



The MINIMAL and MAXIMAL VOCAL TRACT SHAPE VARIABILITY FOR TELUGU VOWELS

R. Ashok Kumar Reddy¹, Dr G.N Kodanda Ramaiah², Dr M.N Giri Prasad³

¹(Research Scholar, Dept. of ECE, JNTU Ananthapuramu),

²(Professor, Dept. of ECE, Kuppam Engineering College, Kuppam),

³(Professor, Dept. of ECE, JNTU Ananthapuramu)

¹ashokreddy02@gmail.com, ²hod_ece@kec.ac.in, ³mahendragiri1960@gmail.com

Abstract

This paper presents vocal tract shape estimation for south Indian vowels based on LPC. This is demonstrated by Acoustic model on vowels and speaker basis using area function approximation to the vocal tract shape. We propose to use area function approximation of a person taken at different times, and in different context, at many instants. These vowel utterances of an adult male, and female, are recorded 30 times and the variability of the resulting shapes, are measured on Intra speaker basis. We have also looked at the spreads for the same vowel of the individual speaker at different occurrences and investigated the variability role, on 'speaker specific recognition applications'. The vocal tract shape variability estimation is implemented using MATLAB.

Index Terms: LPC, Acoustic model, Speaker recognition.

I. INTRODUCTION

Speech is the most accepted and convenient means of communication and 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 uttering. 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.

Speech is a 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. The entire gamut of information is basically conveyed in the traditional telephone bandwidth of 4 kHz. Speech energy 4 kHz reflects audio quality and sensation.

A. Vocal Tract Analysis

The Acoustic tube model, is an accepted model of the vocal tract [Chiba and Kajiyama, 1941, Dunn 1950, Stevens, et, al 1953; Fant 1965, Stevens 1972, & 1989]. The vocal tract is modelled as a coaxial concatenation of lossless, acoustic tubes, of different lengths, and diameters. The cross sectional area of any of the tubes can be varied independently to simulate the changing shape of the vocal tract. The first tube starts at the glottis and the last tube ends at the lips or the nostrils. Most of the acoustic tube models assume that the tract length for an adult male is generally 17cm running from glottis to lips, the cross sectional area being up to 20cm².

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 1. 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 as shown in Figure 1 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.

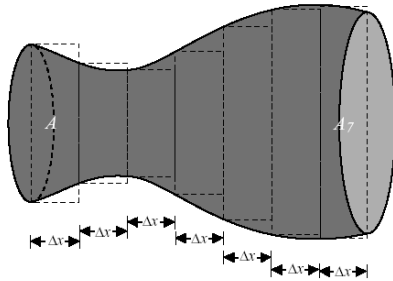


Fig 1: Concatenated tube model.

B. Objective of the Paper

This work models the vowels obtained from the model spectrograms, formants, pitch and vocal tract shape information. Error minimization is carried out using an all pole LPC filter. Analysis is done for a vowel in the above format, to get the vocal tract shape for vowels of males by taking 30 samples of 30 subjects at different times. The Vocal Tract shape arrived at for each subject for 30 sets of data at different instants of times. Using LPC, along with Correlation analysis, we found the vocal tract shape variability of the individual subject. Study of variability of the above Vocal Tract shape among 30 different speakers is highlighted to identify *Intra Speaker* variability. The time averages of the worst and the best patterns for the example of 30 subjects, has been found. The resultant worst pattern and resultant best pattern for a subject of the phoneme have been plotted, for different phonemes for male speaker. The same have been repeated for female speaker.

II. IMPLEMENTATION

A. Implementation of LPC Based Vocal Tract Shape Estimation for Vowels

The Auto regression method for speech analysis (Durbin's Recursive Algorithm) based on linear prediction has been used. This method is identified as LP Modeling and referred as AR Modeling. The model depends only on the previous outputs of the system. The simplest model of a vocal tract consists of co-axial many linked cylindrical tubes producing an all pole

transfer function. Vocal tract shape is estimated from reflection co-efficients obtained from LPC analysis of speech signal, using Wakita's speech analysis model and Durbin's algorithm for optimum inverse filtering. The vocal tract length is assumed to be 17 cm long from the glottis to the lips. Vocal tract area values obtained for the natural vowels for male and female speakers with voluntary participation. Participants chosen spoke standard South Indian vowels without distinct accents, special speech habits, were aged between 18 to 21 years did not suffer from any speech hearing disorder. Each speaker was asked to record the required speech as naturally as possible, and their speech recorded individually, in a speech laboratory, with a portable digital recorder via a small collar microphone, the distance between the Microphone and mouth of speakers was approximately 10 cm, and samples acquired with a sampling frequencies of 11.025 K Hz per second, in 30 ms blocks, 12th order LPC is used. From the speech production model, it is known that the speech undergoes a spectral tilt of -6dB/octave . To counter this and to boost the higher frequencies and flatten the spectrum a pre emphasis filter is used. Pre-emphasis followed by a 6 dB per octave rate. This pre-emphasized speech signal bereft of the ill effects of glottal pulse flattening and lip radiation was hamming windowed and applied as the current block.

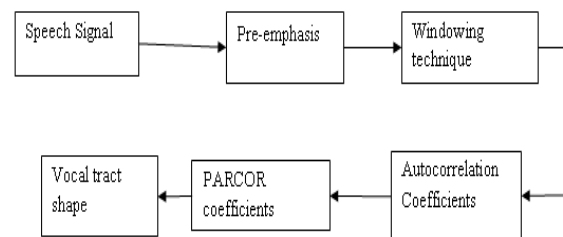


Fig 2: Block diagram for vocal tract shape calculation

B. Intra Speaker Vocal Tract Shape Algorithms

Figure 3 and 4 shows the Block Diagram for Intra and Inter Speaker Vocal Tract Shape Variability for vowels of male speaker. 30 samples of 30 subjects at different times for the vowels are recorded and Vocal Tract Shape arrived for each subject, for 30 sets of subjects' data at different times for predefined set of phonemes. Using LPC and correlation analysis, the vocal tract shape variability of the subject

was found. Study of variability of the above vocal tract shape among 30 different subjects is highlighted to get *Intra Speaker Variability*. The above identified variability can be used as a clue for personal identification and voice print signature, as well as Vocal Tract Signature of an individual. We found the Time averages of the worst and the best patterns of 30 subjects and plotted the resultant worst pattern and resultant best pattern for a subject for south Indian vowels.

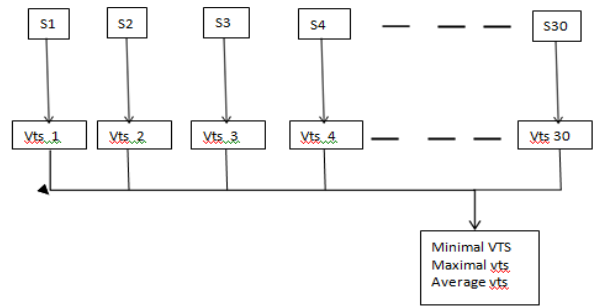


Fig 3: Block Diagram for Intra Speaker Vocal Tract Shape Variability. Here S1-S30: One Subject (Speaker) Samples, VTS ; Vocal Tract Shape

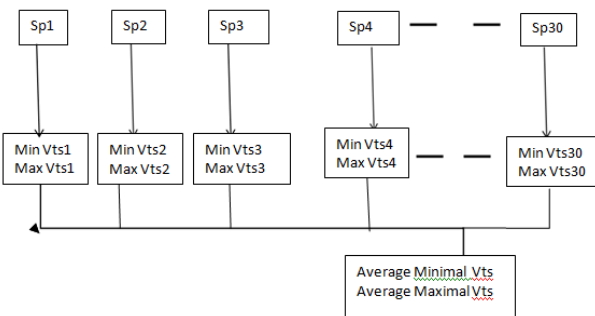


Fig 4: Block Diagram for Time Varying Minimal and Maximal Vocal tract shape Variability for different speakers. Here S1-S30 : Speakers (Subject) vts ; Vocal Tract Shape

The data base of 30 sets of samples of different speakers vocal tract shapes is tabulated and their minimal and maximal Vocal Tract Shapes for English vowels and south Indian vowels are plotted for male speakers and the graphs for English are shown in the figures 5&6.

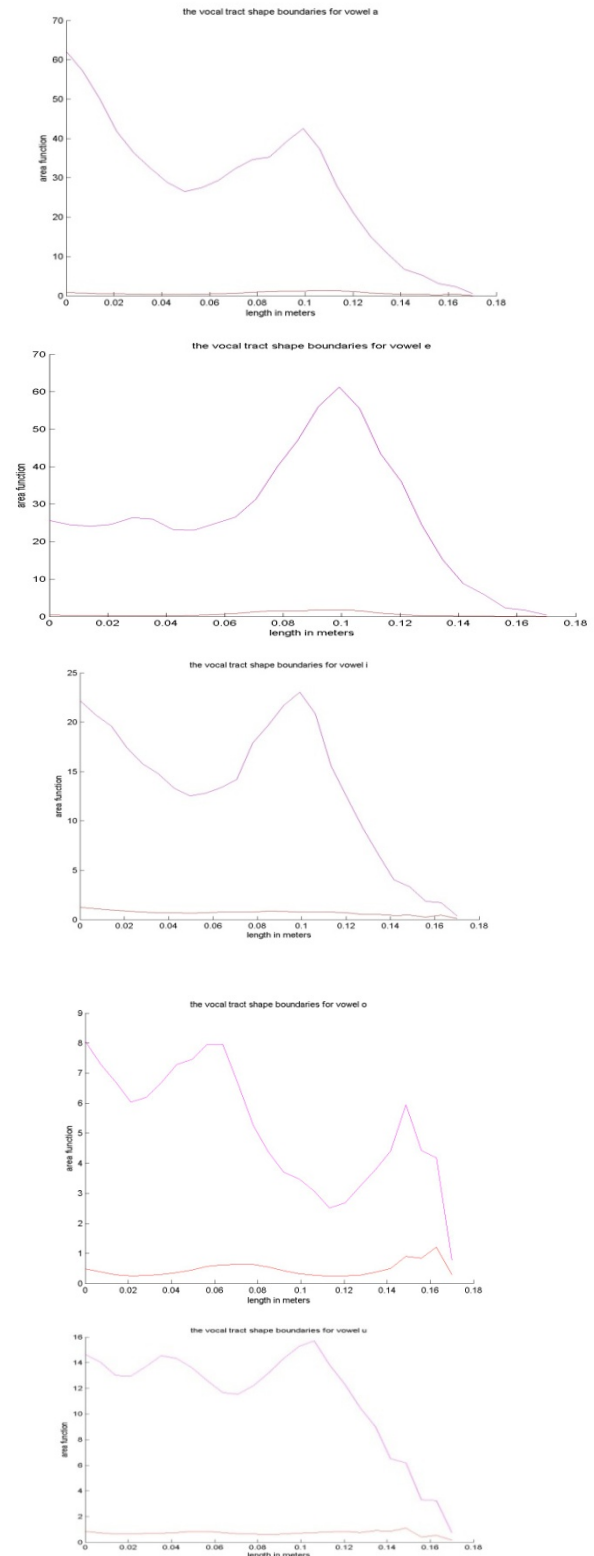


Fig 5: The male model Minimal and Maximal Vocal tract shape with its bounds for the English vowels

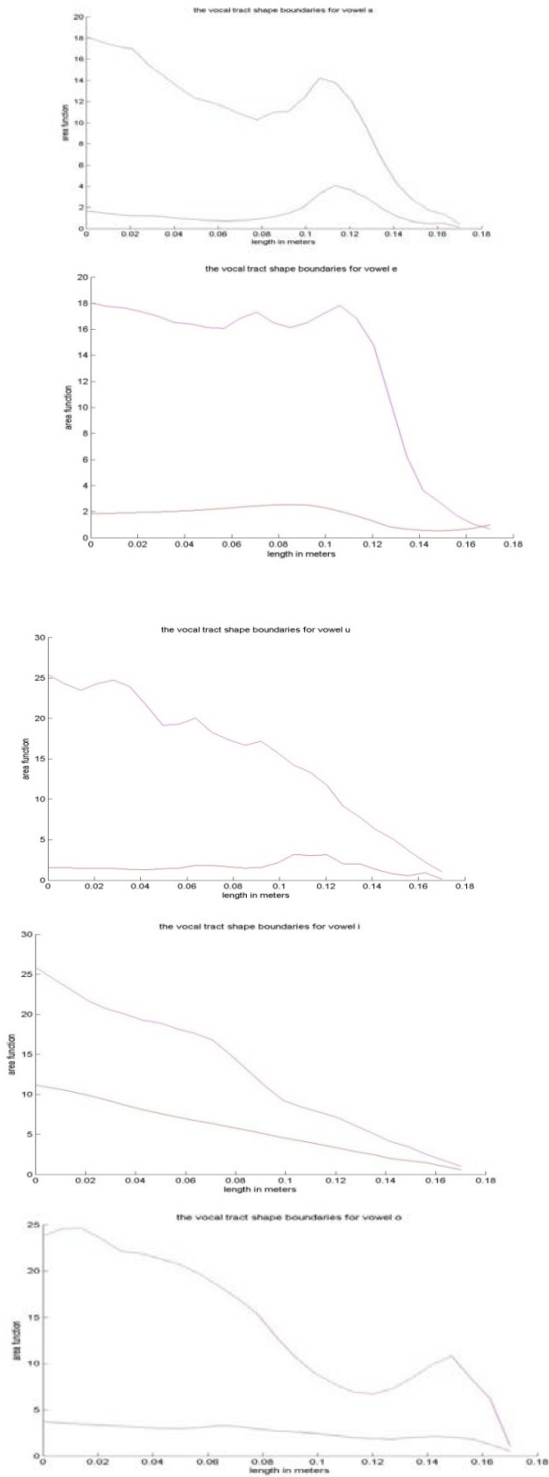


Fig. 6: The female model Minimal and Maximal Vocal tract shape with its bounds for English vowels

The bounds for average maximal and average minimal for specific male and female the vocal tract estimation for phonemes are shown in fig 7 & 8:

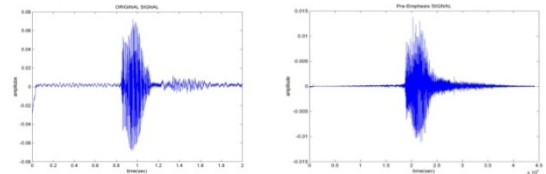
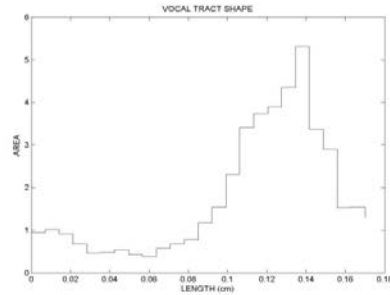
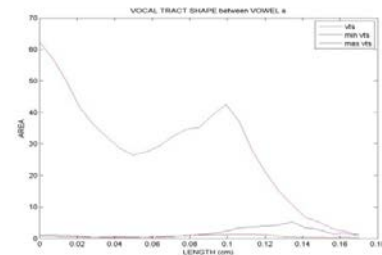


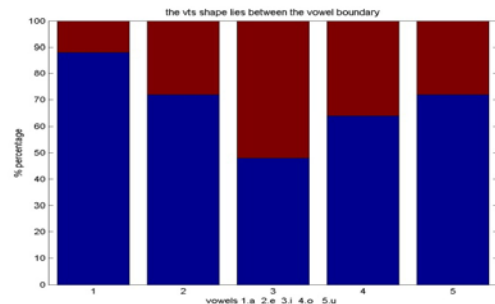
Fig. 7: (a) The speech waveform for vowel /a/ (b) The pre-emphasis signal for vowel /a/



(c) The vocal tract shape (area function) for the vowel /a/



(d) The bounds for average maximal and average minimal for vowel 'a'.



(e) The percentage of matching of vowel /a/ in various phonemes.

@BLUE represents the vowels lies inside the boundary.

@Brown represents the vowels lies outside the boundary.

1. Vowel /a/ in /a/
2. Vowel /a/ in /e/
3. Vowel /a/ in /i/
4. Vowel /a/ in /o/
5. Vowel /a/ in /u/

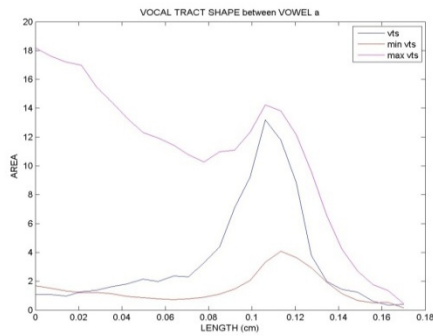


Fig. 8: (a) For girls: The bounds for average maximal and average minimal for vowel ‘a’.

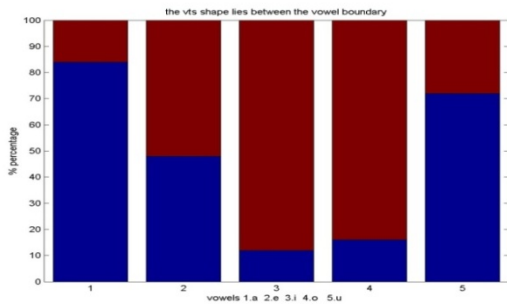


Fig 9: The percentage of matching of vowel /a/ in various phonemes

@BLUE represents the vowels lies inside the boundary.

@Brown represents the vowels lies outside the boundaries.

1. Vowel /a/ in /a/
2. Vowel /a/ in /e/
3. Vowel /a/ in /i/
4. Vowel /a/ in /o/
5. Vowel /a/ in /u/

III. CORRELATION ANALYSIS

The below table shows the vocal tract shape percentage placement of a vowel with itself versus the discrimination provided against other phoneme groups of male.

	A	E	I	O	U
A	96	92	72	76	84
E	92	96	64	64	84

I	72	72	76	64	72
O	76	72	72	96	84
U	88	88	76	60	92

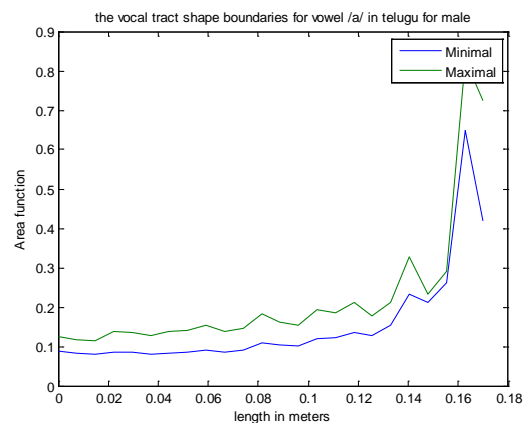
Table 1: The percentage placing in various English vowels for male

The below table shows the vocal tract shape percentage placement of a vowel with itself versus the discrimination provided against other phoneme groups of female.

	A	E	I	O	U
A	84	48	12	16	72
E	20	28	4	20	20
I	20	22	28	18	23
O	70	72	17	76	72
U	70	72	16	70	78

Table 2: The percentage placing in various English vowels for female

Similarly vocal tract shape and its boundaries have been found for the South Indian languages. The following graphs represent the Vocal Tract shapes for Telugu Language.



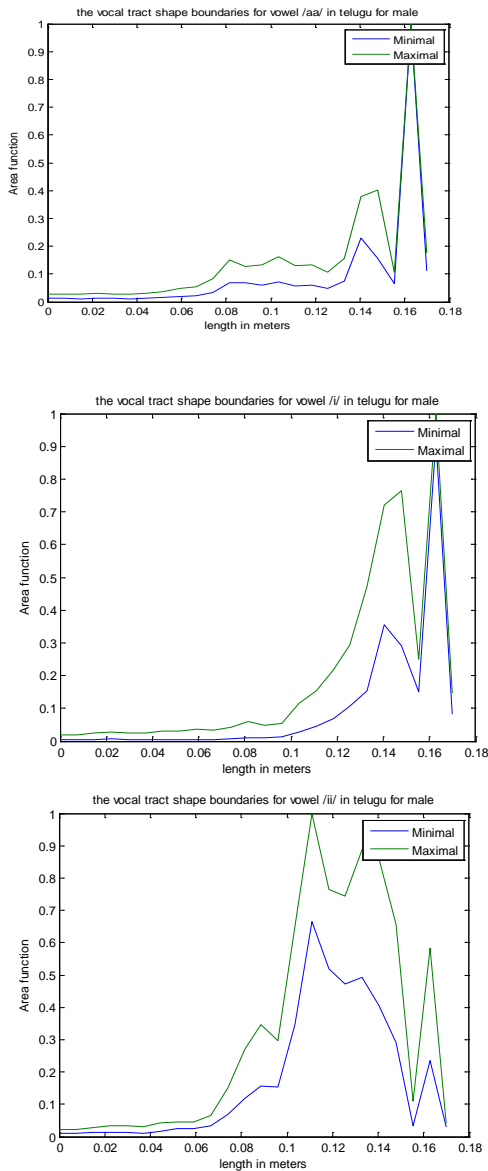
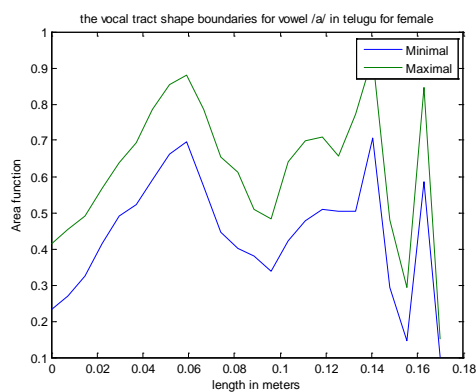


Fig. 10 The male model Minimal and Maximal Vocal tract shape with its bounds for vcv Telugu vowels (intra)

Similarly we can draw the female model Minimal and Maximal Vocal tract shape for the remaining vowels with its bounds for vcv



Telugu vowels (intra) as shown in fig 11.

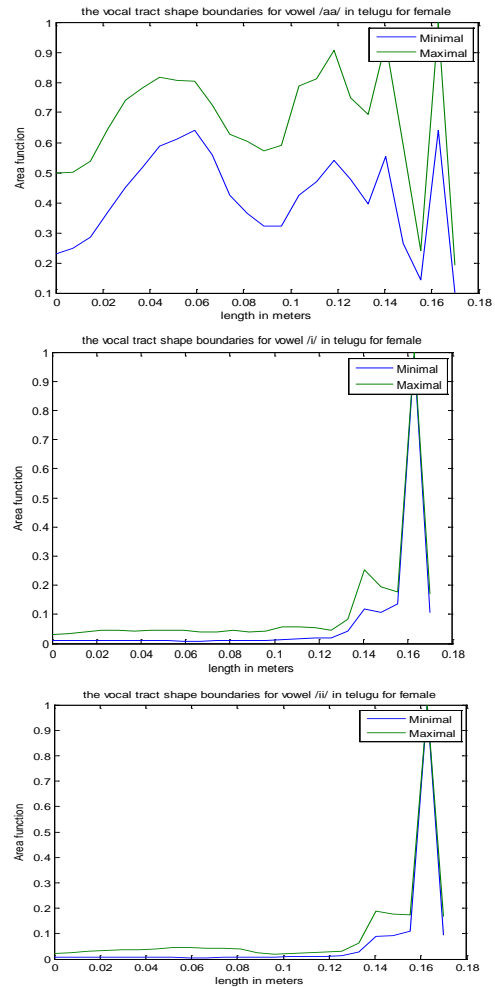
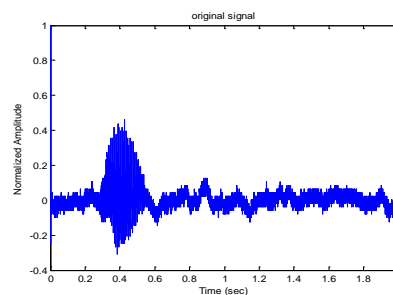


Fig. 11: The female model Minimal and Maximal Vocal tract shape with its bounds for vcv Telugu vowels (intra)

Similarly we can draw the male model Minimal and Maximal Vocal tract shape for the remaining vowels also with its bounds for vcv Telugu vowels.

The bounds for average maximal and average minimal for a specific male speaker the vocal tract estimation for phonemes shown in the below:



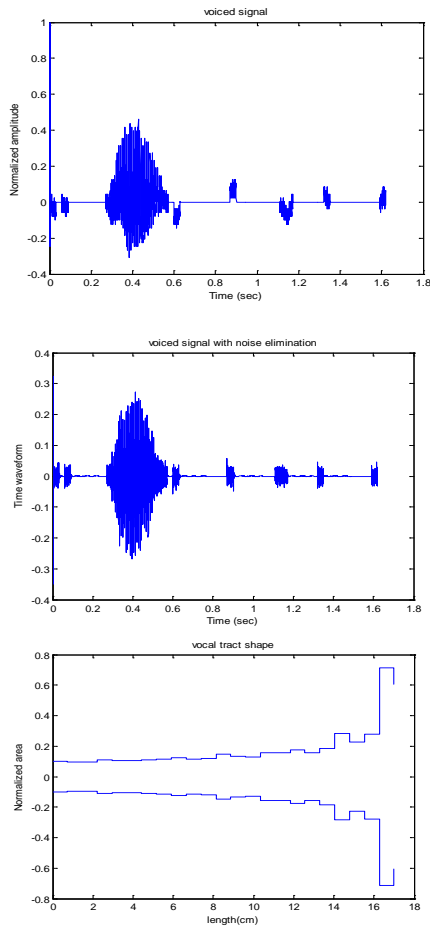


Fig 12: The speech waveform for Telugu vowel / అ /.

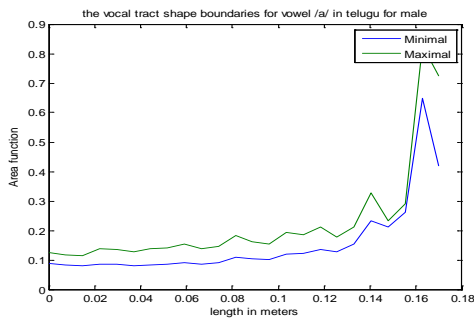


Fig 13: Boundaries for minimal Maximal and average minimal for the vowel / అ /

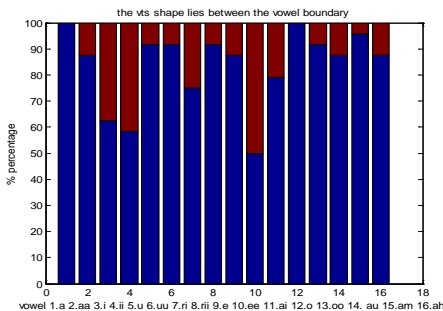


Fig 14: The percentage of matching of Telugu vowel / అ /

@BLUE represents the vowels lies inside the boundary.
 @Brown represents the vowels lies outside the boundary.

1. Vowel / అ / in / అ /
2. Vowel / అ / in / ఆ /
3. Vowel / అ / in / ఇ /
4. Vowel / అ / in / ఈ /
5. Vowel / అ / in / ఊ /
6. Vowel / అ / in / ఊః /
7. Vowel / అ / in / ఋ /
8. Vowel / అ / in / ఋః /
9. Vowel / అ / in / ఎ /
10. Vowel / అ / in / ఏ /
11. Vowel / అ / in / ఐ /
12. Vowel / అ / in / ఒ /
13. Vowel / అ / in / ఓ /
14. Vowel / అ / in / ఔ /
15. Vowel / అ / in / అః /
16. Vowel / అ / in / అ /

For girls the boundaries for average maximal and average minimal for vowel ‘అ’ is shown in the below figure.

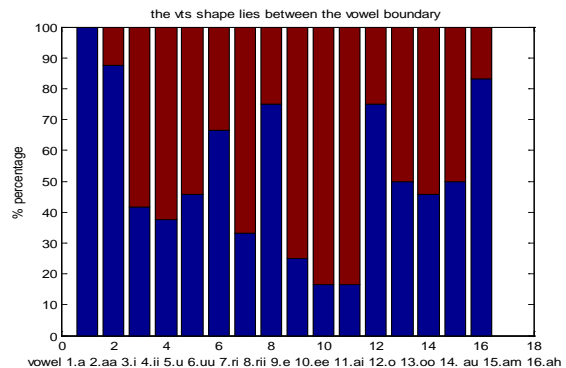


Fig 15: The percentage of matching of Telugu vowel / అ /

@BLUE represents the vowels lies inside the boundary.
 @Brown represents the vowels lies outside the boundary.

1. Vowel / అ / in / అ /
2. Vowel / అ / in / ఆ /
3. Vowel / అ / in / ఇ /
4. Vowel / అ / in / ఈ /
5. Vowel / అ / in / ఊ /
6. Vowel / అ / in / ఊః /
7. Vowel / అ / in / ఋ /

8. Vowel / అ/ in /యూ
9. Vowel / అ/ in /ఎ/
10. Vowel / అ/ in /వ/
11. Vowel / అ/ in /ఐ/
12. Vowel / అ/ in /ఊ/
13. Vowel / అ/ in /ఓ/
14. Vowel / అ/ in /ఔ/
15. Vowel / అ/ in /అం/
16. Vowel / అ/ in /అః/

Here we do the correlation analysis for Telugu vowels then, below table shows the vocal tract shape percentage placement of a vowel with itself versus the discrimination provided against other phoneme groups of male.

The below table shows the vocal tract shape percentage placement of a Telugu vowel with itself versus the discrimination provided against other phoneme groups of male.

	a	ā	i	ī	u	ū	e	ē	ai	au	am	sha				
a	97	87.5	62.5	58.33	91.67	91.67	75	91.67	87.5	50	79.17	86	91.67	87.5	95.83	87.5
ā	91.83	95.83	87.5	83.33	75	79.17	58.33	83.33	91.67	83.33	91.67	87.5	91.67	91.67	58.33	91.67
i	87.5	83.33	100	66.67	75	91.67	83.33	91.67	91.67	50	87.5	87.5	91.67	79.17	87.5	66.67
ī	50	58.33	91.67	95.83	50	70.83	50	75	79.17	79.17	62.5	54.17	70.83	50	54.17	54.17
u	95.83	70.83	87.5	62.5	97.6	94.4	79.17	95.83	91.67	50	95.83	76.2	95.83	95.83	91.67	79.17
ū	75	62.5	70.83	50	92	96.4	87	95.83	83.33	45.83	79.17	83.33	87.5	75	91.67	62.5
e	45.83	33.53	50	54.17	87.5	87.64	91.6	70.83	54.17	45.83	45.83	54.17	58.33	29.17	62.5	20.83
ē	72	59.5	87.83	47	91	94	92	97	80.33	42.83	76.17	80.33	84.5	72	88.67	39.5
ai	84.3	80.13	91	63.47	71.8	88.47	80.13	88.47	93.7	46.8	84.3	84.3	88.47	75.97	84.3	63.47
au	46.5	54.25	72.7	64.3	46.5	65.87	46.5	69.75	73.63	76.1	58.13	50.38	65.87	46.5	50.38	50.38
am	62.29	63.99	77.95	60.84	56.33	73.5	60.3	75.34	79.68	58.52	86.3	64.13	73.5	58.32	64.13	54.21
sha	83.2	74.87	83.2	83.2	87.37	91.53	49.87	91.53	87.37	74.87	87.37	95.6	90	87.37	70.7	69
ā	87.5	79.17	87.5	87.5	91.67	95.83	54.17	95.83	91.67	79.17	91.67	95.83	100	91.67	75	73.3
au	41.67	58.33	66.67	62.5	20.83	20.83	4.17	12.5	41.67	50	16.67	50	74.3	8.33	45.83	
am	82	75	75	58.33	86	95.83	75	91.67	83.33	50	91.67	91.67	87.5	100	75	
sha	93.13	93.13	84.8	80.63	72.3	76.47	55.63	80.63	88.97	80.63	88.97	84.8	88.97	88.97	55.63	97.1

Table 3: The percentage of placing in various Telugu vowels for male

The below table shows the vocal tract shape percentage placement of a Telugu vowel with itself versus the discrimination provided against other phoneme groups of female.

	a	ā	i	ī	u	ū	e	ē	ai	au	am	sha				
a	97.90	87.50	41.67	37.50	45.83	66.67	33.33	75.00	23.00	16.67	16.67	75.00	50.00	45.83	50.00	83.33
ā	88.16	92.00	84.00	80.00	72.00	76.00	56.00	80.00	88.00	80.00	88.00	84.00	88.00	88.00	56.00	88.00
i	77.88	74.16	89.00	59.34	66.75	81.59	74.16	81.59	81.59	44.50	77.88	77.88	81.59	70.46	77.88	59.34
ī	47.50	55.41	87.09	93.04	47.50	67.29	47.50	71.25	75.21	59.38	51.46	67.29	47.50	51.46	51.46	
u	93.91	69.41	85.75	61.25	93.25	92.51	77.59	93.91	89.84	49.00	93.91	74.68	93.91	93.91	89.84	77.59
ū	64.50	53.75	60.91	43.00	79.12	82.90	74.82	82.41	71.66	39.41	68.09	71.66	75.23	64.50	78.84	53.75
e	39.41	28.86	43.00	48.59	75.23	75.37	78.38	60.91	48.59	39.41	39.41	48.59	50.16	23.09	53.75	17.91
ē	69.12	57.12	65.12	43.12	87.36	90.24	68.57	93.13	77.12	43.12	73.12	77.12	81.12	69.12	65.12	57.12
ai	72.50	68.91	78.26	54.58	61.75	76.08	68.91	76.08	80.58	40.25	72.50	72.50	76.08	60.33	72.50	54.58
e	44.64	52.08	69.79	61.73	44.64	63.24	44.64	66.90	70.58	73.06	55.80	48.36	63.24	44.64	48.36	48.36
ē	87.34	48.63	59.24	46.24	42.83	55.86	45.83	57.26	60.56	44.48	60.59	48.74	55.86	44.32	48.74	41.20
ai	65.73	59.15	65.73	65.73	69.02	72.31	39.40	72.31	69.02	59.15	69.02	73.31	71.10	69.02	59.83	34.51
au	69.13	62.54	69.13	69.13	72.42	75.71	42.79	75.71	72.42	62.54	72.42	75.71	79.00	72.42	59.25	57.91
am	32.82	46.08	52.67	49.38	16.46	16.46	3.29	9.88	32.92	39.50	13.37	39.50	39.50	58.70	6.58	36.21
sha	64.78	59.23	59.23	46.08	67.94	75.71	59.23	72.42	65.83	39.50	72.42	72.42	69.13	79.00	59.23	
ā	86.61	86.61	78.86	74.99	67.24	71.12	51.74	74.99	82.74	74.99	82.74	78.86	82.74	82.74	51.74	90.50

Table 4: The percentage of placing in various Telugu vowels for female

IV. CONCLUSION AND FUTURE SCOPE

Vocal tract shape of an unknown male lies within bounds of a given localized, geographical, cultural population. It will have very high correlation with average, normalized behavior of the ensemble. The above correlation will remain statistically true, and time invariant. We established an analytical computational frame work to study and analyses stochastic speech process for the south Indian languages. The analysis of stochastic data has resulted for verification and identification of regional languages (Telugu). The future scope of this concept is to use glides, and nasal phonemes, for vocal tract. These signatures might reduce the uncertainty in recognizing and identification and thus improve the accuracy of identification and recognition by several orders. Utilizing these facts properly and new passwords can be designed for transactional experiment and validation. We believe that the use of regional language with these methodologies will considerably reduce the error in vocal tract signature recognition and identification and further it can be extended to the roots of India (rural areas) for more practical application areas using regional language speech recognition..

V. REFERENCES

- [1] L.R. Rabiner and R.W. Schafer, Digital Processing of Speech Signals, Prentice Hall, Englewood Cliffs, NJ, 1978.
- [2] Thomas F. Quatieri, Discrete-time Speech Signal Processing, Principles and Practice, Pearson Education, 2002.
- [3] Lawrence Rabiner and Biing-Hwang Juang, Fundamentals of Speech Recognition, Pearson Education, 1993.
- [4] J.L. Flanagan, Speech Analysis, Synthesis, and Perception, Second Edition
- [5] G. Fant, Acoustic Theory of Speech Production, Mouton, The Hague, 1970.
- [6] M. M. Sondhi and B. Gopinath, "Determination of Vocal-Tract Shape from Impulse Response at the Lips," J. Acoust. Soc. Am., Vol. 49, No. 6 (Part 2), pp. 1847-1873, June 1971.
- [7] J. D. Markel and A. H. Gray, Jr., Linear Prediction of Speech, Springer-Verlag, New York, 1976.
- [8] G.N.Kodandaramaiah,, R.B. Kulkarni, M.N.Giriprasad, "The Minimal and

- Maximal Vocal tract shape variability for vowels based on LPC” ASECI-2010, Technical Session 1, pp 91-95.
- [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] 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.
- [11] Raxy RR, Rosini RR, Mani Sankar R. A Novel Speech Recognition System using Hidden Markov Model. I
- [12] Sohal B, Sandeep K. A HMM Integrated SVM Model for Hindi Speech Recognition. Indian Journal of Science and Technology. 2016 December; 9(47). Crossref
- [13] 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.
- [14] 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).
- [15] khodai-joopaari M, Clermont F, Barlow M. Speaker variability on a continuum of spectral sub-bands from 297-speakers’ non-contemporaneous cepstra of japans vowels. Proceedings of the 10th Australian international conference on speech science and technology. 2004.
- [16] 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.
- [17] Vallabha G, Tuller B. Choice Of Filter Order In Lpc Analysis Of Vowels. Sound to Sense. 2004 June; at MIT.
- [18] 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
- [19] Raut SV, Panthangi LC, Akhil BG, Syed Faisal Ali, Sanjay HS and 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.