Volume 3, Issue 6, June 2013 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

Lattice Filter Model of Human Vocal Tract

Neha Garg[1], Rakesh Garg[2]
ECE Department, KITM,
Kurukeshtra, India

Abstract- This paper describes the generalized lattice model of human vocal tract for speech analysis, relating it to the all pole type linear prediction. A more natural synthesized speech can be achieved using Auto Regressive (AR) lattice model. Speech analysis is done using linear prediction. The structure of AR lattice model is numerically stable, which is main requirement in speech analysis and synthesis applications. Lattice model uses reflection coefficients; relation of these reflection coefficients with LPC predicted coefficients is given by Levinson Durbin algorithm. Simulation results shows that roots of the polynomial in the denominator of transfer function of lattice model is always guarantee to be inside the unit circle and the stability of the system is achieved.

Keywords: Vocal Tract; Lattice model; Reflection coefficients; Linear Prediction; Levinson algorithm.

I. INTRODUCTION

The vocal tract is the cavity in human beings where sound is produced at the sound source and filtered. Human speech is produced in the vocal tract which can be approximated as a variable diameter tube [1]. The speech mechanism can be modelled as a time-varying filter which acts as the vocal tract excited by an oscillator as the vocal folds, with different outputs as shown in figure 1. where G is the gain, x(n) the input signal, a(n) the amplified signal and s(n) the output signal; n is a digitized time variable. When voiced sound is produced, the filter is excited by an impulse sequence. When unvoiced sound is produced, the filter is excited by random white noise [2].

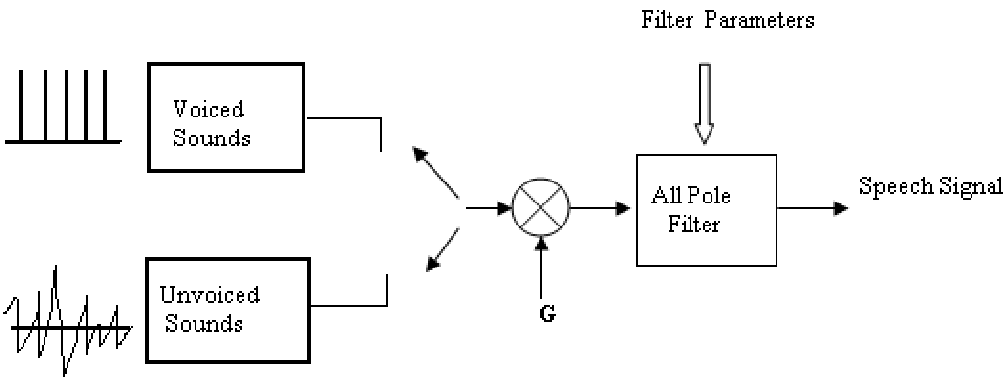

Figure 1- Block diagram of the source-filter model of speech

The linear predictive coding (LPC) model is based on a mathematical approximation of the vocal tract. Vocal tract and glottal excitation can be represented by a time-varying digital filter [3] whose steady state system function is of the form as:

$$H(Z) = \frac{S(Z)}{U(Z)} = \frac{G}{1 + \sum_{k=1}^{p}(a_k)\ Z^{-k}} \quad (1)$$

Generally the all-pole model is preferred for most applications because it is computationally more efficient. Moreover, it is easy to solve an all-pole model [4]. The basic problem of linear prediction here is to determine a set of predictor coefficients directly from speech signal in such a manner as to obtain a good estimate of spectral properties of speech signal..In section 2 we have described the basic principle of linear prediction and LPC analysis steps with brief introduction. LPC predicted coefficients computation is described using Levinson Durbin algorithm in section 3. Lattice implementation is introduced in section 4. At last simulation results is shown in section 5 and at the end conclusion in section 6 and references.

II. LINEAR PREDICTION BASIC PRINCIPLE

Linear predictive coding is defined as a digital method for encoding an analog signal in which a particular value is predicted by a linear function of the past values of the signal [5] . The speech samples s(n) can be given by using simple difference equation (2). Where a_k is the predicted LPC coefficients and G is the gain and u(n) is unknown input signal.

$s(n) = \sum_{k=1}^{p} (a_k\, s(n-k)) + G\, u(n)$ (2)

The signal s (n) can be predicted only approximately from a linearly weighted summation of past samples as in equation (3).

$s'(n) = -\sum_{k=1}^{p} (a_k\, s(n-k))$ (3)

Linear Prediction (LP) residual is the prediction error e(n) obtained as the difference between the predicted speech sample s'(n) and the current sample s(n). This is shown in equation (4).

$e(n) = s(n) - s'(n)$ (4)

$e(n) = s(n) + \sum_{k=1}^{p} (a_k\, s(n-k))$ (5)

In the frequency domain, the equation (5) can be represented as,

$E(Z) = S(Z) + \sum_{k=1}^{p} (a_k\, S(Z))\ Z^{-k}$ (6)

By this we can obtained:

$A(Z) = \frac{E(Z)}{S(Z)} = 1 + \sum_{k=1}^{p} (a_k)\, Z^{-k}$ (7)

Now by comparing the equation (1) and equation (7) by taking gain (G) equals to one, we can obtained LP residual by filtering the speech signal with A (z) as indicated in figure 2.

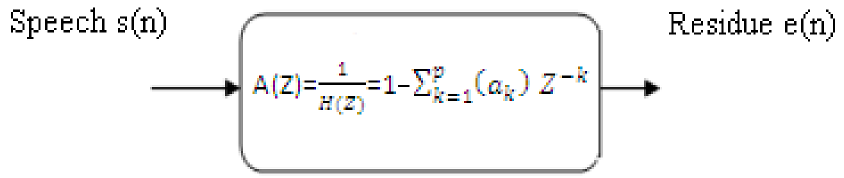

Figure 2: Computing the LP residual by inverse filtering

As A (z) is the reciprocal of H(z), LP residual is obtained by the inverse filtering of speech [14].

A. LPC Analysis Steps

Figure 3 shows a block diagram of the main computational steps required for the LPC front-end processor,

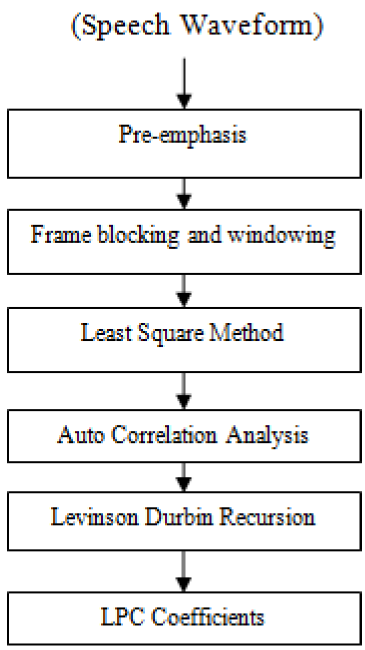

Figure 3: LPC Coefficients Extraction Process

It is combines activities like, pre-emphasis, act as a low-order digital filter, increases the power in the upper frequency bands of the speech signal [8]. The FIR filter is similar to a high pass filter .Frame blocking step blocks the pre-emphasized speech signal into frames. In order to minimize the signal discontinuities at the beginning and end of each frame, windowing step is applied. Autocorrelation analysis identifies the zeroth autocorrelation which is the energy of each frame which is very important parameter for speech detection systems [12]. After auto correlation difference equations are solved using Levinson Durbin algorithm and LPC predicted coefficients a_k and reflection coefficients k_i are obtained. As residual prediction error is represented by equation (5) then the method of least squares is applied to obtained coefficients a_k [7] .

Where E Denote the total squared error by:

$$E = \sum_n e_n^2 = \sum_n \left(s_n + \sum_{k=1}^{p} a_k \; s_{n-k} \right)^2 \tag{8}$$

E is minimized by setting:

$$\frac{\partial E}{\partial a_k} = 0, \quad 1 \le k \le p \tag{9}$$

The set of p equations in p unknowns can be solved for the predictor coefficients which minimize E .The minimum total squared error, denoted by E_p is obtained by:

$$E_p = \sum_n s_n^{\;2} + \sum_{k=1}^{p} a_k \sum_n (s_n \, s_{n-k}) \tag{10}$$

Now auto correlation analysis is applied, here we assume that the error in E is minimized over the infinite duration.

$$\sum_{k=1}^{p} a_k \; R\,(i\text{-}k) = -\,R\,(i), \; 1 \le i \le p \tag{11}$$

$$E_p = R\,(0) + \sum_{k=1}^{p} a_k \, R(k) \tag{12}$$

Where R (i) is the autocorrelation function of the signal s (n):

$$R\,(i) = \sum_{n=-\infty}^{\infty} s_n \;\; s_{n+i} \tag{13}$$

Note that R (i) is an even function of i :-

$$R\,(-i) = R\,(i) \tag{14}$$

III. Computation of Predicted Coefficients

The predicted coefficients can be computed by solving a set of p equations with p unknowns.

$$\begin{bmatrix} R_0 & R_1 & R_2 \cdots & R_{p-1} \\ R_1 & R_0 & R_1 \cdots & R_{p-2} \\ \vdots & \vdots & \vdots & \vdots \\ R_{p-1} & R_{p-2} & R_{p-3} \text{—} & R_0 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{bmatrix} = - \begin{bmatrix} R_1 \\ R_2 \\ \vdots \\ R_p \end{bmatrix}$$

Note that the p X p auto correlation matrix is symmetric and the elements along any diagonal are identical (i.e., a Toeplitz matrix).

This simplification causes the predictor coefficients to be computed efficiently using the Levinson-Durbin recursion as [10]:

$$E^{(0)} = R\,(0) \tag{15}$$

For $1 \le i \le p$ and $1 \le j \le i-1$

$$c_i = \left[R\,(i) - \sum_{j=1}^{i-1} R(i-j) \; a_j^{(i-1)} \right] \tag{16}$$

$$k_i = \frac{c_i}{E^{(i-1)}} \tag{17}$$

$$a_i^i = k_i \tag{18}$$

$$a_j^i = a_j^{(i-1)} - k_i a_{(j-1)}^{(i-1)} \tag{19}$$

$$E^{(i)} = (1 - k_i^2).\, E^{(i-1)} \tag{20}$$

Equations (16) to (20) are solved recursively for i = 1, 2,..., p and finally the predictor coefficients a_j^i and reflection coefficients k_i are calculated.

IV. Lattice Implementation of LPC Filters

Any LPC analysis filter can be implemented as a lattice filter. The relationship between a_k and k_i is given by the Levinson-Durbin recursion. The first order prediction coefficient a11 is the same as the reflection coefficient k1. The mth order linear prediction coefficients are obtained from the (m−1)th order prediction coefficients and the reflection coefficients k_m [11] . Thus, M linear prediction coefficients are equivalent to M reflection coefficients. Quantization of reflection coefficients is easier because of the well-defined range of values that they take on. Note that the absolute value of reflection coefficients is never greater than one [9] .This is the reason why reflection coefficients instead of linear prediction coefficients are often used to represent a vocal tract filter. Figure 4 shows the lattice implementation of LPC analysis filter.

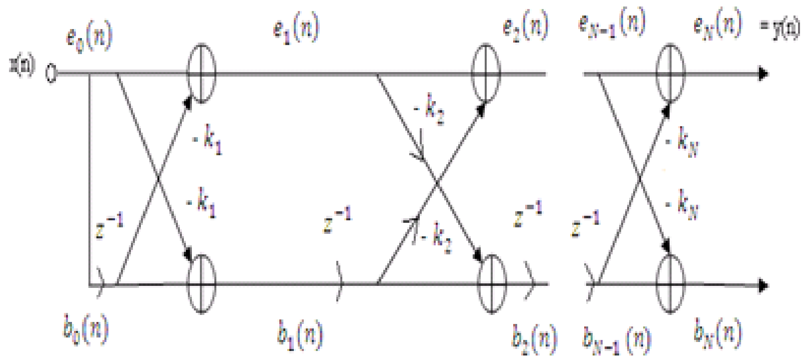

Figure 4: Lattice Implementation of LPC Analysis Filter

From each frame of speech samples, M reflection coefficients are computed. Because important information about the vocal tract model is extracted in the form of reflection coefficients, the output of the LPC analysis filter using reflection coefficients will have less redundancy than the original speech. Thus, less number of bits is required to quantize this so-called residual error. This quantized residual error along with the quantized reflection coefficients are transmitted or stored. To play back, a lattice implementation of the LPC synthesis filter is required. In this case, the input is the residual error and the output is the reconstructed speech.

V. SIMULATION SET UP AND RESULTS

For analysis of speech signal we have used MATLAB software. Digital speech processing tool box is used. Here we have taken a speech signal of duration 5 seconds. We have performed four steps. First is the calculation of residue signal, second is the conversion of predicted LPC coefficients in to reflection coefficients ,third step is transfer function representation of vocal tract filter and fourth is the pole-zero plot.

Technical Details:

TABLE 1
TECHNICAL DETAILS OF SPEECH SIGNAL

Start time: 0 seconds	**NO. of samples: 4085**
End time: 0.51065 seconds	**Sampling period: 0.000125 sec**
Total Duration:0.51065 seconds	**Sampling frequency: 8000 Hz**
Prediction order: 10	**Number of frames: 17**

Figure 5 shows the speech signal waveform and .Figure 6 shows the output waveforms LPC analysis of 1st frame duration [1 : 240] and table 2 shows the reflection coefficients for 17 frames. At last average is taken and transfer function is obtained. Figure 7 shows the pole zero plot of transfer function.

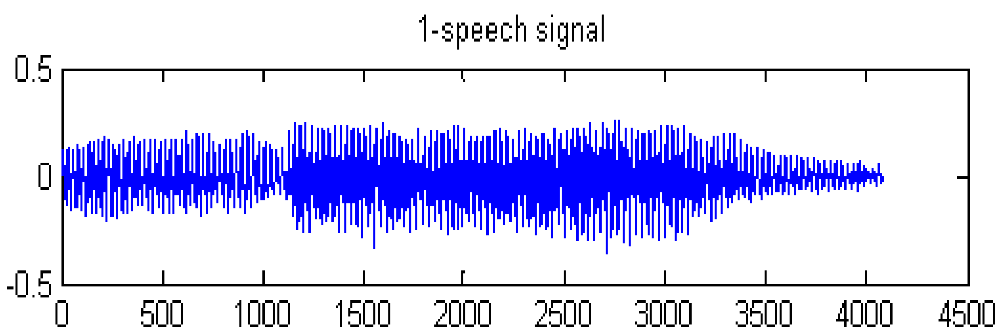

Figure 5: Speech Signal Waveform

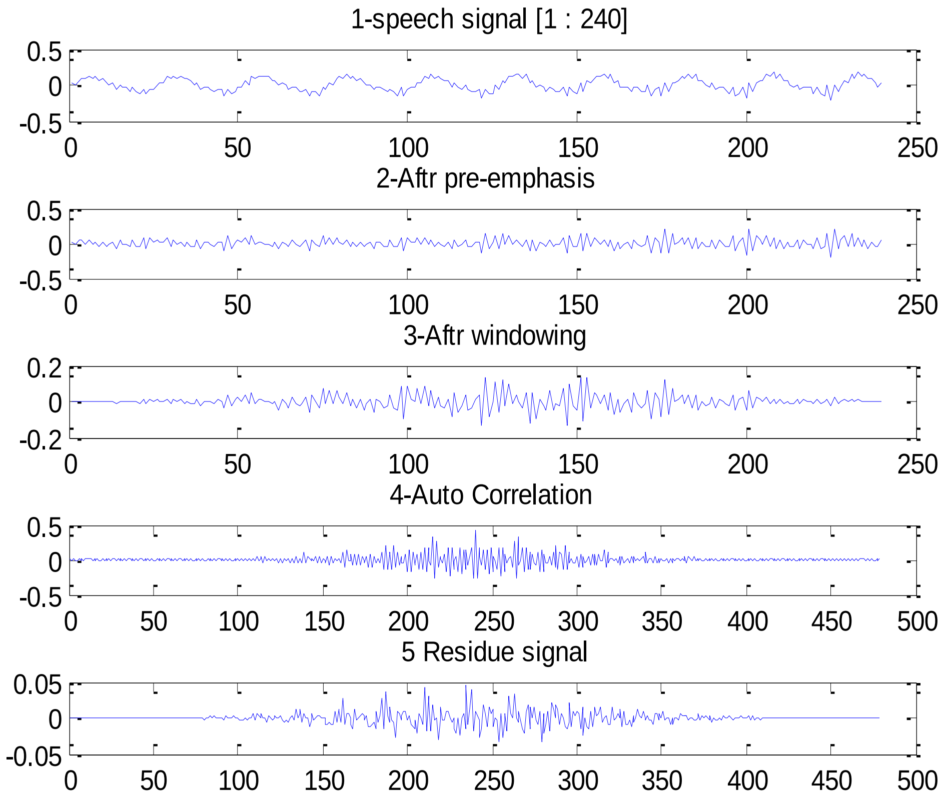

Figure 6: LPC Analysis waveform [1: 240]

TABLE 2
LATTICE MODEL REFLECTION COEFFICIENTS

Frame [1: 240]	**Frame [240:480]**	**Frame [480:720]**	**Frame [720 : 960]**	**Frame[960:1200]**	**Frame [1200 :1440]**
0.8351	0.8738	0.8139	0.2914	-0.5052	0.5885
0.3716	0.6402	0.2057	-0.5556	-0.3078	0.3934
-0.6924	-0.7827	-0.6944	-0.5218	-0.2215	-0.6372
-0.3362	-0.0936	0.0088	0.4719	0.7583	-0.073
-0.1164	0.1041	0.4205	0.5688	0.4712	0.4179
0.8819	0.8474	0.6411	0.4138	0.2438	0.9497
0.1953	-0.3638	0.0279	-0.6419	-0.3654	0.299
0.2137	-0.1353	-0.2632	0.2723	-0.4354	0.4666
-0.6398	-0.244	-0.1942	-0.3856	-0.1858	-0.3098
0.3575	0.1973	-0.0683	0.1004	0.2195	-0.3983

Frame[1440:1680]	**Frame[1680:1920]**	**Frame[1920:2160]**	**Frame[2160:2400]**	**Frame[2400:2640]**	**Frame[2640:2880]**
0.7204	0.7545	0.7615	0.7422	0.7931	0.8162
0.755	0.8866	0.8983	0.9233	0.9345	0.9616
-0.4644	-0.2439	-0.1959	-0.2293	-0.2754	-0.0991
-0.2977	-0.5388	-0.5066	-0.5331	-0.4242	-0.3443
-0.1084	-0.0814	-0.1848	-0.1115	-0.2199	-0.4651
0.9086	0.8799	0.8615	0.9157	0.9164	0.9003
0.5939	0.5439	0.5868	0.6265	0.7076	0.5472
0.3501	0.2343	0.3301	-0.0192	0.4316	0.5573
-0.5766	-0.477	-0.3447	-0.6129	-0.3703	-0.6421
-0.3864	-0.3135	-0.2766	-0.2169	-0.5167	-0.4988

Frame[2880:3120]	**Frame[3120:3360]**	**Frame[3360:3600]**	**Frame[3600:3840]**	**Frame[3840:4080]**	**Average[1:4080]**
0.7395	0.4111	-0.7264	-0.8299	-0.8680	0.3653
0.9488	0.8512	0.3904	0.1315	0.6913	0.5364

-0.5094	-0.4827	-0.3942	0.0059	0.1350	-0.3700
-0.3574	-0.2258	0.3147	0.7205	0.2809	-0.0691
0.0839	-0.1414	0.1915	-0.0871	0.2052	0.0557
0.8446	0.7206	0.7358	0.0936	0.8549	0.7417
0.6379	0.2965	-0.4661	-0.2163	-0.1782	0.1665
0.5890	0.5991	0.0575	0.2136	0.1860	0.2145
-0.8026	-0.5689	-0.4041	-0.5818	-0.2216	-0.4448
-0.0542	-0.3704	-0.0616	0.4282	0.2155	-0.0966

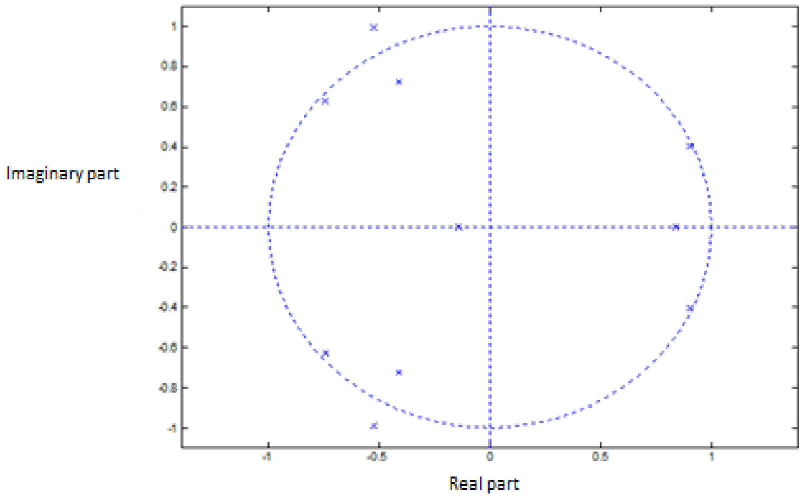

Figure 7: Pole zero plot

VI. CONCLUSION

In this study, we have used the AR lattice model to model vocal tract for glottal source excited speech analysis. Using LPC analysis we calculated the residue signal which is the compressed speech signal and can be represented by fewer bits for transmission purposes. Levinson Durbin algorithm computes the predicted coefficients and reflection coefficients efficiently. Using the AR lattice model, it can be observed from pole zero plot that all poles lies inside the unit circle, which is the essential condition for the system to be stable. Lattice filters are used rather than direct-form filters since the lattice filter coefficients have magnitude less than one and, conveniently, are available directly as a result of the Levinson-Durbin algorithm.

REFERENCES

[1]. G. Fant, “*Acoustic Theory of Speech Production*”. The Hague: Mouton & Co.,1960; Walter de Gruyter, 1970.

[2]. B. S. Atal and M. R. Schroeder, “*Linear prediction analysis of speech based on a pole-zero representation*,” J. Acoust. Soc. Am., vol. 64, no. 5, pp. 1310–1318, 1971.

[3]. E.M. Hofstetter, “An Introduction to the Mathematics of Linear Predictive Filtering as Applied to Speech Analysis and Synthesis,” Technical Note 1973-36, MIT Lincoln Labs, July 1973.

[4]. J.D. Markel and A.H. Gray Jr., “*A Linear Prediction Vocoder Simulation Based Upon the Autocorrelation Method*,” IEEE Transactions on Acoustics, Speech, and Signal Processing, vol. ASSP-22, no. 2, pp. 124-134, April 1974.

[5]. J. Makhoul. “*Linear prediction: A tutorial review*,” Proc. IEEE, vol 63, no. 4, pp. 561-80, Apr. 1975.

[6]. J.D. Markel and A.H. Gray Jr., “*Linear Prediction of Speech*”, Springer-Verlag, Berlin Heidelberg, New York, USA, 1976.

[7]. L.R. Rabiner, B.S. Atal and M.R. Sambur, “*LPC Prediction Error Analysis of Its Variation with the Position of the Analysis Frame*,” IEEE Transactions on Acoustics, Speech, and Signal Processing, vol. ASSP-25, no. 5, pp. 434-442, October 1977.

[8]. P. Thevenaz, and H. Hugli. “*Usefulness of the LPC-residue in text-independent speaker verification,” Speech Commun*., vol. 17, pp.145-57, Aug. 1995.

[9]. Kari Roth And Ismo Kauppinen” *Frequency warped burg's method for AR-modeling*” 2003 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics.

[10] .Ziad A. Al Bawab,”*An Analysis-by- Synthesis Approach to Vocal Tract Modeling for Robust Speech Recognition*”, Ph.D, December 17, 2009.

[11]. loan Tabus” *Interleaved Quantization-Optimization and Predictor Structure Selection for Lossless Compression of Audio Companded Signals*” Proceedings of the 4th International Symposium on Communications, Control and Signal Processing, ISCCSP 2010.

[12]. Thiang and Suryo Wijoyo "*Speech Recognition Using Linear Predictive Coding and Artificial Neural Network for Controlling Movement of Mobile Robot*" 2011 International Conference on Information and Electronics Engineering IPCSIT vol.6 (2011) IACSIT Press, Singapore.

[13]. Chian-Hong Wong 2011 "*Warped Linear Predictive Speech Coding*" IEEE International Conference on Signal and Image Processing Applications (ICSIPA2011).

[14]. Christoph R. Norrenbrock1 "*On the Use of Vocal-Tract Approximations for Instrumental Quality Assessment*" ITG-Fachbericht 236: Sprachkommunikation September 28, 2012.

