NON-LINEAR ACOUSTIC ECHO CANCELLATION USING CASCADED KALMAN FILTERING

Muhammad Z. Ikram

Embedded Processing Systems Lab
Texas Instruments Incorporated, Dallas, TX 75243
mzi@ti.com

ABSTRACT

This paper presents a new paradigm of solving the non-linear acoustic echo cancellation problem. The non-linear echo path is modeled by a memoryless non-linearity followed by a linear FIR filter. The problem is cast into a state-space framework and solved using a cascade of Kalman filters in time domain, one filter adapting to the linear echo path and the other filter adapting to the memoryless non-linearity. It is shown that the proposed method outperforms the existing NLMS-based method in filter convergence and misalignment while enjoying an additional benefit of unsupervised and variable step-size control. Interesting connections will be made between the proposed method and the widely-known NLMS-based echo canceler. Practical recommendations are provided on implementing the proposed method efficiently on a general-purpose processor. Finally, simulation results are presented that exhibit its performance advantages.

Index Terms— Echo cancellation, Kalman filter, Non-linear

1. INTRODUCTION

Acoustic echo cancellation (AEC) has been an area of active research for the past many decades. As AEC is implemented on newer and sophisticated devices, it is expected that its performance is also improved over the devices of prior generation. Various aspects of AEC have been investigated in the past; this includes double-talk, convergence, stereo-echo cancellation, etc [1]. One of the issues that has limited the performance of a practical echo canceler is non-linearity of the echo path. A typical echo canceler is designed assuming that the echo path is linear and modeled by a linear finite impulse response (FIR) filter. Consequently, the echo cancelling filter is also modeled as a linear FIR filter. However, in practical applications, like cell phones, the echo path is typically non-linear because of imperfections introduced by the data converter, amplifier, and the loudspeaker operating close to saturation. Attempting to cancel non-linear echo using a linear FIR filter leaves residual echo in the uplink signal resulting in annoying user experience. Several methods have been proposed in the past to address non-linear echo cancellation. Popular among them are the Volterra filter based methods. These methods, however, suffer from high complexity because of large number of filter parameters to adapt [2]. A large class of methods start with an assumption of the type of non-linearity that the system can introduce. In some methods, memoryless non-linearity is used for smaller loudspeakers used in portable hand-held devices like cell phones. In other methods, non-linearity with memory is generally employed for applications in high-end audio devices like precision audio systems.

In this work, we focus on mobile applications, where the echo path is modeled as a cascade of memoryless non-linearity followed by an FIR filter. In order to compensate for the non-linearity, a set of adaptive filters are placed in parallel to the echo path as shown in Fig. 1. These filters are learned to cancel the echo in the uplink channel.

We deviate from the traditional approach of using conventional least-mean squared (LMS) based adaptive filters and use state-space formulation to model the non-linear echo cancellation problem. By performing various mathematical manipulations we transform the non-linear AEC model to a state-space model consisting of an observation equation and a state equation. We then propose the use of cascaded Kalman filters operating sequentially to update the unknown filter coefficients. We show that by doing so, faster convergence is achieved without the need of controlling the step size. Furthermore, the filter misalignment is improved when compared with the prior state of the art.

The use of Kalman filter to solve the *linear* AEC problem has been suggested in the past by authors in [3] and [4]. In [3], the authors proposed a single-measurement update to learn the state estimates, whereas, a generalized version in [4] used a set of past measurements for Kalman updates. In [5], the authors used a *frequency-domain* Kalman filter to solve the non-linear AEC problem. In this work we will demonstrate the benefits of using cascaded Kalman filters running in time domain to solve the challenging problem of non-linear echo cancellation. Our experimental results show that, in addition to improved convergence and misalignment, the pro-

978-1-4799-2893-4/14/$31.00

posed method exhibits best performance by using only the current and a past measurement at each update step.

2. SYSTEM MODEL

The non-linear echo cancellation problem is depicted in Fig. 1, where the echo path is modeled as a cascade of memoryless non-linearity followed by an FIR filter that models the linear echo path. The output of the non-linear part is given by

$$z(n) = \sum_{p=1}^{P} a_p x^p(n) = \mathbf{a}^T \mathbf{x}(n), \quad (1)$$

where $\mathbf{a} = [a_1\ a_2\ \cdots\ a_P]^T$, $\mathbf{x}(n) = \left[x(n)\ x^2(n)\ \cdots\ x^P(n)\right]^T$, P is order of the non-linearity, and $(\cdot)^T$ denotes conjugate transposition.

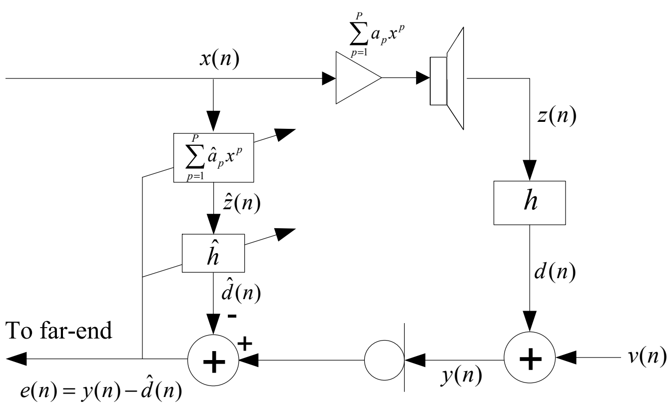

Fig. 1. Nonlinear AEC setup.

The output of the FIR filter is expressed as

$$d(n) = \sum_{k=0}^{K-1} h_k z(n-k) = \mathbf{h}^T \mathbf{z}(n), \quad (2)$$

where $\mathbf{h} = [h_o\ h_1\ \cdots\ h_{K-1}]^T$, $\mathbf{z}(n) = [z(n)\ z(n-1)\ \cdots\ z(n-K+1)]^T$, and K is order of the FIR filter.

Using (1) and (2), the echo input to the microphone is given by

$$d(n) = \mathbf{a}^T \mathbf{X}(n) \mathbf{h} = \mathbf{h}^T \mathbf{X}^T(n) \mathbf{a}, \quad (3)$$

where

$$\mathbf{X}(n) = \begin{pmatrix} x(n) & x(n-1) & \cdots & x(n-K+1) \\ x^2(n) & x^2(n-1) & \cdots & x^2(n-K+1) \\ \vdots & \vdots & \ddots & \vdots \\ x^P(n) & x^P(n-1) & \cdots & x^P(n-K+1) \end{pmatrix}. \quad (4)$$

The microphone input signal consists of echo and the near-end noise $v(n)$, resulting in

$$y(n) = d(n) + v(n). \quad (5)$$

In order to solve the non-linear AEC problem, an adaptive non-linear filter with coefficients $\widehat{\mathbf{a}}(n) = [\widehat{a}_1(n)\ \ \widehat{a}_2(n)\ \ \cdots\ \ \widehat{a}_P(n)]^T$ and an adaptive FIR filter $\widehat{\mathbf{h}}(n) = [\widehat{h}_o(n)\ \ \widehat{h}_1(n)\ \ \cdots\ \ \widehat{h}_{K-1}(n)]^T$ are placed as shown in Fig. 1 such that that they try to mimic the effects of the non-linearity in the amplifier as well as the echo path.

Referring to Fig. 1, the error output is given by

$$e(n) = y(n) - \widehat{d}(n). \quad (6)$$

In [6], the authors proposed joint adaptation of $\widehat{\mathbf{a}}(n)$ and $\widehat{\mathbf{h}}(n)$ via normalized least mean squares (NLMS). As in conventional LMS based adaptive filter, the choice of step size is critical in controlling the convergence rate and filter misalignment. This is even more important in this case, where two filters are jointly adapted and where the output of one filter is dependent on the other. The NLMS is also the choice of adaptation in [7], where a raised cosine function is used to model the non-linearity. The authors in [8] employed orthogonalized power filters to model the memoryless non-linearity and linear echo path. This resulted in an increased number of parameters to be estimated; nevertheless, it was preferred by the authors as it resulted in stable convergence and optimum solution. In [2], the authors modeled the non-linearity using a Wiener-Hammerstein model and proposed its solution using a a set of adaptive filters. These filters were also learned using NLMS-based adaptation.

In this paper, we propose to use state-space framework to model the non-linear AEC. We will see that, in contrast to the existing methods, the task of controlling the step size is minimized and the number of unknown parameters to be estimated is not increased while the convergence rate as well as the filter misalignment are improved significantly.

3. STATE SPACE FORMULATION

In an attempt to solve the problem using Kalman filter, we set up the state and measurement (observation) equations for estimating the non-linear filter parameters $\mathbf{a}$ and the FIR filter parameters $\mathbf{h}$.

Let us first write the output of the two adaptive filters, $d(n)$, as follows

$$d(n) = \left[\mathbf{h}^T \tilde{\mathbf{x}}(n)\ \ \mathbf{h}^T \tilde{\mathbf{x}}^2(n)\ \ \cdots\ \ \mathbf{h}^T \tilde{\mathbf{x}}^P(n)\right] \mathbf{a}, \quad (7)$$

where

$$\tilde{\mathbf{x}}(n) = [x(n)\ \ x(n-1)\ \ \cdots\ \ x(n-K+1)]^T. \quad (8)$$

Using (5), we stack M measurements in a vector to obtain (12), which can be compactly written as follows

$$\mathbf{y}(n) = \mathbf{X}_h(n) \mathbf{a} + \mathbf{v}(n), \quad (9)$$

where $\mathbf{v}(n)$ is a vector of measurement noise with covariance matrix $\mathbf{R}$ of dimension $M \times M$.

The state update equation is written as

$$\mathbf{a}(n) = \mathbf{a}(n-1) + \mathbf{w}_a(n), \tag{10}$$

where $\mathbf{w}_a(n)$ is the process noise with $P \times P$ covariance matrix $\mathbf{Q}_a$, and it is assumed to be uncorrelated with $\mathbf{v}(n)$.

Similarly, in order to solve for the FIR filter coefficients $\widehat{\mathbf{h}}(n)$, we express the output $d(n)$ as follows

$$d(n) = \left[\mathbf{a}^T\mathbf{x}(n)\ \mathbf{a}^T\mathbf{x}(n-1)\ \cdots\ \mathbf{a}^T\mathbf{x}(n-K+1)\right]\mathbf{h}. \tag{11}$$

Again, putting together M measurements in a vector form gives (13), which can be stated as

$$\mathbf{y}(n) = \mathbf{X}_a(n)\mathbf{h} + \mathbf{v}(n). \tag{14}$$

The state update for $\mathbf{h}$ is expressed as

$$\mathbf{h}(n) = \mathbf{h}(n-1) + \mathbf{w}_h(n), \tag{15}$$

where $\mathbf{w}_h(n)$ is process noise with covariance matrix $\mathbf{Q}_h$ of dimension $K \times K$.

Equations (9), (10) and Equations (14), (15) capture the dynamics of the non-linear AEC system, and these sets of equations form the basis of the cascaded Kalman filter.

4. ESTIMATION USING CASCADED KALMAN FILTERS

The Kalman filter consists of a time update and a measurement update [9]. In the time update, the filter coefficients and their error covariances at time n are updated using only the measurements available until the previous time instant $n-1$. Such estimates are referred to as *a priori* estimates and are denoted using $(\cdot)^-$. The time update for $\widehat{\mathbf{a}}(n)$ is given by

$$\widehat{\mathbf{a}}^-(n) = \widehat{\mathbf{a}}(n-1) \tag{16}$$

$$\mathbf{P}_a^-(n) = \mathbf{P}_a(n-1) + \mathbf{Q}_a, \tag{17}$$

where the error covariance is defined as

$$\mathbf{P}_a^-(n) = \mathrm{Cov}\left[\mathbf{a}(n) - \widehat{\mathbf{a}}^-(n)\right]. \tag{18}$$

As the measurement at time n become available, the parameter estimates and their error variance estimates are updated in the measurement update as follows [9].

$$\mathbf{K}_a(n) = \mathbf{P}_a^-(n)\mathbf{X}_{\hat{h}}^T(n)\left[\mathbf{X}_{\hat{h}}(n)\mathbf{P}_a^-(n)\mathbf{X}_{\hat{h}}^T(n) + \mathbf{R}\right]^{-1} \tag{19}$$

$$\widehat{\mathbf{a}}(n) = \widehat{\mathbf{a}}^-(n) + \mathbf{K}_a(n)\left[\mathbf{y}(n) - \mathbf{X}_{\hat{h}}(n)\widehat{\mathbf{a}}^-(n)\right] \tag{20}$$

$$\mathbf{P}_a(n) = \mathbf{P}_a^-(n) - \mathbf{K}_a(n)\mathbf{X}_{\hat{h}}(n)\mathbf{P}_a^-(n), \tag{21}$$

where $\mathbf{K}_a(n)$ is commonly referred to as Kalman gain and the error covariance $\mathbf{P}_a(n)$ is defined by replacing $\widehat{\mathbf{a}}^-(n)$ with $\widehat{\mathbf{a}}(n)$ in (18).

Similar set of time and measurement updates can be developed for estimating $\mathbf{h}$ as follows

$$\widehat{\mathbf{h}}^-(n) = \widehat{\mathbf{h}}(n-1) \tag{22}$$

$$\mathbf{P}_h^-(n) = \mathbf{P}_h(n-1) + \mathbf{Q}_h. \tag{23}$$

The measurement update for $\widehat{\mathbf{h}}(n)$ comprises of the following steps.

$$\mathbf{K}_h(n) = \mathbf{P}_h^-(n)\mathbf{X}_{\hat{a}}^T(n)\left[\mathbf{X}_{\hat{a}}(n)\mathbf{P}_h^-(n)\mathbf{X}_{\hat{a}}^T(n) + \mathbf{R}\right]^{-1} \tag{24}$$

$$\widehat{\mathbf{h}}(n) = \widehat{\mathbf{h}}^-(n) + \mathbf{K}_h(n)\left[\mathbf{y}(n) - \mathbf{X}_{\hat{a}}(n)\widehat{\mathbf{h}}^-(n)\right] \tag{25}$$

$$\mathbf{P}_h(n) = \mathbf{P}_h^-(n) - \mathbf{K}_h(n)\mathbf{X}_{\hat{a}}(n)\mathbf{P}_h^-(n). \tag{26}$$

The data matrices $\mathbf{X}_{\hat{h}}(n)$ and $\mathbf{X}_{\hat{a}}(n)$ are obtained by using the parameter estimates $\widehat{\mathbf{h}}(n)$ and $\widehat{\mathbf{a}}(n)$ in (12) and (13), respectively as well as the downlink data $\mathbf{x}(n)$. These matrices are updated at each iteration. The Kalman filter begins with appropriate initialization of $\widehat{\mathbf{a}}(0)$ and $\widehat{\mathbf{h}}(0)$ as well as for the error-covariance matrices $\mathbf{P}_a(0)$ and $\mathbf{P}_h(0)$. In each iteration, the filter $\widehat{\mathbf{a}}(n)$ is first estimated followed by estimation of $\widehat{\mathbf{h}}(n)$.

Let us review the time and measurement update equations of the proposed method and draw some comparisons with the classical LMS or the recursive least squares (RLS) based adaptation methods. We see in (20) and (25) that the error is weighted by the Kalman gain of (19) and (24), respectively to update the state vectors. This is in a way similar to NLMS processing where the normalized step size is used in filter update. In contrast to the NLMS update, which requires a judicious choice of step size, the Kalman gain is computed as part of the measurement update using the noise covariance and the downlink data. The Kalman filter based non-linear AEC, therefore, works like a variable step size adaptive filter. As we will see, this helps in promoting filter convergence while the filter misalignment is improved.

5. IMPLEMENTATION CONSIDERATIONS

Like in traditional Kalman filter, the choice of process noise $\mathbf{Q}_h$ and $\mathbf{Q}_a$ plays a key role in the cascaded adaptive filter framework. A smaller value of this noise variance implies smaller adaptive filter updates, whereas a larger value results in bigger update steps. Since the implementation in Section 4 is based on using the estimate of $\mathbf{a}(n)$ to update $\mathbf{h}(n)$ and vice-versa, we prefer using a small value of $\mathbf{Q}_h$ and $\mathbf{Q}_a$ to prevent divergence. As suggested in [4], the value of $\mathbf{R}$ was estimated during silence intervals, and this value was periodically updated.

6. EXPERIMENTAL RESULTS

We compared the performance of the proposed cascaded Kalman filter against the dual adaptive filter method of [6],

$$\begin{pmatrix} y(n) \\ y(n-1) \\ \vdots \\ y(n-M+1) \end{pmatrix} = \begin{pmatrix} \mathbf{h}^T\tilde{\mathbf{x}}(n) & \mathbf{h}^T\tilde{\mathbf{x}}^2(n) & \cdots & \mathbf{h}^T\tilde{\mathbf{x}}^P(n) \\ \mathbf{h}^T\tilde{\mathbf{x}}(n-1) & \mathbf{h}^T\tilde{\mathbf{x}}^2(n-1) & \cdots & \mathbf{h}^T\tilde{\mathbf{x}}^P(n-1) \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{h}^T\tilde{\mathbf{x}}(n-M+1) & \mathbf{h}^T\tilde{\mathbf{x}}^2(n-M+1) & \cdots & \mathbf{h}^T\tilde{\mathbf{x}}^P(n-M+1) \end{pmatrix} \mathbf{a} + \begin{pmatrix} v(n) \\ v(n-1) \\ \vdots \\ v(n-M+1) \end{pmatrix} \tag{12}$$

$$\begin{pmatrix} y(n) \\ y(n-1) \\ \vdots \\ y(n-M+1) \end{pmatrix} = \begin{pmatrix} \mathbf{a}^T\mathbf{x}(n) & \mathbf{a}^T\mathbf{x}(n-1) & \cdots & \mathbf{a}^T\mathbf{x}(n-K+1) \\ \mathbf{a}^T\mathbf{x}(n-1) & \mathbf{a}^T\mathbf{x}(n-2) & \cdots & \mathbf{a}^T\mathbf{x}(n-K) \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{a}^T\mathbf{x}(n-M+1) & \mathbf{a}^T\mathbf{x}(n-M) & \cdots & \mathbf{a}^T\mathbf{x}(n-K-M+2) \end{pmatrix} \mathbf{h} + \begin{pmatrix} v(n) \\ v(n-1) \\ \vdots \\ v(n-M+1) \end{pmatrix} \tag{13}$$

where the linear and non-linear filter coefficients were updated using NLMS adaptation. Fig. 2 shows the microphone signal $y(n)$, which was generated by passing $x(n)$ through a $P = 5$-pt non-linear filter in (1) and a $K = 256$-pt FIR filter. As the filter parameters are learned, the performance is measured by computing the misalignment as follows

$$\text{Misalignment of } \mathbf{h}(n) = 20\log_{10}\frac{\|\widehat{\mathbf{h}}(n)-\mathbf{h}\|}{\|\mathbf{h}\|} \tag{27}$$

$$\text{Misalignment of } \mathbf{a}(n) = 20\log_{10}\frac{\|\widehat{\mathbf{a}}(n)-\mathbf{a}\|}{\|\mathbf{a}\|}. \tag{28}$$

For the proposed method, we used $\mathbf{Q}_h = 1\mathrm{e}^{-8}\mathbf{I}$ and $\mathbf{Q}_a = 1\mathrm{e}^{-3}\mathbf{I}$, where $\mathbf{I}$ is an identity matrix of appropriate dimension. The matrix $\mathbf{R}$ was set to $1\mathrm{e}^{-8}\mathbf{I}$ by estimating the noise variance in the silence interval. An initial value of $\widehat{\mathbf{h}}(0) = [1\ 0\ 0\ \cdots\ 0]^T$ and $\widehat{\mathbf{a}}(0) = [0\ 0\ \cdots\ 0]^T$ was used. The matrices $\mathbf{P}_h(0)$ and $\mathbf{P}_a(0)$ were initialized with identity matrices. We ran the experiments for $M = 1$, 2, and 4. For the method of [6], we tuned the adaptive filters that resulted in best performance by setting the step sizes of $\widehat{\mathbf{a}}(n)$ and $\widehat{\mathbf{h}}(n)$ to be equal to 0.2 and 0.5, respectively. The comparative results are shown in Fig. 3. We first note that the performance of the proposed Kalman-filter based non-linear AEC solution is better than the NLMS-based algorithm in both convergence and misalignment for all values of M that we tried. Furthermore, we note that performance of the proposed method is optimal for $M = 2$; i.e., the case where only 2 measurements are stacked in the vectors of (12) and (13). This also implies that with only a moderate increase in complexity over the single-observation model ($M = 1$), significant performance is obtained that is maintained as M is increased.

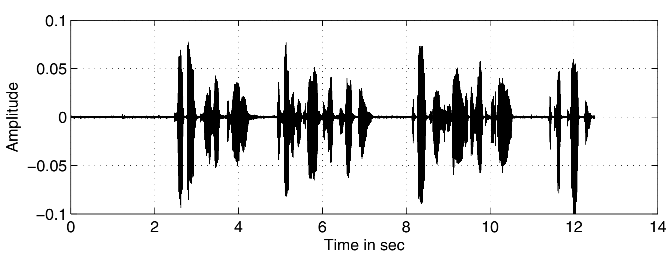

Fig. 2. Microphone signal $y(n)$.

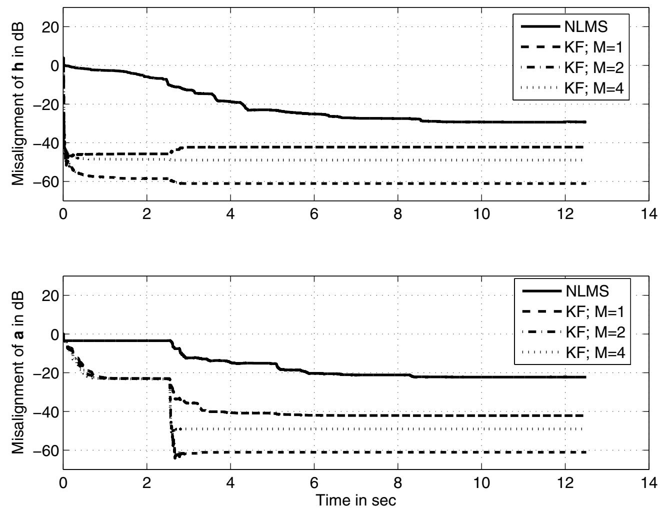

Fig. 3. Comparison of proposed method and the NLMS method.

7. SUMMARY

We developed a state-space framework to solve the non-linear acoustic echo cancellation problem. Both the linear as well as non-linear blocks of the echo model are assumed unknown and adapted using a set of properly designed Kalman filters running back to back on a per-sample basis. With such an arrangement, improvements in convergence and misalignment are obtained over NLMS-based adaptation. We suggested practical guidelines on using the proposed algorithm and showed that with only a minor increase in complexity significant improvement is obtained over the single-snapshot Kalman filter implementation. Going forward, we plan to analyze the cascaded Kalman filter with other forms of non-linearities, including non-linearities with memory as well as non-linearities of varying orders.

8. REFERENCES

[1] S. Gay and J. Benesty, Eds., *Acoustic Signal Processing for Telecommunication*, Norwell, MA: Kluwer Academic, 2000.

[2] B. S. Nollett and D. L. Jones, "Nonlinear echo cancellation for hands-free speakerphones," in *Proc. IEEE NSIP*, 1997.

[3] C. Paleologu, J. Benesty, and S. Ciochină, "Study of the optimal and simplified Kalman filters for echo cancellation," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, 2013, pp. 580–584.

[4] C. Paleologu, J. Benesty, and S. Ciochină, "Study of the general Kalman filter for echo cancellation," *IEEE Trans. Audio, Speech, Language Processing*, vol. 8, no. 21, pp. 1539–1549, 2013.

[5] S. Malik and G. Enzner, "State-space frequency domain adaptive filtering for nonlinear acoustic echo cancellation," *IEEE Trans. Audio, Speech, Language Processing*, vol. 20, no. 7, pp. 2065–2079, 2012.

[6] A. Stenger, W. Kellermann, and R. Rabenstein, "Adaptation of acoustic echo cancellers incorporating a memoryless nonlinearity," in *Proc. Int. Workshop Acoustic Echo and Noise Control*, 1999.

[7] H. Dai and W. Zhu, "Compensation of loudspeaker nonlinearity in acoustic echo cancellation using raised-cosine function," *IEEE Trans. Circuits Systems–II*, vol. 53, no. 11, pp. 1190–1194, 2006.

[8] F. Kuech, A. Mitnacht, and W. Kellermann, "Nonlinear acoustic echo cancellation using adaptive orthogonalized power filters," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, 2005, pp. 105–108.

[9] E. W. Kamen and J. K. Su, *Introduction to Optimal Estimation*, London: Springer-Verlag, 1989.