TIME DELAY ESTIMATION AND SIGNAL RECONSTRUCTION USING MULTI-RATE MEASUREMENTS

Omid Jahromi and Parham Aarabi

Artificial Perception Laboratory,
Department of Electrical and Computer Engineering, University of Toronto
Toronto, ON, Canada, M5S 3G4
Email: omidj@control.toronto.edu, parham@ecf.toronto.edu

ABSTRACT

This paper considers the problem of fusing two low-rate sensors (e.g., microphones) for reconstructing one high-resolution signal when time delay of arrival (TDOA) is present as well. We show that under certain conditions the phase of the cross-spectrum-density of low-rate measurements becomes independent of the signal in the high-rate front end of the system. We then utilize this fact to demonstrate that it is possible to extend a class of TDOA estimation techniques known as *the generalized cross correlation technique* to linear-phase multi-rate sensor systems. Finally, we illustrate how the combination of the theory of linear-phase multi-rate filter banks and TDOA estimation can result in a practical, multi-sensor signal reconstruction system.

1. INTRODUCTION

Extensive research has been conducted on multi-rate signal processing [1]. However, the classic theory of multi-rate signal processing does not deal with delayed multi-rate signals. In practical situations, multiple observations of a signal will inevitably have a delay associated with them. For example, in the case of multiple cell phones being fused to provide greater bandwidth, their distance to the speaker may be different which will result in TDOA differences for the different cell phones. Furthermore, because the internal sampling of the phones is unlikely to be synchronized, there will be a further delay added to the system.

2. TDOA ESTIMATION

TDOA estimation arises in a variety of fields, including speech localization and processing using microphone arrays [2, 3, 4]. As a result, various algorithms have been developed for the estimation of TDOAs between two signals. The general discrete-time model can be stated as follows:

$$u_0(n) = x(n) + s_0(n) \tag{1}$$

$$u_1(n) = x(n-D) + s_1(n) \tag{2}$$

where $u_0(n)$ and $u_1(n)$ are the two signals at the observation points (i.e. microphones), $x(n)$ is the signal of interest that is referenced (zero time-delay) according to the first channel and will have a delay of D by the time it arrives at the second channel, and $s_0(n)$ and $s_1(n)$ are the (possibly dependent) noises of the first and second channels, respectively. The goal is to estimate D from a segment of observed data from the microphones, without prior knowledge regarding the source signal $x(n)$ or the noises. The most common solution to the above problem is the generalized cross correlation technique [3, 4], defined below:

$$\tilde{D} = \arg\max_D \int_\omega W(e^{j\omega}) U_0(e^{j\omega}) U_1^*(e^{j\omega}) e^{-j\omega D} d\omega \tag{3}$$

where $U_0(e^{j\omega})$ and $U_1(e^{j\omega})$ are the discrete-time Fourier transforms of the signals $u_0(n)$ and $u_1(n)$ respectively and $W(e^{j\omega})$ is a cross-correlation weighting function. Various weighting functions have been proposed in the past [4], but the most common solution is a whitening filter which results in the following cross-correlation form:

$$\tilde{D} = \arg\max_D \int_\omega \cos\left(\omega D - (\angle U_0(e^{j\omega}) - \angle U_1(e^{j\omega}))\right) d\omega \tag{4}$$

This is known as the PHAse Transform (PHAT).

3. THE MULTIRATE SENSOR FUSION PROBLEM

Consider the model shown in Fig. 1. Here, $x(n)$ represents the reference (first microphone's) signal which we model by a zero-mean wide-sense stationary (WSS) random process. We depict by D the unknown time delay (in number of samples) between the signals received by the two microphones. In each transmitter, the microphone signal is processed using a linear filter and then down-sampled to generate a low-rate signal $v_i(n)$, $i = 0, 1$. The signals $v_i(n)$ are transmitted to a central receiver. Note that the sampling rate (bandwidth) assumed for the input signal is twice the symbol rate (bandwidth) that each transmitter is allowed to transmit at.

The problem is to design the transmitter filters $H_0(z)$ and $H_1(z)$ and the receiver such that the receiver may reconstruct $x(n)$ from the low-rate signals $v_0(n)$ and $v_1(n)$ without a prior knowledge of D.

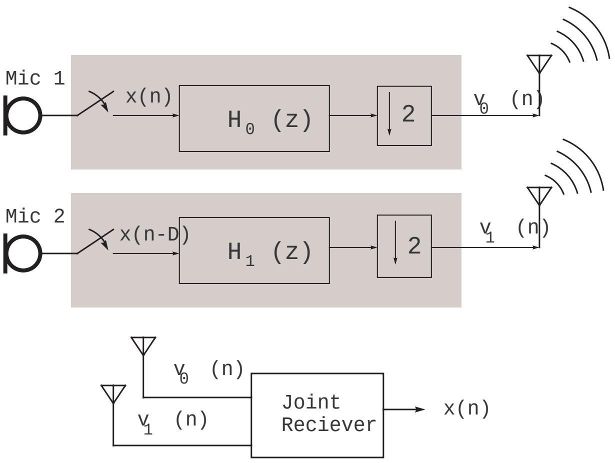


Figure 1: The multirate microphone array fusion problem.

We deal with the above stated problem in two steps. First, we study the possibility of estimating D at the receiver and then consider the problem of reconstructing $x(n)$ when D is specified.

4. ESTIMATING D FROM LOW-RATE RECEIVED SIGNALS

In this section we show that under certain conditions, the unknown time delay D can be estimated by examining the phase of the cross spectral density (CSD) of the low-rate signals $v_0(n)$ and $v_1(n)$. The main result is summarized in the theorem below.

Theorem 1 *If the difference between the phase response of the filters $H_0(z)$ and $H_1(z)$ used in the transmitters in Fig. 1 is a constant, then $\angle P_{v_0v_1}(e^{j\omega})$ is independent of the input signal statistics. Furthermore, $\angle P_{v_0v_1}(e^{j\omega}) = -\omega\frac{D}{2}+c$ where $c \stackrel{\triangle}{=} \angle H_0(e^{j\omega})-\angle H_1(e^{j\omega})$.*

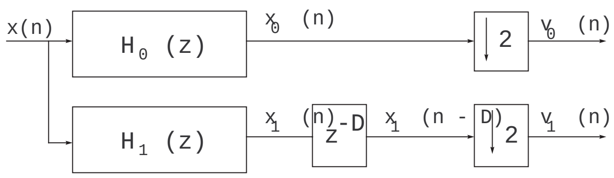


Figure 2: Representing the transmitters in Fig. 1 as a two-channel analysis filter bank with a delay in the lower branch.

Proof: Consider the equivalent block diagram shown in Fig. 2. It is straightforward to verify that the output signals $v_0(n)$ and $v_1(n)$ are jointly wide-sense stationary. Thus, the cross-correlation function $R_{v_0v_1}(k)$ defined by

$$R_{v_0v_1}(k) \stackrel{\triangle}{=} E\{v_0(n)v_1(n+k)\} \qquad (5)$$

exists. The signals $v_0(n)$ and $v_1(n)$ are down-sampled versions of $x_0(n)$ and $x_1(n-D)$. That is, $v_0(n) = x_0(2n)$ and $v_1(n) = x_1(2n-D)$. Thus we have

$$\begin{aligned} R_{v_0v_1}(k) &= E\{x_0(2n)x_1(2n+2k-D)\} \\ &= R_{x_0x_1}(2k-D). \end{aligned} \qquad (6)$$

The above equation allows us to express the CSD $P_{v_0v_1}(e^{j\omega})$ of the low-rate signals in terms of the CSD $P_{x_0x_1}(e^{j\omega})$ associated with $x_0(n)$ and $x_1(n)$:

$$\begin{aligned} &P_{v_0v_1}(e^{j\omega}) \stackrel{\triangle}{=} \textstyle\sum_{k=-\infty}^{\infty} R_{v_0v_1}(k)e^{-j\omega k} \\ &= \textstyle\sum_{k=-\infty}^{\infty} R_{x_0x_1}(2k-D)e^{-j\omega k} = \frac{1}{2}e^{-j\omega\frac{D}{2}} \times \\ &\begin{cases} P_{x_0x_1}(e^{j\frac{\omega}{2}}) + P_{x_0x_1}(e^{j\frac{\omega-2\pi}{2}}) & D \text{ even} \\ P_{x_0x_1}(e^{j\frac{\omega}{2}}) - P_{x_0x_1}(e^{j\frac{\omega-2\pi}{2}}) & D \text{ odd} \end{cases} \end{aligned} \qquad (7)$$

In the last step of the above derivations we used the following properties of the discrete-time Fourier transform:

$$x(n) \stackrel{\mathcal{F}}{\rightleftharpoons} X(e^{j\omega}) \Rightarrow \begin{cases} x(2n) & \stackrel{\mathcal{F}}{\rightleftharpoons} & \frac{X(e^{j\frac{\omega}{2}})+X(e^{j\frac{\omega-2\pi}{2}})}{2} \\ x(n-D) & \stackrel{\mathcal{F}}{\rightleftharpoons} & e^{-j\omega D}X(e^{-j\omega}) \end{cases}$$

It is straightforward to show that, for the setup in Fig. 2,

$$P_{x_0x_1}(e^{j\omega}) = H_0(e^{j\omega})H_1^*(e^{j\omega})P_{xx}(e^{j\omega}), \qquad (8)$$

where $P_{xx}(e^{j\omega})$ is the power spectral density (PSD) of the input signal.

Recall that the PSD of a real-valued WSS process is real. Thus, if $\angle H_0(e^{j\omega}) - \angle H_1(e^{j\omega}) = c$, it follows from (8) that $\angle P_{x_0x_1}(e^{j\omega})$ will be equal to c. Then, (7) shows that $\angle P_{v_0 v_1}(e^{j\omega}) = -\omega \frac{D}{2} + c$ which is independent of $P_{xx}(e^{j\omega})$. **Q.E.D.**

An interesting observation is that if $H_0(z)$ and $H_1(z)$ are linear phase and FIR (with the same length N), then the condition of the above theorem is satisfied.

Remark 1 *The results of the above theorem remain valid when* independent *noise components $s_0(n)$ and $s_1(n)$ are added to the signals $x(n)$ and $x(n-D)$, respectively. However, if the noise sources are correlated, an extra term (which depends on the cross-correlation between the two noise signals) will be added to the right hand side of (6). This will introduce additional terms in the phase of $P_{v_0 v_1}(e^{j\omega})$ and, hence, bias in the estimation of D.*

Linear phase FIR filters prove to be very viable from a signal synthesis point of view as well. In fact, it is possible to design FIR and linear-phase filters $H_0(z)$ and $H_1(z)$ such that $x(n)$ is recovered from $v_0(n)$ and $v_1(n)$. We will explore this issue in the next section.

5. RECONSTRUCTING THE REFERENCE SIGNAL FROM LOW-RATE RECEIVED SIGNALS

5.1. Case A: Assuming D is even

When D is even, the diagram shown in Fig. 2 can be treated as a standard two-channel analysis filter bank. Analysis filter banks for which it is possible to reconstruct the input signal using the down-sampled outputs are known as *perfect reconstruction* (PR) filter banks [1]. Two-channel PR filter banks whose analysis filters are both linear-phase and FIR have been studied by Nguyen and Vaidyanathan [5]. In the following, we briefly review some of the results in [5] and define the class $\mathcal{P}_N$ of filters that we suggest for use in the transmitters in Fig. 1.

Define $\mathbf{h}(z) \triangleq [H_0(z)\ H_1(z)]^T$. Now, $\mathbf{h}(z)$ can be factored as

$$\mathbf{h}(z) = \mathbf{E}(z^M)\mathbf{e}(z) \tag{9}$$

where $\mathbf{e}(z) \triangleq [1\ z^{-1}]^T$ and $\mathbf{E}(z)$ is the *polyphase matrix* associated with the analysis filter bank $\mathbf{h}(z)$ [1].

Definition 1 *We denote by $\mathcal{P}_N$ the class of 2-channel analysis filter banks for which the following conditions are satisfied:*

1. *The filters $H_0(z)$ and $H_1(z)$ are of length $N \triangleq 2(K+1)$, where $K \in \mathbb{Z}^+$ is fixed. In other words, $\mathbf{E}(z)$ is FIR of order K.*
2. *The matrix $\mathbf{E}(z)$ has the factorization*

$$\mathbf{E}(z) = \mathbf{A}_K \mathbf{D}(z) \mathbf{A}_{K-1} \mathbf{D}(z) \cdots \mathbf{D}(z) \mathbf{A}_0 \tag{10}$$

where

$$\mathbf{D}(z) \triangleq \begin{bmatrix} 1 & 0 \\ 0 & z^{-1} \end{bmatrix}$$

and

$$\begin{aligned} \mathbf{A}_i &\triangleq \begin{bmatrix} 1 & \theta_i \\ \theta_i & 1 \end{bmatrix} \quad 0 \le i \le K-1, \\ \mathbf{A}_i &\triangleq \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \qquad i = K. \end{aligned} \tag{11}$$

One can verify that the construction introduced in the above definition generates linear phase $H_0(z)$ and $H_1(z)$. The filters in $\mathcal{P}_N$ allow perfect reconstruction. To achieve this, one has to design a synthesis system using the (adjoint) polyphase matrix

$$\mathbf{R}(z) = \mathbf{A}_0^T \mathbf{C}(z) \mathbf{A}_1^T \mathbf{C}(z) \cdots \mathbf{C}(z) \mathbf{A}_K^T \tag{12}$$

where

$$\mathbf{C}(z) \triangleq \begin{bmatrix} z^{-1} & 0 \\ 0 & 1 \end{bmatrix}.$$

Then, two synthesis filters $F_0(z)$ and $F_1(z)$ are calculated from

$$\begin{bmatrix} f_0(z) & f_1(z) \end{bmatrix} = \mathbf{e}^T(z)\mathbf{R}(z^M). \tag{13}$$

The block diagram of the receiver using the filters $F_0(z)$ and $F_1(z)$ calculated as above is shown in Fig. 3.

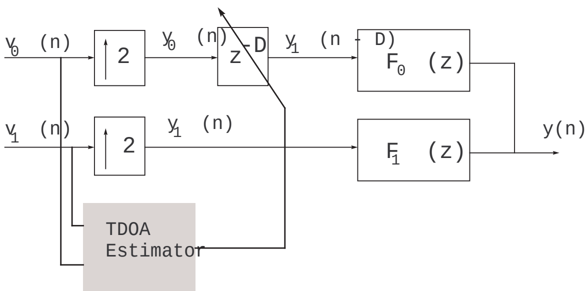


Figure 3: Block diagram of the suggested receiver.

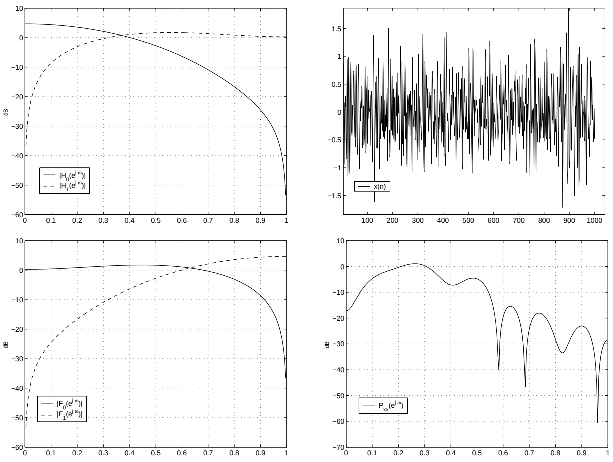
5.2. Case B: Assuming D is odd

When D is odd, the low-rate signal $v_1(n)$ is not a delayed version of $x_1(2n)$. As a result, it is not possible, in general, to achieve perfect reconstruction by adjusting the delay in the upper branch of the receiver shown in Fig. 3. Nevertheless, the reconstruction error may be minimized by optimizing the parameters θ_0 to θ_{K-1} that specify the analysis and synthesis filters[1].

[1]Length limitations preclude a thorough discussion of optimal synthesis methods in this paper.

6. AN ILLUSTRATIVE EXAMPLE

Here, we present a simple simulation example to illustrate the results in sections 4 and 5. For simplicity, we choose the analysis filters to be in the class $\mathcal{P}_4$. This means $N = 4$ and $K = 1$ so we have only one parameter θ_0 to select. Choosing $\theta_0 = 4.012$ leads to the analysis filters whose frequency response is shown in Fig. 4(a). The filters $H_0(z)$ and $H_1(z)$ designed above



(a) (b)

Figure 4: (a) Frequency response of the analysis and synthesis filters used in the example. (b) The reference signal $x(n)$ and its power spectral density.

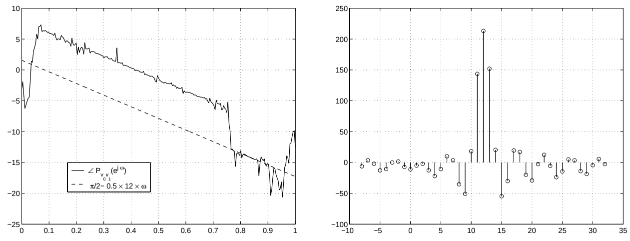
have symmetric and anti-symmetric impulse responses, respectively. This means $\angle H_0(e^{j\omega}) - \angle H_1(e^{j\omega}) = \frac{\pi}{2}$.

The reference microphone's signal $x(n)$ and its power spectral density $P_{xx}(e^{j\omega})$ are shown in Fig. 4(b). The delayed signal at the second microphone is assumed to be $x(n-12)$. Fig. 5(a) shows the estimated phase of $P_{v_0 v_1}(e^{j\omega})$ plotted along with the reference line $\frac{\pi}{2} - \frac{D}{2}\omega$ for $D = 12$, i.e., the correct delay. As can be seen from this plot, $\angle P_{v_0 v_1}(e^{j\omega})$ closely follows the linear path predicted by Theorem 1 within an additive $2k\pi$ ambiguity. Once an estimation of $\angle P_{v_0 v_1}(e^{j\omega})$ is calculated by the receiver, it can infer the TDOA by finding the value of D that maximizes the PHAse Transform integral

$$\int_{\omega} \cos\left(\angle P_{v_0 v_1}(e^{j\omega}) - (-\omega\frac{D}{2} + \frac{\pi}{2})\right) d\omega. \qquad (14)$$

The value of this integral for various values of D is plotted in Fig. 5(b). It is clearly seen that the PHAse Transform integral reaches its peak when $D = 12$. Thus, the time delay is estimated correctly.

Since the time delay in this example is even, the receiver can reconstruct the signal $x(n)$ perfectly using



(a) (b)

Figure 5: (a) Phase of $P_{v_0 v_1}(e^{j\omega})$ estimated using 1000 samples of $v_0(n)$ and $v_1(n)$. (b) The PHAse Transform plotted as a function of D.

the synthesis filters $F_0(z)$ and $F_1(z)$ obtained from (12) and (13) by setting $K = 1$ and $\theta_0 = 4.012$. Frequency response of the synthesis filters is shown in Fig. 4(a).

7. CONCLUDING REMARK

The synthesis system in Fig. 3 is not, in general, optimal for approximating $x(n)$ when the filters $H_0(z)$ and $H_1(z)$ are not perfect reconstruction. Specifying an optimal set of analysis and synthesis systems for this case is an open problem.

8. REFERENCES

[1] P. P. Vaidyanathan, *Multirate Sytems and Filter Banks*, Prentice-Hall, Upper Saddle River, NJ, 1993.

[2] P Aarabi and M Mahdavi, "The relation between speech segment selectivity and time-delay estimation accuracy," *Proceedings of the IEEE Conference on Acoustics, Speech and Signal Processing*, May 1996.

[3] M.S. Brandstein and H. Silverman, "A robust method for speech signal time-delay estimation in reverberant rooms," in *Proceedings of the IEEE Conference on Acoustics, Speech, and Signal Processing*, May 1996.

[4] C. H. Knapp and G. Carter, "The generalized correlation method for estimation of time delay," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. ASSP-24(4), pp. 320–327, Aug. 1976.

[5] T. Q. Nguyen and P. P. Vaidyanathan, "Two-channel perfect-reconstruction FIR QMF structures which yield linear-phase analysis and synthesis filters," *IEEE Transactions on Signal Processing*, vol. 37, no. 5, May 1989.