Global Journal of Computational Intelligence Research.
ISSN 2249-0000 Volume 1, Number 2 (2011), pp. 103-110

http://www.ripublication.com/gjcir.htm

English Vowel Recognition using Area Function Approximation by Linear Predictive Analysis

[1]Prabhakar V. Hunagund, [2]M.B. Manjunatha[#], [2]G.N. Kodandaramaiah[*] and [2]S.A.K. Jilani

[1]Dept. of P.G. Studies and Research in Applied Electronics, Jnana Ganga, Gulbarga University, Gulbarga, India
[2]ECE Department, Madanapalle Institute of Technology, Madanapalle, India
E-mail: kodandaramaiah@yahoo.com

Abstract

This Paper presents vocal tract shape for non contextual vowels based on LPC. Phonetic distinctiveness and speaker individuality are deeply ingrained in the vocal tract shapes estimated from the vowels. This is demonstrated by Acoustic Articulator model on vowels and speaker basis using area function approximation to the vocal tract shape. Here we proposed a new technique to approximate an area function of a person at different times, and in different context. This vowel utterance of an adult male are recorded 35 times and variability of the resulting vocal tract shapes are measured on Intra and Inter speaker basis and investigate the variability role on speaker recognition applications by using the Euclidian distance method. The vocal tract shape variability estimation is implemented using MATLAB

Keywords: LPC, Euclidian distance.

Introduction
Vocal tract Analysis
The Hearing Impaired have trouble perceiving speech as they generally use lip reading as a tool for the same. They do not have access to the internal articulators that differentiate twoutterances like /aba/ and /apa/. These utterances seem the same from the lip reading point of view. The process of acoustic - to - articulatory mapping involves displaying the vocal tract (articulator) shape from the speech (acoustic) signal. This is further used to assist in speech training for the hearing impaired.

The vocal tract shape generation methods can be broadly classified into two types:

the direct and the indirect method. The direct methods involve exposure of the human body to different waves. Older techniques like X Ray and newer techniques like Magnetic Resonance Imaging (MRI), Electromagnetic Articulography (EMA), Optopalatography (EPG) have been used for getting the vocal tract shape directly. These methods are not feasible to be used in training though they serve a purpose of validation of results recovered from the indirect methods.

LPC is the most convenient as real time processing is possible as well as the ease of providing the input speech signal.

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 losses, 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 shape. The first tube starts at the glottis and the last tube ends at the lips and/or the nostrils. Most of the acoustic tube models assume the (Schroeter and Sondhi 1994). The vocal tract area of an average adult male is approximately 17.5 cm long and about 4 cm in diameter. An adult female vocal tract is approximately 15 cm long (Dew and Jensen 1977). The vocal tract articulators like tongue, soft palate, hard palate, hypoid bone and lips modify the voice sounds are produce recognizable words. The resonance frequencies of the vocal tract are called formant frequencies or simply formants, and they depend on the shape and dimensions of the vocal tract. Various estimation techniques like measurement of acoustic impedance at the lips, measurement of formant. Frequencies and LPC based analysis have been used for the estimation of vocal tract shapes.

Objective of the Paper

This works models the solo vowels, non-contextually, obtains from the model spectrograms, formants, pith and vocal tract shape information. Error minimization is carried out using an all pole LPC filter. Analysis is done for a vowel 'ae' of males by taking 35 samples of 35 subjects at times. The vocal tract shape arrived at for each subject for 35 sets of data at different instants of times for predefined set of phonemes namely ae, ei, iy, oa, uh. 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 35 different speakers is highlighted to identify Intra Speaker Variability. This identified variability, can be used as a cue for personal Identification in speaker specific recognition. They can also be used as Vocal tract signature of an individual, in forensic/ other applications. The time averages of the worst and the best patterns for the example of 35 subjects, has been plotted, for different phonemes for male speaker. The same have been repeated for female speaker.

Implementation

Implementation of LPC based Vocal Tract Shape Estimation for Vowels

The Auto regression method fro speech analysis (Durbin's Recursive Algorithm) based on linear prediction has been used. This method is identified as LP Modelling and referred as AR Modelling. The model depends only on the previous outputs of the system. The simplest model of the vocal tract consists of co-axial many linked cylindrical tubes producing an all pole transfer function. Vocal tract shape is estimated from reflection co-efficient obtained from LPC Analysis of speech signal, using Wakita's speech analysis model and Durbin's algorithm for optimum inverse filtering. 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 16 K Hz per second, in 30 ms blocks, 25^{th} 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, a pre-emphasis filters, to boost the higher frequencies and flatten the spectrum . Pre-Emphasis speech signal bereft of the ill effects of glottal pulse flattening and lip radiation was hamming windowed and applied as the current block. By using the vocal tract analysis for the hamming window signal we get the vocal tract area coefficients. Finally by using these vocal tract area coefficients we get the vocal tract shape of the speech signal

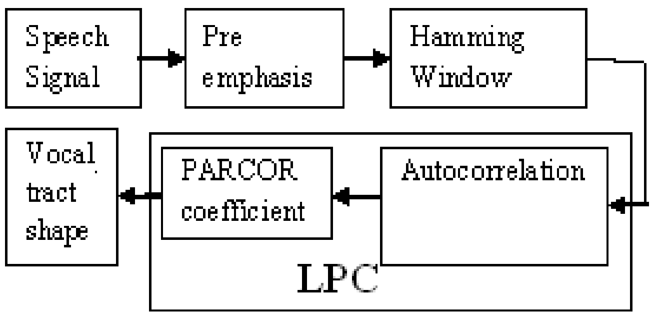

Figure 1: Block diagram for Vocal Tract Shape Calculation.

Intra Speaker Vocal Tract Shape Algorithms

Fig 2 shows the block diagram for Intra Speaker Vocal tract shape variability for vowels of male, female and boys speaker. 35 samples of 35 subjects at different times for the vowels a, e, i, o, u are recorded and vocal tract shape arrived for each subject, for 35 set 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 35 different subjects is highlighted to get Intra speaker variability. The above identified variability can be used as a cue 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 35 subjects and plotted the resultant worst pattern and resultant best pattern for a subject for the phoneme 'ae'. Further we repeated the above steps for the remaining phonemes 'ei', 'iy', 'ou', & 'uh'.

The data base of 35 set of samples of different speakers. The average vocal tract values for the vowels /ae/, /ei/, /iy/, /oa/, /uh/ is tabulated.

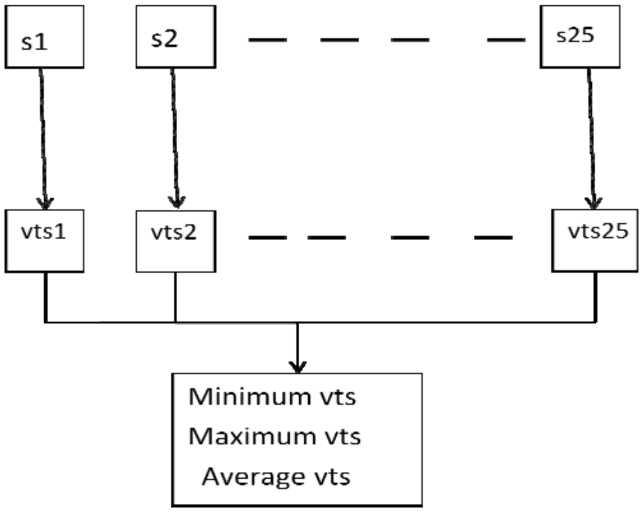

Figure 2: Block diagram of Intra Speaker Vocal Tract Shape Variability: Here S1-S30: One Subject (Speaker) Sample, Vts; Vocal Tract Shape.

Correlation of Trained Vowel /ae / by Euclidian distance method for male speaker

There are varieties of ways for studying phonetics and their distinctive features or characteristics of the phonemes. For our purpose of study, it is sufficient to consider an acoustic characterization of the various sounds, including the place and manner of articulation, wave forms, and role of back reflection co-efficient in modeling the vowels. Vowels are produced by exciting a fixed vocal tract with quasi periodic pulses of air caused by the vibration of the vocal cords. The cross- sectional area varies along the vocal tract that determines the resonant frequencies of the vocal tract, and thus the sound is produced. The dependence of the cross-sectional area upon the distance along the tract is called the area functions of the vocal tract.

Results

Table 4.12: Shows the vocal tract correlation graph of a vowel superimposed on itself verses the discrimination provided against other vowels pronounced by Male speakers.

Vowels	*/ae/*	*/ei/*	*/iy/*	*/oa/*	*/uh/*
/ae/	***96***	*72*	*84*	*82*	*78*
/ei/	*82*	***97***	*80*	*70*	*62*
/iy/	*79*	*76*	***96***	*82*	*75*
/oa/	*75*	*66*	*79*	***87***	*66*
/uh/	*29*	*75*	*52*	*25*	***96***

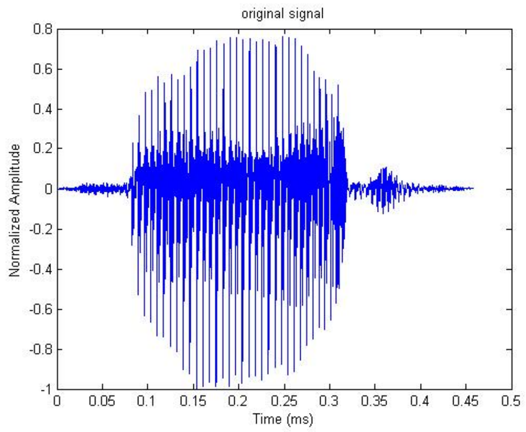

Figure 1: The speech signal for vowel /ae/ Male test sample.

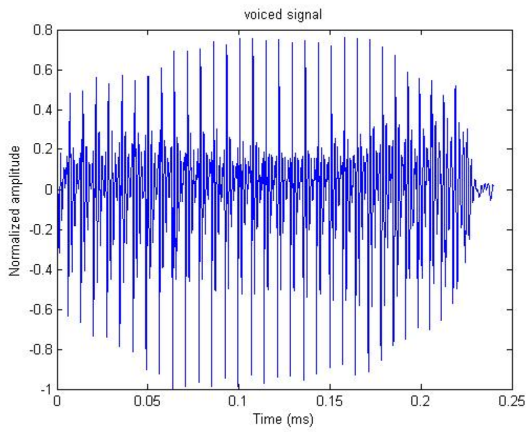

Figure 2: The Voiced signal for vowel /ae/ Male test sample.

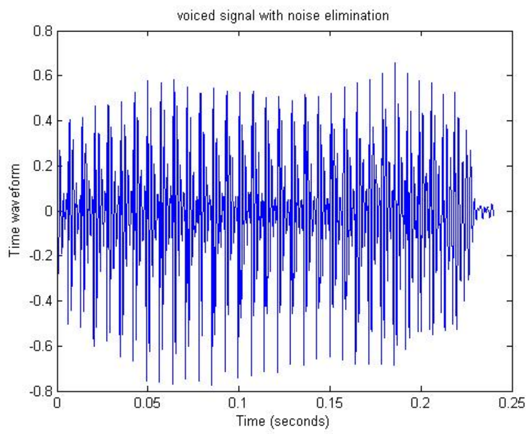

Figure 3: The Normalized speech signal with noise elimination for vowel /ae/ Male test sample

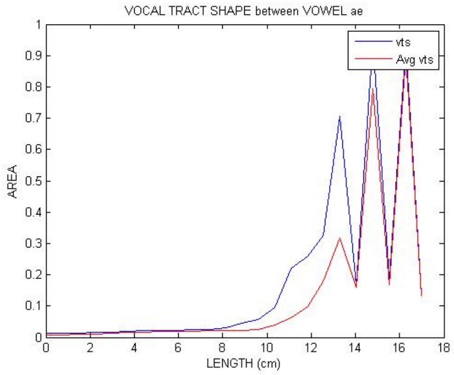

Figure 4: The approximate vocal tract shape for Vowel /ae/. For single Male speaker test sample. Where V_{ts}= test sample, Avg V_{ts} = Average Vocal tract shape of vowel /ae/

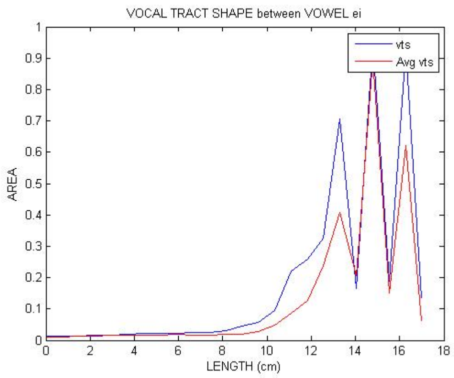

Figure 5: The approximate vocal tract shape for Vowel /ae/. For single Male speaker test sample. Where V_{ts}= test sample, Avg V_{ts} =Average Vocal tract shape of vowel /ei/

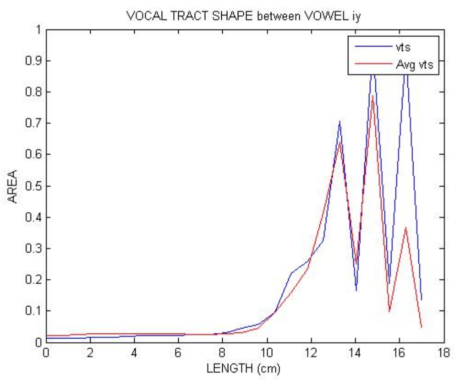

Figure 6: The approximate vocal tract shape for Vowel /ae/. For single Male speaker test sample. Where V_{ts}= test sample, Avg V_{ts} = Average Vocal tract shape of vowel /iy/

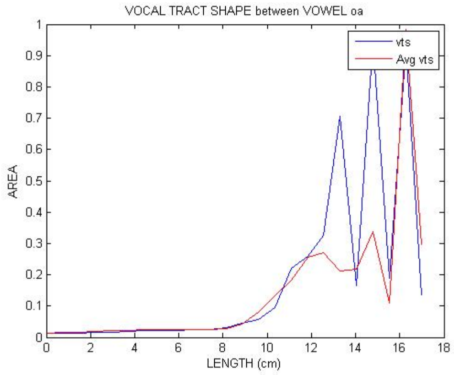

Figure 7: The approximate vocal tract shape for Vowel /ae/. For single Male speaker test sample. Where V_{ts}= test sample, Avg V_{ts} = Average Vocal tract shape of vowel /oa/

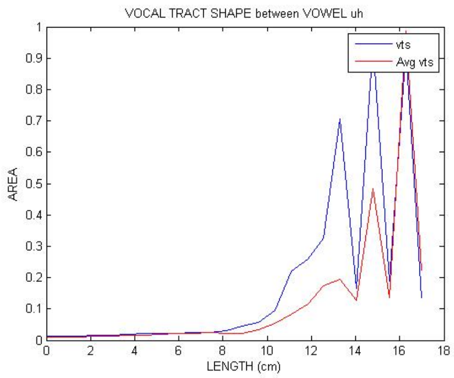

Figure 8: The approximate vocal tract shape for Vowel /ae/. For single Male speaker test sample. Where V_{ts}= test sample, Avg V_{ts} = Average Vocal tract shape of vowel /uh/

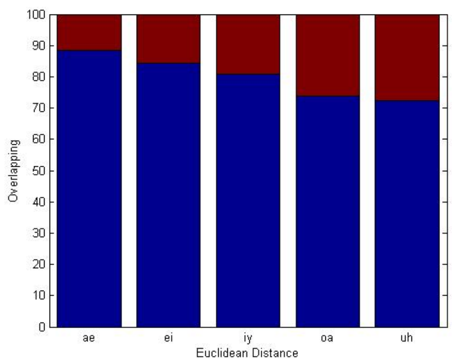

Figure 9: percentage of matching of vowel /ae/ in various vowels for single female speaker test sample where BLUE represents the vowel that lies inside the boundary ,brown represents the vowel that lies outside the boundary, 1 Vowel /ae/ in /ae/, 2 Vowel /ae/ in /ei/,3 Vowel /ae/ in /iy/,4 Vowel /ae/ in /oa/,5 Vowel /ae/ in /uh/

Conclusions

It was an attempt presents to the standard method for vocal tract shape estimation has been the basis for many successful automatic speech recognition (ASR) systems. Here we describe a “standard” approach for classification of vowels based on formants. We achieved 80 to 95 percentage of speaker recognition using Euclidean distance measure.

Acknowledgment

We would like to thanks the Management, Principal of Madanapalli Institute of Technology and Science, Madanapalli, A.P., for their Cooperation and Encouragement

References

[1] R.Rabiner and R.W.Schafer, Digital processing of Speech signals, Droling Kindersly(india)pvt.Ltd.,licensees of pearson eduction in south asia, 1978, PP. 54-101,412-460.
[2] Thomas F. Quatieri, Discrete time speech signal processing principles and practice,2002, pp 56-59.
[3] R. S. Nikerson, “Characteristics of the speech of deaf persons,” *Volta Rev.*, vol. 77, 1975, pp. 342–362; reprinted in: H. Levitt, J. M. Pickett, and R. A. Houde *Sensory Aids for the Hearing Impaired*, (Eds.). New York: IEEE Press, 1980, pp. 540–545.
[4] H.Kinngara, H.Kamata & Y.Ishida. Spoken word Recognition using Vocal Tract shapes,1993,pp.133-136.
[5] P.Rose,Long-and short-term within-speaker differences in the formants of Australian hello, j.Int. Phonetic. Assoc. 29(1) (1999) 1-31.
[6] Lawrence Rabiner and Biing-Hwang Juang, Fundamentals of Speech Recognition, Pearson Education, 1993.
[7] J. L. Flanagan, Speech Analysiss, Sythesis, and Perception, Second Edition.
[8] G Fant, Acoustic Theory of speech Production, Mouton, The Hague, 1970.
[9] M. M. Sondhi and B.Gopinath,”Determination of vocal tract shape from Impulse Response at the Lips,” J.Acoust. Soc. Am.,49(6) (part 2), pp. 1847-1873, june 1971.
[10] J. D. Markel and A. H. Gray, Jr., Linear Prediction of speech, Springer-Verlag, New York, 1976.