Analysis of Vocal Tract Shape Variability based on Formant Frequency Ratio at Various Conditions of Vowels for Indian English Speakers

Anil Kumar Chandrashekar\(^1\) and M. B. Manjunatha\(^2\)

\(^1\)Electronics Engineering, Jain University, Bangalore – 560069, Karnataka, India; canilkumarc22@gmail.com
\(^2\)Akshaya Institute of Technology, Tumkur – 572106, Karnataka, India; manju.kari29@gmail.com

Abstract

The paper presents the notability of variation of vocal tract shape and formant frequency with its adjacent ratios, for the Indian English speakers for the five vowels of the English language. we have estimated the vocal tract shape of the Indian English speakers. For the estimation we have incorporated Autoregressive Model and the same for the formant frequency. The speech samples are considered with three various utterances namely the consumption of Ice cold water, with time relaxation of five minutes compared with normal recordings of the vowels namely /a/, /e/, /I/, /o/ and /u/. These utterances are recorded for 20 individual iterations. the vocal tract shape estimation and formant frequency estimation are done on the Matlab platform. the vocal tract shape of first vowel /a/ comes to normal shape rapidly with time lapse of five minutes. the vocal tract shape of vowel /e/ shrinks after consumption of ice cold water and slowly attains the normal shape. The vowel /I/ and /or/ vocal tract shape changes linearly for all the conditions considered, whereas the vocal tract shape of /u/ compresses for ice cold water and retains to be same even after a time lapse of five minutes. All the above variations are done by considering the vocal tract length of 17cm and it is modelled according to lossless uniform tube model. these outcomes are used for observation of vocal tract infection among the speech disorder patients, it can be adopted for the user authentication system by considering it as a vocal tract signature.

Keywords: English Vowels, Formant Frequency, Indian English, Vocal Tract Shape, Vowels

1. Introduction

That speech is the most accepted and convenient means of communication is well known and recognized. The narrow concept of speech is that it is just a sequence of sounds punctuated by abrupt changes happening from one to another or some signals that are ignored and go into oblivion soon after utterance. Speech is not just an information signal it is actually a complex wave and acoustic output arising as a result of the speaker\'s effort.\(^3\) Speech is signal with information galore exploring frequency modulated, amplitude modulated and time-modulated carriers (example: resonance movements, harmonics, noise, pitch, power, duration).

Speech analysis is synonymous with feature extraction of speech. Speech sounds are sensations of air pressure variations produced by exhaled air and later modulated and shaped by vibration of glottal cords and the resonance of the vocal tract during the time air is pushed out through the lips and the nose. The entire gamut of information is basically conveyed in the large of the traditional telephone bandwidth of 4 kHz. Speech energy 4 kHz reflects audio quality and sensation.\(^2\)
2. Speech Production

Speech is meant for communication. It has the distinct feature as a signal that carries a message or information. It is known to be an acoustic waveform that carries a message or information. It is known to be an acoustic waveform that carries temporal information from the speaker to the listener. Efficiency underlies acoustic transmission and reception of any speech. But this is applicable only for transmission over a short distance. There is a spread of radiated acoustic energy at frequencies that are used by the vocal tract and ear. But this gets reduced in intensity rapidly. Even on occasions when the source is able to produce substantial volume of acoustic power, there is a support of only a fraction thereof by the medium without any distortion, while the rest of the it gets squandered in air dust particles molecular disturbance. It also results in getting over aeromolecular viscosity. The ambient acoustic noise places a restriction or limit on the sensitivity of the ear. Physiological noses to play this role in and around the ear drum. Voluntary, formalized motions of the respiratory and masticators apparatus have the speech as the acoustic end product. The closed loop has the ability to develop, control, maintain and correct it. Acoustic feedback of the hearing mechanisms and the kinesthate feedback of the speech musculature too have a role here. The central neurons system organizes and coordinates information from the senses which is then used for directing the function as also for delivering the descend, linguistically dependant, vocal articulator motion and acoustic speed.

2.1 Problem Statement

The vowel speech database considered in our paper is by considering all five vowels of English at three different conditions. A database is created by us at normal recording room environment, vowels (/a/, /e/, /I/, /o/, /u/) to be analyzed with respect to Formant frequency ratio with respective vocal tract shape variability.

2.2 The Speech Communication Pathway

Figure 1. gives a simplified view of the speech communication route starting from the speech and reaching the listener. An idea or a concept originates at the linguistic level of communication, in the speaker's mind. It is the stimulated longitudinal acoustic wave propagation in air starting from the speaker and terminating with the listener.

2.3 Formant Frequency

Excitement of a fixed vocal tract produce vowels with quasi-periodic pulses of air, forced through the vibrating vocal cords. A quasi-periodic puff of air flow is the source, acting through vibrating vocal folds at a definite basic frequency. The term “quasi” is used considered with perfect periodicity. An formant frequency is the accumulation of acoustic energy in the particular frequency in the input speech wave. Every formants each at a different frequency with the bandwidth of roughly 1000Hz. Every formant corresponds to resonance in turn results in par-cor coefficients in the respective vocal tract, where F1 represents width of the pharynged cavity and position of the tongue, F2 indicates the length of oral cavity and position of the tongue, F3 denotes the shape of the lip of speaker and F4 is the additional resonant peak. Hence first two formant frequencies carry the maximum information.
about the speaker information and the next two formants holds the necessary information for speaker identity essential for Automatic Speaker Recognition process. Hence F2/F1 ratio shows the versatility, where as F3/F2, F4/F3 ratio showing less variation.

### 2.4 Vocal Tract Shape

It is possible to model the vocal tract as an acoustic tube resonance referred to as formants and anti-resonances and as a chain of cylinders of cross sectional area with varying features. Alteration in the form of the acoustic tube is caused by movement of the articulators of vocal tract resulting in changes in frequency response. Poles are caused by resonances while anti-resonances arise from zeroes of frequency response. Articulators are structures in the vocal tract, which moves in the production of sounds of varying decibels. The important articulators are the lips and the tongue. They can be classified as direct and indirect the latter are the jaw and mandible helping the positioning of the tongue and the lips for a variety of sounds. Such positioning of the tongue creates a number of vocal tract shapes that are needed for the production of speech sounds.

Vocal tract shape is the combination of oral cavity and pharynged cavity which is directly represented by the first two formant frequencies namely \( F_1 \) and \( F_2 \) in terms of Hertz. The tract length for an adult male is generally 17cm running from glottis to lips, the cross sectional area being up to 20cm\(^2\).

### 2.5 Concatenated Tube Model of Vocal Tract

The widely used model for speech production is based on the assumption of the capacity of the vocal tract to represent as a concatenation of small cylindrical tubes figure 3. An independent variation of the cross sectional area of any tube for stimulating the changing shape of the vocal tract is seen. Digital signal processing (DSP) techniques find use in modelling, use of the speech signal converted into its discrete time sequence, on the assumption of all cylindrical segments being small are shown in Figure 2. as of equal length. This variation in the shape and length of the tract at different points along its length, ultimately causes production of different sounds.

### 2.6 Speech Database

20 samples of 20 subjects (male speakers aged about 18-25) at different times and at different conditions i.e., normal sample, by the instant of consuming ice cold water and with time lapse of 5 minutes were recorded using table top mic make of i-ball Model No M27 with sensitivity -58dB ±3dB, frequency response of 100Hz to 16kHz, with sampling frequency of 22,100Hz in MAT LAB and the samples were normalized to have amplitude normalization by considering English vowels by Indian English speakers (a,e,i,o,u). five Indian English vowels are categorized into three categories namely rounded vowels, spread vowels and unrounded vowels i.e. /a/, /e/ are unrounded, /i/ spread vowel and /o/, /u/ are rounded vowels.

### 2.7 Forward-Backward (FB) auto regressive model for Formant Estimation

Formant Frequency Estimation Algorithm for the speech analysis based on Forward-Backward (FB) in auto regressive model is shown in Figure 3. Figure depicts the estimation process for formant frequency estimation. By the model of speech production, it is analyzed that the speech encounters a spectral tilt of –20dB/decade, hence to match this, pre-emphasis filter is used to flatten the spectrum and boost the higher frequencies, followed by Hamming window, by obtaining the AR Model and Vocal Tract transfer function. From the frequency spectrum of the Vocal Tract transfer function, the formants are extracted.
3. Block Diagram for Vocal Tract Shape Calculation

The figure 4. shows the various steps involved in the estimation of the vocal tract shape by providing speech signal as input and with various co-efficient calculation and its description as follows:

Pre-emphasis: the area function obtained through use of reflection coefficients cannot be the area function of the human vocal tract. In the event of the pre-emphasis being used prior to linear predictive analysis for removal of defects arising out of glottal pulse and radiation, the area function that emerge as results are often similar to vocal tract configurations that find use in human speech. The speech production model discloses the speech undergoing a special tilt of -6dB/octave. The use of pre-emphasis in displaying the spectrogram if speech signals is rather common.\textsuperscript{12,13}

3.1 Window Technique

The window function \( w(k - n) \) is a real window sequence that finds use in isolation of isolate the portion of the input sequence analyzed at a particular time index \( k \). Many window functions have been generated to improve upon the basic rectangular window design, such as Hamming, Hanning, Bartlett, Blackman, Kaiser, etc., each having different specification with regards to its frequency response. In this paper, LP analysis was performed on frames weighted with the hamming window. This window, \( w(n) \) was chosen as it provides a good balance between its main lobe width and side lobe attenuation.\textsuperscript{14}

3.2 Autocorrelation Method

The most widely used method to Linear Predictive Analysis is called the Autocorrelation method. It is the most popular method of short-term LP analysis. This method provides the most computationally efficient manner in determining the LP parameters with guaranteed stability. It avails of the Toeplitz property possessed by the autocorrelation matrix.\textsuperscript{16,17}

3.3 Parcor Coefficients

PARCOR coefficients are bounded ±1. This has been seen earlier. This features has given them the facility of being the attractive parameter for quantization.\textsuperscript{18} With a set of PARCOR coefficients, it is first possible to use them at step (2) of the Levinson Durbin algorithm followed by the getting of an algorithm that can help conversion, which can be computed from a given set of predictor coefficients by working backward through the Levinson Durbin algorithm.\textsuperscript{19,20}

4. Intra Speaker Vocal Tract Shape Estimation Algorithm

1. Using Auto Regressive model of speech analysis determining the formant frequency and vocal tract shape variability of an individual subject.
2. Study of variability of the above vocal tract shape among 30 different subjects to high light and identify Intra Speaker Variability.
3. The above identified can be adopted for personal identification similar to signature. It can also be named as Vocal Tract Signature of an individual.
4. Find the formant frequency variability for each speaker (subject).
5. Finding the formant frequency ratio and its respective graph with vocal tract shape is plotted.
6. Plot the resultant pattern for a subject for the vowel.
5. Results and Discussion

![Figure 5](image-url)  
Figure 5. Vocal tract shape for vowel /a/ for all the three conditions.

Table 1. Four different formant frequency of vowel /a/

|        | \(f_1\) (Hz) | \(f_2\) (Hz) | \(f_3\) (Hz) | \(f_4\) (Hz) |
|--------|--------------|--------------|--------------|--------------|
| Normal | 612.18209    | 1517.1281    | 2144.2687    | 2970.6265    |
| Ice cold| 460.34875    | 1689.3834    | 2097.3323    | 2751.231     |
| After five min | 514.05292 | 1736.4056    | 2265.3495    | 2993.7702    |

![Figure 6](image-url)  
Figure 6. Comparative graph of formants for the vowel /a/.

The vocal tract variation for the vowel /a/ is as shown in figure 5, for all the three conditions, which showing the constriction of vocal tract for ice cold water compared with normal condition resulting in higher \(F_2\) as shown in table 1 and comparative graph of all four formants as indicated in figure 6, impacts on the increased length of oral cavity.

![Figure 7](image-url)  
Figure 7. Vocal tract shape for vowel /e/ for all the three conditions.

Table 2. Four different formant frequency of vowel /e/

|        | \(f_1\) (Hz) | \(f_2\) (Hz) | \(f_3\) (Hz) | \(f_4\) (Hz) |
|--------|--------------|--------------|--------------|--------------|
| Normal | 291.32943    | 1702.7908    | 2292.883     | 2878.1543    |
| Ice cold| 482.99677    | 2078.1335    | 2512.4837    | 3137.3099    |
| After five min | 351.10009 | 1984.8818    | 2382.0561    | 3042.5215    |

![Figure 8](image-url)  
Figure 8. Comparative graph of formants for the vowel /e/.

The vocal tract variation for the vowel /e/ is as shown in figure 7, for all the three conditions, the width of the pharynged cavity is less and increased oral cavity for the normal condition, vice versa for the ice cold water and continues to be same for the time lapse of 5 minutes. Table 2 shows the formant frequency tabulation and figure 8 is comparative graph of it.

![Figure 9](image-url)  
Figure 9. Vocal tract shape for vowel /i/ for all the three conditions.

Table 3. Four different formant frequency of vowel /i/

|        | \(f_1\) (Hz) | \(f_2\) (Hz) | \(f_3\) (Hz) | \(f_4\) (Hz) |
|--------|--------------|--------------|--------------|--------------|
| Normal | 647.53608    | 1657.1006    | 2306.1086    | 2961.7258    |
| Ice cold| 830.62063    | 1664.4509    | 2138.7426    | 2952.5652    |
| After five min | 687.52753 | 1657.8592    | 2271.0133    | 3037.5383    |

The vocal tract variation for the vowel /i/ is as shown in figure 9, for all the three conditions, as \(F_1\) is high due to decreased oral cavity and constant pharynged cavity as /i/ is spread vowel \(F_1\) and \(F_2\) are varied proportionally. Table 3 shows the formant frequency tabulation and figure 10 is comparative graph of it.
Analysis of Vocal Tract Shape Variability based on Formant Frequency Ratio at Various Conditions of Vowels for Indian English Speakers

Figure 10. Comparative graph of formants for the vowel /i/.

Figure 11. Vocal tract shape for vowel /o/ for all the three conditions.

Table 4. Four different formant frequency of vowel /o/

|          | \( f_1 \) (Hz) | \( f_2 \) (Hz) | \( f_3 \) (Hz) | \( f_4 \) (Hz) |
|----------|----------------|----------------|----------------|----------------|
| Normal   | 495.06655      | 968.4073       | 1739.9349      | 2511.3353      |
| ice cold | 469.64243      | 934.61768      | 1632.054       | 2566.8026      |
| five min | 487.47977      | 1020.8186      | 1664.5661      | 2694.705       |

The vocal tract variation for the vowel /o/ is as shown in figure 11. For all the three conditions, the length of oral cavity is double as pharynged cavity first two formants are varied linearly. Table 4 shows the formant frequency tabulation and figure 12. is comparative graph of formants.

Figure 12. Comparative graph of formants for the vowel /o/.

Figure 13. Vocal tract shape for vowel /u/ for all the three conditions.

Table 5. Four different formant frequency of vowel /u/

|          | \( f_1 \) (Hz) | \( f_2 \) (Hz) | \( f_3 \) (Hz) | \( f_4 \) (Hz) |
|----------|----------------|----------------|----------------|----------------|
| Normal   | 426.10873      | 1436.3321      | 2080.3389      | 2674.8642      |
| ice cold | 760.19553      | 1541.877       | 2246.105       | 3028.9773      |
| five min | 736.92003      | 1545.9753      | 2178.9476      | 3071.0632      |

The vocal tract variation for the vowel /u/ is as shown in figure 13. For all the three conditions, as \( F_1 \) is less due to decreased pharynged cavity which \( F_2 \) is higher for normal condition but even after ice cold water consumption with time lapse of 5 minutes variation of vocal tract shape will not come to normal condition. Table 5 shows the formant frequency tabulation and figure 14. shows the comparison of formants.

Figure 14. Comparative graph of formants for the vowel /u/.

Table 6. \( f_2/f_1 \) of all the five vowels

|          | A    | E    | I    | o    | u    |
|----------|------|------|------|------|------|
| Normal   | 2.478| 5.85 | 2.56 | 1.956| 3.37 |
| ice cold | 3.67 | 4.302| 2    | 1.99 | 2.028|
| after five min | 3.377| 5.653| 2.411| 2.094| 2.097|
Table 6 and figure 15 provides the ratio of first two formants and its graphical representation for all the vowels and three different conditions for /a/ the ratio will increase from 2.478 to 3.67 for ice cold water, for /e/ the ratio decreases to 4.302 from 5.85, for /i/ ratio falls from 2.56 to 2, for /o/ ratio remains almost similar with minute variation and for /u/ ratio rolls from 3.37 to 2.028.

Table 7. \( f_3/f_2 \) of all the vowels

|        | A     | E     | I     | o     | u     |
|--------|-------|-------|-------|-------|-------|
| Normal | 1.4133| 1.3465| 1.3916| 1.796 | 1.44  |
| ice cold| 1.2414| 1.209 | 1.2849| 1.7462| 1.456 |
| after five min | 1.3046| 1.2   | 1.3698| 1.6306| 1.4094|

Figure 16. Comparative graph of \( f_3/f_2 \) for all the vowels.

Table 8. \( f_4/f_3 \) of all the vowels

|        | A     | E     | I     | o     | u     |
|--------|-------|-------|-------|-------|-------|
| Normal | 1.383 | 1.2552| 1.2842| 1.4433| 1.2857|
| ice cold | 1.3177| 1.2486| 1.3805| 1.5747| 1.3485|
| after five min | 1.3215| 1.2772| 1.3375| 1.6188| 1.4094|

Figure 17. Comparative graph of \( f_4/f_3 \) for all the vowels.

6. Conclusion

Using AutoRegressive modelling of speech processing the formant frequency’s are calculated and by extracting PAR-COR co-efficients from the speech database considered. The estimation of vocal tract shape was done and the effect of \( f_2/f_1 \), \( f_3/f_2 \), and \( f_4/f_3 \) were studied for the vowels (/a/,/e/,/i/,/o/,/u/) for individual speakers. The input speech signal considered is divided into frame length of 30msec with an overlapping of 10msec frame length by keeping the sampling rate of 22,100Hz by 663 samples in each frame. By autoregressive model of vocal tract shape estimation from reflection co-efficient, along with respective format frequency are done to calculate their ratio corresponding to its adjacent formant frequency. The variation of vocal tract shape of unrounded vowels /a/ and /e/ are in same passion independently for all the three conditions, the response of vocal tract shape of spread vowel /i/ is linear to formant frequency with ice cold water, time lapse of five minutes and the variability of vocal tract shape for rounded vowels /o/ and /u/ will shrink and sustain to be same even after ice cold water consumption and time lapse of five minutes. These observations will serve as the basic parameter setting for many user/speaker authentication and speech operated devices.
7. References

1. Anil Kumar C, Shiva Prasad KM, Manjunatha MB, KodandaRamaiah GN. Basic Acoustic Features Analysis of Vowels And C-V-C of Indian English Language. ITSI-TEE. 2015; 3(1): 20–3.

2. Shiva Prasad KM, Anil Kumar C, Manjunatha MB, KodandaRamaiah GN. Gender based Acoustic Features and Spectrogram analysis for kannada Phonetics. ITSI-TEE. 2015; 3(1): 16–9.

3. Anil Kumar C, Manjunatha MB, Kodanda Ramaiah GN. Formant Frequency Based Analysis of English vowels for various Indian Speakers at different conditions using LPC and default AR modelling. International Journal of Computer Science and Information Security (IJCSIS). 2016; 14 Special Issue.

4. Powen Ru, Taishih Chi, and S Shamma. The synergy between speech production and perception. The Journal of Acoustical Society of America. 2003 Jan; J. Makhoul, "Linear Prediction: A Tutorial Review," Proc. IEEE, Vol. 63, pp 561-580,1975.

5. Rexy RR, Rosini RR, Mani Sankar R. A Novel Speech Recognition System using Hidden Markov Model. Indian Journal of Science and Technology. 2015 Nov; 8(32).

6. Shiva Prasad KM, Kodanda Ramaiah GM, Manjunatha MB. Backend Tools for Speech Synthesis in Speech Processing. Indian Journal of Science and Technology. 2017 January; 10(1). Crossref

7. Anil Kumar C, Shiva Prasad KM, Manjunatha MB, KodandaRamaiah GN. Vocal Tract shape estimation of Vowels and C-V-V-C for diversified Indian English Speakers. IEEE International Conference on Electrical, Electronics, Signals, Communications and Optimization. 2015 Jan; 1–7. Crossref

8. Shiva Prasad KM, Anil Kumar C, KodandaRamaiah GN, Manjunatha MB. Speaker based vocal tract shape estimation for kannada vowels. IEEE International Conference on Electrical, Electronics, Signals, Communications and Optimization. 2015; 1–6. Crossref

9. Wankhede NS, Shah MS. Investigation on optimum parameters for LPC based vocal tract shape estimation. IEEE-Emerging Trends in Communication, Control, Signal Processing and Computing Applications C2SPCA. 2013; 1–6.

10. Sohal B, Sandeep K. A HMM Integrated SVM Model for Hindi Speech Recognition. Indian Journal of Science and Technology. 2016 December; 9(47). Crossref

11. Scordillis M, Gowdy JN. Effects of the vocal tract shape on the spectral tilt of the glottal pulse wave form. IEEE Transactions. 1990; 86–9.

12. Sajeer K, Rodrigues P. Novel Approach of Implementing Speech Recognition using Neural Networks for Information Retrieval. Indian Journal of Science and Technology. 2015 December; 8(33).

13. khodai-joopaari M, Clermont F, Barlow M. Speaker variability on a continuum of spectral sub-bands from 297-speakers'non- contemporaneous cepstra of japon's vowels. Proceedings of the 10th Australian international conference on speech science and technology. 2004.

14. Dong-Il Kim, Byung-Cheol Kim. Speech Recognition using Hidden Markov Models in Embedded Platform. Indian Journal of Science and Technology. 2015 December; 8(34). Crossref December 2015.

15. Shiva Prasad KM, Anil Kumar C, Manjunatha MB, KodandaRamaiah GN. Various front end tools for digital speech. Processing in IEEE. 2015; 905–11. Print ISBN: 978-9-3805-4415-1.

16. Shiva Prasad KM, Kodanda Ramaiah GN, Manjunatha MB. Speech Features Extraction Techniques for Robust Emotional Speech Analysis/Recognition. Indian Journal of Science and Technology, 2017 January; 10(3): . Crossref

17. Kodandaramaiah GN, Giriprasad MN, Mukundarao M. Implementation of LPC based vocal tract shape estimation for vowels. The Technology world quarterly journal. 2010 March-April; 90(2): 97–102. ISSN 2180-1614.

18. Vallabha G, Tuller B. Choice Of Filter Order In Lpc Analysis Of Vowels. Sound to Sense. 2004 June; at MIT.

19. Shah MS, Pandey PC. Estimation of vocal tract shape for VCV syllables for a speech training aid. Proceedings of 27th International Conference on IEEE Engineering Medicine and Biology Society. 2005; 6642–5. Crossref

20. Raut SV, Panthangi LC, Akhil BG, Syed Faisal Ali, Sanjay HS a5nd

21. Bhargavi S. Classification of Sex based Speech Differentiation in Healthy Human Beings based on Voiced and Unvoiced Components. Indian Journal of Science and Technology. 2017 January; 10(1). Crossref