International Journal of Applied Information Systems (IJAIS) – ISSN : 2249-0868
Foundation of Computer Science FCS, New York, USA
International Conference on Communication Computing and Virtualization 2016

Methods Used For Vocal Tract Shape Estimation and There Applicability for Children

Veena S
Departmentof Electronics
and Telecommunication
Fr. C. Rodrigues
Institute of Technology
Navi Mumbai 400703,
Maharashtra, India
FCRIT, Vashi

Nilashree S Wankhede
Department of Electronics
and Telecommunication
Fr. C. Rodrigues
Institute of Technology
Navi Mumbai 400703,
Maharashtra, India
FCRIT, Vashi

Milind S Shah
Department ofElectronics
and Telecommunication
Fr. C. Rodrigues
Institute of Technology
Navi Mumbai 400703,
Maharashtra, India
FCRIT, Vashi

ABSTRACT

Vocal tract shape estimation is essential for development of speech training aids for hearing impaired. Speech therapists have concluded from their research that if speech disorders are detected at an early stage then the rectification of those becomes faster and to a great accuracy. In case of adults, the vocal tract shape estimation can be performed by using techniques like X-Ray, MRI or using articulatory analysis by synthesis.Due to certain difficulties present in the methods mentioned like radiation effect, loud sound impulses produced by rapid switching of magnetic fields during MR scanning is not considered suitable for children. Thus there is lack of availability of vocal tract shape data for children due to inability to use direct estimation methods for children. So it becomes necessary to carry out comparative study of vocal tract shape estimation techniques of children using indirect vocal tract shape estimation techniques. This paper aims to make available a considerable amount of data to researchers for study towards children’s vocal tract shape using implemented techniques.

The final objective of research work is to perform a detailed study on techniques available for vocal tract shape estimation, to implement and further compare the covariance and lattice methods for children using MATLAB software. Currently the VTSE based autocorrelation method and lattice method is implemented and the results obtained are presented.

General Terms

Vocal Tract Shape estimation, Autocorrelation method, Lattice method, Covariance method, Acoustic Pulse Reflectometry.

Keywords

VTSE; LPC; APR; speech training.

1. INTRODUCTION

The most natural form of communication used by humans is verbal communication. Lungs act as pressure source for speech by providing airflow. Vocal folds modulate airflow to create sound. The most important component of speech production is the vocal tract. The major parts of vocal tract are the pharynx, the oral cavity, the nasal cavity, the velumand the lips. Some major functions of vocal tract are to modify the spectral distribution of energy in glottal sound and contribution to generation of sound [1].Thus articulation is a result of the vocal tract that modifies their position and shape during air expulsion, producing different sounds and acoustic representations. Vocal tract length is defined as the curvilinear distance along the midline of the tract starting at the glottis to the intersection with a line drawn tangentially to the lips. It has been found to vary from about 6-8cm in infants to 15-18cm in adult females and male respectively [2]. Researchers have used two different methods to determine the shape of the vocal tract: direct methods based on geometrical measurements of the vocal tract; and indirect methods based on acoustic inversion [3]. The direct measurement methods are several imaging methods like X-ray Radiography, Computed Tomography or Magnetic Resonance Imaging. It is difficult to measure soft tissue structures by using X-Rays. Repeated use can cause harm to subjects. For a large scale study on speech production, collecting data using these techniques is not suitable. Indirect methods on the other hand determine the vocal tract shape from a speech signal. Indirect methods studied in the paper are Linear Predictive Coding (LPC), analysis by synthesis, acoustic pulse reflectometry, application of dynamic constraints and measurement of lip area.Several other methods like Mel Frequency cepstum coefficients (MFCC), Particle Swarm optimization algorithm (PSO) are also used. This paper mainly focuses on indirect approach of vocal tract shape estimation (VTSE) techniques like LPC method. Also importance will be given to children’s speech and their vocal tract shape estimation as speech is learnt usually in early childhood.

As the research work is focused on VTSE of children it is very essential to have insight on children’s vocal tract. Children have shorter vocal tract and shorter vocal folds. This makes their fundamental and formant frequencies higher. Speech development in children is predicated partly by the growth and anatomic restructuring of the vocal tract. One observation presented includes that the children use different articulatory position than adults for producing speech sounds [5].

The final objective of research work is to perform a detailed study on techniques available for vocal tract shape estimation, to implement and further compare the methods suitable for children using MATLAB software.

2. METHODS USED FOR VOCAL TRACT SHAPE ESTIMATION

Schroder [6] proposed an attractive approach of estimation of vocal tract shape from speech signal as it offers new perspectives for speech processing. Two types of researchers are interested in this kind of estimation. One who likes to determine properties of the tract for development of speech training aids while the others are those interested in synthesizing best quality speech for speech coding and automatic speech recognition. Other possible applications are for language acquisition and second language learning. In the field of phonetics, it enables to know how sounds were articulated without requiring medical imaging or other measurement techniques. Various direct and indirect methods used for vocal tract shape estimation (VTSE), are presented in this section.

X- Ray is a classical technique first used by phoneticians to study speech production mechanism soon after the discovery of X-rays in 1895.Due to potential side effects of radiation exposure inflicted on subjects this technology cannot be considered safe especially for children [7].Another imaging technique used is Computed Tomography. This imaging technique allows making the distinction between bones, soft tissues and air, but does not allow for discriminating different soft tissues. It requires significant amount of ionizing radiation doses, and for this reason, is not at all safe for children. Among the imaging techniques used to study the vocal tract's shape, MRI has been the most commonly accepted. Its advantages include the quality and resolution of soft-tissues and the use of nonionizing radiation. In addition, it allows for morphologic measurements in static as well as dynamic studies. So this technique seems to be apt for obtaining vocal tract shapes and is suitable for adults. However, collecting vocal tract information from children during speech production is difficult to obtain [8].This is largely because the data collection environment in particular the MR scanner can be intimidating for children. Now it becomes essential to move on to indirect method of estimating vocal tract shape.

2.1 Analysis by Synthesis

McGowan et al.[8] made use of an articulatory synthesizer and a vocal tract model to infer vocal tract shape of adults as well as children. Certain adjustments of dynamic parameters of vocal tract model were carried by using stochastic optimization algorithm. The algorithm matches the formant frequency obtained from the articulatory synthesizer with the format frequency of data vowel.

2.2 Measurement of lip area and formant frequency

Bunton et.al [9] proposed a technique to measure vocal tract shape estimation of children on the basis of Measuring lip area. An analysis of video frames was carried out in Matlab for measuring lip area by determining the number of pixels enclosed in the frames. Later on, formant frequencies were extracted. Next step was to generate an initial area function by imposing a constriction of length *Le* and cross-sectional area *Ae* on a uniform tube of area *At*. The area function for a given vowel is estimated by perturbation algorithm. This is described in detail in the paper. Before applying this method on a child speaker it was tested on an adult speaker so as to test the accuracy of method.MRI images were used for the purpose of comparing obtained VTSE which were observed to be much similar to derived shape. During research it was found that the cross sectional area of oral cavity of child speaker was matching with adult. So a further study on cross-sectional area of oral cavity was suggested by author using an independent imaging technique for much accurate estimation of lip area of speaker.

2.3 Application of dynamic constraints

Kuc et al.[10]have made use of Memelstein's vocal tract model which is specified by using six articulatory parameters.Generation of synthetic speech was carried with the help of model.The author concluded that the parameter set with minimum movement of articulators was able to produce suitable vocal tract shape. The constraint used is presented in eq.1

$$D = \min_{(i,j)} \left\{ \sum_{i=1}^{Ns} \sum_{j=1}^{Np} \alpha_j \left[P_{i,j} - \bar{P}_j \right]^2 \right\} \tag{1}$$

N_s: number of segments.

N_p : Number of parameters

α_j : Weights determined by inertia of j^{th}articulatory parameter.

$P_{i,j}$: j^{th} articulatory parameter at time i.

$$\bar{P}j = \frac{1}{N_s} \sum_{i=1}^{Ns} P_{i,j} \tag{2}$$

2.4 Acoustic Pulse Reflectometry(APR)

Calum's [4] work focused on capturing vocal tract parameters during articulation of vowel sounds using acoustic pulse reflectometer. This method is considered very safe as it only uses audio signals for measurement purpose. Subjects place the wavetube in their mouth and are supposed to hold that position for few seconds till the sonic pulses are sent down the vocal tract and measurement takes place. Holding breath can be very tiring for children and hence cannot be used for estimating vocal tract shape.

2.5 Method suitable for estimating VTSE of children

2.5.1 LPC method (Autocorrelation) for vocal tract shape estimation

The method makes use of recorded voice samples to extract the LPC coefficients and generates the vocal tract shape from obtained autocorrelation coefficients which are converted to reflection coefficients. The speech analysis model used by Wakita[11] is shown in Fig. 1.The filtering process of inverse filter model was analyzed. The basic aim of this analysis was to minimize the error between the output of inverse filter and input impulse train. He also regarded the vocal tract as an acoustic tube with varying cross sectional area and divided the tube into M sections of equal length Δl.Fig.1 shows a non-uniform acoustic tube model of vocal tract.

m :section number from lips to glottis.

$u_m^+(t,d)$:Component of volume velocity $(u_m(t,d)$ due to sound wave travelling from glottis to lips.

$u_m^-(t,x)$:Component of volume velocity $(u_m(t,x)$ due to sound wave travelling from lips to glottis.

t: time variable

d: distance variable

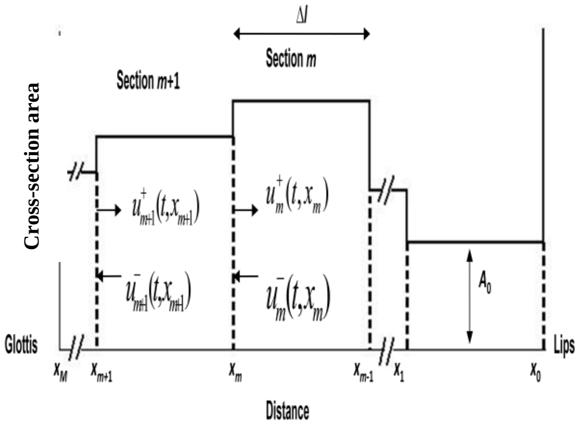


Fig1: A non-uniform acoustic tube model of vocal tract [17]

Wankhede et al [12] have presented a method of estimating vocal tract shape of children by utilizing autocorrelation method based on LPC analysis and then investigated optimum parameters. The three optimum parameters reported are the sampling frequency, vocal tract length and LPC order.

The algorithm used is presented in Fig. 2.Intially the recording of the speech signal was done in Praat software with a suitable frequency. Resampling of speech signal was done according to the table for particular age group. The speech signal was divided in to frames with every frame having duration of 20 ms and 50% overlapping between adjacent frames. The function of windowing is to smooth the estimated power spectrum and to avoid abrupt transitions in the frequency response of adjacent frames. In this study Hamming window of 240 samples equal to the frame length is used. So, windowing is applied and every frame is multiplied with a window function w(n) of length N.

A downward spectral tilt of 12 dB/ octave is introduced by glottal source and radiation at lips results in 6 dB/octave upward tilt. Its aim is to compensate effect of spectral tilt present in speech signal. The value of LPC order would be varied with the age group of child to acquire a proper vocal tract shape. For determining autocorrelation coefficients Levinson-Durbin algorithm was used [13].

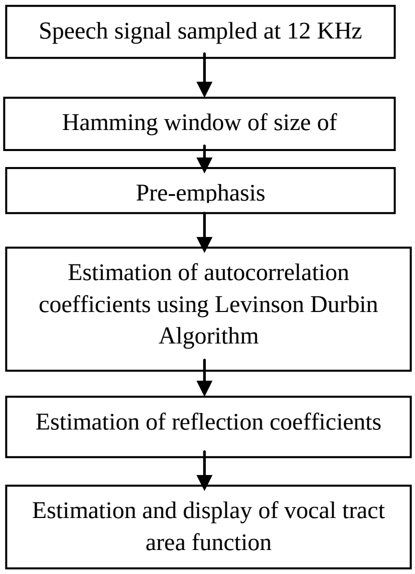

Speech signal sampled at 12 KHz
Hamming window of size of
Pre-emphasis
Estimation of autocorrelation coefficients using Levinson Durbin Algorithm
Estimation of reflection coefficients
Estimation and display of vocal tract area function

Fig.2. Block diagram for VTSE based on autocorrelation method [6]

The autocorrelation coefficients are converted to reflection coefficients so as to find the area values. The area value at glottis is kept to be 1 and the area values of other sections have to be determined from glottis to lips from the following equation

$$A_m = \frac{1+\mu_m}{1-\mu_m} A_{m+1} \tag{3}$$

A_m: area of mth section of vocal tract

μ_m: Reflection coefficient between m and m+1

A_{m+1}: area of (m+1) th section of vocal tract

The table 1 presents the optimum value of three parameters including sampling frequency, vocal tract length and LPC order to be used for obtaining proper VTSE. These values changes within different age groups.

Table 1. Optimum values to be used for VTSE of children of each age group [12]

Age group in years	Average vocal tract length	LPC order (M)	Sampling frequency(F_s)
2 to 5	10	8	14
6 to 9	12	8 or 9	14
10 to 12	13	9 or 10	13.5
13 to 16	14.7	10 or 11	13
17 to 21	16	11 or 12	12
Above 21(adult)	17	12	11

3. METHOD TO BE USED FOR ESTIMATING VOCAL TRACT SHAPE OF CHILDREN

The overall objective of this research work is to implement lattice as well as covariance method and carry out comparative study among the methods Autocorrelation method of linear prediction assumes that the signal is defined for all time such that it is identically zero outside a portion of the signal *N* samples long, where *N* is some positive integer [5]. The speech signal is weighted by a finite window of length *N*. This windowing causes unwanted spectral distortion which is more for smaller values of *N*. Thus if the signal can be defined for all time without windowing, such spectral distortion will not occur. The advantage of the covariance method over the autocorrelation method is that no windowing of data is required in the formation of the autocorrelation estimates. As a result, for short data records the covariance method generally produces higher resolution spectrum estimates than the autocorrelation method [14].

3.1 Lattice method

In case of autocorrelation method described previously we have to determine the correlation from the observed speech waveform and then perform the matrix operation in order to determine the LPC coefficients. Now in case of lattice formulation these two approaches would be treated in an integrated manner [1]. That means to there is no explicit computation of the correlation but rather the LPC coefficients will be determined in terms of the prediction errors. The lattice method does not apply a window to data. The estimates of the auto regressive parameters are more accurate than those obtained with the autocorrelation method [14].By making use of this method the reflection coefficients can be obtained directly from speech signal. An implementation of Burg's method was carried in Matlab software to obtain the reflection coefficients. These values were used to obtain vocal tract shape [17].

3.2 Covariance Method

Windowing of speech segment is not carried as signal values are taken in the interval $-p \le n \le N-1$, such that p samples before the interval are taken to predict the samples at the beginning of the interval [16]. The resulting matrix is minimized using cholesky decomposition method [14] [17].

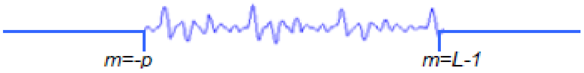


Fig 3: Speech segment in interval $-p \le n \le N-1$ [16]

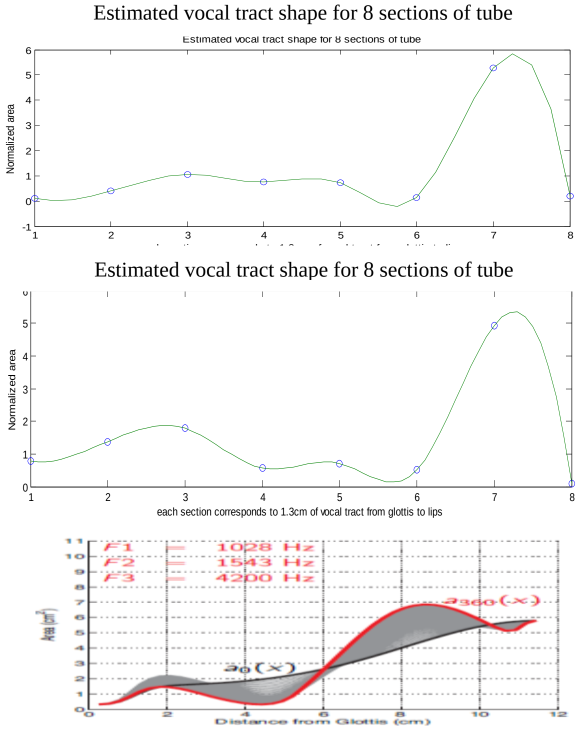
4. RESULTS OBTAINED USING AUTOCORRELATION AND COVARIANCE METHODS

VTSE using autocorrelation method and lattice method was carried. Recording of vowels /a/, /e/, /i/, /o/ and /u/ were carried out for children of age group starting from 3 to 17 years. The result of girl child aged 6 year uttering vowel /i/ and /a/ are shown in Fig 4 and Fig.5 respectively.

Figure 4 shows the results based on autocorrelation for vowel /a/ of female child speaker aged 6 years. Part (a) of Fig.4 shows vocal tract area function using autocorrelation method. Part (b) shows vocal tract area function using lattice method (c) shows the vocal tract area function obtained by Bunton et al.[9] Since the speaker is of 6 years the optimum parameters for the child according to Table 1 are

1. Average vocal tract length (l) =12cm
2. LPC order M = 8
3. Sampling frequency (Fs) =14 KHz.

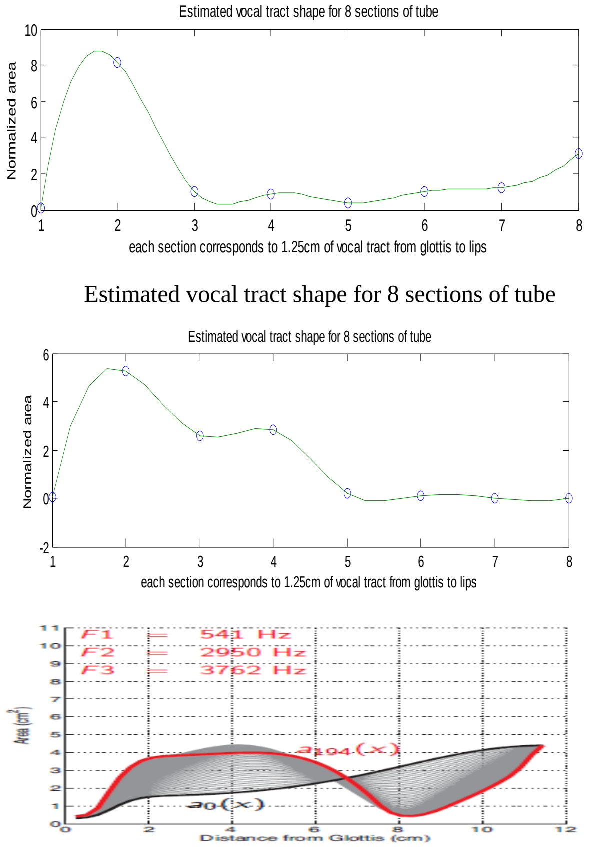
In fig.4 x-axis function of vowel /a/ obtained after 360 iterations. The obtained results in fig 4 (a) and (b) using autocorrelation represents distance from glottis to lips and y-axis represents the normalized cross-sectional area. Fig 4 (c) shows the lip termination method described earlier in this paper by Bunton et al [9]. Here the black curve represents the initial seed area function and the red curve shows final vocal tract area and lattice are compared with the final area function of Fig.4 (c).Two peaks observed near the pharyngeal region and at middle part of oral cavity in fig 4 (a) and (b) are almost in the same position in x-axis as the one in fig 4 (c).


Estimated vocal tract shape for 8 sections of tube

Estimated vocal tract shape for 8 sections of tube

Fig 4: Results for vowel /a/ of female child speaker(a)vocal tract area function using autocorrelation method(b) vocal tract area function using lattice method (c) VTAF results obtained by Bunton et al. [9]

In fig.5 x-axis represents distance from glottis to lips and y-axis represents the normalized cross-sectional area. Fig 5(c) shows the lip termination method described earlier in this paper by Bunton et al [9]. Here the black curve represents the initial seed area function and the red curve shows final vocal tract area function of vowel /i/ obtained after 194 iterations. The obtained results in fig 5 (a) and (b) using autocorrelation and lattice are compared with the final area function of Fig.5 (c).In this case a steep rise of area values is obtained near to the glottis end and thereafter a gradual decrease of area values near to lip end is present.



Estimated vocal tract shape for 8 sections of tube

Fig 5 : Results for vowel /i/ of female child speaker(a)vocal tract area function using autocorrelation method(b) vocal tract area function using lattice method (c) VTAF results obtained by Bunton et al. [9]

5. CONCLUSION

In this paper a short description of estimating vocal tract shape is done. Indirect method of estimation is observed to be suitable for children as they are noninvasive. Result obtained by implementing the VTSE of autocorrelation and lattice is presented. Validation of the result is done by using the results obtained by researcher Bunton et al [9]. Slight variations are expected as the speech signal used for comparison purpose are different from the one used in the autocorrelation and lattice method. The future plan is to implement the implement covariance method and then carry out comparative study among them.

6. REFERENCES

[1] L.R. Rabiner, R.W. Schafer, Digital Processing of Speech Signals, Prentice-Hall, 1978.

[2] D. O. Shaughnessy, Speech Communication: Human and Machine. Reading, Massachusetts: Addison-Wesley, 1987.

[3] M. J. Vascolenes, "Dissertation report on Computational algorithms for image analysis: Applications on human vocal tract and silhouette", Faculdade university, pp no. 1-25, 2015.

[4] D. Calum, **"**Acoustic pulse reflectometry for measurement of the vocal tract", pp no: 65-74, PhD thesis, University of Edinburgh, 2005.

[5] M. S. Shah and P.C. Pandey "Estimation of Vocal tract shape for VCV syllables for a speech training aid" in proc 27th annual conference of the IEEE Engineering in Medicine and Biology society ,pp no .6642-6645,2005.

[6] J. Schroeter and M. M. Sondhi, "Techniques for estimating vocal-tract shapes from the speech signal", *IEEE Trans. Speech Audio Process.*, vol. 2, no. 1, pp.133 -150, 1994.

[7] N. Hanna.," Investigations of the acoustics of the vocal tract and vocal folds in vivo, ex vivo and in vitro. Human health and pathology, University de Grenoble, pp .3 -12, 2014.

[8] "Perception of synthetic vowel exemplars of 4 year old children and estimation of their corresponding vocal tract shapes",*Journal of the Acoustical society of America, vol. 120*, pp. 2850-2858, 2006.

[9] K. Bunton, B.H. Story, I.Titze ,"Estimation of vocal tract area functions in children based on measurement of lip termination area and inverse acoustic mapping", *in Proc. of Meetings on Acoustics*, Vol. 19,060054 Montreal, , pp no. 1-8,2013.

[10] R. Kuc, F. Tuteur and J.R. Vaisnys, "Determining vocal tract shape by applying dynamic constraints", *IEEE International conference on ICASSP*, vol. 10, pp.1101-1104, 1985.

[11] H. Wakita, "Direct Estimation of the Vocal Tract Shape by Inverse Filtering of Acoustic Speech Waveforms", *IEEE Trans. Audio Electroacoust.* , vol. 21, pp. 417- 427, 1973.

[12] N. S. Wankhede and M. S. Shah "Investigation on optimum parameters for LPC based vocal tract shape estimation", *Proc. Int. Conf. Emerging trends Commun. Control, Signal Process. & Comput. Appl.*, pp.1 -6, 2013.

[13] H. Wakita, "Direct Estimation of the Vocal Tract Shape by Inverse Filtering of Acoustic Speech Waveforms", *IEEE Trans. Audio Electroacoust.* , vol. 21, pp. 417- 427, 1973.

[14] M. Raifel and F.A. Flomen ,"Split Burg and Covariance Lattice Algorithms", *IEEE Trans on Signal Processing*, vol. 42, no. 5, pp.1279 -1281 1994 .

[15] N. S. Wankhede, "Dissertation report on Vocal tract shape estimation", Mumbai university, 2013.

[16] Available:http://www.ece.ucsb.edu/Faculty/.../Lecture%2013winter20126tp.pdf.

[17] Veena S, Nilashree S Wankhede and Milind S Shah "Study of Vocal Tract Shape Estimation Techniques for Children" *Proceedings of International Conference on Communication, Computing and Virtualization (ICCCV)*,vol.79,2016,pp.270-277.