Progress In Electromagnetics Research, PIER 61, 89–110, 2006

SYSTEM IDENTIFICATION OF ACOUSTIC CHARACTERISTICS OF ENCLOSURES WITH RESONANT SECOND ORDER DYNAMICS

S. M. Chaudhry

University of Engineering and Technology
Taxila, Pakistan

A. M. Chaudhry

Department of Electronics
Federal Urdu University of Arts Sciences and Technology
Islamabad, Pakistan

Abstract—This research concerns offline identification of acoustic characteristics of enclosures with second-order resonant dynamics and their modeling as linear dynamic systems. The applied models can be described by basis function expansions. The practical problem of acoustic echo in enclosures is used as the target problem to be addressed. It has been found out that the classical filters are ineffective filter structures for approximating an echo generating system, due to their many required parameters. In order to reduce the number of estimated parameters, alternative methods for modeling the room impulse response need to be investigated. Out of various available techniques impulse response identification is utilized. With the help of given experimental data, the enclosures' impulse response is modeled using special orthonormal basis functions called Kautz functions. As another improved approximation, hybrid multistage system identifiers have been used in which the simplicity of classical filter structures and fast convergence of orthonormal structures is utilized as an advantage.

1. INTRODUCTION

Sounds originate when air particles are induced to vibrate around their mean position with consequent changes in the air pressure with respect to a static value. This results in a *soundfield* in which variations of air density and pressure are function of time and space and propagate as *acoustic waves* [1].

A simplified but realistic hypothesis in room acoustics, is to assume that air is a homogeneous medium at rest. The speed c of acoustic wave propagation in air then depends only on temperature and is given by the following formula:

$$c = 331.45\sqrt{T/273}\ \text{m/s}$$

where T is the temperature in kelvins.

Basic laws of physics [2], such as conservation of momentum and conservation of mass, and the ideal-gas equations, lead to a relationship between pressure and density of a fluid in a general differential equation governing sound propagation (Kinsler et al., 1982; Morse and Ingard, 1986). This relationship is the so-called *wave equation*:

$$\nabla^2 p = \frac{1}{c^2}\frac{\partial^2 p}{\partial t^2}$$

where p denotes sound pressure, t denotes time and ∇ is the laplacian operator.

The wave equation is a linear equation and complex exponentials are solutions of it. Since any function with a convergent Fourier integral can be expressed as a weighted superposition of complex exponentials, it is possible to assert that, thanks to linearity, any signal described by this type of function satisfies the wave equation. As a consequence of the superposition principle, any sound field can be thought of as resulting from the superposition of elementary plane or spherical waves. Waves propagate undistorted (under the hypothesis of nondispersive, lossless and homogeneous medium) and carry along the information generated by distant sources.

1.1. Room Acoustics

Generally when we listen to someone talking in a room, unless we are face to face with her/him, most of the acoustic energy captured by our ears is carried by indirect sounds arising from reflection at various surfaces inside the room.

The original signal produced by the speaker undergoes an alteration of its perceptual features such as level, timbre and spatial impression, due to reflections from surfaces and diffusion and diffraction by objects inside the room.

This phenomenon, known as *reverberation* [2], is the most noticeable acoustical phenomenon directly perceivable in an enclosure.

For this reason the most important objective measure in room acoustics is *reverberation time.* Reverberation time, T_{60}, of a room is defined as the time needed for the acoustic power of a received signal to decay by 60 dB from a steady state value after the sound source is abruptly stopped. The reverberation time is nearly independent from the listening position since it is almost constant in a given enclosure [3]. Reverberation times up to 1s (measured for frequencies between 500 and 1000 Hz) do not cause any loss in speech intelligibility.

The space separating a sound source from a receiving transducer can be seen as a transmission channel in which the acoustic signal is conveyed and modified, according to the acoustic characteristics of the environment. The *impulse response h(t)* of the acoustic channel between source and sensor in a reverberant room represents all the multiple reflections from the surrounding surfaces that reach the sensor in addition to the direct sound. Reflections from walls and objects produce a variety of paths between the source and the sensor. Since paths have different length, the propagation delay differs from one path to another and several replicas of the radiated signal reach the sensor after the direct wave front. Every time a wavefront hits a surface, it is partially absorbed and partially reflected. All the reflections contribute to create the complicated impulse response typical of an enclosure, as shown in Figure 1. The effect is more pronounced when the source is placed further from the sensor: the original waveform is distorted by constructive and destructive interference at various frequen-cies. It should be noted that, in ordinary reverberation conditions ($T_{60} <$ 0.7 s), the listener is generally unaware of the distortion of the direct signal caused by reflections [3, 16].

The reflected wavefronts are perceived by human ears only when their delay from the direct sound is longer than 50 ms. Reflections

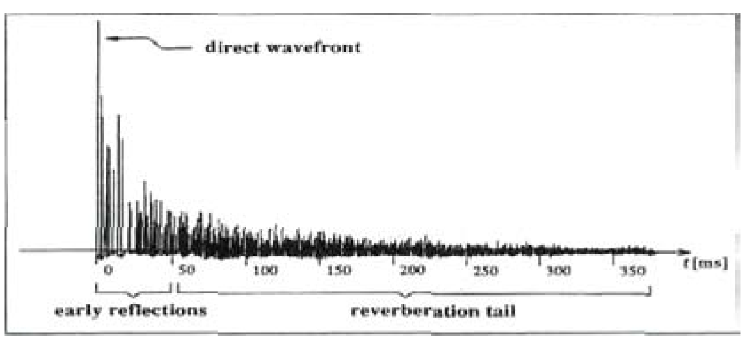

Figure 1. The phenomenon of reverberations in an cnloslure.

exceeding this threshold generally have a negative effect on speech intelligibility, because they result in a merging of basic speech sounds.

On the contrary, for delays shorter than 50 ms, reflections ("early reflections") are perceived as part of the direct sound. Their effect on sound perception is an increase in loudness and a spectral alteration, known as "coloration", that characterizes the "acoustics" of each room.

This research paper is organized as follows. Section 1 gives an introduction to acoustics in general and room acoustics in particular. Section 2 discusses a convenient method of modeling the acoustic echo generating system. Section 3 gives an introduction to linear model structures. Section 4 describes orthonormal functions and rationale behind using them in modeling. A novel multistage hybrid model is also suggested. Section 5 gives the optimization techniques to be implemented on orthonormal models to reduce the number of estimation parameters. Section 6 describes the experimental setup and Section 7 contains some concluding remarks.

2. MODEL OF THE ACOUSTIC ECHO GENERATING SYSTEM

The echo generating system can be approximately described by the linear model structure. To deal with the problem of estimating the acoustic echo path, we first assume that the echo path is linear and stationary for a short time interval. Although there exist some nonlinearity characteristics presented in the system, the assumption of linearity still gives rise to satisfied results depending on an amount of non-linearity of the system. It is difficult to gain a deep insight about the echo generating system for each specific purpose. Consequently, it is suitable to model the system by using a black box model approach which is mainly based on observed input signal $u(t)$ and output signal $y(t)$. The task is to estimate the impulse response of the unknown system (echo path), since if the system is linear, its impulse response will then completely represent it [4].

Supposing that the measured output signal is given by the echo $r(t)$ corrupted by the near-end disturbance $x(t)$, i.e.,

$$y(t) = r(t) + x(t), \quad \text{where} \quad r(t) = H_e * u(t)$$

where * denotes the convolution operation, and H_e is the transfer function of the true echo path. The echo generating system is shown in Fig. 2.

Since one cannot know the true echo signal, the purpose of modeling is to find out the model that can properly reproduce the true echo signal, given the input signal. It is not possible to exactly obtain

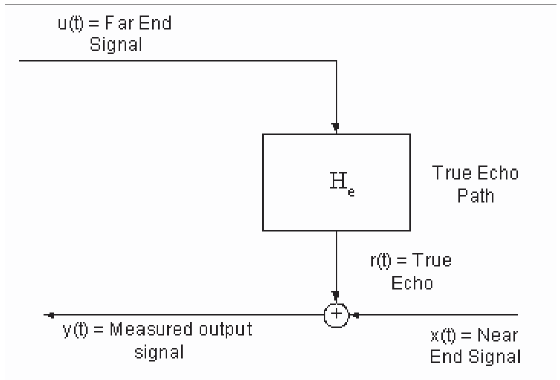

Figure 2. The echo generating system.

the model of true system H_e. Hence, it is appropriate to introduce modeling error, $\varepsilon_m(t)$, in an estimated echo signal, $\hat{r}(t)$ to represent the true echo signal as shown in Fig. 3. This modeling error can be used to determine how fit the estimated model, H, could be. If $\varepsilon_m(t)$ is equal to zero, then we have succeeded in modeling the true echo path.

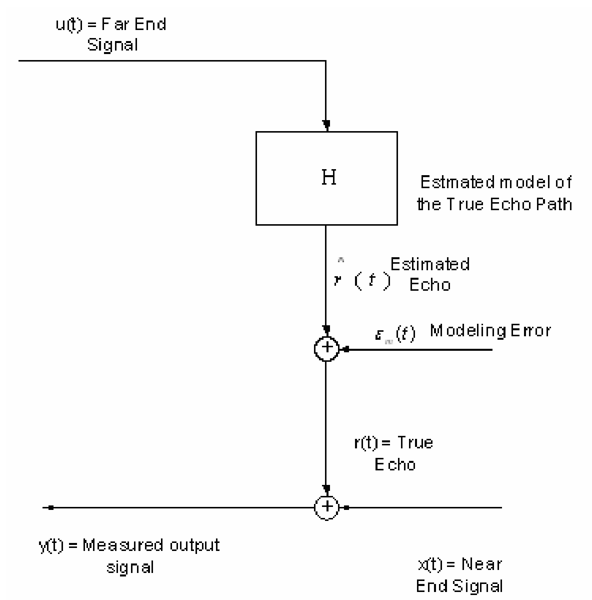

Figure 3. An estimated model H with modeling error.

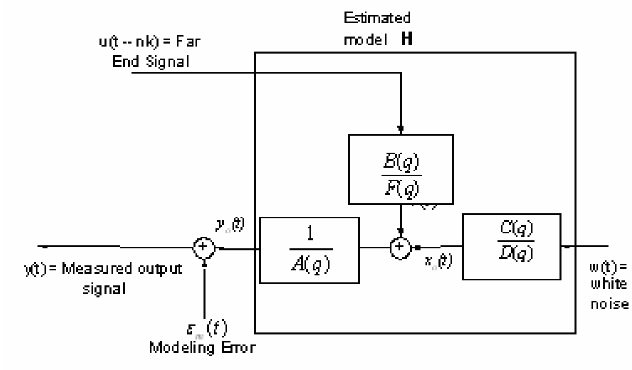

Figure 4. A general linear model structure.

3. LINEAR MODEL STRUCTURES

Again, the stable LTI system can be adequately modeled by a linear black box model structure. Its graphical representation associated with the echo generating system is depicted in Fig. 2, where $y_0(t)$ is considered to be the optimum estimated output that can be obtained within a specific model class and a fixed model size [4]. Note that, $A(q)$ represents the poles that are common between the input and the noise. Similarly, $F(q)$ corresponds to the poles that are unique for the input, whereas $D(q)$ determines the poles that are unique for the noise. The advantage of this model is that the effect of disturbance is also taken into consideration to obtain a more accurate result. That is, since the actual near-end disturbance $x(t)$ is unknown, it is often more convenient to model $x(t)$ as being obtained by filtering a white noise source $w(t)$ through a linear filter $[C(q)/(D(q)A(q))]$. The modeling error $\varepsilon_m(t)$, represents the inability of the proposed model structure to correctly describe the relationship between the input and the output. Herein, three basic linear model structures FIR, ARX and OE are presented.

3.1. The Finite Impulse Response Model Structure (FIR)

FIR model is a widely used model structure to describe the system of interest due to its simplicity and stability [5]. It is a special case of Eq. (1).

When $A(q) = C(q) = D(q) = F(q) = 1$ which can be expressed as

$$y_0(t) = B(q)u(t - nk) + w(t)$$

$$= \sum_{k=1}^{nb} b_k q^{-k} u(t-nk) + w(t)$$
$$= \left(b_1 q^{-1} + b_{2q} q^{-2} + \cdots + b_{nb} q^{-nb}\right) u(t-nk) + w(t)$$
$$= b_1 u(t-nk-1) + b_2 u(t-nk-2) + \cdots + b_{nb} u(t-nk-nb) + w(t)$$

where the parameters, b_k for $k = 1, 2, \ldots, nb$, represent the magnitude of truncated system impulse response, and the signal $w(t)$ is assumed to be uncorrelated with the input signal $u(t)$. Fig. 5 illustrates the FIR model.

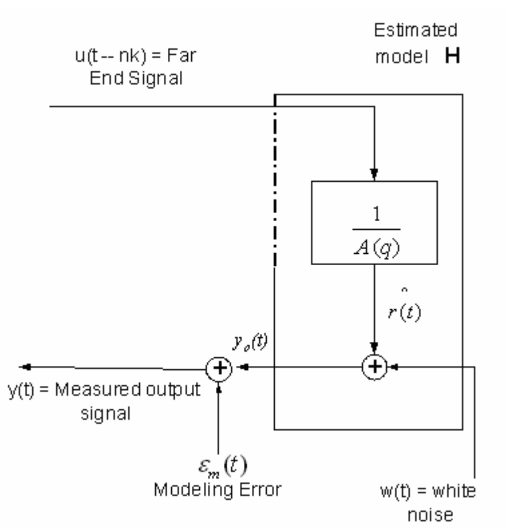

Figure 5. The FIR model structure.

Observe that, the signal $w(t)$ can be perceived as the nearend speaker's speech signal and/or unknown disturbance.

In the view of the prediction error, the predicted signals are mainly based on the old inputs. They are usually called the regressors of the model and can be collected in a regression vector, $\varphi(t) = [u(t-nk-1) \;\; u(t-nk-2) \;\; \cdots \;\; u(t-nk-nb)]^T$. The parameter vector to be estimated can be written as $\theta = [b_1 \;\; b_2 \;\; \cdots \;\; b_{nb}]^T$.

Obviously, FIR model is suitable to represent the echo path because it does not model the signal $w(t)$ which is considered as the desired signal in this thesis project. However, it leads to a large number of estimated parameters in order to obtain an accurate approximation.

4. ORTHONORMAL MODELS

In theory, the number of parameters employed in the filter depends upon the echo path delay and the length of the impulse response of an echo path [9]. For the echo canceller to work suitably, the number of parameters should have a length greater than the combined effect of the true echo path's impulse response length and the echo path delay. Let T_s be the sampling period of the digitized speech signal, M be the number of parameters used in the filter, and τ be the combined effect to be considered. Consequently, one obtain

$$MT_s > \tau$$

Since the typical value of T_s is $125\,\mu$s for a standard digital communication system, it is obvious that, if $\tau = 500\,\mathrm{ms}$, $M > 4000$ parameters are required in the filter. The traditional approaches, as discussed earlier, for modeling the system of interest lead to the approximation of very high order in case of rapid sampling and/or dispersion in time constant, which is closely related to the dominating pole of a true system. Such a high model order cannot be acceptable in practice due to some difficulties in terms of performance and hardware complexity. By exploiting a priori information about the dominating pole of the system, more appropriate series expansions related to the use of orthonormal basis functions are proposed [4].

These orthonormal functions are constructed by orthonormalizing a given set of exponential functions [10]. They are orthogonal in $L_2(0,\infty)$, and form a complete set in $L_2(0,\infty)$, and $L_1(0,\infty)$. Laguerre and two-parameter Kautz (or, more popularly, just "Kautz") functions are all special cases of orthonormal bases. Laguerre function is suitable for the system with dominant first-order dynamics, whereas Kautz function is appropriate for the system with dominant second-order resonant dynamics [10, 11]. It has been found out [11] that the model order can be substantially reduced when the dominating pole is chosen suitably. In this section, Kautz functions are presented. The Kautz model can be easily implemented with FIR and ARX structures by just replacing a traditional delay operator of these structures with Kautz function. These functions can be generalized, to a family of rational orthonormal basis functions for the Hardy space H_2 of stable linear dynamic systems.

Consider a stable LTI system which is given by

$$y(t) = G(z,\theta)u(t-nk) + H(z,\theta)w(t)$$

where $y(t)$, $u(t)$, and nk are the output, input, and delay respectively [5]. The noise $w(t)$ is assumed to be a stationary process with zero mean and unit variance. $G(z, \theta)$ and $H(z, \theta)$ are the set of transfer function parameterized by the parameter vector θ for the input and the noise, where z is the unit step delay $(z^{-1}u(t) = u(t-1))$. Given N samples of observed data, by t we means the time at the sampling instants $t = kT$, for $k = 1, 2, \ldots, N$, and T is assumed to be equal to 1 for simplicity [7].

By assuming that $H(z, \theta)$ is minimum phase to guarantee that $H(z, \theta)^{-1}$ is stable, and normalized $(H(\infty, \theta) = 1)$, the optimal one-step ahead predictor of $y(t)$ is thus defined as

$$y(t, \theta)[H(z, \theta)^{-1}G(z, \theta)]u(t - nk) + [1 - H(z, \theta)^{-1}]y(t) \qquad (1)$$

We shall assume that $H(z, \theta)$ and $G(z, \theta)$ have the same unstable poles. It then follows that

$$H(z, \theta)^{-1}G(z, \theta) = \sum_{k=1}^{\infty} b_k \beta(z) \qquad (2)$$

$$1 - H(z, \theta)^{-1} = \sum_{k=1}^{\infty} b_k \beta(z) \qquad (3)$$

where the basis function $\beta_k(z) = z^{-k}, \ |z| \geq 1$, known as the delay operator [7]. For a stable LTI system, it is required that

$$\sum_{k=1}^{\infty} |b_k| < \infty \quad \text{and} \quad \sum_{k=1}^{\infty} |a_k| < \infty$$

Practically, since both b_k and a_k tend to zero as $k \to \infty$, we may truncate these expansions at $k = n$ to adequately approximate the system with the finite number of estimated parameters.

Hence, Eq. (1) can be rewritten in a truncated series expansion as

$$\hat{y}(t, \theta) = \sum_{k=1}^{n} b_k \beta_k(z) u(t - nk) + \sum_{k=1}^{n} a_k \beta_k(z) y(t) \qquad (4)$$

By taking the parameter vector $\theta = [a_1 \ a_2 \ \cdots \ a_n \ b_1 \ b \ \cdots \ b_n]^T$, the well-known ARX model structure is derived [8]. Setting $a_k = 0$, for all k, leads to FIR model structure, or by taking $b_k = 0$, for all k, AR model structure is realized. In theory, the usefulness of the estimate is limited by how fast the sums in Eq. (2) and Eq. (3) converge, i.e., the rate of the error terms

$$\sum_{k=n+1}^{n} |\beta_k| \quad \text{and} \quad \sum_{k=n+1}^{n} |a_k|$$

tend to zero as $n \to \infty$. The rate of convergence is basically determined by the location of the poles of $G(z, \theta)$ and the zeros of $H(z, \theta)$. Poles and zeros close to unit circle imply a slow rate of convergence. If $G(z, \theta)$ and $H(z, \theta)$ are obtained by sampling a continuous time system, using a sampling interval T, the continuous time poles and zeros, λ, are approximately mapped to the discrete time poles and zeros at $\{e^{\lambda T}\}$ for small T. Since most digital applications require a high sampling rate, i.e., $T \to 0$, one could get a serious problem of estimating this system due to a very slow rate of convergence because the discrete time poles and zeros approach one.

Additionally, because the variance of an estimated model is proportional to the number of estimated parameters, it is advantageous to use as few parameters as possible but still guarantee an useful model. These problems have motivated to use an alternative operator which is less sensitive to the location of poles and zeros. This is a consequence of the fact that the delay operator has too short memory (only one sampling step). By introducing the operator with longer memory, the number of estimated parameters necessary to obtain an accurate approximation can then be reduced [10]. This operator can be chosen as

$$\beta_k(z) = \frac{1}{z - \xi_k} = z^{-1} + \xi_k z^{-2} + \xi_k^2 z^{-3} + \xi_k^3 z^{-4} + \cdots$$

where the poles ξ_k are chosen according to a priori knowledge of the true system. It is desirable to construct the orthonormal basis function based on this operator. The procedure of a unifying construction of orthonormal bases can be seen in [8] and the result is given by

$$\beta_k(z) = \left[\frac{\sqrt{1 - |\xi_k|^2}}{z - \xi_k}\right] \prod_{j=1}^{k-1} \left[\frac{1 - \overline{\xi}_j z}{z - \xi_j}\right], \quad \text{for} \quad k \geq 1 \tag{5}$$

where a variety of poles at $\{\xi_1,\ \xi_2\ \cdots\ \xi_k\}$ are incorporated. These basis functions are orthonormal with respect to the following inner product [8].

$$\langle \beta_i(z), \beta_j(z) \rangle = \frac{1}{2\pi j} \oint_{\Gamma} \beta_i(z) \beta_j(z) \frac{dz}{z} = 0, \quad \text{for} \quad i \neq j$$

and have unit normalization

$$\|\beta_k(z)\|^2 = 1, \quad \text{for } \forall k$$

Several advantages are attained from employing the orthonormal basis functions in system identification problem [13]. Firstly, they have transforms that are rational functions with a very simple repetitive

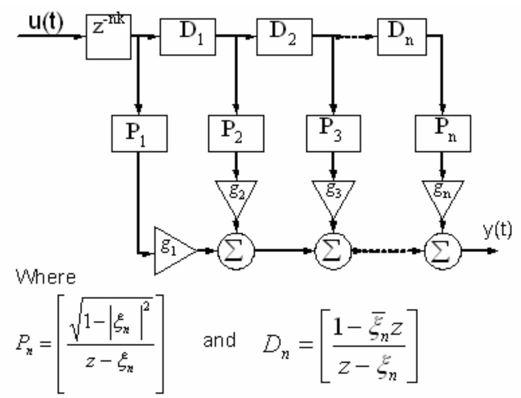

Figure 6. Discrete-time orthonormal networks of order n.

form. This allows their practical realization with concatenated blocks as illustrated in Fig. 6. Secondly, the solution of transfer function estimation problems leads to the normal equation having a diagonal structure. However, it holds only if the input is white sequences. Furthermore, if the normal equation has a Toeplitz structure [6], much is gained in terms of numerical algorithm, sensitivity, etc. Finally, as seen in Eq. (5), the orthonormal basis function consists of a first order low-pass term and $k-1$ all-pass terms. Such all-pass filters are favorable in terms of numerical sensitivity, and they are often recommended to use in filter design. FIR, Laguerre and Kautz functions are all special cases of orthonormal basis functions [12]. Manifestly from Eq. (5), FIR function is realized by choosing $\xi_k = 0$ for all k. Kautz function is generalized when the pole is chosen such that $\xi_k = \xi$, where ξ is complex valued and $|\xi| < 1$, for all k.

System identification deals with the problem of finding the estimate of $G^0(z)$ from the experimental data of $\{y(t),\ u(t)\}$, where $t = 1, 2, \ldots, N$. Theoretically, $G^0(z)$ can be represented by an infinite number of given basis functions. In practice, a truncated series expansion, say the nth order, is used to estimate $G^0(z)$ with the result that there is a truncation error. This error can be minimized by a proper choice of basis function. In addition, the identification problem can be simplified to a linear regression estimation problem if the model structure is a priori linear in the parameters, i.e., if the model can be expressed as

$$G(z) = \sum_{k=1}^{n} g_k \beta_k(z)$$

where $\beta_k(z)$ is a set of given basis functions and gk are the parameters to be estimated. The least squares estimation method is then applied to find the optimal values of the model parameters of lower order terms. Since the proper choice of basis function, $\beta_k(z)$, will give rise to a considerable increase in the rate of convergence and a significant decrease in the asymptotic variance of the estimated transfer function, this leads to accurate approximation with less number of estimated parameters. Herein, we shall a special case of orthonormal basis functions, namely (two-parameter) Kautz functions.

The problem in system identification using Kautz function has been studied in [12]. Since the system in real applications is much more complicated than modeling it by using one real-valued pole, it is necessary to use a complex-valued pole to characterize the resonant part of the system. Since as soon as one pole, say ξ_j is chosen as a complex pole, then the impulse responses for the $\beta_k(z)$ for $k > j$ become complex, and this is physically unacceptable in a system identification problem. To avoid this problem, one needs to include its complex conjugated pole automatically whenever one complex-valued pole is chosen in Kautz basis function. The detail of deriving Kautz basis function can be seen in [8]. Given the nth model order, the truncated Kautz series expansion can be written as

$$G(z) = \sum_{k=1}^{n/2} \{g_{2k-1}\Psi_{2k-1}(z) + g_{2k}\Psi_{2k}(z)\}$$

where the Kautz basis function is given by

$$\Psi_{2k-1}(z) = K_k^{(1)} \prod_{i=1}^{k-1} \Lambda_i(z) \tag{6}$$

$$\Psi_{2k}(z) = K_k^{(2)} \prod_{i=1}^{k-1} \Lambda_i(z) \tag{7}$$

where

$$K_k^{(1)} = \frac{\sqrt{(1-\gamma_k^2)\,(z-\alpha_k)}}{z^2 - \alpha_k(\gamma_k+1)z + \gamma_k} \quad \text{and} \quad K_k^{(2)} = \frac{\sqrt{(1-\gamma_k^2)\,(z-\alpha_k^2)}}{z^2 - \alpha_k(\gamma_k+1)z + \gamma_k},$$

$$\Lambda_i(z) = \left(\frac{\gamma_i z^2 - \alpha_i(\gamma_i+1)z + 1}{z^2 - \alpha_i(\gamma_i+1)z + \gamma_i}\right), \quad \text{for } i = 1, 2, \ldots, k-1$$

$$\alpha_k = \frac{\xi_k + \overline{\xi}_k}{1 + |\xi_k|^2} \text{ and } \gamma_k = |\xi_k|^2, \text{ for which } |\alpha_k| < 1 \text{ and } |\gamma_k| < 1,$$

and ξ_k is a complex-valued pole. In this research paper, we shall consider only two-parameter Kautz functions where ξ_k is chosen equal to (complex-valued) ξ, for all k [14, 15]. On account of an orthonormal property, the Kautz parameters $\{g_{2k-1},\ g_{2k}\}$ can be determined by the projection on the basis functions, i.e., $g_{2k-1} =< G(z), \Psi_{2k-1}(z) >$ and $g_{2k} =< G(z),\ \Psi_{2k}(z) >$. The discrete Kautz network can be simply represented in concatenated blocks as illustrated in Fig. 7.

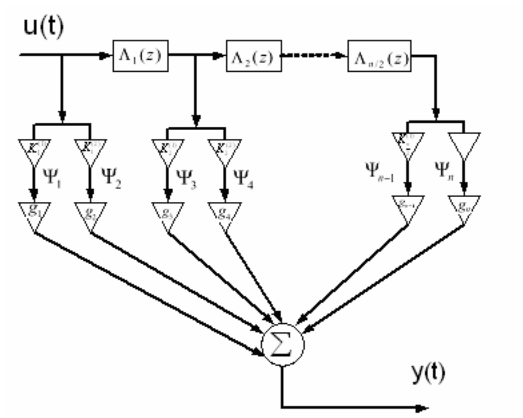

Figure 7. Discrete-time Kautz network of order n.

Orthonormal model is extended and generalized to Kautz model which can cope with several different possible complex poles [15].

4.1. Two-Stage Impulse Response Model

The typical echo impulse response can be decomposed into 2 parts, namely the first part which normally has a rapid time variation, and the second part known as the tail of the impulse response which is slowly decaying towards zero.

By exploiting the characteristics of such an echo impulse response, an approach of using a two-stage echo canceller is introduced [16]. Its purpose is to reduce the number of estimated parameters as well as the computational complexity of an echo canceller algorithm, particularly in case of the echo response with a long tail. This can be done by dividing the structure of an echo canceller into two stages. The first stage is a conventional transversal FIR filter that spans for the first few parameters of the echo impulse response, while the second stage approximates the remainder or tail of the response by a linear combination of orthonormal functions. In other words, the tail may be well approximated by combining a few parameters, whereas a

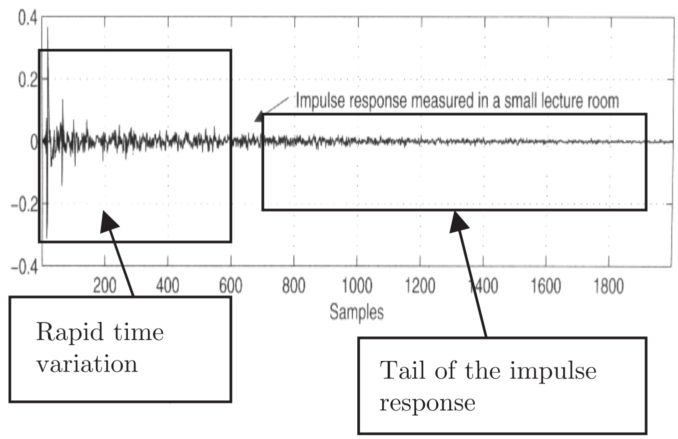

Figure 8. Dissection of impulse respone in two parts.

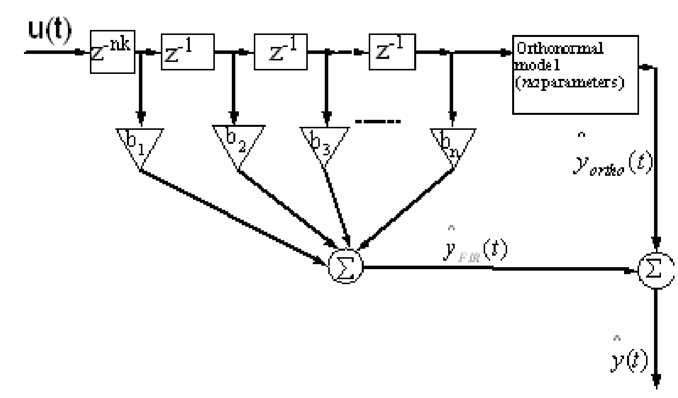

Figure 9. Hybrid two stage model for system identification.

conventional transversal FIR filter representation will involve far more parameters. As a result, if the number of estimated parameters in an orthonormal model is not large, a substantial reduction in the model order and the computational complexity can be achieved [17].

Fig. 9 shows the two-stage echo canceller which is divided into two parts. The first is a conventional transversal FIR filter with nb tap parameters, b_k for $k = 1, 2, \ldots, nb$, which approximate the first nb of the first part of the echo impulse response. The second is an orthonormal model with na parameters, g_k for $k = 1, 2, \ldots, na$, which are used to approximate the tail portion of the echo response.

5. OPTIMIZATION

Any stable LTI system can be modeled by an infinite series of orthonormal functions which involve a free parameter [19], closely related to the dominating pole. Theoretically, when infinitely many parameters are employed in the expansion, the choice of a dominating pole is somewhat arbitrary. In practice, however, a truncated series expansion is used and results in the truncation error. This error is basically a function of the model order and its dominating pole. For a fixed model order, there exists an optimal dominating pole that minimizes the truncation error [19].

The input-output relation of an estimated model of order n be represented by

$$\hat{y}(t) = \left(\sum_{k=1}^{n} g_k \beta_k(z)\right) u(t-nk) = \varphi^T(t)\theta \tag{8}$$

where $\beta_k(z)$ corresponds to a set of chosen basis function, i.e., Kautz function,

$$\begin{aligned} \varphi(t) &= [\beta_1(z)u(t-nk)\beta_2(z)u(-nk)\cdots\beta_n(z)u(t-nk)]^T \text{ and} \\ \theta &= [g_1\ g_2\ \cdots\ g_n]^T. \end{aligned}$$

Consequently, the least squares solution of $\hat{\theta}$ defined as the parameter $\hat{\theta}$ that by minimizing the loss function as given by

$$v(\theta) = \frac{1}{N}\sum_{t=1}^{N}(y(t) - \varphi^T(t)\theta)^2 \tag{9}$$

Now let us introduce the following notation:

$$\begin{aligned} \hat{d}_j(\xi) &= \frac{1}{N}\sum_{t=1}^{N} y(t)\beta_j u(t-nk), \quad \text{for } 1 \le j \le n \\ \hat{\gamma}(\xi) &= \left(\hat{d}_1(\xi), \hat{d}_2(\xi), \ldots, \hat{d}_n(\xi)\right)^T \\ \hat{\Gamma}_n(\xi) &= \frac{1}{N}\sum_{t=1}^{N} \varphi(t)\varphi^t(t) \end{aligned}$$

Given a stable LTI system, the loss function will be a function of the model order and its dominating pole. For a fixed model order, there exists an optimal dominating pole that minimizes the loss function.

The optimal dominating pole is therefore chosen as the one that minimizes the Eq. (7), which is simply equivalent to [19]

$$\xi_{opt} = \underset{\xi}{\operatorname{argmin}} \left[\frac{1}{N} \sum_{t=1}^{N} y^2(t) - \hat{\gamma}_n(\xi)^T \hat{\Gamma}_n(\xi)^{-1} \hat{\gamma}_n(\xi) \right] \tag{10}$$

or equally

$$\xi_{opt} = \underset{\xi}{\operatorname{argmax}} \left[\hat{\gamma}_n(\xi)^T \hat{\Gamma}_n(\xi)^{-1} \hat{\gamma}_n(\xi) \right] \tag{11}$$

The solution in equations (10) and (11) can be either a real value or a complex value. Accordingly, this method will be applied to find the optimal dominating pole for both Laguerre and Kautz models. Nevertheless, the drawback of this method is that we need to use the search method to find the optimal dominating pole at each model order. It is done by varying the value of ξ within unit circle in order to retain the stability of a model, and choosing the pole that yields the maximum value in Eq. (11). Since the matrix size of each element in Eq. (10) will be proportional to the number of estimated parameters in orthonormal model, if one is required to find the optimal dominating pole at very high order, it will take long time for solving Eq. (10). Therefore, such wasted time can be viewed as the computational complexity, i.e., the larger the matrix size, the longer the wasted time, and then the higher the computational complexity [17]. Evidently, this method requires much higher computational complexity than the first method. As a result, to avoid the computational complexity of calculating the optimal dominating pole at very high order, one need to subdivide the whole interval of the pole into smaller subintervals of more manageable size.

6. PROCESS OF GENERATING AND COLLECTING THE DATA

Since the real acoustic echo data employed in this research paper has been collected from the system that is made as near stationary as possible in order to be able to use the offline method to estimate the values of the model parameters, it is noteworthy to explain the process of generating and collecting the real acoustic echo data. Such a process can be represented in Fig. 10. The microphone is located approximately 80 cm from the loudspeaker of high quality. The environment in which the recording is made is a hard acoustical room.

The dimension of the room is approximately 30 m by 20 m and the height is 20 m.

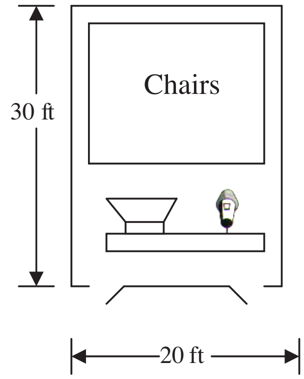

Figure 10. Settings for the experiment to collect real acoustic data in Lecture Room #3 on ground floor of EED,UET, Taxila.

Model Order	FIR	Orthonormal Model	
		Fit of errorr	DB Improvement
2	0.2094	0.2094	0
10	0.1463	0.1371	0.5641
5	0.1195	0.1192	0.0218
100	0.1174	0.1172	0.0148
200	0.11	0.1098	0.0158
300	0.1049	0.1042	0.0582
400	0.0991	0.0985	0.0527
500	0.0928	0.0925	0.0281

Kautz		
Fit of error	Pole	DB improvement
0.1895	0.80+0.50i	0.8673
0.1345	0.14+0.30i	0.7304
	0.53+0.45i	0.073
0.1145	0.56+0.59i	0.2173
0.1063	0.44+0.64i	0.2972
0.0994	0.41+0.50i	0.4678
0.0931	0.35+0.39i	0.5425
0.0882	0.35+0.31i	0.4416

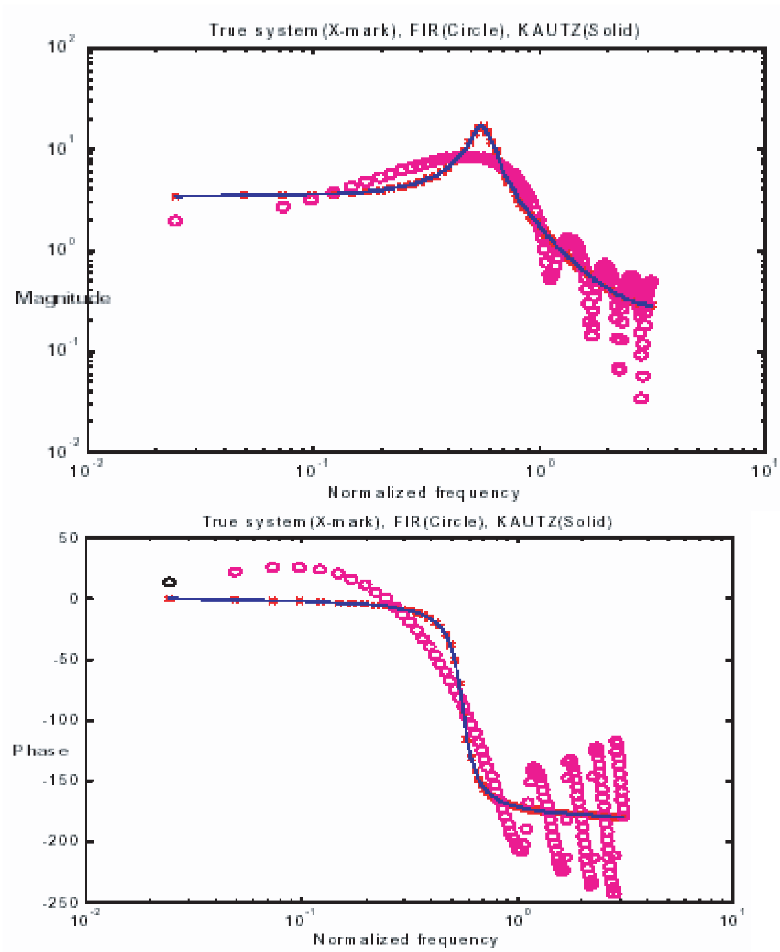

Figure 11. Bode plots (Magnitude and Phase) of transfer function approximations, where True system (X-mark), FIR (Circle) and Kautz (Solid).

The signal (white noise) of 10 ms is fed to the amplifier before being collected at the recorder as the input signal, $u(t)$, and transmitted to the loudspeaker. The signal is propagated in the room and then reflected back to the microphone as acoustic echo which is collected by the same recorder as the output signal, $y(t)$. All signals collected at the recorder are sampled at 12 kHz. Nothing is moved in this room.

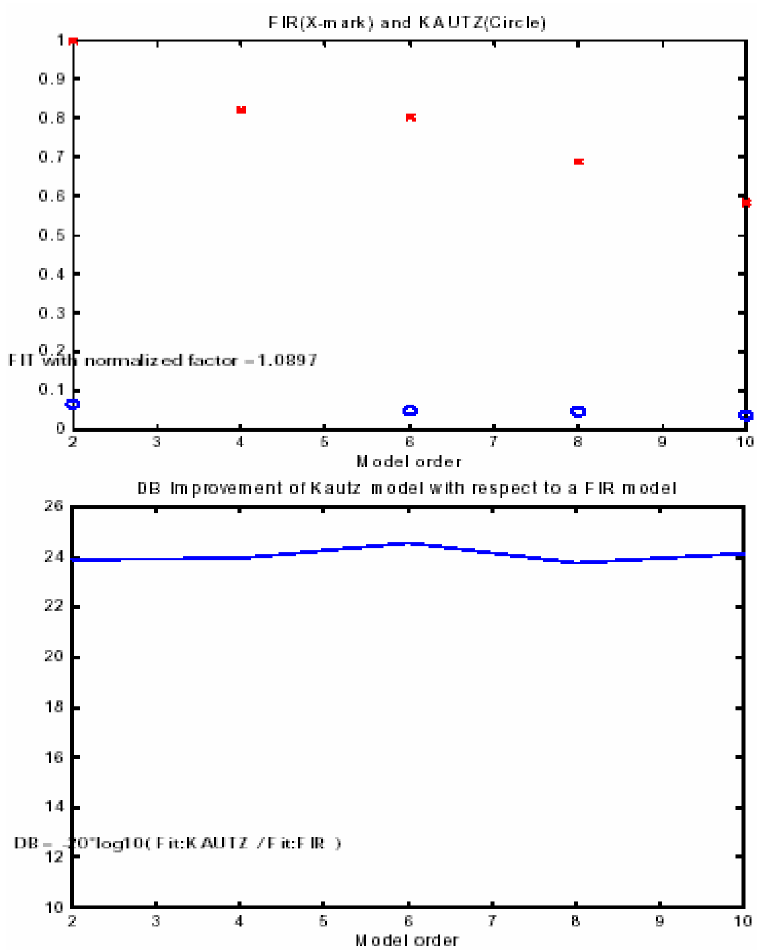

Figure 12. Left: Fits of error between FIR model (X-mark) and Kautz model (Circle). Right: The DB improvement of Kautz model with respect to a FIR model at each model order.

Firstly we shall investigate how the segmented data used for estimation and validation affects the resulting optimal dominating pole. Three different segments of data are randomly chosen. We shall use the segment of estimation data at $ZE = 10000 : 17999$ and validation data at $ZV = 20000 : 27999$ as the reference set.

Fits of error as the reference set when $ZE \neq ZV$ and the data length of 8000 samples.

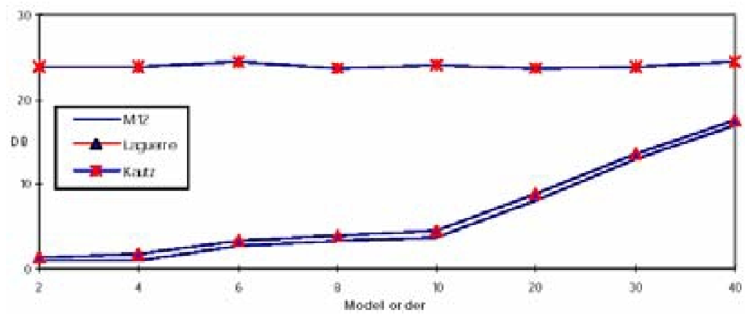

Figure 13. The DB improvement of the proposed models with respect to a FIR model at each model order.

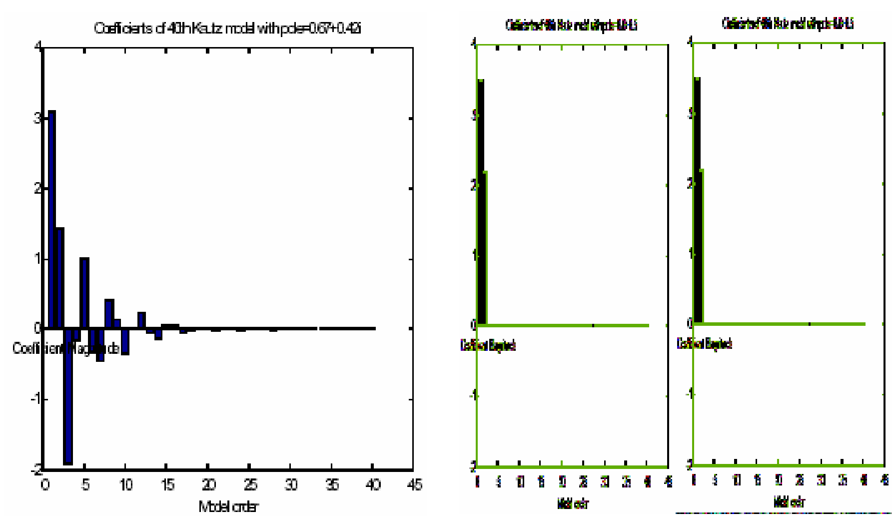

Figure 14. The coefficient magnitude of the Kautz model with a pole of 0.67+0.42i, and the Kautz model with a pole of 0.8+0.5i.

7. CONCLUSIONS

The effect of acoustic echo is quite complicated in a closed room. FIR model is widely used to model this room impulse response due to simplicity and stability. However, it leads to the approximation of very high order, probably in the order of 4000 or may be more with larger reverberation T_{60} time. By exploiting a priori information about the dominating pole of the system, an approximation of the room impulse response by means of the Kautz functions is proposed. We have also presented methods to find such an optimal dominating pole.

The changes in the echo path impulse response cause an increase in the residual echo error signal. This forces the acoustic echo canceller to start adapting to the new impulse response and it can even diverge, if the changes are fast or abrupt. In the installed phone, the speaker and microphone should not be directed to the path that is subject to fast changes. It is usually better to direct the speaker and microphone towards the ceiling since this echo path changes rarely. So the offline system identification techniques can be employed because the system in these condition is practically time invariant.

REFERENCES

1. Beranek, L. L., *Acoustics*, Acoustical Society of America, Woodbury, NY, 1993.
2. Morse, P. M. and K. U. Ingard, *Theoretical Acoustics*, Princeton University Press, Princeton, NJ, 1987.
3. Cremer, L. and H. A. Muller, *Principles and Applications of Room Acoustics*, Vols. 1 and 2, Applied Science Publishers, Essex, Great Britain, 1982 (translated by T. J. Schultz).
4. Gritton, C. W. K. and D. W. Lin, "Echo cancellation algorithms," *IEEE Acoustics, Speech and Signal Processing Magazine*, Vol. 1, No. 2, 30–38, April 1984.
5. Haneda, Y., S. Makino, and Y. Kaneda, "Common acoustical pole and zero modeling of room transfer functions," *IEEE Transactions on Acoustics, Speech and Signal Processing*, Vol. 2, No. 2, 320–328, April 1994.
6. Golub, G. H. and C. F. Van Loan, *Matrix Computations*, The Johns Hopkins University Press, Baltimore, MD, 1996.
7. Ljung, L., *System Identification: Theory for the User*, Prentice-Hall, Englewood Cliffs, NJ, 1999.
8. Mourjopoulos, J. and M. A. Paraskevas, "Pole and zero modeling of room transfer function," *Journal of Sound and Vibration*, Vol. 146, No. 2, 281–302, April 1991.
9. Kuttru, H., *Room Acoustics*, Elsevier Applied Science, London, Great Britain, 1991.
10. Ninness, B. and F. Gustafsson, "A unifying construction of orthonormal bases for system identification," *IEEE Transactions on Automatic Control*, Vol. 42, No. 4, 515–521, April 1997.

11. Davidson, G. W. and D. D. Falconer, "Reduced complexity echo cancellation using orthornormal functions," *IEEE Transactions on Circuits and Systems*, Vol. 38, No. 1, 20–28, Jan. 1991.
12. Wahlberg, B. and P. M. Makila, "On approximation of stable linear dynamical systems using Laguerre and Kautz functions," *Automatica*, Vol. 32, No. 5, 693–708, May 1996.
13. Heuberger, P. S. C., P. M. J. Van den Hof, and O. H. Bosgra, "A generalized orthonormal basis for linear dynamical systems," *IEEE Transactions on Automatic Control*, Vol. 40, No. 3, 451–465, March 1995.
14. Makila, P. M. and J. R. Partington, "Laguerre and Kautz shift approximations of delay systems," *International Journal of Control*, Vol. 72, No. 10, 932–946, July 1999.
15. Silva, T. O., "Optimality conditions for truncated Kautz networks with two periodically repeating complex conjugate poles," *IEEE Transactions on Automatic Control*, Vol. 40, No. 2, 342–346, Feb. 1995.
16. Kaneda, Y., "A study of non-linear effect on acoustic impulse response measurement," *Journal of Acoustical Society of Japan*, Vol. 16, No. 3, 193–195, May 1995.
17. Carayannis, G., D. G. Manolakis, and N. Kalouptsidis, "A fast sequential algorithm for least-squares filtering and prediction," *IEEE Transactions on Acoustics, Speech and Signal Processing*, Vol. 31, No. 6, 1394–1402, Dec. 1983.
18. Tanguy, N., R. Morvan, P. Vilbe, and L. C. Calvez, "Improved method for optimum choice of free parameter in orthogonal approximations," *IEEE Transactions on Acoustics, Speech and Signal Processing*, Vol. 47, No. 9, 2576–2578, Sep. 1999.
19. den Brinker, A. C., "Optimality conditions for a specific class of truncated Kautz series," *IEEE Transactions on Circuits and Systems-II*, Vol. 43, No. 8, 597–600, Aug. 1996.
20. Murano, K., S. Unagami, and F. Amano, "Echo cancellation and applications," *IEEE Communication Magazine*, Vol. 28, No. 1, 49–55, Jan. 1990.