International Journal of Computer Architecture and Mobility
(ISSN 2319-9229) Volume 1-Issue 7, May 2013

Carnatic Swara Synthesizer (CSS) Design for different Ragas

Shruti Iyengar, Alice N Cheeran

***Abstract*— Carnatic music is one of the oldest forms of music and is one of two main sub-genres of Indian Classical Music. Carnatic Swara Synthesizer(CSS) has been designed to generate Carnatic Swaras at various pitches for male and female voice. The synthesizer has huge applications in music tutoring, assisted learning and a quick reference point for Carnatic music connoisseurs. We have been able to synthesize the seven swaras for both male and female at different pitches as required for a linear raga such as "Mayamalavagowla Raga". This can be further extended to synthesize any other linear or non-linear raga. Spectrogram analysis comparing male and female synthesized sounds and comparison between different synthesized vowels have been performed. The results are presented as a Mean Opinion Score calculated based on feedback provided by Carnatic music teachers and students.**

***Keywords*— Carnatic Music, raga, swaras, shruti, formant, synthesizer, filter**

I. INTRODUCTION

Carnatic music and Hindustani music are traditional Indian systems of music and are very different from the traditional Western system of music. Carnatic music system is a *just tempered* system of music compared to the *even tempered* system of Western music. This system of music gives the singer the flexibility to start a particular song at any frequency as the fundamental frequency. An additional important difference is that in Carnatic music an octave has 22 intervals as against the 12 intervals of an octave system in Western music and Hindustani music [2]. There are essentially seven notes otherwise called "*swaras*" in Carnatic music called S, R, G, M, P, D, N which is synonymous to C, D, E, F, G, A, B in the Western music. The ascending order of the arrangement of *swaras* is called *Arohanam* and the descending order of the arrangement of *swaras* is called *Avarohanam*. *Śruti* is musical pitch which ordinarily refers to frequency. It can be said as a group of frequencies with varying amplitudes. Before the arrival of the harmonium in India, musicians used either a tambura or a specific pitch reference instrument, such as the nadaswaram, to produce the drone (pitch or Śruti). Śruti Box is a device designed first by "Radel System", is now used by the carnatic music practitioners from novice, trained to expert to match their pitch. A Śruti box" has a set of pitches at intervals which are pre-recorded and played in loop to provide a reference to the musical exponent.

This paper attempts to generate the swaras synthetically and thereby provide a highly customizable and simple interface to replace or compliment the Śruti box. This paper discusses in detail, how various swaras have been synthesized for a particular raga- "Mayamalavagowla Raga" at customized pitches for male and female. At present there are no techniques to synthesize ragas at desired pitches. A lot of work has been done for fundamental frequency estimation of carnatic music singers [2]. CSS would greatly facilitate carnatic music tutors to generate and demonstrate various ragas at desired pitches for training musicians from novice to expert, which is otherwise not possible.

Speech sounds are characterized by a number of different articulations or vocal tract formants. Thus the vocal tract is a complex filter, and the formants are in the vocal tract's filter function. Different vocal tract configurations yield different filters. Different fundamental frequencies (pitches) change the harmonic spacing (and thus the resolution of the spectrum), but the shape of the spectrum is constant. The frequencies of the source and the frequencies of the filter are independent. In Carnatic music fundamental frequency refers to frequency of the middle octave 'S'. Carnatic Swara is the combination of the fundamental frequency which is the source and the formant frequency which acts as the filter together. Formant frequencies are a set of natural frequencies in vocal tract, having unique physical properties associated to each vowel. Vocal tract modifies the signal causing formant (pole) and anti- formant (zero) frequency.

There are twelve Swaras in Carnatic Music denoted famously as - s, r, R, g, G, m, M, p, d, D, n and N. Frequencies assigned to each Swara is not fixed, but is relative. The frequency of all Swaras depends on the frequency of the base Swara "s". That is, the twelve Swaras are positioned in geometric progression on the frequency line with basis of base Swara "s". Ragas are another characteristic of Indian music, be it Carnatic or Hindustani genre. Permutations and combinations amongst the twelve Swaras yield a *Raga*. The scope of the paper is limited to the generation of seven basic swaras of Mayamalavagowla Raga.

This paper briefly describes how a Carnatic swara synthesizer has been designed to produce swaras for various ragas at customized desired pitches for male and female. The system model has been described in detail along with distinguishing features and differences in male and female voices and there characteristics. Spectrogram analysis and comparison has been done to illustrate the difference between synthesized male and female voice as well as synthesized vowels.

II. SYSTEM MODEL

A. Approach

Mayamalavagowla Raga is a linear raga with seven swaras distributed in frequency line as depicted in Figure 1.

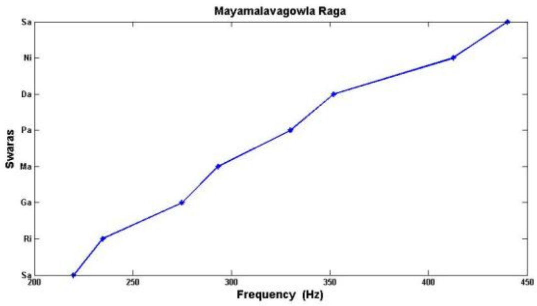

Fig 1 Plot depicting the seven swaras distributed in frequency line for Mayamalavagowla raga

The respective frequencies and ratio of the same is provided in Table 1. The ratio of the upper swara 'S' and the lower swara's' is always 2 depicting an octave

TABLE I
RELATION BETWEEN THE FREQUENCY AND RATIO OF SWARAS WITH THE BASE SWARA "SA"

Swara	**Ratio**	**Frequency**
S	1	220
R	16/15	234.7
G	5/4	275
M	4/3	293.3
P	3/2	330
D	8/5	352
N	15/8	412.5
S	2	440

A comparison of a linear and complete raga like Mayamalavagowla and a Non-linear (vakra) raga like Bhairavi is provided in figure 1 and 2. The current work does not provide results for non-linear, in-complete ragas but the approach does not restrict the scope of the same for extension. As a proof of concept, we have limited ourselves to Mayamalavagowla Raga for the simple reason for it being complete and linear.

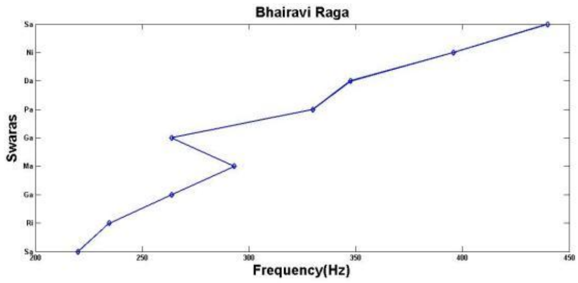

Fig.2 Plot depicting the seven swaras distributed in frequency line for 9 Swaras for Bhairavi Raga

The approach for generating Ragas have been kept intentionally simple to enable the adoption of the approach in embedded synthesizers with low memory requirements and low processing power.

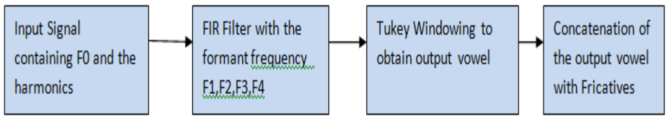

Fig. 3 Block diagram of the system model.

The inputs for the system are the pitch at which the Swaras are to be synthesized and the gender for generation i.e. male or female voice. The pitch/Śruti as explained is the fundamental frequency on basis of which other swaras are generated. Input signal as described in the Figure 3 is the combination of sinusoids of fundamental frequency and harmonic frequencies which are integral multiples of this fundamental frequency.

As per the "source filter theory"[10] our vocal fold acts as the source producing the fundamental frequency and harmonics. Thereby the first block in Figure 3 is designed to mimic the vocal fold vibration. The spectrum of this wave contains energy at the fundamental frequency of laryngeal vibration at integral multiples of the fundamental frequency representing the harmonics.

This source signal is passed through an FIR Filter which is characterized by the formant frequencies .The vocal tract (or actually, air in the vocal tract) has certain resonances. We call these formants. Thus the vocal tract can be modelled as a complex filter, and the formants are peaks in the vocal tract‘s transfer function. Formant frequency depends on the vowel produced. Swaras S, G, M, P, D are all combination of fricatives with the vowel "*ah*". Formant frequencies for these are designed accordingly with F1 = 700, F2 = 1220, F3 = 2600, F4 = 3500 [3]. The swaras R, N are combination of fricatives with vowel "*ee*". Formant frequencies for these are designed with F1 = 300, F2 = 2285, F3 = 2508, F4 = 4565 [3]. Source (input) frequencies close to a resonant frequency are amplified other frequencies are attenuated. Thereby formants can be generated with the help of Block 2 in Figure 3 which is the FIR filter|. Figure[4] provides the time and frequency domain representation of the vowel "*ah*" and Figure [5] provides the same for vowel "*ee*".

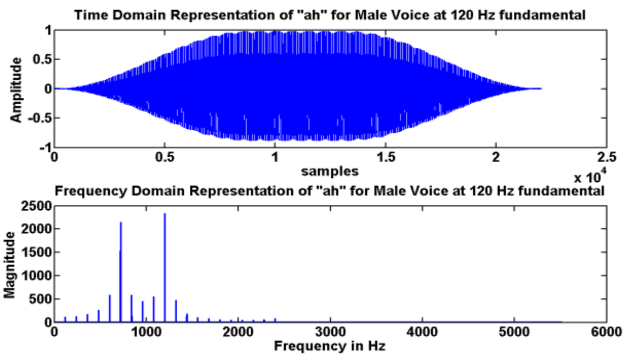

Fig 3 Time/Frequency Domain Representation of —ahǁ for Male Voice

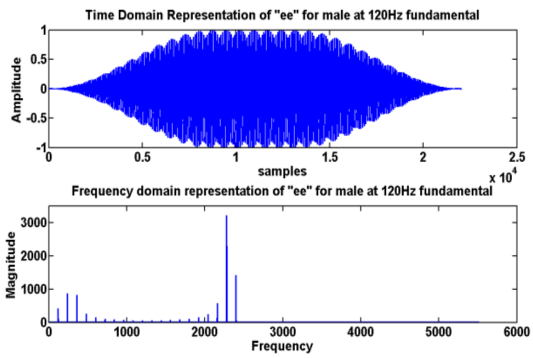

Fig 4. Time/Frequency Domain Representation of " ee" for Male Voice

B. *Wave Shaping*

Post the synthesis of swaras, it is imperative to shape the waveform to simulate the modulation of speech in human. Windowing and wave shaping techniques are employed to achieve the same. With several windowing methodologies available for use, with the current application in mind, we have chosen Tukey window to serve the purpose.

The time and frequency domain representations of the same are presented in Figure[5]. The primary advantage of employing a tukey window comes from the fact that the ascent and descent of the function in time domain is gradual and can be controlled and there is a sustenance period where there is no change in amplitude. A generation of swara by human is also similar in the sense that the volume of the swara is modulated and held constant for a considerable period of time before the demise of the same.

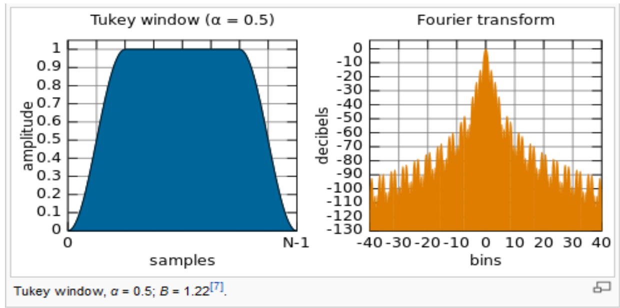

Fig 5 Illustrating the Tukey Windowing

Voice of a singer, whether it is a male or female largely depends on the fundamental and formant frequencies. Male vocal folds are larger than female by approximately 60% and thus yielding a lower fundamental frequency in males. Male vocal tract is 15% longer than females corresponding to lower formants in females .As per the Acoustic Analysis done by Surendra Shetty et al[1] ,the ratios of mean formant frequencies F1, F2, F3 were calculated and a composite formant frequency scale factor (female/male) was calculated by averaging the three ratios , it was found to be 1.2 .

By the described method, swara can be synthesized at given pitch for a male/female voice

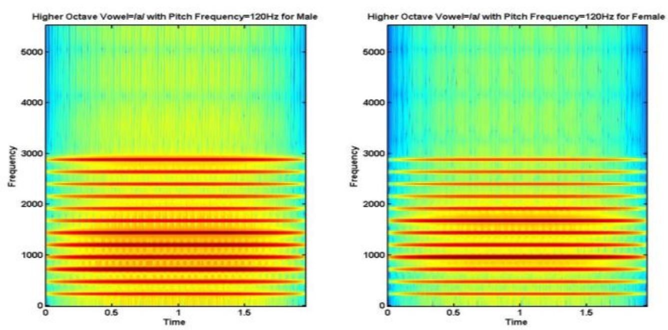

Fig 6 :Spectrogram Comparison for male and female at 120 Hz for the vowel

/a

Figure 6 indicates the spectrogram comparison for male and female for vowel "*ah*" at same pitch frequency 120 hz female differ by a certain factor, the spectrogram clearly indicates the difference. Dark red bands which indicate the formants are close to range of 750 Hz to 1750 Hz for males while for females they are at a higher range from 1000Hz to 2000 Hz. Energy concentration for females is more for higher frequency range than for females. First formant F1 for female =1000 Hz while for male s it is around 750 Hz. The pitch frequency being the same for both we can clearly see the acoustic spectrum is spread over the same fundamental frequency and the respective harmonics but the formants act as the filter. This filter shifts the energy of some harmonics out while boosting others. The dark red bands indicate the formants with high energy.

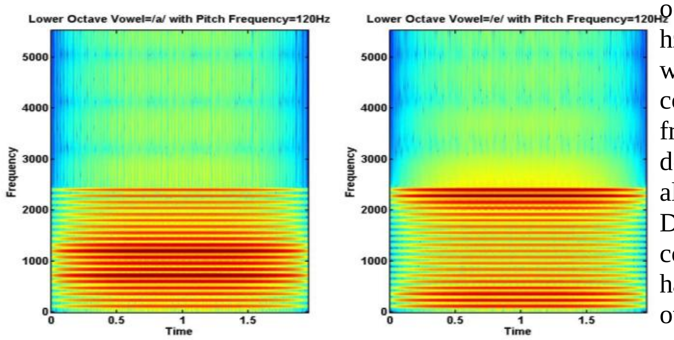

Fig 7:Spectrogram Comparison of vowels /a and /e for pitch frequency 120hz

Figure 7 indicates the spectrogram comparison for vowels /a and /e. Acoustic spectrum is spread over same fundamental frequency and harmonics for both since the pitch frequency is same. While the formant frequency largely depends on the vowel and the movement of tongue, which differs for /a and /e. For the vowel /a the first formant F1 is around 750, the formant energy is concentrated between 700 and 1700 .While for vowel /e the first Formant F1 is around 300 , the formant energy indicated by dark bands are concentrated in the lowest frequencies around 300 -500 hz and highest frequency 2000 to 2500 hz.

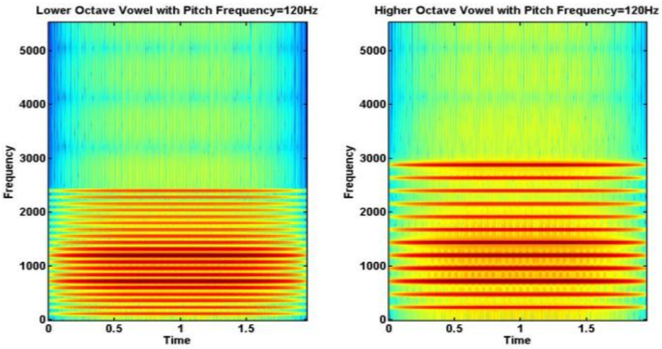

Fig 8 Spectrogram comparison for lower octave vowel at 120 hz and higher octave vowel at 240 Hz

Figure 8 indicates comparison of spectrogram for lower octave vowel /a at 120 hz and higher octave vowel /a at 240 hz. Acoustic Spectrum for the higher octave vowel is clearly widely spaced and spread over higher frequency range compared to the lower octave vowel as the fundamental frequency is doubled. Since the fundamental frequency is double in case of higher octave the harmonic frequencies are also larger and it is spread over a larger frequency range. Difference in pitch frequency does not affect the energy concentration or the formant frequencies. We can see, both have an F1 of around 700 hz and formant energy is spread over a range 700 hz to 1700 hz for both

III. .RESULTS

The algorithm was implemented with the aid of Matlab. Most of voice/speech synthesis applications are evaluated through the MOS (Mean Opinion Score). We adopt the same methodology to perform the same. The evaluators consisted of 10 Carnatic musicians and tutors who invariably resort to classical methods of teaching. There experience with the current approach has been encouraging and motivating. The results of the same are explained in Table 2

TABLE II
ILLUSTRATING THE MOS FOR MALE AND FEMALE VOICES AT 3 DIFFERENT FUNDAMENTAL FREQUENCIES

Gender	**Fundamental Frequency**	**Mean Opinion Score**
Male	120	2.75
	140	2.9
	160	2.8
Female	200	2.95
	220	3
	240	2.8

IV. CONCLUSIONS AND FUTURE WORK

In this paper, the attempt has been to keep the implementation simple to enable adoption of the same to embedded systems with low processor power and low memory. Raga synthesis generating as a result of this paper can be a powerful tool for Carnatic Music tutors to provide insight into various Ragas. Ensuring a non-robotic rendering of the swaras has been one of the focus of the papers.

Future work would involve research on generation of non-linear (vakra) and in-complete ragas. More challenging would be the generation of ragas with gamakas (noticeable modulation in a given swara). The approach taken in this paper and its extension to gamakas would be a significant force to reckon with in music tutoring and perhaps music synthesis. This perhaps would be the single most motivation for us to carry this work forward.

REFERENCES

[1] *Surendra Shetty, Raga Mining of Indian Music using Arohana Avarohana Pattern International Journal of Recent Trends in Engineering. Vol1,No 1,May 2009*

[2] *Rajeswari Sridhar, Fundamental Frequency Estimation of Carnatic Music Songs Based on the Principle of Mutation, IJCSI Vol7,Issue 4,No 7,July 2010*

[3] *Ramya S, Automatic Music Note Transcription Systems Using Artificial Neural Networks, ICEDSP Special Issue (0975-8887)*

[4] *J.Chandrasekaran, "Spectral Analysis of Indian Music" Indian Journal Traditional Knowledge Vol.4(2),April 2005*

[5] *Lorin F .Wilde, "Analysis and Synthesis of Carnatic Music" Phd Thesis .Massachusetts Institute of Technology 1996*

[6] *S.Hariram , Real time adaptive note and Swara Recognition using HMM and Neural Network, International EUROSIM Congress June 26-29 2001*

[7] *B.Tarakeswara Rao, Automatic Melakarta Raaga Identification System Carnatic Music, IJARAI Vol 1, No 4 2012.*

[8] *K Priya, Data Mining Techniques for Automatic Recognition of Carnatic Raga Swaram Notes. International Journal of Computer Applications Volume 52-No 10,August 2012*

[9] *Gopal K Koduri, Charachterization of Intonation in Carnatic Music by Parametrizing Pitch Histograms, ISMIR 2012.*

[10] *Siu-Fung Poon, Contribution of Voice Fundamental Frequency and Formants to the Identification of Speaker, ICPhS XVII ,17-21 Aug 2011.*