Estimation of acoustic resonances for room transfer function equalization

Pepe Gil-Cacho,
Toon van Waterschoot
and Marc Moonen.
Katholieke Univeristeit Leuven
ESAT-SCD, Kasteelpark Arenberg 10
B-3001 Leuven, Belgium.

Søren Holdt Jensen.
Aalborg University
Dept. Electronic Systems
Niels Jernes Vej 12, DK-9220.
Aalborg, Denmark

***Abstract*—Strong acoustic resonances create long room impulse responses (RIRs) which may harm the speech transmission in an acoustic space and hence reduce speech intelligibility. Equalization is performed by cancelling the main acoustic resonances common to multiple room transfer functions (RTFs), i.e., common-acoustical-poles, in the room. This paper discusses the utilization of different norms (i.e., 2-norm and 1-norm) and models (i.e., all-pole and pole-zero) for RTF modelling and then equalization. Acoustic resonances may be modelled by means of the poles of the RTF. In the literature, however, it is not clear what model (i.e., all-pole or pole-zero) generates pole estimates that perform better in RTF equalization. Furthermore, least squares error (i.e., 2-norm) minimization is typically employed for the estimation of the poles. In this paper a least absolute error (i.e., 1-norm) minimization is further proposed for pole estimation. A comparative evaluation of these different norms and models in terms of their residual RTF and residual RIR (i.e., the residuals after equalization) is provided.**

I. Introduction

The sound transmission characteristics between a loudspeaker and a microphone are described in the frequency domain by the room transfer function (RTF) or in the time domain by the room impulse response (RIR) which depends on the loudspeaker-microphone position. In some audio applications (e.g., sound reproduction systems in train stations or other large spaces) where speech intelligibility is an issue, an equalization filter is commonly used to compensate for the frequency response of the room. Such a filter may remove, by inverse filtering, acoustical artefacts present in spaces with long RIR originating from strong acoustic resonances. The equalization performance then depends on the model from which the inverse filter is derived. The most common model for the RTF is an *all-zero* model [1]. It represents the physics of room acoustics where a microphone signal is a weighted sum of discrete reflections of the loudspeaker signal. Its drawback is that any change in the loudspeaker-microphone or obstacle position inside the room will change all the coefficients of the model. The equalization filter is then a single point inverse filter which is only valid for a single point in the room and therefore a recalculation of every coefficient will be needed at each and every other point in the room [6].

One of the alternatives is to use a *pole-zero* model for the RTF [1]. This model represents the physics of room acoustics by also including the modelling of the acoustic resonances by means of the poles of the transfer function. Poles can represent long impulse responses caused by resonances, while zeros represent time delays and antiresonances [7]. Yet another alternative model is the *all-pole* model, which represents only the acoustic resonances in an effort to model the RTF spectral envelope [1]. This provides an appropriate strategy to cancel only the main resonances discarding the cancellation of the zeros. Moreover, the concept of common acoustical poles, first introduced in [7], can be applied to these two alternative models. The underlying idea is that the acoustic resonances in a room depend on the dimensions and shape of the enclosure and not on the loudspeaker-microphone position. Each RTF in the room may then be expressed using a common set of poles and different zeros. Hence an inverse filter that cancels the common acoustical poles can be created that equalizes the physically common main resonances in multiple points in the room [6]. Finding the poles of the transfer function may be seen as an optimization problem in which the set of poles that minimizes the error between the model and the measurements (i.e., the actual impulse response) is sought for. Different models and error criteria will obviously render different pole estimates and consequently different equalization filters and residual RTFs. Typically a least squares error criterion (2-norm minimization) is employed. In the literature, however, it is not clear which of these models generate pole estimates that perform better in the RTF equalization contest.

In this paper the use of a least absolute error criterion (1-norm minimization) for RTF pole estimation is proposed. In [2] sparse linear prediction of speech signals is used to obtain sparse residuals with minimum number on non-zeros elements. With this idea in mind, 1-norm minimization is proposed here to calculate the poles of an all-pole model, so that the inverse filtering will render a residual impulse response having discrete separated reflection rather than a dense impulse response. On the other hand, it is found that the poles calculated from a pole-zero model using 1-norm minimization gives results similar to the 2-norm pole-zero approach and therefore this will not be considered any further. The aim of this paper is hence to make a comparative evaluation involving 2-norm minimization using an all-pole and a pole-zero model and 1-norm minimization using an all-pole model. The comparative evaluation will be presented both in time and frequency domain. Several questions on how the choice of a model and norm affects the residual RTF and RIR will be addressed.

The paper is organized as follows. In section 2, the mathematical formulation is presented for each model. In section 3, equalization results using the presented techniques on real measured room impulse

responses are presented. Finally, section 4 concludes the paper.

II. Pole estimation using different norms

Although the RTFs are different for each loudspeaker-microphone position, all RTFs in a room share the same resonance frequencies. These resonance frequencies may be visible as spectral peaks in the RTFs [1]. If only the zeros cause RTF variation then the RTFs can be expressed using a common denominator for all and a different numerator for each of them. This can be represented by either common poles, $p(k)$, and distinct zeros, $z_i(k,t)$, or in polynomial form using common autoregressive (AR), $a(k)$, and distinct moving average (MA), $b_i(k,t)$, coefficients [7],

$$H_i(q,t) = \frac{\prod_{k=1}^{Q}(1 - z_i(k,t)q^{-1})}{\prod_{k=1}^{P}(1 - p(k)q^{-1})} = \frac{\sum_{k=1}^{Q} b_i(k,t)q^{-k}}{1 - \sum_{k=1}^{P} a(k)q^{-k}} \quad (1)$$

where Q and P are the order of numerator and denominator respectively, $i = 1, ..., M$ the number of RTFs and where q denotes the time shift operator, i.e., $q^{-k}u(t) = u(t-k)$.

Mathematically the class of problems considered in this paper can be cast into one general optimization problem associated with finding the filter coefficient vector $\mathbf{x}$ from a set of measured impulse responses cast in $\mathbf{v}$ and $\mathbf{W}$, so that the error $\mathbf{e} = \mathbf{v} - \mathbf{W}\mathbf{x}$ is minimized

$$\min_{\mathbf{x}} \|\mathbf{v} - \mathbf{W}\mathbf{x}\|_p^p \quad (2)$$

where $\|\cdot\|_p$ is the p-norm defined for $p > 1$ as $\|\mathbf{x}\|_p = \left(\sum_{t=1}^{N} |x(t)|^p\right)^{1/p}$. Matrix $\mathbf{W}$ and vector $\mathbf{v}$ involved in the minimization problem (2) depend on whether the model for pole estimation is all-pole or pole-zero and whether common acoustical poles are considered. In the pole-zero model with common acoustical poles case, matrix $\mathbf{W}$ and vector $\mathbf{v}$ are formed as

$$\begin{aligned}
\mathbf{v} &= [\mathbf{h}_1, \mathbf{h}_2, ..., \mathbf{h}_M]^T \\
\mathbf{h}_i &= [h_i(0), h_i(1), ..., h_i(N-1), 0, 0, 0]^T \\
\mathbf{x} &= [\mathbf{a}, \mathbf{b}_1, ..., \mathbf{b}_M]^T \\
\mathbf{a} &= [a(1), a(2), ..., a(P)]^T \\
\mathbf{b}_i &= [b_i(0), b_i(1), ..., b_i(Q)]^T \\
\mathbf{W} &= \begin{bmatrix} \mathbf{W}_1 & \mathbf{D} & 0 & 0 & 0 \\ \mathbf{W}_2 & 0 & \mathbf{D} & 0 & 0 \\ \vdots & \cdots & 0 & \ddots & 0 \\ \mathbf{W}_M & \cdots & \cdots & \cdots & \mathbf{D} \end{bmatrix} \\
&\Rightarrow \text{size } [M(N+P-1) \times (P + M(Q+1))]
\end{aligned} \quad (3)$$

$$\mathbf{D} = \begin{bmatrix} 1 & & & \\ & 1 & 0 & \\ & & \ddots & \\ & 0 & & 1 \\ 0 & & & 0 \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ 0 & \cdot & \cdot & 0 \end{bmatrix}$$

$$\Rightarrow \text{size } [M(N+P-1) \times (P + M(Q+1))]$$

$$\mathbf{W}_i = \begin{bmatrix} 0 & 0 & \cdots & 0 \\ h_i(0) & 0 & \cdots & 0 \\ h_i(1) & h_i(0) & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & h_i(0) \\ h_i(N-1) & 0 & \cdots & \vdots \\ 0 & h_i(N-1) & \cdots & \vdots \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & h_i(N-1) \end{bmatrix} \quad (4)$$

$$\Rightarrow \text{size } [(N+P-1) \times P]$$

where h_i is the i^{th} length-N measured impulse response. If an all-pole model is considered, matrix $\mathbf{D}$ and vectors $\mathbf{b}_j$ are just set to zero and vector $\mathbf{x}$ will only consist of AR coefficients (i.e., $\mathbf{x} = [\mathbf{a}]$). Conversely, when a pole-zero model is considered, vector $\mathbf{x}$ will consist of both AR and MA coefficients (i.e., $\mathbf{x} = [\mathbf{a}, \mathbf{b}_1, ..., \mathbf{b}_M]^T$). The set of $\mathbf{a}$ coefficients is estimated using either 2-norm or 1-norm minimization and forms the filter polynomial $A(q) = 1 - a(1)q^{-1} - ... - a(P)q^{-P}$. The residual $\tilde{B}_i(q)$ is the result of multiplying the actual RTF with the inverse filter $\tilde{A}(q)$.

$$\tilde{A}(q)H_i(q,t) = \tilde{A}(q)\frac{B_i(q)}{A(q)} = \tilde{B}_i(q) \quad (5)$$

where $\tilde{A}(q) \cong A(q)^{-1}$. The description of $\tilde{B}_i(q)$ in both time and frequency domain (i.e., residual RIR and residual RTF respectively) is directly affected by how $\tilde{A}(q)$ is calculated. In 2-norm (i.e., least squares error) minimization the optimal filter coefficient vector may be given in closed-form as

$$\mathbf{x} = \left(\mathbf{W}^T\mathbf{W}\right)^{-1}\mathbf{W}^T\mathbf{v} \quad (6)$$

However in 1-norm (i.e., least absolute error) minimization there exists no closed-form solution and therefore the filter coefficient vector is calculated as a solution to a convex optimization problem, which can be solved efficiently, e.g. using `CVX` [8]. It is found, experimentally, that similar ARMA coefficients estimates are extracted from a pole-zero model using 1-norm or 2-norm minimization, so only 2-norm pole-zero model will be considered in the sequel.

III. Results from measured impulse responses

In this section results obtained from two measured room impulse responses are presented. The three methods (i.e., 1-norm all-pole and 2-norm all-pole and pole-zero) are compared by inspecting both the residual RTF and the residual RIR. Matlab computer simulations were performed at $fs = 16kHz$. The impulse responses, shown in Fig. 1, (h_1 and h_2) of length $N = 2001$ samples were measured in a rectangular room of about $5 \times 3 \times 3$ m. The order was chosen as $P = 500$ and in the ARMA case also $Q = 500$. This order was found to be sufficient to accurately model the measured impulse response in the ARMA case, which served as a quality reference. Two objective measures are employed:

- Spectral Flatness (SF)

The spectral flatness is calculated by dividing the geometric mean of the power spectrum by the arithmetic mean of the power spectrum, i.e.

$$SF = \frac{\sqrt[N]{\prod_{f=0}^{N-1} P(f)}}{\frac{\sum_{f=0}^{N-1} P(f)}{N}} \quad (7)$$

where $P(f)$ represents the magnitude of the f^{th} bin, with $N = 512$.

- Sparseness Degree (SD)

The sparseness degree is the number of elements in the residual RIR that have an absolute value smaller than some threshold close to zero, i.e.

$$SD = \{R(n) : |R(n)| < threshold\} \quad (8)$$

where $R(n)$ represents the residual RIR and the threshold is set to $2 \cdot 10^{-6}$.

Two different cases for each method are presented: In the first case, the coefficients of the inverse filter are calculated from one impulse response h_1 (shown in Fig. 1 a). This filter is used to equalize h_1 so as to clearly observe the residual RTF and residual RIR. In the second case, the coefficients of the inverse filter to equalize h_1 are calculated using the set of two measured impulse responses (i.e., with common acoustical poles). The frequency response of h_1 is shown in Fig. 1 b.

- First case

When $\tilde{A}(q)$ is calculated from the all-pole model using 2-norm minimization (Fig. 2(a), 2(d)) the flattest residual RTF is achieved. On average, the magnitude difference between peaks and dips is very small; however, it exhibits a long noise-like residual RIR because the 2-norm minimization shapes the residual into coefficients that exhibit Gaussian like features [2].

When $\tilde{A}(q)$ is calculated from the pole-zero model using 2-norm minimization (Fig. 2(b), 2(e)) a highly coloured residual RTF is achieved, which is a perceptually undesirable characteristic in RTF equalization [5]. However, the short residual RIR may be desirable in other audio applications such as acoustic echo or feedback cancellation [3], [4].

When $\tilde{A}(q)$ is calculated from the all-pole model using 1-norm minimization (Fig. 2(c), 2(f)), the residual RTF has been flattened with respect to the true RTF and, in addition, the main low-frequency resonances have been cancelled. The residual RIR exhibits a sparse distribution of non-zero coefficients, which implies that the acoustic reflections modelled by the residual RIR are forced to be more spaced in time. The residual RIR from the 1-norm all-pole model shows the largest number of zero coefficients, i.e., the highest degree of sparseness. These results are summarized in Table I, where it can be seen that the 1-norm all-pole model and 2-norm all-pole model present the highest SD and the highest SF respectively.

- Second case

Fig. 3 shows that although in the common-acoustical-poles case the overall performance has been deteriorated with respect to the single impulse response case, the same main features can be observed on the residuals.

TABLE I: SF and SD for the 1-norm all-pole, 2-norm all-pole and pole-zero case

	2-norm all-pole	pole-zero	1-norm all-pole
SF	0.7	0.06	0.2
SD	5	43	503

IV. Conclusions

In this paper, results from RTF equalization have been presented. Equalization is achieved by cancelling the poles associated with the main resonances common to multiple RTFs in a room. Poles were estimated by means of three different methods. Poles estimated using 2-norm minimization and an all-pole model offered a residual RTF having the flattest response. Poles estimated using 2-norm minimization and a pole-zero model offered a large reduction in the main low-frequency acoustic resonances. However the residual RTF exhibited a highly coloured response, while the residual RIR was shorter in time. Finally, poles estimated using 1-norm minimization and an all-pole model offered a flat residual RTF with its main resonances cancelled and a sparse residual RIR. This means that the residual RIR will represent sparsely distributed discrete reflections. This feature may be desirable in speech applications although a deeper study taking into account perceptual considerations would be needed.

V. Acknowledgements

This research work was carried out at the ESAT Laboratory of Katholieke Univeristeit Leuven, in the frame of the EC FP6 project SIGNAL: 'Core Signal Processing Training Program' Marie-Curie Fellowship program (http://est-signal.i3s.unice.fr) under contract No. MEST-CT-2005-021175, the Concerted Research Action GOA-MaNet and the Belgian Programme on Interuniversity Attraction Poles initiated by the Belgian Federal Science Policy Office IUAP P6/04 (DYSCO, 'Dynamical systems, control and optimization', 2007-2011). The scientific responsibility is assumed by its authors.

References

[1] J. Mourjopoulos and M. A. Paraskevas, "Pole and zero modeling of room transfer functions", *J. Sound and Vibration*, vol. 146, no. 2, pp. 281-302, April 1991.

[2] J. A. Cadzow, "Minimum l_1, l_2, and l_∞ Norm Approximate Solutions to an Overdetermined System of Linear Equations", *Dig. Sig. Proc.*, vol. 12, no. 4, pp. 524-560, Oct. 2002.

[3] T. van Waterschoot and M. Moonen, "Adaptive feedback cancellation for audio applications," *Signal Processing*, vol. 89, no. 11, pp. 2185-2201, Nov. 2009.

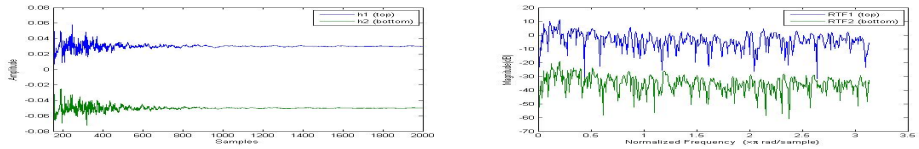
[4] P. S. R. Diniz, *Adaptive Filtering: Algorithms and Practical Implementations*. Springer, Boston, MA., 2008.

[5] D. M. Howard and J. A. S. Angus, *Acoustics and Psychoacoustics*. Focal Press, Elsevier, 2006.

[6] Y. Haneda, S. Makino and Y. Kaneda, "Multiple-Point Equalization of Room Transfer Functions by Using Common Acoustical Poles," *IEEE Trans. Speech Audio Process.*, vol. 5, no. 4, pp. 325-333, July 1997.

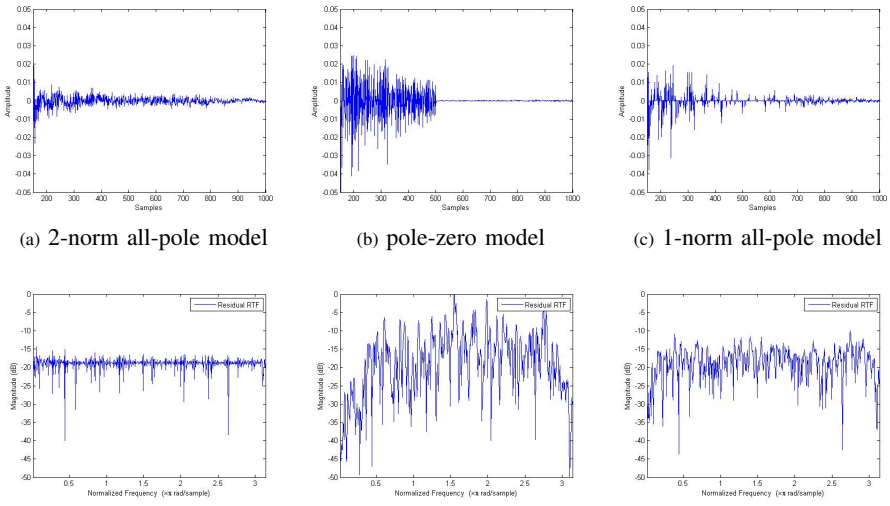
[7] Y. Haneda, S. Makino and Y. Kaneda, "Common Acoustical Pole and Zero Modeling of Room Transfer Function," *IEEE Trans. Speech Audio Process.*, vol. 2, no. 2, pp. 320-328, Apr. 1994.

[8] M. Grant and S. Boyd, CVX: Matlab software for disciplined convex programming (web page and software). http://cvxr.com/cvx, April, 2010.



(a) Time evolution of h_1 and h_2 (b) Spectrum of h_1 and h_2 (RTF1 and RTF2)

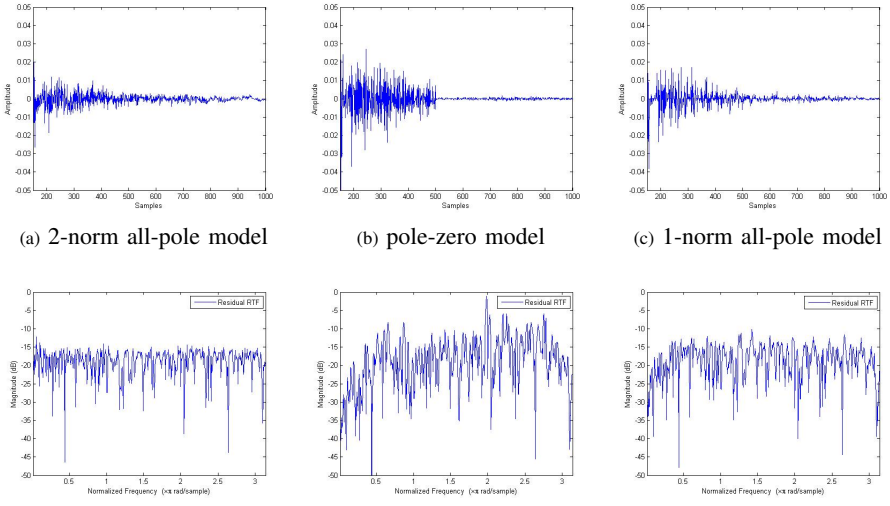
Fig. 1: Frequency response function and time evolution of h_1 and h_2



(a) 2-norm all-pole model (b) pole-zero model (c) 1-norm all-pole model

(d) 2-norm all-pole model (e) pole-zero model (f) 1-norm all-pole model

Fig. 2: a,b,c) h_1 residual RIRs and d,e,f) h_1 residual RTFs, form single impulse response estimation



(a) 2-norm all-pole model (b) pole-zero model (c) 1-norm all-pole model

(d) 2-norm all-pole model (e) pole-zero model (f) 1-norm all-pole model

Fig. 3: a,b,c) h_1 residual RIRs and d,e,f) h_1 residual RTFs, from common-acoustical-poles calculation