B. TECH. PROJECT REPORT

On

FUNDAMENTAL FREQUENCY DETERMINATION FOR SPEECH SIGNALS OF VEDIC MANTRAS

BY

Anmay Kumar

DISCIPLINE OF ELECTRICAL ENGINEERING

INDIAN INSTITUTE OF TECHNOLOGY INDORE

December 2019

FUNDAMENTAL FREQUENCY DETERMINATION FOR SPEECH SIGNALS OF VEDIC MANTRAS

A PROJECT REPORT

Submitted in partial fulfillment of the requirements for the award of the degrees

of
BACHELOR OF TECHNOLOGY
in

ELECTRICAL ENGINEERING

Submitted by:
Anmay Kumar
(160002007)

Guided by:
Dr. Ram Bilas Pachori
Professor
Discipline of Electrical Engineering

INDIAN INSTITUTE OF TECHNOLOGY INDORE
December 2019

DECLARATION

I hereby declare that the project entitled **"Fundamental Frequency Determination for Speech Signals of Vedic Mantras"** submitted in partial fulfillment for the award of the degree of Bachelor of Technology in Electrical Engineering completed under the supervision of **Dr. Ram Bilas Pachori, Professor, Discipline of Electrical Engineering,** IIT Indore is an authentic work.

Further, I declare that I have not submitted this work for the award of any other degree elsewhere.

Signature and name of the student(s) with date

CERTIFICATE by BTP Guide(s)

It is certified that the above statement made by the students is correct to the best of my knowledge.

Signature of BTP Guide(s) with dates and their designation

<u>Preface</u>

This report on "**Fundamental Frequency Determination for Speech Signals of Vedic Mantras**" is prepared under the guidance of **Dr. Ram Bilas Pachori**, **Professor, Discipline of Electrical Engineering, IIT Indore**.

Through this report I have tried to give a detailed explanation on how to extract pitch for Vedic mantras and to compare two utterances of a specific mantra. The proposed method can be used to train people on how to pronounce Vedic mantras correctly. This method can also be used to create an application which could suggest improvements in the way of pronunciation. The proposed method can also be applied to speech signals in languages different from Sanskrit. Thus, two speech signals in any language can be compared with the suggested method.

I have tried to the best of my abilities and knowledge to explain the proposed method, figures and results in a lucid manner. I have also added figures and results to make it more illustrative.

Anmay Kumar
B.Tech. IV Year
Discipline of Electrical Engineering
IIT Indore

Acknowledgements

I am extremely grateful to my supervisor Prof. Ram Bilas Pachori for providing inspiration, guidance, motivation and support throughout my research work. I am grateful for having the opportunity to work under such an esteemed professor.

I would like to sincerely thank Dr. Srinivas Kannan, Chairman, Arts Culture India Foundation, Chennai, India for his coordination and cooperation and for providing all the data for the analysis. I would also like to thank him for always encouraging me to complete the project. This project would not have reached completion without his support.

I would also like to thank my family and batch mates for constantly helping and motivating me during the course of this project. Lastly, I would like to thank Lord Krishna for enlightening my mind and guiding me in the right direction during the project.

Anmay Kumar
B.Tech. IV Year
Discipline of Electrical Engineering
IIT Indore

Abstract

Instantaneous fundamental frequency (IFF) or pitch detection is a very necessary step in analyzing any speech signal. It has a number of applications like language recognition, speech recognition, vocal activity recognition, etc. The accurate detection of IFF is essential for any analysis to be done on speech signals. Since the past many decades many signal processing algorithms have been developed but, these techniques to the best of our knowledge have not been studied for Vedic mantra speech signals. The proposed method uses an advanced signal processing algorithm named as variational mode decomposition (VMD) in an iterative way to extract IFF from the speech signals. This way the speech signals of Vedic mantras can be compared on the basis of pitch. In the proposed method, VMD is applied on the signal iteratively and when the algorithm converges, the resultant signal is used to find out the voiced regions in the signal. After this all the voiced regions are divided into segments of length 400 samples. Then, VMD is applied to each of these signals and the obtained signals are concatenated to get the final mono-component signal. Finally, Hilbert transform is used to calculate the IFF of the whole input signal. Using the proposed method, I have tried to study the pitch variation of four Vedic mantras namely Gananamtva and three verses of – Krishna Yajur Veda Taittiriya Sakha Purusha Suktam. I have tried to compare the learner's chants of all the above mentioned mantras with the reference chants. The IFF of both the signals is compared and the learner can clearly identify the regions where different fundamental frequencies take place.

Table of Contents

List of Figures

Chapter 1

Introduction

1.1 Literature review

Vedic scriptures are considered to be one of the oldest scriptures in the history of human civilization. Vedic chanting techniques are vast and tough to master. The notes used in chanting, known as Swaras are strictly bound by rules. One has to be very cautious about the duration and pitch of the Swara being pronounced [1].

Vedic chanting requires good memory and attention. From generation to generation, the Vedic knowledge was passed in form of oral recitations without the aid of writing. This helped to utilize memory capabilities to the maximum extent. Various studies have shown that Vedic chanting has many positive effects on our personality. For example it improves our memory and attention capabilities [2].

Vedic chants are also capable of regulating anxiety levels and blood pressure (BP), heart rate (HR), and oxygen saturation [3]. Vedas having so many applications are a wide area of research but, not much technical research has been done on them. Some other methods to compare two Vedic chants include wavelet transforms [1].

Chanting of Vedic mantras such as the maha mantra can reduce stress and cure depression. The effect of chanting the maha mantra was studied on the *three gunas-sattva (enlightenment), rajas (passion), and tamas (inertia)*—described in the Vedas as the basis of human psychology [4].

Research studies on mantra have provided insight into how mantra may work to empower the mind and benefit physical and mental health. Effects of chanting Vedic mantras and chanting any other words were compared and it was found that Vedic mantras indeed help to reduce stress and are different than normal words [5].

1.2 About the data

All the data files are composed of four swaras. Vedic matras can have more than four swaras but we are considering only the mantras which have four swaras. To understand what swaras are, they can be seen as English equivalent of tones. The sampling rate of the speech signals is 44100 samples per second which is scaled down to 8820 samples per second to reduce the time taken by the algorithm to converge.

The dataset includes speech signals of Gananamtva and the first three verses of Krishna Yajur Veda Taittiriya Sakha Purusha Suktam**.**

Gananamtva contains four swaras namely, Anudatta (low pitch), Udatta (high pitch), Svarita (mix of high and low pitch) and Prachaya (near high pitch). The swara composition of Gananamtva is as follows: ‘Ga’ is Anudatta, ‘NA’ is Udatta,‘ nAm’ is Swarita and ‘tvA’ is Prachaya. Krishna Purusha Suktam also contains the aforementioned swaras. Since the Krishna Purusha Suktam is very large, we have studied it in form of verses. For the purpose of this project we are considering only the first three verses.

1.3 Motivation

Many speech signal processing algorithms have been developed in the literature. But, a very less work has been done for the analysis of speech signals corresponding to Sanskrit language for example Vedic mantras. The main motivation for this work is to explore recently developed instantaneous fundamental frequency (IFF) determination method for speech signals for determining IFF of Vedic mantras.

Also the proposed method can be used to compare two speech signals for similarity in the fundamental frequency. So a person can use the proposed method to compare the spoken speech signal with the reference speech signal.

Chapter 2

Variational mode decomposition and Hilbert transform

2.1 Variational mode decomposition

It decomposes a real signal *u(t)* into a discrete number of sub-signals. It is assumed that each mode is centered around a center frequency ω, which is to be determined along with the decomposition [6, 7].

In order to obtain these modes and center frequencies, this method formulates the following constraint optimization problem:

$$min_{\{u_k\}\{\omega_k\}} \left\{ \sum_k \left\| \partial_t \left[(\delta(t) + \frac{j}{\pi t}) * u_k(t)) \right] e^{-j\omega_k t} \right\|_2^2 \right\} \qquad (2.1)$$

$$\text{such that} \quad \sum_k u_k = f$$

Where u_k and ω_k are notations for the set of modes and center frequencies and

* denotes the convolution operator.

The Lagrangian multiplier transforms the current problem into an unconstrained optimization problem, which can be expressed as follows:

$$L\ (\{uk\}, \{\omega k\}, \lambda) := \alpha \sum_k \left\| \partial_t \left[(\delta(t) + \frac{j}{\pi t}) * uk(t)) \right] e^{-j\omega_k t} \right\|_2^2$$

$$+ \|f(t) - \sum_k u_k(t)\|_2^2 + \langle \lambda(t), f(t) - \sum_k u_k(t) \rangle \qquad (2.2)$$

The estimated components and the center frequencies in each iteration can be calculated as follows:

$$u_k^{n+1}(\omega) = \frac{f(\omega) - \sum_{i \neq k} u_i(\omega) + \frac{\lambda(\omega)}{2}}{1 + 2\alpha(\omega - \omega_k)^2} \tag{2.3}$$

$$\omega_k^{n+1} = \frac{\int_0^\infty \omega |u_k(\omega)|^2 d\omega}{\int_0^\infty |u_k(\omega)|^2 d\omega} \tag{2.4}$$

Where $u_k(\omega)$ and ω_k denote the decomposed modes and center frequencies during the iteration respectively.

The variational mode decomposition (VMD) [8] has been used to enhance the speech signals. It has also been used for automatic seizure detection for EEG signals [9] and for automatic sleep staging [10]. Electrocardiogram (ECG) feature extraction is done using VMD [11]. Filter bank properties and its applications of VMD have been studied [12].

2.2 Hilbert transform

The Hilbert transform is a linear operator which transforms a function into different function by rotating it by 90 degrees. Mathematically, it can be expressed as the convolution of the given function with 1/(πt).

$$H(u(t)) = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{u(\tau)}{t - \tau} d\tau \tag{2.5}$$

In signal processing the Hilbert transform is used to obtain an analytic signal whose real part is the function itself and the imaginary part is the Hilbert transform of the function [13]. The analytic signal can be expressed in the polar form which can provide information about the magnitude, phase and frequency. Hilbert transform find application in speech demodulation [14].

Mathematically, the conversion of real signal into analytic signal can be expressed as follows:

$$u_a(t) = u(t) + j.H(u(t)) \tag{2.6}$$

For a sinusoidal signal u(t) = cos(ωt+φ), the analytic signal can be written as follows:

$$u_a(t) = u_m(t).e^{j(\omega t+\varphi)} \tag{2.7}$$

where, u_a(t) is the analytic signal representation of real signal *u(t)*.

The phase of the analytic signal can be used to determine the frequency of the signal as follows:

$$\omega(t) = \frac{d\varphi(t)}{dt} \tag{2.8}$$

The analytic signal previously obtained can also be used to compute the envelope [15] of the speech signal ***u(t).*** It involves taking the absolute value of the analytic signal and smoothing the obtained signal using a moving average filter [16].

The computation for envelope of input signal can be done as follows:

$$\mathrm{E(t)} = \sqrt{[u(t)]^2 + [H(u(t))]^2} \tag{2.9}$$

Chapter 3

IFF determination method

The proposed method for IFF determination of Vedic mantra speech signals can be summarized broadly in the following two steps:

Step 1 : Detection of voiced/non-voiced regions (V/NV) [17]: First, the fundamental component of the Vedic mantra is determined and using this component the envelope of the signal is detected. Once the envelope is extracted, the V/NV regions are extracted.

Step 2 : Obtaining IFF [18, 19]: The signal in previous step is divided into segments and VMD is applied to get the desired IFF.

3.1 Detection of voiced/non-voiced regions

The algorithm to extract V/NV regions is shown in Figure 3.1.

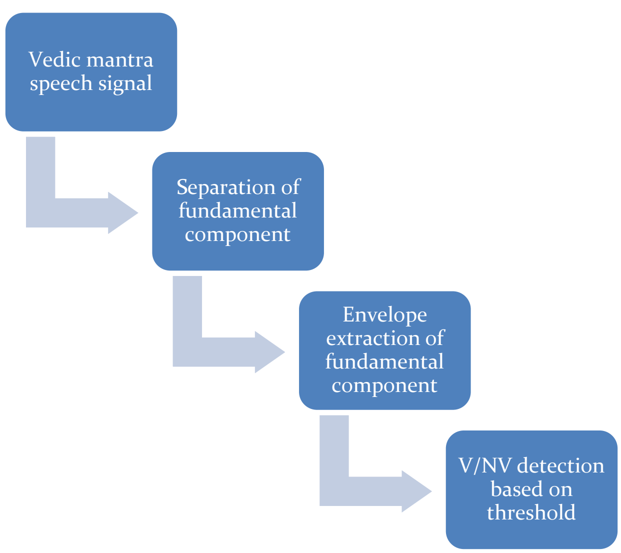


Fig 3.1 Flowchart showing the process of V/NV method for Vedic mantras.

Detailed explanation of the method is provided in the following sections.

3.1.1 Separation of fundamental frequency component

The input speech signal is filtered into low frequency range (LFR) i.e. 50-500 Hz [17]. Then VMD with certain parameters [6] is applied in an iterative way to obtain the fundamental frequency component.

In each iteration, the signal is decomposed into two signals and the component with lower frequency is chosen for the next iteration. This continues until convergence is achieved.

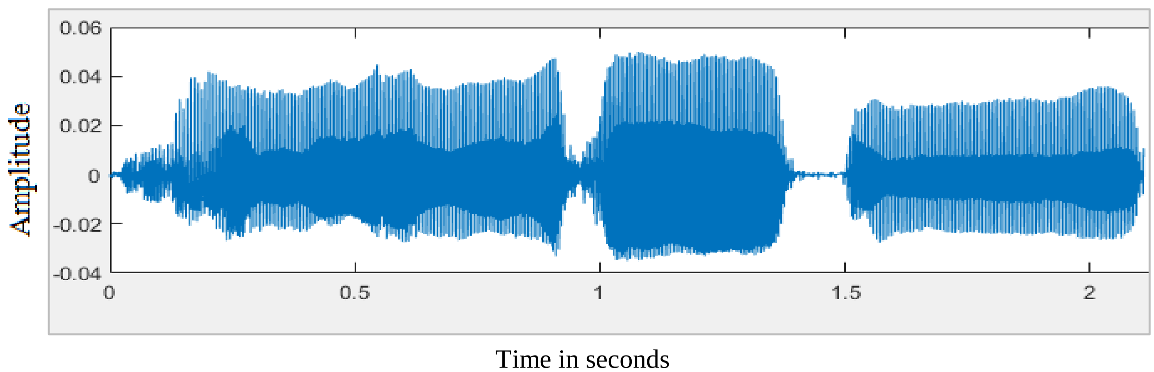
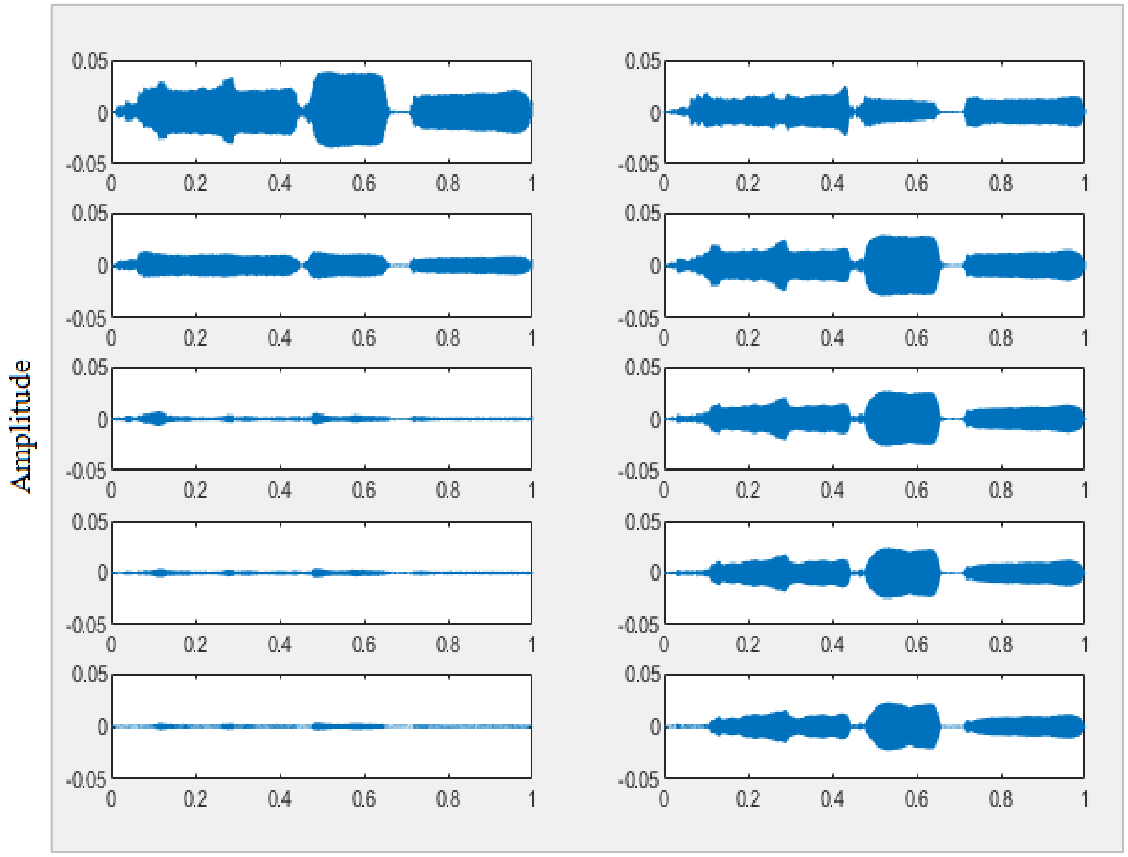


Fig 3.2 Gananamtva speech signal

Figure 3.2 shows the input speech signal for Gananamtva. The y-axis shows the amplitude and x-axis shows the time in seconds.

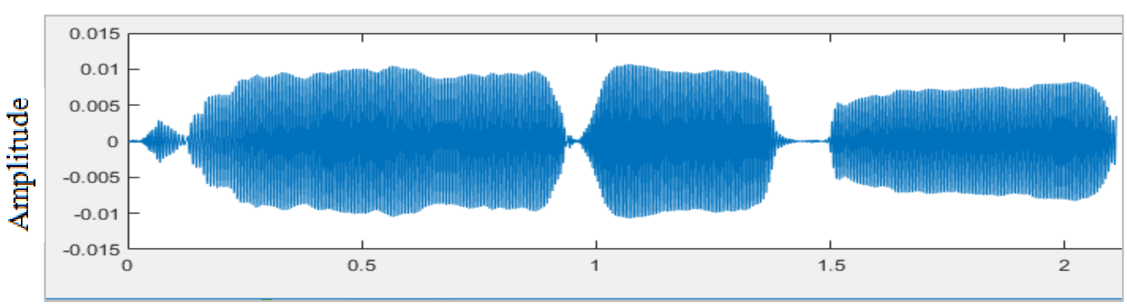


Normalized time in seconds

Fig 3.3 Plot showing the decomposed signals for five iterations

From Figure 3.3 we can clearly see the signal which is selected for each iteration. For first iteration the signal with lower frequency is selected and VMD is applied to it. We can see that after third iteration only one component has significant energy. Hence the algorithm has converged and we have obtained the fundamental component. The unit of x-axis is time divided by total time hence giving scaled time between 0 and 1.

Figure 3.4 shows the extracted fundamental frequency component.



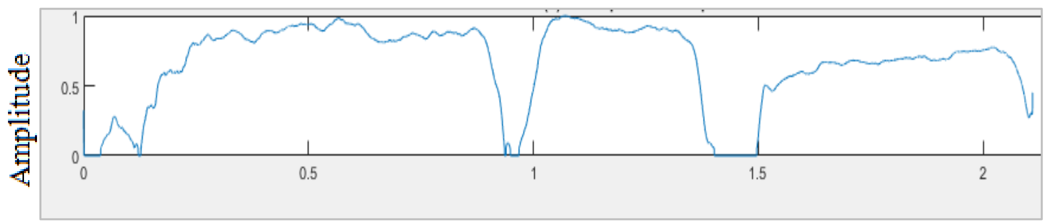
Time in seconds

Fig 3.4 Fundamental component for Gananamtva

3.1.2 Envelope detection

Envelope detection involves creating an analytic signal of the input signal *x(t)* using Hilbert transform [15].

The envelope of a signal *x(t)* is defined as the magnitude of the analytic signal. It can be obtained as per equation (2.9).



Time in seconds

Fig 3.5 Envelope of extracted fundamental component of Gananamtva

Figure 3.5 shows the envelope obtained from the fundamental component.

3.1.3 Extraction of voiced/non-voiced regions

For detecting V/NV regions, an appropriate threshold was applied to the envelope calculated in the previous step.

Value of the threshold (t_h) = 0.07×max(Fundamental component).

Where, max operation gives the maximum value in the fundamental component vector.

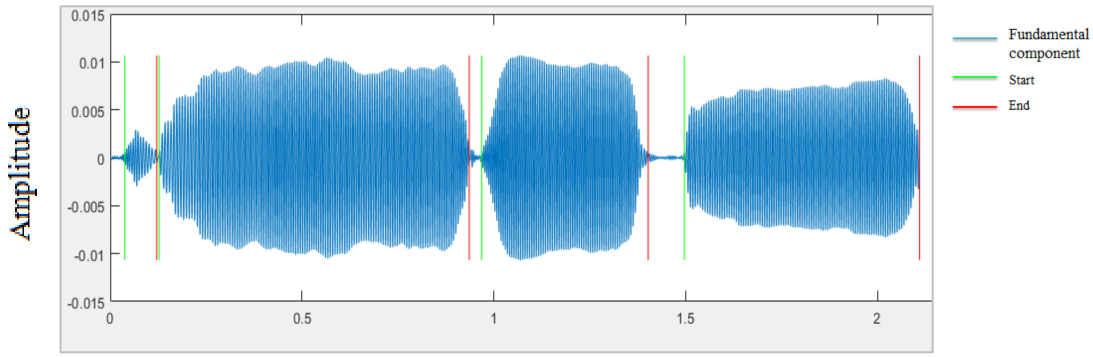


Fig 3.6 Extracted voiced/non-voiced regions in Gananamtva

Figure 3.6 shows the extracted V/NV regions for the speech signal. The start of a voiced region is marked with a green vertical line and end of the region is marked with a red vertical line.

3.2 Determination of IFF

Once the voiced regions are detected, the following operations are performed on the resultant signal:

1. The resultant signal is filtered into low frequency region (50-500Hz) [19].
2. The low frequency region filtered voiced signal is divided into segments of length 400 samples each.
3. Now, for each segment VMD is applied in an iterative way to extract fundamental frequency components.
4. Finally, all the fundamental frequency components are concatenated for obtaining the fundamental frequency component for the whole signal.
5. Further, Hilbert transform is applied on the fundamental frequency component [13, 20] to obtain the analytic signal according to equation (2.7).

   Now, from this analytic signal the IFF is obtained by differentiating its phase as per equation (2.8) [13, 18, 21].
6. A moving average filter is applied to remove abrupt fluctuations from the obtained IFF [16].

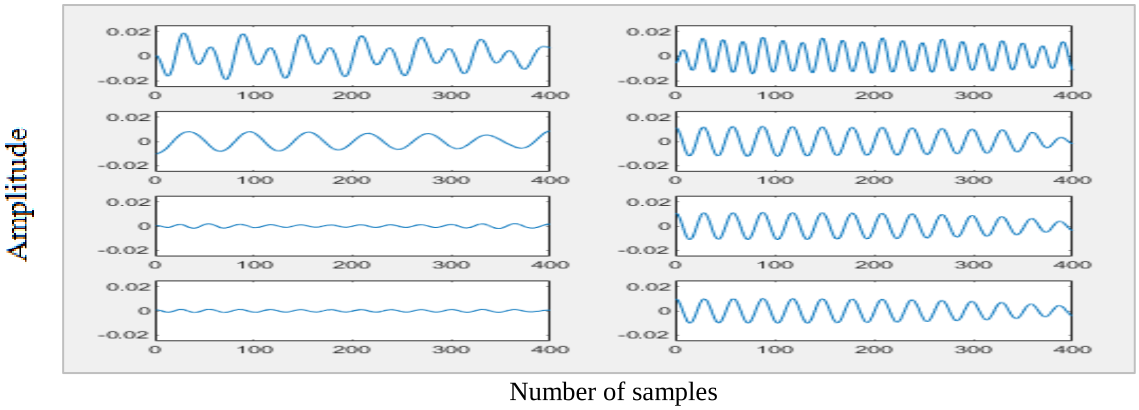


Fig 3.7 Four iterations of VMD on segments of length 400 samples.

From Figure 3.7 we can infer that for first iteration the signal with lower frequency is selected and VMD is applied to it. We can also see that after second iteration only one component has significant energy. Hence the algorithm has converged and we have obtained the fundamental component.

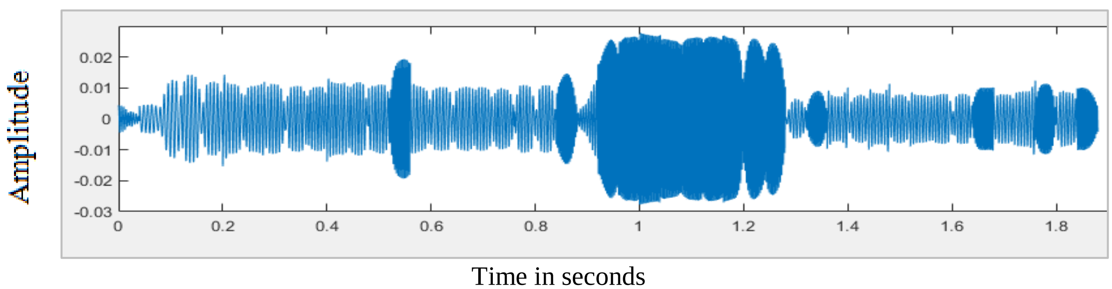


Fig 3.8 Concatenated fundamental frequency component for Gananamtva

Figure 3.8 shows the final fundamental component obtained by the concatenation of all fundamental frequency components of length 400 samples. Figure 3.9 shows the final detected IFF for Gananamtva speech signal.

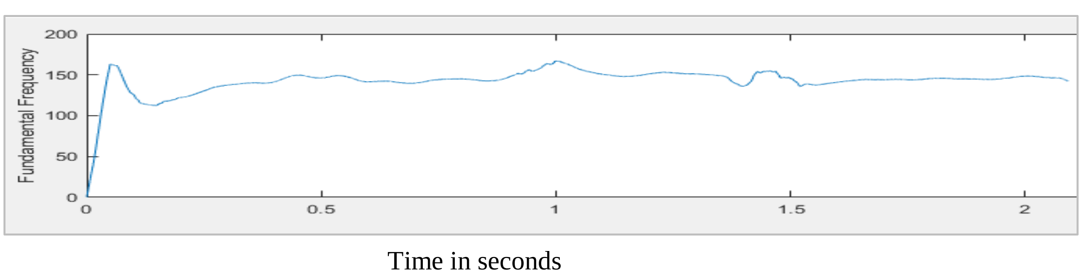


Fig 3.9 Extracted IFF for Gananamtva speech signal

Chapter 4

Results and Discussion

All the results were obtained on utterances of Gananamtva and some verses of Krishna Purusha Suktam. For all the observations, two speech signals were compared with each other on the basis of fundamental frequency. One signal was of the learner and the other was the reference signal.

The IFF of both the signals are plotted on the same figure to compare them. The results for various Vedic mantras are summarized below.

4.1 Gananamtva

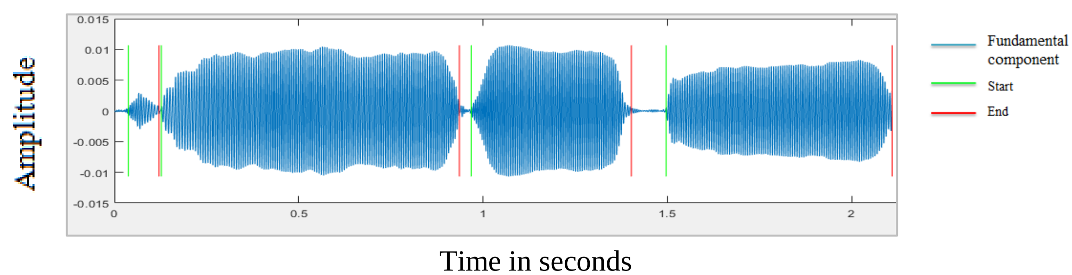


Fig 4.1 V/NV regions of reference signal for Gananamtva

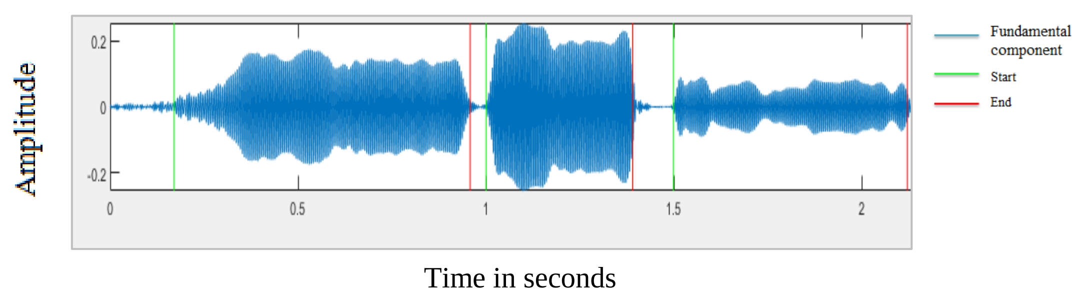
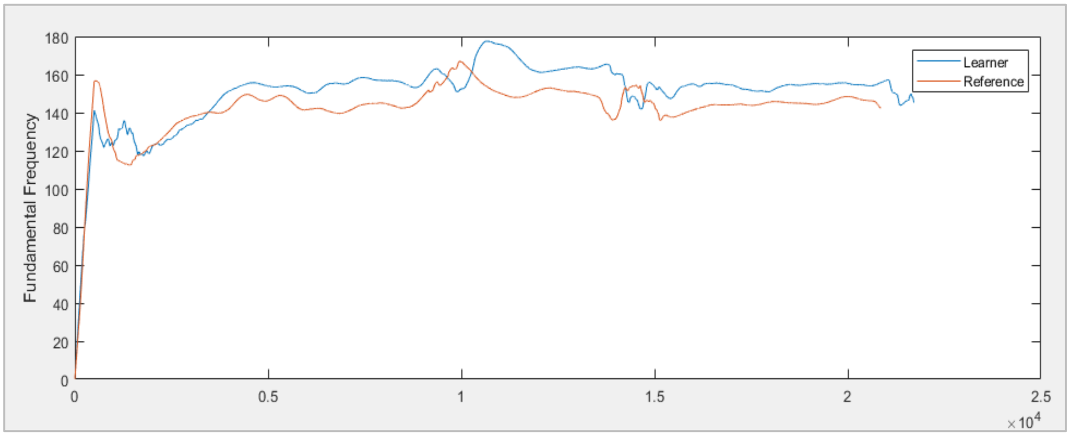


Fig 4.2 V/NV regions of learner’s signal for Gananamtva



Time in seconds (× sampling rate)

Fig 4.3 Plot of IFF for reference and learner's signals for Gananamtva.

Figure 4.1 shows the fundamental component for the reference signal. Here we can see the different voiced regions. Figure 4.2 shows the fundamental component for the learner's signal. From figure 4.3 we can see the IFF of both the signals, we can clearly see that the learner has pronounced the mantra quite well. The reference signal is almost constantly below the learner signal, so we can suggest the learner to lower his pitch little bit.

Now, we are going to look at the obtained results for first three verses of Krishna Purusha Suktam.

4.2 First Verse: सहस्रशीर्षा पुरुषः ।

The first verse can be written in English as ‘sahasrashirsha purushah’.

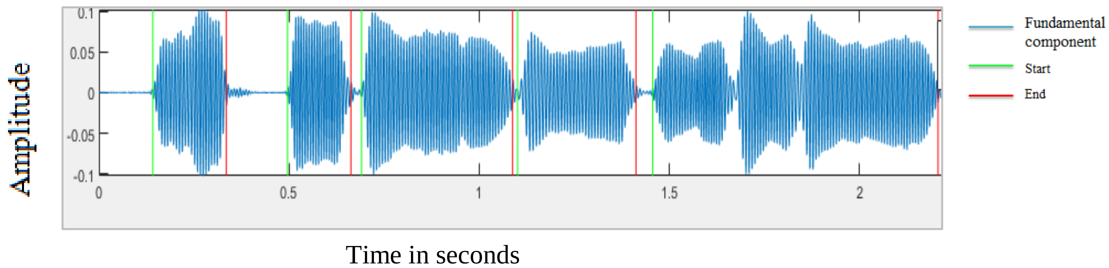


Fig 4.4 V/NV regions of reference signal for first verse

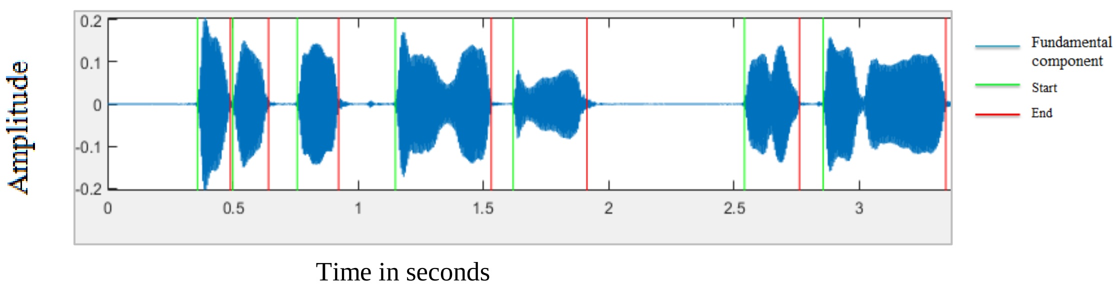


Fig 4.5 V/NV regions of learner’s signal for first verse

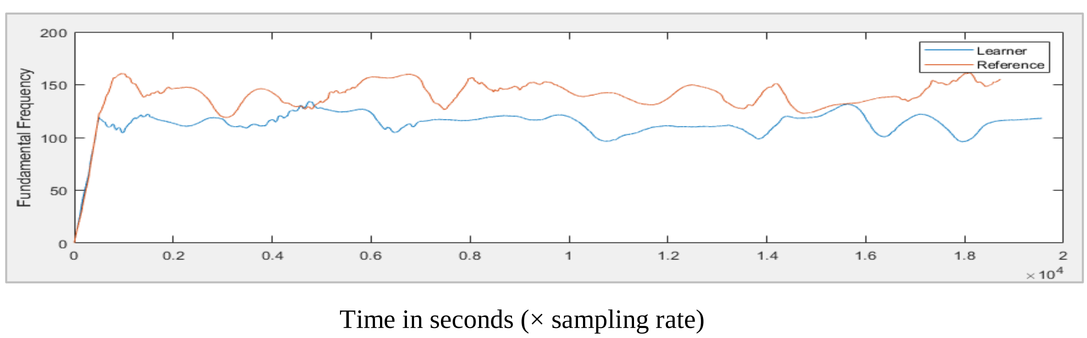
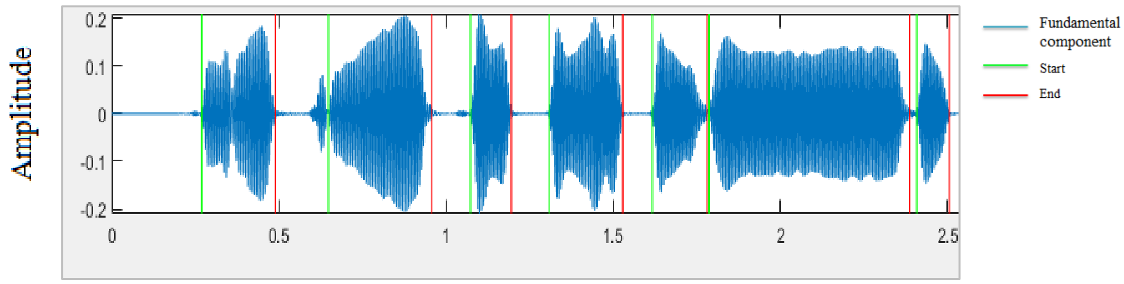


Fig 4.6 Plot of IFF for reference and learner’s signals for first verse.

Figure 4.4 shows the fundamental component for the reference signal. Here we can see the different voiced regions. Figure 4.5 shows the fundamental component for the learner’s signal. We can see the different voiced regions and also the difference between this signal and the reference signal. Figure 4.6 plots the IFF of both the signals. It can be seen that the learner’s fundamental frequency is considerably less than that of reference signal. So, the learner can now improve the pronunciation by increasing the fundamental frequency of spoken voice.

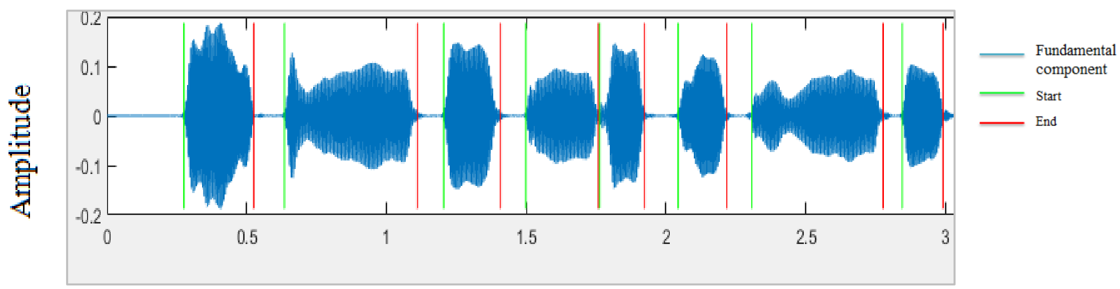
4.3 Second Verse: सहस्राक्षः सहस्रपात् ।

The second verse can be written in English as ‘sahasrakshas sahasrapat’.



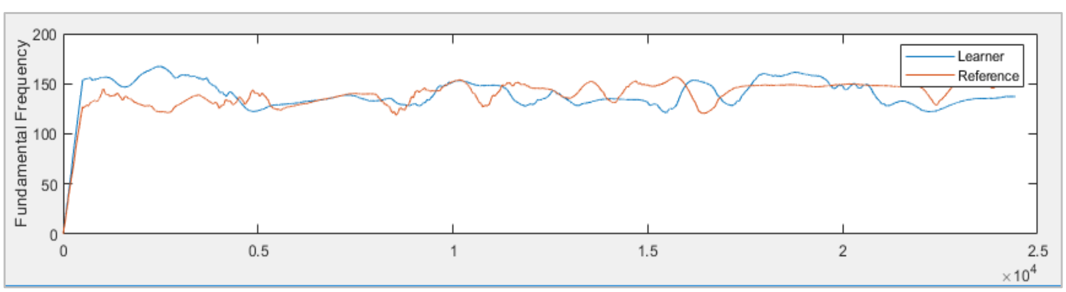
Time in seconds

Fig 4.7 V/NV regions of reference signal for second verse



Time in seconds

Fig 4.8 V/NV regions of learner’s signal for second verse



Time in seconds (× sampling rate)

Fig 4.9 Plot of IFF reference and learner’s signals for second verse.

Figure 4.7 shows the fundamental component for the reference signal. Here we can see the different voiced regions. Figure 4.8 shows the fundamental component for the learner’s signal. We can see the different voiced regions and also the difference between this signal and the reference signal. Figure 4.9 plots the IFF of both the signals. It can be seen that the learner’s fundamental frequency is very close to that of reference signal except for the region at start.

4.4 Third Verse: स भूमिं विश्वतों वृत्वा ।

The third verse can be written in English as ‘sabhumim vishvato vrata’.

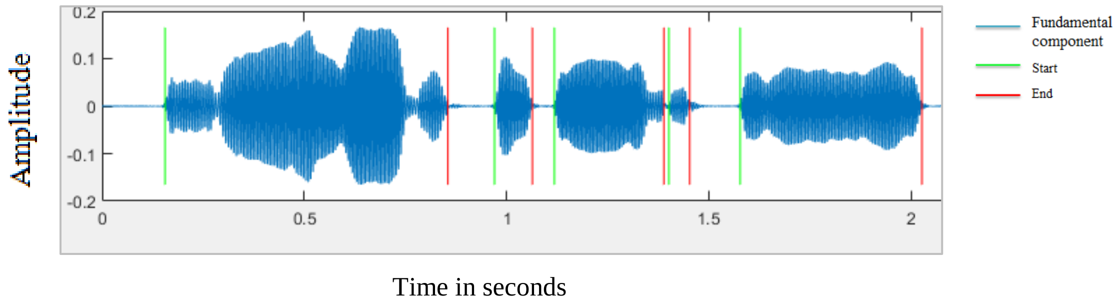


Fig 4.10 V/NV regions of reference signal for third verse

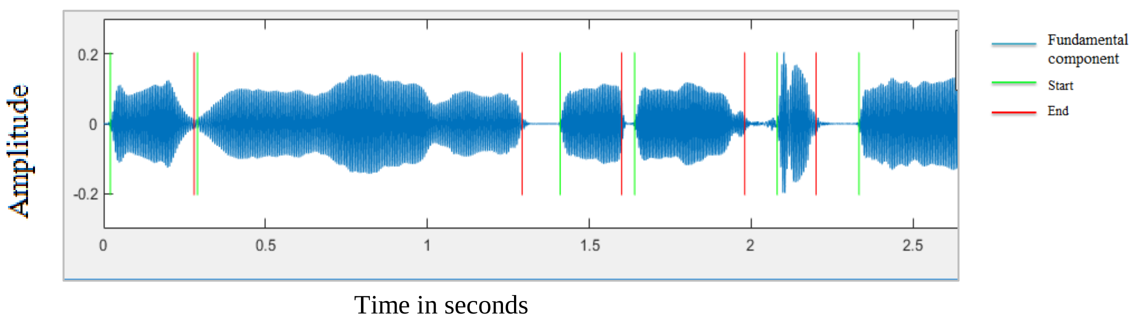


Fig 4.11 V/NV regions of learner’s signal for third verse

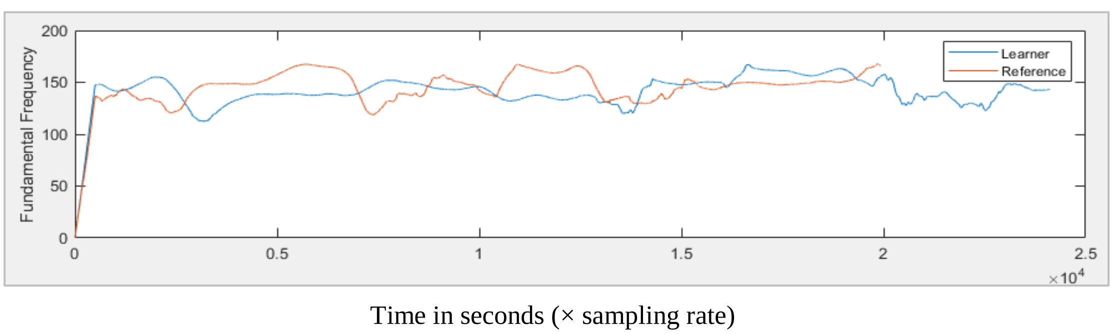


Fig 4.12 Plot of IFF for reference and learner’s signals for third verse.

Figure 4.10 shows the fundamental component for the reference signal. Here we can see the different voiced regions. Figure 4.11 shows the fundamental component for the learner’s signal. We can see the different voiced regions and also the difference between this signal and the reference signal. Figure 4.12 plots the IFF of both the signals. It can be seen that the learner’s signal is occasionally below the reference signal. So, the learner can adjust his pitch by just looking at the IFF plot.

Chapter 5

Conclusions and Future Work

In this project, we were able to devise a method to compare two utterances of the same Vedic mantra and see the regions where the learner needs to change his way of pronunciation. The approach of the method involved extracting voiced/non-voiced regions in the speech signal. Once this is done, the resultant signal is divided into small segments of 400 samples and VMD is applied to each of them to compute the fundamental frequency component for the whole signal. The fundamental frequency component is used to calculate the IFF using Hilbert transform. In the method, we have concatenated only the voiced regions and neglected the non-voiced for calculating the IFF. So, proper shifting is required to correlate the obtained IFF with the original signal.

This algorithm can be used to develop an automated system for providing suggestions for the learner so that he/she can adjust his pitch and match it close to the reference. This method needs to be extended on a wide range of Vedic mantras before using it for learning purposes.

References

[1] S. Tapas, S. Kaustav, Vedic swara recognition system: A move towards vedic chanting, *IOSR journals*, vol. 20, no. 4, 2018.

[2] S. Ghaligi, H. R. Nagendra, R. Bhatt. Effect of Vedic chanting on memory and sustained attention, *Indian Journal of Traditional Knowledge*, vol. 5, no. 2, pp 177-180, 2006, http://nopr.niscair.res.in/handle/123456789/6832.

[3] A. Padam, N. Sharma, O. S. K. S. Sastri, S. Mahajan, R. Sharma, D. Sharma, Effect of listening to Vedic chants and Indian classical instrumental music on patients undergoing upper gastrointestinal endoscopy: A randomized control trial, *Indian Journal of Psychiatry*, vol -59,no. 2, pp 214-218, 2017.

[4] D.B. Wolf , N. Abell, Examining the effects of meditation techniques on psychosocial functioning, *Research on Social Work Practice*, vol. 13, no. 1, pp 27-42, 2003.

[5] J. K. Staples. The science of the mantra. *Science for the Yoga Therapist*, available at: http://www.awarenesstechnologies.net/wp-content/uploads/2018/06/YTT-Mantra-article-Staples.pdf.

[6] K. Dragomiretskiy, D. Zosso. Variational mode decomposition, *IEEE Trans. Sig Process*, vol. 62, no. 3, pp 531-544, 2014.

[7] A. Upadhyay, R.B. Pachori. New methods based on variational mode decomposition for speech signal analysis, Ph.D Thesis, Discipline of Electrical Engineering, IIT Indore, 2017.

[8] A. Upadhyay, R.B. Pachori. Speech enhancement based on mEMD-VMD method. *Electronics Letters*, vol. 53, no. 7,pp 502-504, 2017.

[9] T.Zhang,W. Chen, M. Li. AR based quadratic feature extraction in the VMD domain for the automated seizure detection of EEG using random forest classifier, *Biomedical Signal Processing and Control*, vol. 31, pp 550-559, 2017.

[10] Ş.Yücelbaşa, C. Yücelbaş, G. Tezel, S. Özşen, Ş. Yosunkaya. Automatic sleep staging based on SVD, VMD, HHT and morphological features of single-lead ECG signal, *Expert Systems with Applications*, vol. 102, pp 193-206, 2018.

[11] A. Mert. ECG feature extraction based on the bandwidth properties of variational mode decomposition, *Physiol Meas.* ,vol. 37, no. 4,pp 530-43, 2016.

[12] Y. Wang, R. Markert, Filter bank property of variational mode decomposition and its applications, *Signal Processing*, vol. 120, pp 509-521,2016.

[13] H. Huang, J Pan, Speech pitch determination based on Hilbert-Huang transform, *Signal processing*, vol. 86, no. 4, pp 792-803, 2005.

[14] A. Potamianos, P. Maragos. A comparison of the energy operator and the Hilbert transform approach to signal and speech demodulation, *Signal Processing*, vol. 37, no 1, pp 95-120, 1994.

[15] U. Timothy. Envelope calculation from the Hilbert transform, available at https://www.researchgate.net/publication/257547765_Envelope_Calculation_from_the_Hilbert_Transform, 2006.

[16] P. Jain, R.B. Pachori, Marginal energy density over the low frequency range as a feature for voiced/non-voiced detection in noisy speech signals, *Journal of the Franklin Institute*, vol. 350, no. 4,pp 698-716, 2013.

[17] A. Upadhyay, R.B. Pachori. Instantaneous voiced/non-voiced detection in speech signals based on variational mode decomposition, *J Franklin Inst*, vol. 352, no.7, pp 2679-2707,2015.

[18] A. Upadhyay, M Sharma, R.B. Pachori. Determination of instantaneous fundamental frequency of speech signals using variational mode decomposition, *Computers & Electrical Engineering*,vol. 62,no. 1, pp 630-647, 2017.

[19] A. Upadhyay, R.B. Pachori. A new method for determination of instantaneous pitch frequency from speech signals. *IEEE Signal Processing and Signal Processing Education Workshop*, Salt Lake City, UT, 2015, pp. 325-330, 2015.

[20] S. R. Kharvatkar, M. Sharma, D. G. Khairnar and I. C. Naik, "Detection of pitch frequency of Indian classical music based on Hilbert-Huang transform for automatic note transcription," , *Fourth International Conference on Computing Communication Control and Automation (ICCUBEA)*, Pune, India, pp. 1-4,2018.

[21] P. Jain, R.B. Pachori. Event-based method for instantaneous fundamental frequency estimation from voiced speech based on eigenvalue decomposition of the Hankel matrix. *IEEE/ACM Trans Audio Speech Lang Process*, vol. 22, no. 10, pp 1467-1482,2014.