, qualitative features, spectral features and TEO (Teager Energy Operator) based features (El Ayadi, Kamel and Karray, 2011). Energy, zero crossing rate, pitch, formants are called traditional speech signal features while energy and pitch are prosody features within low-frequency domain and formants and zero crossing rate are high frequency features (Chen et al., 2012).

### **3. METHODOLOGY**

In the analyzing process, Samples of *Angulimala Sutta* recited by male monk chanters were recorded under high precision conditions and 15 samples 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.

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 1.

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 *linear predictive coding (LPC)* and *auto correlation* to extract the formant values. In vowel analysis, frequency values regarding first formant,  $F_1$  and second formant,  $F_2$  were extracted and the vowel distribution was obtained by plotting  $F_1$  against  $F_2$ . In the analysis of vowel distribution, primary cardinal vowels introduced by Daniel Jones were used as a reference. (Jones, 1953).

### **4. RESULTS & 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. Figure 2 shows the vowel distribution for the all 15 samples. Percentage distribution is shown in Figure 3, while a further analysis of denser areas is indicated by Figure 4 and Figure 5. Figure 6 offers a comparison of vowel distribution with primary cardinal vowels.

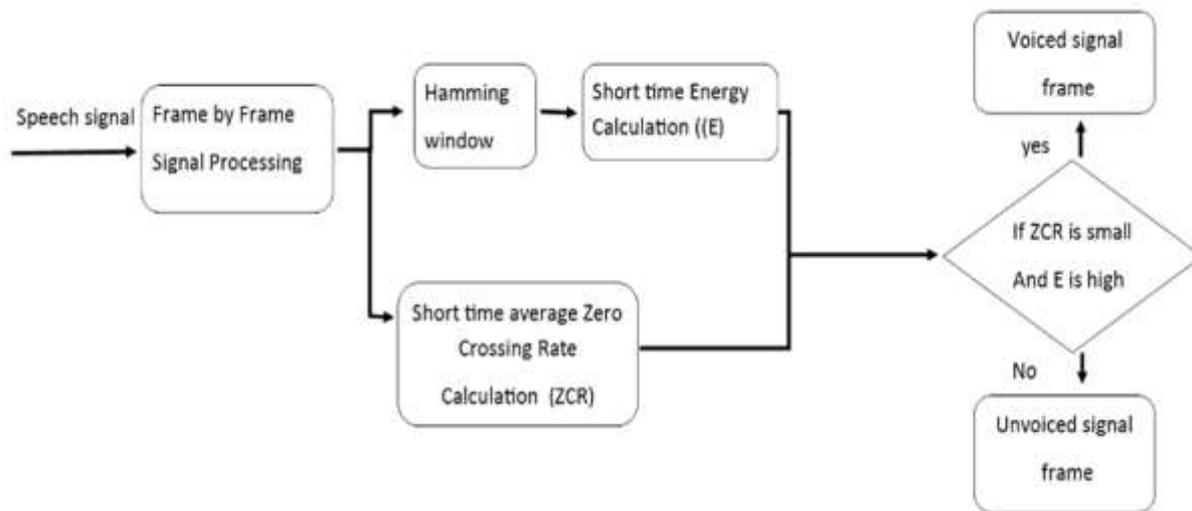


Figure 1- The algorithm to recognize Voiced to Unvoiced ratio

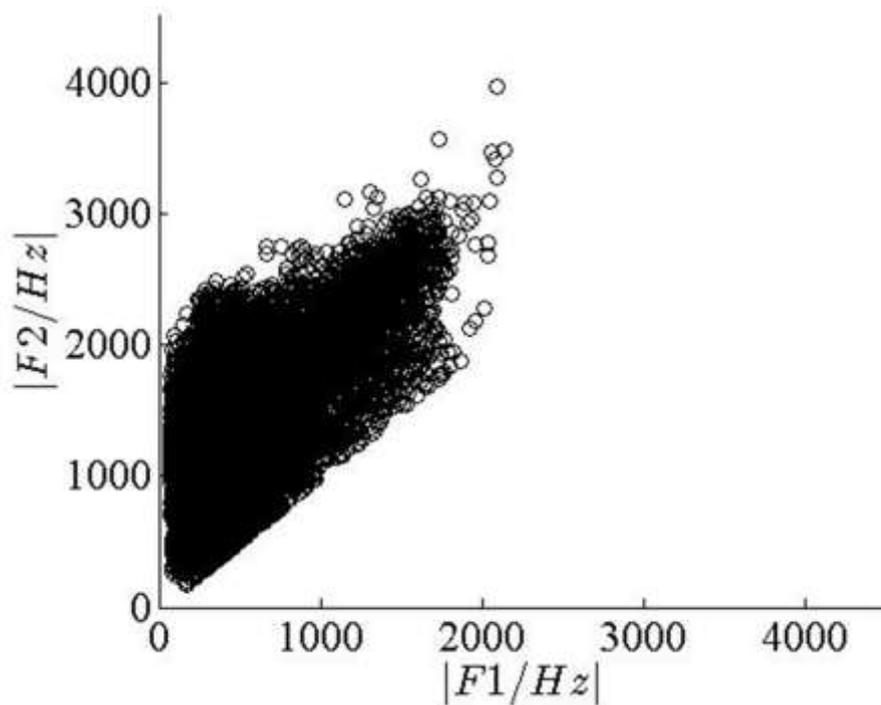


Figure 2- First two formants,  $f_1$  and  $f_2$ , variation for Angulimala Sutta

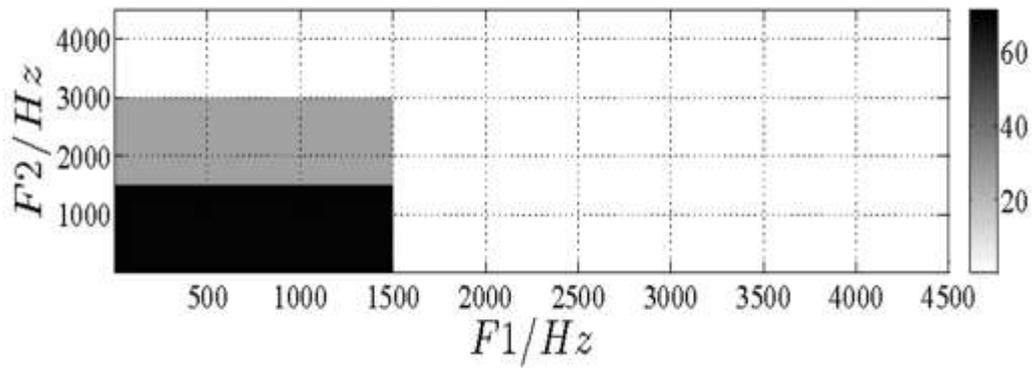


Figure 3 - Vowel map for Angulimala Sutta representing  $f_{1,0} - 4500 \text{ Hz}$  to  $f_{2,0} - 4500 \text{ Hz}$  area

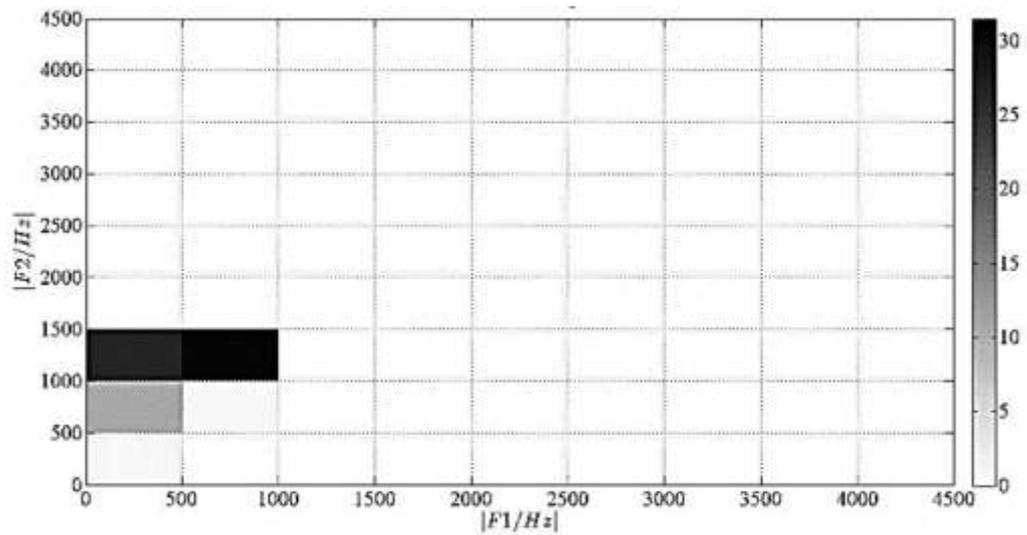


Figure 4 - Further analysis of  $f_{1,0} - 1500 \text{ Hz}$  to  $f_{2,0} - 1500 \text{ Hz}$  area which shows 71.55% of the total vowel distribution

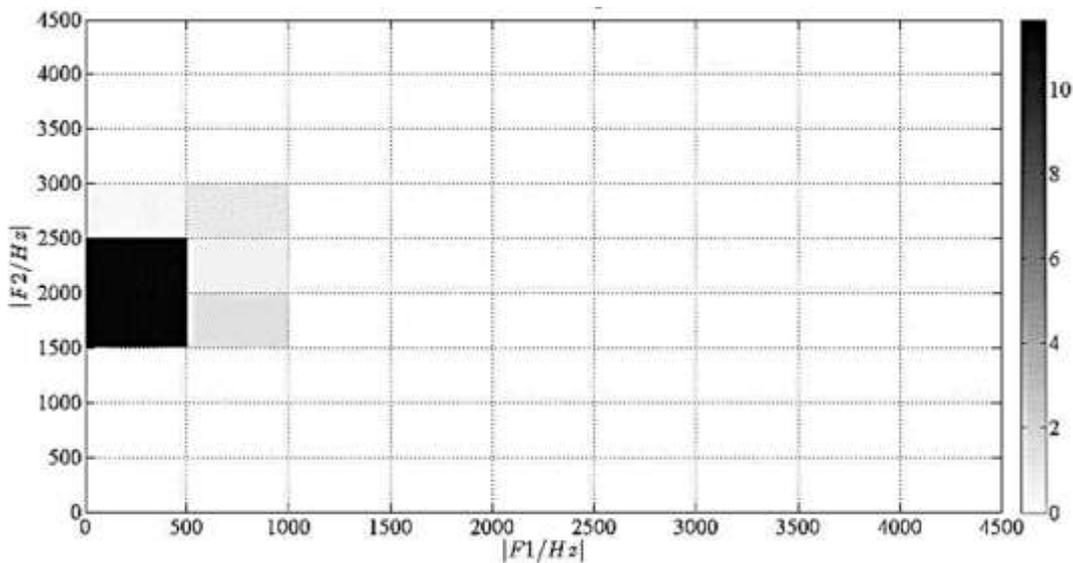


Figure 5- Further analysis of  $f_1, 0 - 1500 \text{ Hz}$  to  $f_2, 1500 - 3000 \text{ Hz}$  area which shows 15% of the total vowel distribution

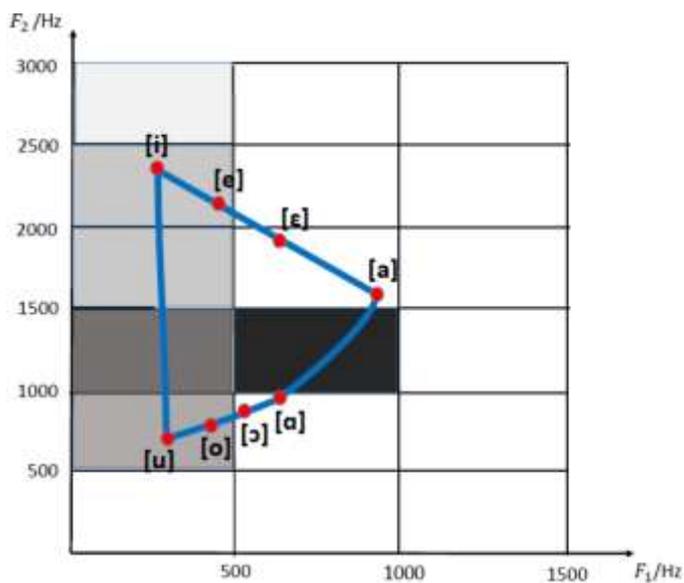


Figure 6 - Further analysis of  $f_1, 0 - 1500 \text{ Hz}$  to  $f_2, 0 - 1500 \text{ Hz}$  area which shows 71.55% of the total vowel distribution

Table 1 - Vowel distribution in Angulimala Sutta in percentage values

Area		Vowel Percentage
F1 range (Hz)	F2 range (Hz)	
0-500	500-1000	10.7
0-500	1000-1500	26.88
0-500	1500-2000	11.62
0-500	2000-2500	11.65
500-1000	1000-1500	31.50

Table 2 – Description of Daniel Jones’s 8 primary cardinal vowels using slanted symbols presented in International Phonetic Alphabet(IPA)

	Tongue position front Lips unrounded	Tongue positioned back Lips rounded
Tongue positioned high	[i]	[u]
Tongue positioned upper mid	[e]	[o]
Tongue positioned lower mid	[ɛ]	[ɔ]
Tongue positioned low	[a]	[ɑ]

Voiced to unvoiced ratio calculation, when combined with zero-crossing rate and energy content, demonstrated 95% of frames as voiced while 5% of frames as unvoiced. Further, it demonstrates a clear tendency to pronounce vowels in chanting *Angulimala sutta*.

A previous research work on formant frequency tuning in professional Byzantine chanters shows clear evidence that chanters have special ability to use personal formant tuning at chanting [2]. In this analysis, vowel distribution shows most

common area for all chanters as shows in figure 2. The calculation of the percentage values indicated that 71.55% of vowels concentrate around the frequency range of  $F_1, 0 - 1500$  Hz and  $F_2, 0 - 1500$  Hz as shown in Figure 3. Further analysis demonstrated the area bounded by,  $F_1, 0 - 500$  Hz ;  $F_2, 1000 - 1500$  Hz and  $F_1, 500 - 1000$  Hz;  $F_2, 1000 - 1500$  Hz represent 26.88% and 31.50% of vowel distribution respectively as showing in Table 1.

## 5. CONCLUSION

The *Angulimala sutta* is rich with vowels as it shows approximately 19:1 voiced to unvoiced ratio. Analysis of these vowels suggest that 71.55% of vowels concentrate around the frequency range of  $F_1, 0$  Hz ~1500 Hz and  $F_2, 0$  Hz ~ 1500 Hz showing high amount of low back unrounded vowels.

When comparing the results with Cardinal vowel chart, the densest vowel area shows the qualities of cardinal vowel [a] and [ɑ] as shown in Table 2.

It can be concluded as when chanting the *Angulimala sutta*, the arrangement is inclined to be the tongue is low positioned in back levels while lips shaped unrounded. Less number of vowels is represented by cardinal vowels [i] and [u], showing high front unrounded vowel and high back rounded vowels respectively.

## ACKNOWLEDGMENT

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

- BACHU R. KOPPARTHI S. & ADAPA B. Voiced/Unvoiced Decision for Speech Signals Based on Zero-Crossing Rate and Energy. *Advanced Techniques in Computing Sciences and Software Engineering*. Springer, Dordrecht. 2010; pp.279-282.
- CHRYSOCHOIDIS G & KOUROUPETROGLOU G. Formant Frequency Tuning in Professional Byzantine Chanters. *Recent Advances in Electrical and Computer Engineering*. 2014.
- FANT G. *Speech Sounds and Features*. Cambridge, MA: MIT Press. 1973.
- GUNAWARDANA MACP & GAMAGE SSN. 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 Jayewardanapura, Sri Lanka, 2017. p.19.
- GURJAR A. SIDDHARTH A. & AJAY P. Analysis of Acoustic of "OM" Chant To Study Its Effect on Nervous System. *IJCSNS International Journal of Computer Science and Network Security*, [online] 9(1), pp.363-3687. Available at: [http://paper.ijcsns.org/07\\_book/200901/20090151.pdf](http://paper.ijcsns.org/07_book/200901/20090151.pdf) . Accessed 01 May 2018.
- JAYARATHNE C. Science behind the benefits of listening to Buddhist paritta (Pirit) Chants. In: *International Buddhist Conference on How to Practice and Protect Buddhism in this Millennium*. Colombo. 2007.
- LADEFOGED P. *Preliminaries to linguistic phonetics*. Chicago [etc.]: The University of Chicago Press. 1981.
- LEE J. & YOO C. 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*. 2003; pp.4149-4156.
- STEVENS K & HOUSE A. Development of a Quantitative Description of Vowel Articulation. *The Journal of the Acoustical Society of America*, 27(3). 1955; pp.484-493.
- JONES D. *An English pronouncing dictionary*. London: J.M. Dent. New York.
- Bell, A. (1867). *Visible Speech, the Science of universal Alphabetic* . London. 1953.
- BRUCE H. *Linguistics 103 - Bruce Hayes - UCLA*. [online] [Linguistics.ucla.edu](http://linguistics.ucla.edu). Available at: <http://linguistics.ucla.edu/people/hayes/103/index.htm> [Accessed 26 Dec. 2017].
- EL AYADI M. KAMEL M. & KARRAY F. Survey on speech emotion recognition: Features, classification schemes, and databases. *Pattern Recognition*. 2011; 44(3), pp.572-587.
- CHEN L. MAO X. XUE Y & CHENG L. Speech emotion recognition: Features and classification models. *Digital Signal Processing*, 22(6). 2012; pp.1154-1160.