Estimation of Vocal Tract Shape of Vowels for Children

Veena S, Nilashree S Wankhede and Dr. Milind S Shah
Department of Electronics and Telecommunication
Fr. C. Rodrigues Institute of Technology
Navi Mumbai 400703, Maharashtra, India
veena.sivan@rediffmail.com, nilashreew@hotmail.com, milind05in@yahoo.co.in

***ABSTRACT*— This paper presents the methods used for vocal tract shape estimation i) Covariance method ii) Lattice Method. The study aims to estimate vocal tract shape of vowels /a, e, i, o, u/ uttered by children and covers the children in the age group of 4 to 18 years. In addition images of children articulating vowels were taken to obtain the lip area of child. A comparison of vocal tract shape obtained with lip area assumed to be 1cm^2 is done with the shape obtained by actual lip area. In case of vowel /u/ the normalized area values obtained were found to be reduced for most of subjects when accurate lip area was given. This could be due to value of lip area less than 1 cm^2.Vocal tract shape of 6 year old child is presented.**

Index Terms — Cholesky decomposition method , Burg Method, Articulatory Models, VTSE.

I. INTRODUCTION

Determination of vocal tract shape can be done via are direct and indirect methods. Direct method is based on measurements of vocal tract and indirect method is based on the speech signal analysis [1].Vocal tract shape estimation has many applications like speech coding, speaker and speech recognition [2].

This paper presents the indirect method of vocal tract shape estimation based on covariance and lattice method for children by obtaining accurate lip area of the subjects. Results obtained using these methods are discussed on the basis of previous studies by Story et al [3]. Section II discusses some of the indirect methods of determining vocal tract shape. Section III presents the implementation details to estimate the lip area and the shape of the vocal tract, followed by Section IV which discusses about the results obtained. Conclusion and future scope is presented in section V.

II. INDIRECT METHODS USED FOR VTSE

A. VTSE using measurement of pressure at lips

Aktosun [4] has suggested a method of determining the shape of vocal tract by posing a direct and indirect problem. Direct problem is described as determination of pressure present at lips when the vocal tract shape is known and the inverse problem is recovering the VTS from the absolute pressure at lips known at all frequencies by using Gel'fand-Levitan method.

$$A(x) = \frac{c\mu\left[1 + \int_0^x dyh\ (x, y)\right]^2}{P_\infty\left[1 + \int_0^l dyh\ (l, y)\right]} \quad (1)$$

$A(x)$ = area of vocal tract.
c = speed of sound.
μ=air density
x = location from glottis to lips.
$h(x,y)$ = Gel'fand-Levitan solution
P_∞ = asymptotic value of absolute pressure

B. Articulatory and Simulation Models used for VTSE

Birkholz et al. [5] developed a 3D model of the vocal tract that can change its anatomy and articulation for children between 1 and 20 years. Children use slightly different articulatory configurations than adults. An observation by author was lower tongue positions than adults for the articulation of low and mid vowels. In a similar study carried by Dang et al. VTSE was done by using a physiological articulatory model consisting of nine parameters the tongue and jaw (in both x and y direction), kinematical parts of the lips (in both x and y direction) and glottis height [6].

Shah et al. [7] developed a speech analysis package for speech training aid which uses areagram for displaying the cross sectional area of vocal tract. VTSE is based on LPC analysis of speech signal. Goldstein (1980) [8] implemented a model of vocal-tract shape based on anatomical data on

children. The sizes of the articulatory structures at various ages affect vocal tract length and cross-sectional area. It was shown in the study that vocal tract of one year old child could form articulatory configuration which could produce the vowels.

C. Formant Frequency

Ladgefoged et al. [9] determined VTSE using formant frequency by choosing appropriate parameters. In this case he initially found formant frequency and used multiple regression technique to find parameters which are correlated with formant frequency. Tongue positions to be relevant parameter in recovering VTS. Displacement present in the tongue position from its reference position was found for each person uttering a vowel sound by using eq.2

$$\hat{d}_{ijk} = (t_{1i} v_{1i} s_{1k}) + (t_{2i} v_{2j} s_{2k}) \quad (2)$$

v_{1i} and v_{2j} : weights of vowel j.

s_{1k} and s_{2k} :scaling constants for k[th] vocal tract with respect to constants t1 and t2

Now amount of front raising and back raising part of tongue were found by using eq.

$$w_1 = c_1\left(\frac{F2}{F3}\right) + c_2\left(\frac{F1}{F3}\right) + c_3\left(\frac{F3}{F1}\right) + c_4 \quad (3)$$

w_1= front raising component of tongue

c_1 =2.3, c_2= 2.1, c_3=0.17, c_4= -2.1

$$w_2 = c_5\left(\frac{F1}{F2}\right) + c_6\left(\frac{F2}{F1}\right) + c_7\left(\frac{F3}{F1}\right) + c_8 \quad (4)$$

w_2 = back raising component of tongue

c_5=-1.9, c_6=-0.24, c_7=0.18, c_8=0.58

F1, F2 and F3 represents the first, second and third formant frequency. Other important parameter used was distance between lips.VTS was generated with the help of these values.

McGowan et al. have derived VTS of 4 year old children by using formant frequency data [10].The representation of results was done by using a five tube model.VTS were obtained by two different sources one is by scaling the adult formant frequency values for vowels / a, e , i , u/ in accordance with the values given by Kent and Forner. Now comparison of the results was done with a listening test in which synthesized vowels were generated by using a Klatt synthesizer by deriving the first three formant frequencies by using genetic algorithm.

D. Autocorrelation Method

Vocal tract shape of children was estimated by utilizing autocorrelation method based on LPC analysis by Wankhede et al. Three optimum parameters for VTSE reported are the sampling frequency, vocal tract length and LPC order. For determining autocorrelation coefficients Levinson-Durbin algorithm was used [11] [14]. The autocorrelation coefficients are converted to reflection coefficients so as to find the area values. The area value at lips is kept 1 and the area values of other sections were determined from lips to glottis from the following Eq. 5.

$$A_m = \frac{1+\mu_m}{1-\mu_m} A_{m+1} \quad (5)$$

A_m: area of m^{th} 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

III. IMPLEMENTATION

This study was carried on children in the age range of 3-18 years by categorizing them on age basis. The first division consisted of children from 3-5yrs, second division from 6-9 yrs, third from 10-12,fourth division from 13-15 and the last from 16-18yrs.Subjects were asked to articulate vowels /a/,/i/,/e/,/o/ and /u/ for five seconds. The recording of vowels was done by using Praat software. Images of children articulating vowels were captured and read into Matlab to determine the lip area [3].

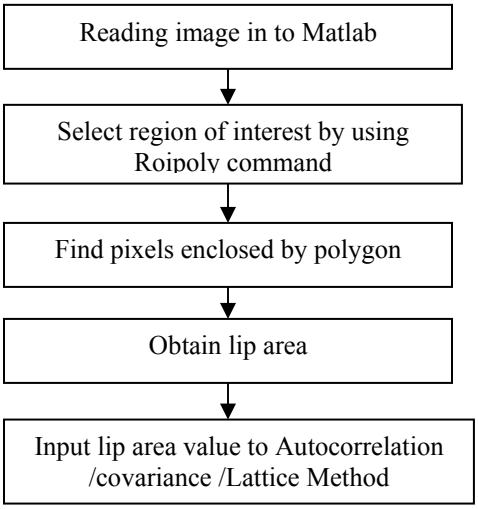


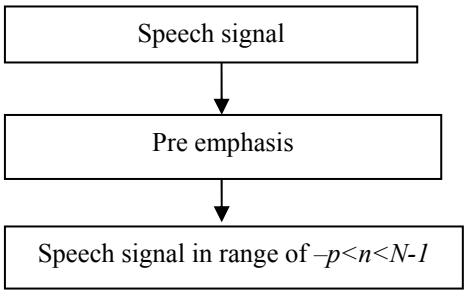
Fig.1 Block diagram for determining lip termination area

Matlab function roipoly was used to select the region of interest (opening of lips) by a series of mouse clicks. The lip termination area is computed by determining the number of pixels enclosed by the resulting polygon. Fig.1 shows the algorithm used. The red region of the video frame in Fig. 2 shows the lip termination area of a 6 yr old female talker producing the vowel [a]. The area in this case is found to be 5.2cm^2.

Fig.2. Lip area of a 6 year old female talker

A. *Covariance Method*

Covariance method is used for defining the speech segment S_n *(m)* and the limits on the sums, namely fix the interval over which the mean-squared error is computed. In this method windowing of segment is not needed as signal values are taken in the interval $-p \le n \le N-1$ [10], such that p samples before the interval are taken to predict the samples at the beginning of the interval. Resampling of speech samples according to the age of the child are done in accordance with data provided by Wankhede et al. [11].Now to obtain the LPC coefficients the Cholesky decomposition method is applied [1]. Now in order to obtain the vocal tract shape a further conversion of LPC coefficients to reflection coefficients is performed [13]. Finally the area values are retrieved The lip area of subject is given as input and the area values of other sections were determined from lips to glottis from the following Eq. 5.Comparison of the results obtained for 6 year old children in the second age group with the results of Bunton et al.[3] was done. Table I gives results of 3 subjects.



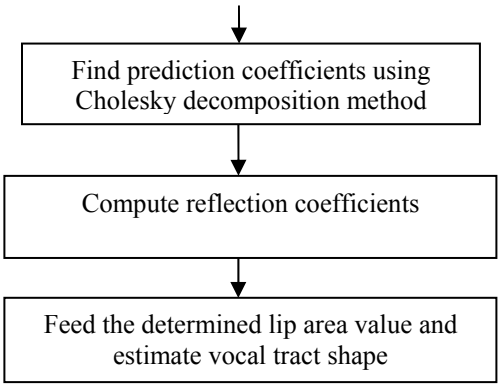


Fig.3. Block diagram of Covariance method [1] [12]

B. *Lattice Method*

In the covariance 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 [1] [13]. Now in case of lattice formulation these two approaches are treated in an integrated manner. Implementation of Burg's method [17] in Matlab was carried. Accordingly the minimization of the prediction errors yields reflection coefficients. The lip area of subject is given as input and the area values of other sections were determined from lips to glottis from the following Eq. 5.Comparison of the results obtained for 6 year old children in the second age group with the results of Bunton et al. was done. Table I gives results of 3 subjects.

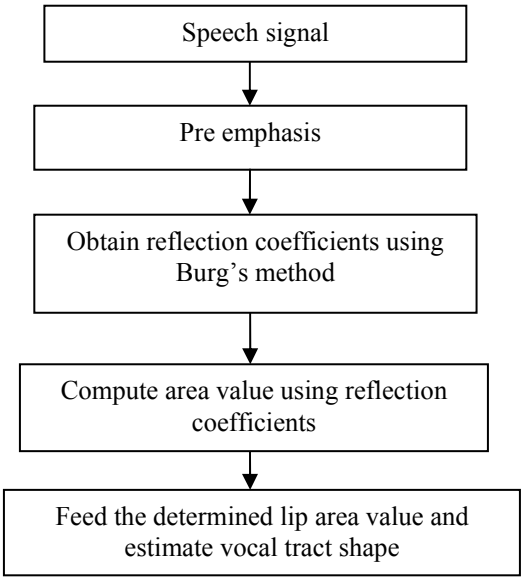


Fig.3. Block diagram of Covariance method [1] [12]

IV. RESULTS

The parameters used for evaluation of the obtained vocal tract shape of vowels are [12]

- Position of the pharyngeal and oral cavity (X_b and X_f).
- Area of the pharyngeal and oral cavity (A_b and A_f).
- Location of constriction (X_c).
- Amount of lip opening.

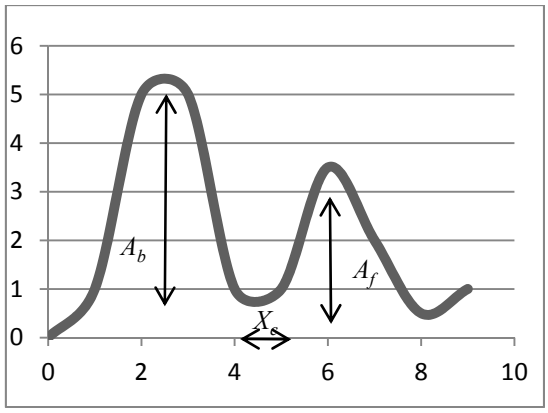


Fig.5. Area function of vocal tract representing parameters [4]

Validation is done by comparing the vocal tract shape obtained from this study with vocal tract profile of Bunton et al [3] obtained using inverse acoustic mapping of children aged 6 years.

Vowel [a]
The results obtained from the three methods and the results by using acoustic mapping have good amount of resemblance. It contains a narrow pharyngeal cavity at glottis and a large oral Cavity [11].The results obtained for speaker 1 are shown in Fig. 6.The figures in the left side indicate the area functions obtained by assuming lip area to be 1 cm^2.While the figures in the right hand side indicate area function obtained by actual measurement of the lip area. In this case the value is 5.2 cm^2.
Vowel [e]
Notable feature in this case is presence of large cavity at back of mouth.
Vowel [i]
Large pharyngeal cavity was observed with a narrow oral constriction.
Vowel [o]
A large mouth cavity formed by a low back tongue position and raised jaw. In this case features are matched with adult database due to lack of child database.
Vowel [u]
Two large cavities were linked by a narrow constriction was observed, and a second constriction at the lips. The size of the cavities and position of constriction matched with the results of previous study.

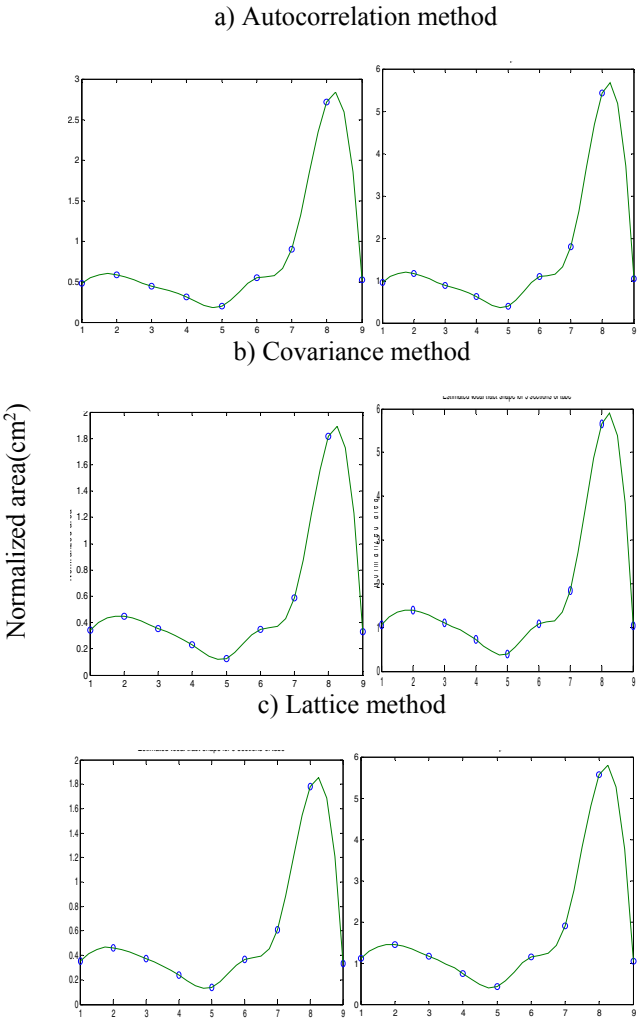


Fig.6. Results for vowel /a/ of female child speaker(a) Aurocorrelation Method(b)Covariance Method(c)Lattice method Fs = 14KHz,M = 9,l = 12cm

Table I summarizes the vocal tract parameters obtained by using autocorrelation, covariance and lattice methods for three speakers aged 6 years.

TABLE I. AREA VALUES OF VOWELS USING AUTOCORRELATION, COVARIANCE AND LATTICE METHODS OF SPEAKERS AGED 6 YEAR

Speaker	**Vow-el**	**Lip area (cm^2)**	**Aurocorrelation**					**Covariance**					**Lattice**				
			X_b (cm)	X_f (cm)	A_b (cm^2)	A_f (cm^2)	X_c (cm)	X_b (cm)	X_f (cm)	A_b (cm^2)	A_f (cm^2)	X_c	X_b (cm)	X_f (cm)	A_b (cm^2)	A_f (cm^2)	X_c (cm)
1	a	5.3	3	8	2	7	0.2	3	8.5	2	7.5	0.2	3.5	8	1.8	7.5	0.2
	e	4.3	NA	8	NA	6.5	0.2	NA	8	NA	6	0.2	NA	8	NA	5.5	0.2
	i	3.38	5	NA	7	NA	1	5	NA	6.5	NA	0.5	5	NA	7	NA	0.5
	o	1.58	4	8	1	8.5	0.4	4	8	0.5	5.5	0.4	4	8	1	8	0.4
	u	0.8	3	7	2.5	5	1	3	7	3.5	5	0.6	3	7	2	5	1
2	a	6.1	2	8	1.2	6.5	0.4	2	8	1.6	6.8	0.4	2	8	1.2	7	0.2
	e	4.3	NA	8	NA	7.5	0.4	0.5	8	NA	8	NA	NA	8	NA	7.5	1
	i	3.75	3	NA	3.5	NA	0.5	2	NA	2.5	NA	0.5	2	NA	2	NA	0.2
	o	2.43	4	8	1	4.5	NA	4	8	1	6.5	NA	4	8	1.5	5	NA
	u	0.79	2	8	4	4.8	1	2	8	5	2	0.2	2	8	5	4.5	0.5
3	a	5.6	5	8	3	9	0.2	4	8	1	8	0.2	4	8	3	9	0.6
	e	2.85	NA	8	NA	6.5	NA	NA	8	NA	6	NA	NA	8	NA	5.5	NA
	i	3.2	6	NA	6	NA	1.2	6	NA	8	NA	1	6	NA	5.5		1
	o	1.48	4	8	1.5	5.5	0.8	4	8	1	4.75	0.5	4	8	1	4.5	0.8
	u	1.2	2.5	8	1.5	2.2	0.2	2	8	2	2.5	0.2	2	8	1.5	2.5	0.2

*NA-Not Applicable

V.CONCLUSION

Vocal tract shape was estimated for vowels by using the autocorrelation, covariance and lattice methods with lip area assumed to be $1cm^2$ and also by using proper lip termination area details of the subjects obtained from their images. Much more accurate area values were obtained by using actual lip area values.

In future speech training systems capable of displaying vocal tract shapes could be developed for hearing impaired children which would assist them in language acquisition. Further suitability of any other method for determination of vocal tract shape estimation can be carried out which gives even better estimates of parameters.

REFERENCES

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

[2] G. Richard, M. Goirand, D. Sinder, J. Flanagan "Simulation and visualization of articulatory tracjectories estimated from speech signals" *Proceedings of the International Symposium on Simulation, Visualization and Auralization for Acoustic Research and Education (ASVA97),* 1997.

[3] K. Bunton, B.H. Story and 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 , pp. 1-8,2013.

[4] T Aktosun, "Determining the shape of human vocal tract from pressure measurement at lips", Technical Report, University of Texas, 2007, pp. 1-10.

[5] P. Birkholz and B.J.Kroger, "Simulation of vocal tract growth for articulatory speech synthesis", *in Proc. 16th International Congress of Phonetic Sciences*, pp.377-380, 2007.

[6] J. Dang and K. Honda, "Estimation of vocal tract shapes from speech sounds with a physiological articulatory model", *Journal of phonetics*, pp. 511-532, 2002.

[7] M.S. Shah and P.C. Pandey, "Areagram display for investigating the estimation of vocal tract shape for a speech training aid", in *Proc. Symposium on Frontiers of Research on Speech and Music, IIT Kanpur*, 2003, pp.121-124.

[8] 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.

[9] P. Ladefoged, R. Harshman, L. Goldstein, and L. Rice, "Generating vocal tract shapes from formant frequencies," *J. Acoustic Soc. Am*., vol. 64, part. 4, pp. 1027-1035, 1978.

[10] R.S. McGowan, **"**Perception of synthetic vowel exemplars of four year-old children and estimation of their corresponding vocal tract shapes," *J. Acoust. Soc. Am*., pp. 2850-8,2006.

[11] 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.

[12] Available:*http//*afshin.sepehri.info/ADSP/LinearPrediction/linearprediction.html.

[13] A.S. Patil and M.S. Shah, "Comparison of vocal tract shape estimation techniques based on formant frequencies, autocorrelation, covariance and lattice", *Int. Conf. on Nascent Technologies in Engineering Field*, pp. 1-6, 2015.

[14] 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.

[15] D. Calum, **"**Acoustic Pulse Reflectometry for Measurement of the Vocal tract", PhD thesis, University of Edinburgh, 2005.

[16] K. Jeevakumar, "Joint Estimation of Vocal Tract and Source Parameters of a Speech Production Model", Ph.D. Thesis, Dublin City University, pp.15-30, 1993.

[17] 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 .