Chapter 11

MICROPHONE ARRAYS FOR VIDEO CAMERA STEERING

Yiteng (Arden) Huang
Center for Signal and Image Processing (CSIP)
School of Electrical and Computer Engineering
Georgia Institute of Technology
arden@ece.gatech.edu

Jacob Benesty
Bell Laboratories, Lucent Technologies
jbenesty@bell-labs.com

Gary W. Elko
Bell Laboratories, Lucent Technologies
gwe@bell-labs.com

Abstract In this chapter, we consider the problem of passively estimating the acoustic source location by using microphone arrays for video camera steering in real reverberant environments. Within a two-stage framework for this problem, different algorithms for time delay estimation and source localization are developed. Their performance as well as computational complexity are analyzed and discussed. A successful real-time system is also presented.

Keywords: Time Delay Estimation, Acoustic Source Localization, Reverberation, Eigenvalue Decomposition, Least Squares, Real-Time Implementation

1. INTRODUCTION

In current video-conferencing environments, a set of human-controlled cameras have to be set up in different locations to provide video of the active talker as he or she contributes to the discussion. Usually, this tedious task needs full

involvement of professional camera operators. Alternatively, acoustic and image object tracking techniques can be used to locate and track an active talker automatically in 3D space - determine his or her range, azimuth, as well as elevation. Thereafter, the need for one or more human camera operators can be eliminated and the practicability will be increased.

There are several possible choices for tracking an active talker. Broadly they can be dichotomized into the class of visual tracking and the class of acoustic tracking, depending on what particular information (visual or acoustic cues, respectively) is applied. Even though visual tracking techniques have been investigated for several decades and have had good success, acoustic source localization systems have some advantages that are not present in vision-based tracking systems. They receive acoustic signals omni-directionally and can act in the dark. Therefore they are able to detect and locate sound sources in the rear or sources that are hiding or occluded.

Humans, like most vertebrates, have two ears which form a microphone array, mounted on a *mobile* base (head). By continuously receiving and processing the propagating acoustic signals with such a binaural auditory system, we can accurately and instantaneously gather information about the environment, particularly about the spatial positions and trajectories of sound sources and about their states of activity. However, the extraordinary performance features demonstrated by our binaural auditory system form a big technical challenge for engineers mainly because of room reverberation. Microphone array processing is a rapidly emerging technique and, we believe, will play an important role in a practical solution.

Existing source localization methods can be loosely divided into three categories: steered beamformer-based, high-resolution spectral estimation-based, and time delay estimation-based locators [1]. With continued investigation over the last two decades, the time delay estimation-based locator has become the technique of choice, especially in recent digital systems. Here, we will concern ourselves strictly with the time delay estimation-based source localization techniques and the primary focus is on obtaining an optimal (in the sense of accuracy, robustness, and efficiency) source location estimator which can be implemented in real-time with a digital computer.

Time delay estimation-based localization systems determine the location of acoustic sources in a two-step process. In the first step, a set of time delay of arrivals (TDOAs) among different microphone pairs is calculated. For the second step, this set of TDOA information is then employed to estimate the acoustic source location with the knowledge of the microphone array geometry. Within such a system, errors are introduced in both steps. The errors incurred in the first step, such as those due to reverberation and quantization, will be propagated through the second step and reduce the overall accuracy of the whole system.

Extensive literature exists on the topic of time delay estimation. In Section 2, two different approaches suitable for real-time implementation will be developed. The first is the conventional approach which is based on cross-correlation calculation [2]. Even though the generalized cross-correlation (GCC) algorithms can be improved to achieve good performance in the presence of noise [3, 4], it has a fundamental drawback of inability to cope well with reverberation as shown clearly in [5]. Therefore, in Section 2.3, a more realistic solution is developed that tries to directly determine the relative delay between the direct paths of two estimated channel impulse responses. This approach, known as the adaptive eigenvalue decomposition algorithm (AEDA) [6, 7], performs reasonably well in a reverberant environment.

Locating point sources using measurements or estimates from passive, stationary sensor arrays has numerous applications in navigation, aerospace, and geophysics. Algorithms for radiative source localization have been studied for nearly 100 years, particularly for radar and underwater sonar systems. Many processing techniques have been proposed, with different complexity and restrictions. These methods, which are described in Section 3, include the maximum likelihood [8], triangulation [9], constrained least-squares, spherical intersection [10], spherical interpolation [11], and the one-step least-squares method [12].

Even though many algorithms for time delay estimation and subsequent source localization have been proposed, few real-time systems [9, 13, 14] have been presented. In Section 4, a successful real-time passive acoustic source localization system for video camera steering recently developed by the authors is presented. The system is based on the adaptive eigenvalue decomposition and one-step least-squares algorithms, and demonstrated the desired features of robustness, portability, and accuracy.

2. TIME DELAY ESTIMATION

Time delay estimation (TDE) is concerned with the computation of the relative time delay of arrival between different microphone sensors. It is a fundamental technique in microphone array signal processing and the first step of our passive acoustic source localization system. In general, there are two steps in developing a time delay estimation algorithm. The first is to choose an appropriate parametric model for the acoustic environment. In Section 2.1, two parametric acoustic models for TDE problems, namely ideal free-field and real reverberant models will be described in detail.

Once the model has been selected, the next step is to estimate the model parameters (here, the TDOAs) that provide *minimum* errors according to the received microphone signals. In this chapter, two classes of different approaches to TDE, suitable for real-time implementation, will be developed. The first,

presented in Section 2.2, is a generalized cross-correlation (GCC) method that selects the time delay that maximizes the cross-correlation function between signals of two distinct microphones, as its estimate. As we will see, this approach assumes the ideal free-field model, and performs poorly in a reverberant environment. In Section 2.3, a solution is derived from directly estimating the acoustic channel impulse responses and demonstrates more immunity to room reverberation.

2.1 ACOUSTIC MODELS FOR THE TDE PROBLEM

2.1.1 Ideal Free-Field Model. In an anechoic open space as shown in Fig. 11.1(a), the given source signal $s(n)$ propagates radiatively and the signal acquired by the i-th ($i = 1, 2$) microphone can be expressed as follows:

$$x_i(n) = \alpha_i s(n - \tau_i) + b_i(n), \tag{11.1}$$

where α_i is an attenuation factor due to propagation loss, τ_i is the propagation time, and $b_i(n)$ is additive noise. The time delay of arrival between the two microphone signals 1 and 2 is defined as,

$$\tau_{12} = \tau_1 - \tau_2. \tag{11.2}$$

Further assume that $s(n)$, $b_1(n)$, and $b_2(n)$ are zero-mean, uncorrelated, stationary Gaussian random processes. In this case, a mathematically clear solution for τ_{12} can be obtained from the ideal model that is widely used for the classical TDE problem.

2.1.2 Real Reverberant Model. The ideal free-field model is simple and only few parameters need to be determined. But unfortunately, in a real acoustic environment as shown in Fig. 11.1(b), we must take into account the reverberation of the room and the ideal model no longer holds. Then, a more complicated but more realistic model for the microphone signals $x_i(n)$, $i = 1, 2$, can be expressed as follows:

$$x_i(n) = g_i * s(n) + b_i(n), \tag{11.3}$$

where $*$ denotes convolution and g_i is the acoustic impulse response of the channel between the source and the i-th microphone. Moreover, $b_1(n)$ and $b_2(n)$ might be correlated, which is the case when the noise is directional, e.g., from a ceiling fan or an overhead projector.

For the real reverberant model, we do not have an "ideal" solution to the TDE problem, as for the previous model, unless we can accurately (and blindly) determine the two impulse responses, which is a very challenging problem.

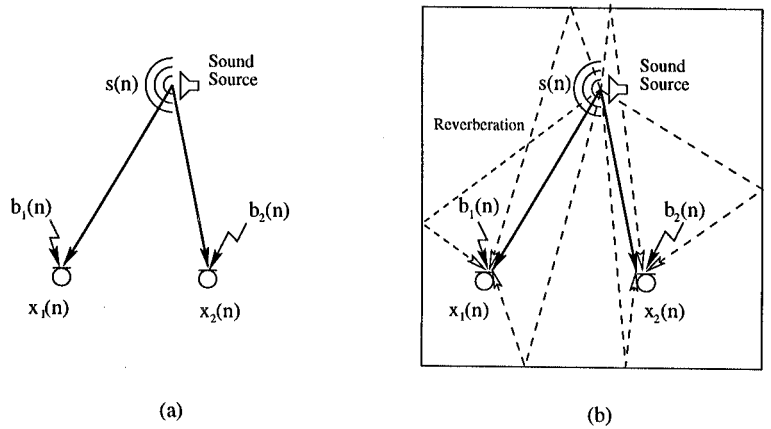

Figure 11.1 Acoustic models for time delay estimation problems. (a) Ideal free-field model. (b) Real reverberant model.

2.2 THE GCC METHOD

In the GCC technique, which is based on the ideal free-field model, the time-delay estimate is obtained as the value of τ that maximizes the generalized cross-correlation function given by,

$$\begin{aligned}\psi_{x_1x_2}(\tau) &= \int_{-\infty}^{+\infty} \Phi(f) S_{x_1x_2}(f) \mathrm{e}^{j2\pi f\tau} df \\ &= \int_{-\infty}^{+\infty} \Psi_{x_1x_2}(f) \mathrm{e}^{j2\pi f\tau} df, \end{aligned} \tag{11.4}$$

where $S_{x_1x_2}(f) = E\left\{X_1(f)X_2^*(f)\right\}$ is the cross-spectrum, $\Phi(f)$ is a weighting function, and

$$\Psi_{x_1x_2}(f) = \Phi(f) S_{x_1x_2}(f) \tag{11.5}$$

is the generalized cross-spectrum. The GCC TDE may be expressed as:

$$\hat{\tau}_\phi = \arg\max_\tau \psi_{x_1x_2}(\tau). \tag{11.6}$$

The choice of $\Phi(f)$ is important in practice. The classical cross-correlation (CCC) method is obtained by taking $\Phi(f) = 1$. In the noiseless case, knowing that $X_i(f) = \alpha_i e^{-j2\pi f\tau_i} S(f)$, $i = 1, 2$, we have:

$$\begin{aligned}\Psi_{x_1x_2}(f) &= \Psi_{cc}(f) = E\{X_1(f)X_2^\star(f)\} \\ &= \alpha_1\alpha_2 e^{-j2\pi f\tau_{12}} E\{|S(f)|^2\}. \end{aligned} \tag{11.7}$$

The fact that $\Psi_{cc}(f)$ depends on the source signal can be problematic for TDE.

It is clear by examining (11.6) that the phase rather than the magnitude of cross-spectrum provides the TDOA information. Thereafter, the cross-correlation peak can be sharpened by pre-whitening the input signals, i.e. choosing $\Phi(f) = 1/|S_{x_1x_2}(f)|$, which leads to the so-called phase transform (PHAT) method [2], [15]. Also in the absence of noise, the cross spectrum,

$$\Psi_{x_1x_2}(f) = \Psi_{\text{pt}}(f) = e^{-j2\pi f\tau_{12}} \tag{11.8}$$

depends only on the τ_{12} and thus can, in general, achieve better performance than CCC, especially when the input signal-to-noise ratio (SNR) is low.

GCC is simple and easy to implement but will fail when the reverberation becomes significant because the fundamental model assumptions are violated.

2.3 ADAPTIVE EIGENVALUE DECOMPOSITION ALGORITHM

The adaptive eigenvalue decomposition algorithm (AEDA) focuses directly on the channel impulse responses for TDE and assumes that the system (room) is linear and time invariant. By following the real reverberant model (11.3) and the fact that (see Fig. 11.2):

$$x_1 * g_2 = s * g_1 * g_2 = x_2 * g_1, \tag{11.9}$$

in the noiseless case, we have the following relation at time n [16]:

$$\mathbf{x}^T(n)\mathbf{u} = \mathbf{x}_1^T(n)\mathbf{g}_2 - \mathbf{x}_2^T(n)\mathbf{g}_1 = 0 \tag{11.10}$$

where, $\mathbf{x}_i$ is the signal picked up by the i-th ($i = 1, 2$) microphone, ^{T} denotes transpose, and

$$\begin{aligned}
\mathbf{x}_i(n) &= [x_i(n), x_i(n-1), \cdots, x_i(n-M+1)]^T, \\
\mathbf{g}_i &= [g_{i,0}, g_{i,1}, \cdots, g_{i,M-1}]^T, \quad i = 1, 2 \\
\mathbf{x}(n) &= \left[\mathbf{x}_1^T(n), \mathbf{x}_2^T(n)\right]^T, \\
\mathbf{u} &= \left[\mathbf{g}_2^T, -\mathbf{g}_1^T\right]^T,
\end{aligned}$$

and M is the length of the impulse responses.

Multiplying (11.10) by $\mathbf{x}(n)$ and taking expectation yields,

$$\mathbf{R}(n)\mathbf{u} = \mathbf{0}, \tag{11.11}$$

where $\mathbf{R}(n) = E\{\mathbf{x}(n)\mathbf{x}^T(n)\}$ is the covariance matrix of the microphone signals $\mathbf{x}(n)$. This implies that the vector $\mathbf{u}$ (containing the two impulse responses) is in the null space of $\mathbf{R}(n)$. More specifically, $\mathbf{u}$ is the eigenvector of the covariance matrix corresponding to the eigenvalue equal to 0. If noise is present, $\mathbf{u}$ is

estimated by minimizing $\mathbf{u}^T\mathbf{R}(n)\mathbf{u}$ subject to $\|\mathbf{u}\| = 1$. This is equivalent to finding $\mathbf{u}$ by computing the normalized eigenvector of $\mathbf{R}(n)$ corresponding to the smallest eigenvalue.

Before developing the adaptive eigenvalue decomposition algorithm to solve (11.11), it is worthwhile to expose the hidden factors that determine whether the acoustic channel impulse responses can be estimated, or if the acoustic channels are *identifiable*. From observation, $\mathbf{u}$ can be uniquely determined by solving (11.11) if and only if the covariance matrix $\mathbf{R}(n)$ is rank deficient by 1. Equivalently speaking, the information provided by the microphone signals is sufficient to uniquely identify the two acoustic channels if the following conditions hold [17]:

(I1) the polynomials $\mathbf{g}_i(m)$, $i = 1, 2$, $m = 0, ..., M-1$, are co-prime, or the channel transfer functions $G_i(z)$, $i = 1, 2$ do not share any common zeros,

(I2) the autocorrelation matrix of the source signal $s(n)$ is of full rank.

To avoid an ill-conditioned acoustic system, the two microphones can not be placed too close to each other (due to I1) and the analysis window should be long enough (due to I2). Both of the foregoing conditions are assumed in the AEDA and theoretically the unit-norm, non-zero solution to the normal equation $\mathbf{R}(n)\mathbf{u} = \mathbf{0}$ is uniquely equal to the desired acoustic channel impulse responses.

However, in practice, accurate estimation of the vector $\mathbf{u}$ is not trivial due to the nature of speech, the length of the impulse responses, the background noise, etc. However, for this application we only need to find an efficient way to detect the direct paths of the two impulse responses.

In order to efficiently estimate the eigenvector (here $\hat{\mathbf{u}}$) corresponding to the minimum eigenvalue of $\mathbf{R}(n)$, a constrained LMS algorithm [18] is often used. The error signal as illustrated in Fig. 11.2 is,

$$e(n) = \frac{\hat{\mathbf{u}}^T(n)\mathbf{x}(n)}{\|\hat{\mathbf{u}}(n)\|}, \tag{11.12}$$

and the constrained LMS algorithm may be expressed as,

$$\hat{\mathbf{u}}(n+1) = \hat{\mathbf{u}}(n) - \mu e(n)\nabla e(n), \tag{11.13}$$

where μ, the adaptation step, is a positive small constant, and

$$\nabla e(n) = \frac{1}{\|\hat{\mathbf{u}}(n)\|}\left[\mathbf{x}(n) - e(n)\frac{\hat{\mathbf{u}}(n)}{\|\hat{\mathbf{u}}(n)\|}\right]. \tag{11.14}$$

Substituting (11.12) and (11.14) in (11.13) and taking expectation after convergence gives:

$$\mathbf{R}\frac{\hat{\mathbf{u}}(\infty)}{\|\hat{\mathbf{u}}(\infty)\|} = E\{e^2(n)\}\frac{\hat{\mathbf{u}}(\infty)}{\|\hat{\mathbf{u}}(\infty)\|}, \tag{11.15}$$

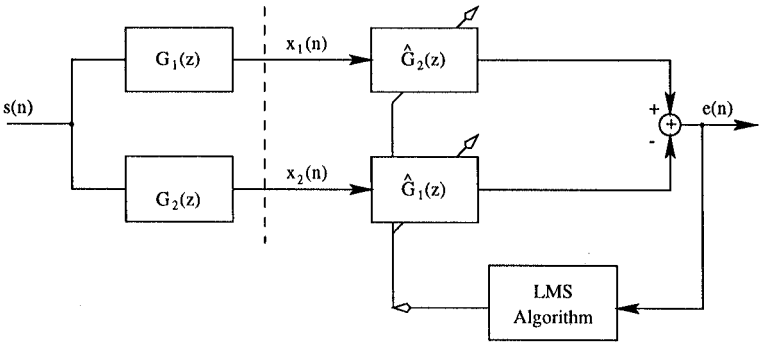

Figure 11.2 An adaptive filter for eigenvalue decomposition algorithm.

which is the desired result: $\hat{\mathbf{u}}$ converges in mean to the eigenvector of $\mathbf{R}$ corresponding to the smallest eigenvalue $E\{e^2(n)\}$. To avoid roundoff error propagation, normalization is imposed on the vector $\hat{\mathbf{u}}(n+1)$ after each update step. Finally the update equation is given by:

$$\hat{\mathbf{u}}(n+1) = \frac{\hat{\mathbf{u}}(n) - \mu e(n)\nabla e(n)}{\|\hat{\mathbf{u}}(n) - \mu e(n)\nabla e(n)\|}. \tag{11.16}$$

Note that if this normalization is used, then $\|\hat{\mathbf{u}}(n)\|$ (which appears in $e(n)$ and $\nabla e(n)$) can be removed, since we will always have $\|\hat{\mathbf{u}}(n)\| = 1$. If the smallest eigenvalue is equal to zero, which is the case here, the algorithm can be simplified as follows:

$$e(n) = \hat{\mathbf{u}}^T(n)\mathbf{x}(n), \tag{11.17}$$

$$\hat{\mathbf{u}}(n+1) = \frac{\hat{\mathbf{u}}(n) - \mu e(n)\mathbf{x}(n)}{\|\hat{\mathbf{u}}(n) - \mu e(n)\mathbf{x}(n)\|}. \tag{11.18}$$

Since the goal here is not to accurately estimate the two impulse responses g_1 and g_2 but rather the time delay, only the two direct paths are of interest. In order to take into account negative and positive relative delays, we initialize $\hat{u}_{M/2}(0) = 1$ which will be considered as an estimate of the direct path of g_2 and, during adaptation, keep it dominant in comparison with the other $M-1$ taps of the first half of $\hat{\mathbf{u}}(n)$ (containing an estimate of the impulse response g_2). A "mirror" effect will appear in the second half of $\hat{\mathbf{u}}(n)$ (containing an estimate of the impulse response $-g_1$): a negative peak will dominate which is an estimate of the direct path of $-g_1$. Thus the relative sample delay will be simply the difference between the indices corresponding to these two peaks.

To take advantage of the FFT, the filter coefficients are updated in the frequency domain in our real-time implementation using the unconstrained frequency-domain LMS algorithm [19]. The AEDA can be seen as a generalization of the LMS TDE proposed in [20].

3. SOURCE LOCALIZATION

In the TDE-based passive acoustic source localization system, the second step employs the set of TDOA estimates to calculate the location of the acoustic source. Since the equations describing source localization issues are highly nonlinear, the estimation and quantization errors introduced in the TDE step will be magnified more with some methods than others. In this section we first define the source localization problem. We then present the non-closed-form maximum likelihood locator, several closed-form locators, and their sensitivity to errors in TDOA. Detailed derivations of each localization method are developed and their performance features as well as computational complexities are analyzed.

3.1 SOURCE LOCALIZATION PROBLEM

The problem addressed here is the determination of the location of an acoustic source given the array geometry and relative TDOA measurements among different microphone pairs. The problem can be stated mathematically as follows:

A microphone array of $N+1$ microphones is located at,

$$\mathbf{r}_i \triangleq (x_i, y_i, z_i)^T, \ i = 0, ..., N \tag{11.19}$$

in Cartesian coordinates (see Fig. 11.3). The first microphone ($i = 0$) is regarded as the reference and placed at the origin of the coordinate system, i.e. $\mathbf{r}_0 = (0, 0, 0)$. The acoustic source is located at $\mathbf{r}_s \triangleq (x_s, y_s, z_s)^T$. The distances from the origin to the i-th microphone and the source are denoted by R_i and R_s, respectively.

$$R_i \triangleq \|\mathbf{r}_i\| = \sqrt{x_i^2 + y_i^2 + z_i^2}, \ i = 1, ..., N \tag{11.20}$$

$$R_s \triangleq \|\mathbf{r}_s\| = \sqrt{x_s^2 + y_s^2 + z_s^2}. \tag{11.21}$$

The distance between the source and the i-th microphone is denoted by,

$$D_i \triangleq ||\mathbf{r}_i - \mathbf{r}_s||_2 = \sqrt{(x_i - x_s)^2 + (y_i - y_s)^2 + (z_i - z_s)^2}\,. \tag{11.22}$$

The distance difference between microphones i and j from the source is given by,

$$d_{ij} \triangleq D_i - D_j, \quad i, j = 0, ..., N. \tag{11.23}$$

The difference is usually termed as the *range difference* and is proportional to the time delay of arrival τ_{ij} with the speed of sound c,

$$d_{ij} = c \cdot \tau_{ij}, \tag{11.24}$$

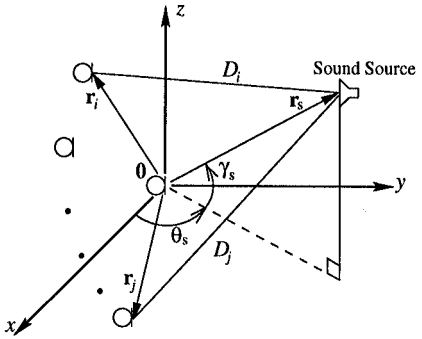

Figure 11.3 Spatial diagram illustrating notation defined in the source localization problem.

where, the speed of sound (in meters per second) can be estimated from the air temperature t_{air} (in degrees Celsius) according to the following approximate (first-order) formula,

$$c \approx 331 + 0.610 \times t_{\text{air}} \,. \tag{11.25}$$

The localization problem is then to estimate $\mathbf{r}_s$ given the set of $\mathbf{r}_i$ and τ_{ij}. Note that there are $\frac{(N+1)N}{2}$ distinct TDOA estimates τ_{ij} which exclude the case $i = j$ and count the $\tau_{ij} = -\tau_{ji}$ pair only once. However, in the absence of *noise*, the space spanned by such TDOA estimates is of N dimensions. Any N linearly independent TDOAs determine all the rest. In a noisy environment, the TDOA redundancy can be utilized to improve the accuracy of the localization algorithms, at the expense of increasing the computational complexity of the corresponding source localization algorithm. For simplicity and also without loss of generality, we choose $\tau_{i0}, i = 1, ..., N$ as the basis of such $\mathbb{R}^N$ space in this chapter.

3.2 IDEAL MAXIMUM LIKELIHOOD LOCATOR

Given a set of TDOA estimates $\tau = (\tau_{10}, \tau_{20}, ..., \tau_{N0})^T$, the likelihood function of a source location is the joint probability of these TDOA estimates conditioned on the source location, i.e.

$$\mathcal{L}(\mathbf{r}_s) = \Pr(\tau|\mathbf{r}_s) = \Pr(\tau_{10}, \tau_{20}, ..., \tau_{N0}|\mathbf{r}_s). \tag{11.26}$$

The maximum likelihood solution is then given by,

$$\hat{\mathbf{r}}_{s,\text{ML}} = \arg\max_{\mathbf{r}_s} \Pr(\tau|\mathbf{r}_s) = \arg\max_{\mathbf{r}_s} \Pr(\tau_{10}, \tau_{20}, ..., \tau_{N0}|\mathbf{r}_s). \tag{11.27}$$

By assuming that the additive noise in TDOAs is zero mean and jointly Gaussian distributed, the joint probability density function of τ conditioned on $\mathbf{r}_s$ is given

by,

$$f(\tau|\mathbf{r}_s) = \mathcal{N}(\tau;\ \tau_m(\mathbf{r}_s),\ \Sigma) = \frac{\exp\left\{-\frac{1}{2}[\tau - \tau_m(\mathbf{r}_s)]^T\ \Sigma^{-1}[\tau - \tau_m(\mathbf{r}_s)]\right\}}{\sqrt{(2\pi)^N \det(\Sigma)}}, \quad (11.28)$$

where, $\tau_m(\mathbf{r}_s)$ and Σ are respectively the mean vector and the covariance matrix of the estimated TDOAs, and det denotes determinant. In these parameters, only the mean vector $\tau_m(\mathbf{r}_s)$ is a function of the source location $\mathbf{r}_s$. The covariance matrix depends only on the noise signal.

Since Σ is positive definite, the maximum likelihood solution is equivalent to that minimizing the error function defined as,

$$\mathcal{E}(\mathbf{r}_s) \triangleq [\tau - \tau_m(\mathbf{r}_s)]^T\ \Sigma^{-1}[\tau - \tau_m(\mathbf{r}_s)]. \quad (11.29)$$

Direct estimation of the minimizer is not practicable. If the noise is assumed uncorrelated, the covariance matrix is diagonal:

$$\Sigma = \mathrm{diag}(\sigma_1^2, \sigma_2^2, ..., \sigma_N^2) \quad (11.30)$$

and the error function (11.29) can be rewritten as,

$$\mathcal{E}(\mathbf{r}_s) = \sum_{i=1}^{N} \frac{\left[\tau_{i0} - \tau_{i0,m}(\mathbf{r}_s)\right]^2}{\sigma_i^2}. \quad (11.31)$$

The steepest descent algorithm can be used to find $\hat{\mathbf{r}}_{s,ML}$ iteratively by,

$$\hat{\mathbf{r}}_s(n+1) = \hat{\mathbf{r}}_s(n) - \frac{1}{2}\mu \bigtriangledown \mathcal{E}(\hat{\mathbf{r}}_s(n)), \quad (11.32)$$

where μ is the step size.

The foregoing maximum likelihood locator is optimal in this problem only if the two assumptions made above hold. However, this is not the case in practice. The noise in the TDOA estimates comes from two sources: time delay estimation and quantization errors. The TDE error can be modeled as an additive Gaussian noise. For the quantization error, a uniformly distributed noise in $U[-T_s/2, T_s/2]$, where T_s is the sampling period, can certainly not be modeled as Gaussian noise. Moreover, in order to avoid local minima, we need to select a good initial guess of the source location, which is difficult to do in practice, and convergence of the iterative algorithm to the desired solution can not be guaranteed. Therefore, closed-form source localization methods have been studied.

3.3 TRIANGULATION LOCATOR

Of the closed-form source locators, triangulation is the most straightforward method. Consider the given set of TDOA estimates $\tau = (\tau_{10}, \tau_{20}, ..., \tau_{N0})^T$ and convert these into a range difference vector $\mathbf{d} = (d_{10}, d_{20}, ..., d_{N0})^T$ by using (11.24). Each range difference d_{i0} in turn defines a hyperboloid by,

$$D_i - D_0 = \|\mathbf{r}_i - \mathbf{r}_s\| - \|\mathbf{r}_0 - \mathbf{r}_s\| = d_{i0} \,. \tag{11.33}$$

All points on such hyperboloid have the same range difference d_{i0} to the two microphones 0 and i. Since the acoustic source location is intended to be determined in $\mathbb{R}^3$ space, there are three unknowns and $N = 3$ (4 microphones) will make the problem possible to solve. Then the acoustic source lies on three hyperboloids and satisfies the set of hyperbolic equations:

$$\begin{cases} \|\mathbf{r}_1 - \mathbf{r}_s\| - \|\mathbf{r}_0 - \mathbf{r}_s\| &= d_{10} \\ \|\mathbf{r}_2 - \mathbf{r}_s\| - \|\mathbf{r}_0 - \mathbf{r}_s\| &= d_{20} \\ \|\mathbf{r}_3 - \mathbf{r}_s\| - \|\mathbf{r}_0 - \mathbf{r}_s\| &= d_{30} \end{cases} \tag{11.34}$$

This equation set is highly nonlinear. If microphones are arbitrarily arranged, a closed-form solution may not exist [21] and numerical methods must be used.

From a geometric point of view, the triangulation locator finds the source location by intersecting three hyperboloids. This approach is easy to implement. However, determinant (the number of unknowns is equal to that of equations) characteristics of the triangulation algorithm makes it very sensitive to noise in TDOAs. Small errors in TDOA can deviate the estimated source far from the true location.

3.4 THE SPHERICAL EQUATIONS

In order to avoid solving the set of hyperbolic equations (11.34) whose solution is very sensitive to noise, the source localization problem can be reorganized into a set of spherical equations. The acoustic source is potentially located on a group of spheres centered at the microphones.

Consider the distance from the i-th microphone to the acoustic source. From the definition of the range difference (11.23) and the fact that $D_0 = R_s$, we have:

$$D_i = R_s + d_{i0}. \tag{11.35}$$

From the Pythagorean theorem, D_i^2 can also be written as,

$$D_i^2 = \|\mathbf{r}_i - \mathbf{r}_s\|^2 = R_i^2 - 2\mathbf{r}_i^T\mathbf{r}_s + R_s^2 \,. \tag{11.36}$$

Substituting (11.35) into (11.36) yields,

$$(R_s + d_{i0})^2 = R_i^2 - 2\mathbf{r}_i^T\mathbf{r}_s + R_s^2 \,, \tag{11.37}$$

or equivalently,

$$\mathbf{r}_i^T \mathbf{r}_s + d_{i0} R_s = \frac{1}{2}(R_i^2 - d_{i0}^2), \quad i = 1, ..., N . \tag{11.38}$$

Putting the N equations together and writing them into a matrix form,

$$\mathbf{A}\,\theta = \mathbf{b}, \tag{11.39}$$

where,

$$\mathbf{A} \triangleq [\mathbf{S} \mid \mathbf{d}], \quad \mathbf{S} \triangleq \begin{bmatrix} x_1 & y_1 & z_1 \\ x_2 & y_2 & z_2 \\ & \vdots & \\ x_N & y_N & z_N \end{bmatrix},$$

$$\theta \triangleq \begin{bmatrix} x_s \\ y_s \\ z_s \\ R_s \end{bmatrix}, \quad \mathbf{b} \triangleq \frac{1}{2}\begin{bmatrix} R_1^2 - d_{10}^2 \\ R_2^2 - d_{20}^2 \\ \vdots \\ R_N^2 - d_{N0}^2 \end{bmatrix}.$$

By introducing one supplemental variable, the source range R_s, the source localization problem is *linearized.* The spherical equations are linear in $\mathbf{r}_s$ given R_s and vice versa. This set of the spherical equations can be expanded into an over-determined system with ease. Therefore, the locators that try to solve (11.39) can take advantage of redundant microphone signals to achieve more accurate estimation performance without dramatically increasing their computational complexity.

3.5 CLS AND SPHERICAL INTERSECTION (SX) METHODS

In Section 3.4, we have reorganized the source localization problem into a linear system by introducing a supplemental variable. Therefore, the source location can be obtained by solving the spherical equations (11.39) subject to $\|\mathbf{r}_s\| = R_s$. This is in fact a constrained least-squares (CLS) problem. A Lagrangian can be written as,

$$J_c \triangleq (\mathbf{b} - \mathbf{A}\,\theta)^T(\mathbf{b} - \mathbf{A}\,\theta) + \lambda\left(\theta^T \Lambda\,\theta\right), \tag{11.40}$$

where $\Lambda = \text{diag}[1, 1, 1, -1]$ is a diagonal matrix. Taking the gradient with respect to θ and equating the result to zero yields:

$$\theta_c = [\mathbf{A}^T\mathbf{A} + \lambda\,\Lambda]^{-1}\mathbf{A}^T\mathbf{b}, \tag{11.41}$$

from which the quadratic constraint $\theta^T \Lambda \theta = 0$ becomes,

$$\mathbf{b}^T\mathbf{A}[\mathbf{A}^T\mathbf{A} + \lambda \Lambda]^{-1} \Lambda[\mathbf{A}^T\mathbf{A} + \lambda \Lambda]^{-1}\mathbf{A}^T\mathbf{b} = 0. \tag{11.42}$$

This quadratic equation in the unknown λ can be solved only numerically using iterative methods which are computationally intensive in real-time implementation.

The SX locator solves the problem in two steps. It first finds the least-squares solution for $\mathbf{r}_s$ in terms of R_s,

$$\mathbf{r}_s = \mathbf{S}^\dagger(\mathbf{b} - R_s\mathbf{d}), \tag{11.43}$$

where,

$$\mathbf{S}^\dagger = (\mathbf{S}^T\mathbf{S})^{-1}\mathbf{S}^T.$$

Then, substituting (11.43) into the constraint $R_s^2 = \mathbf{r}_s^T\mathbf{r}_s$, yields the quadratic equation,

$$R_s^2 = \left[\mathbf{S}^\dagger(\mathbf{b} - R_s\mathbf{d})\right]^T\left[\mathbf{S}^\dagger(\mathbf{b} - R_s\mathbf{d})\right]. \tag{11.44}$$

After expansion,

$$aR_s^2 + bR_s + c = 0, \tag{11.45}$$

where,

$$a = 1 - \|\mathbf{S}^\dagger\mathbf{d}\|^2,\ b = 2\mathbf{b}^T\mathbf{S}^{\dagger T}\mathbf{S}^\dagger\mathbf{d},\ c = -\|\mathbf{S}^\dagger\mathbf{b}\|^2.$$

The positive real root is taken as an estimate of the source range $\hat{R}_s$ and is then substituted into (11.43) to calculate the SX estimate of the source location $\hat{\mathbf{r}}_{s,SX}$.

In the SX procedure, the solution of the quadratic equation (11.45) for the source depth R_s is required. This solution must be a positive value by all means. If no real positive root is available, the SX solution does not *exist*. On the other hand, if both of the roots are real and greater than 0, then the SX solution is not *unique*. In both cases, the SX locator fails to provide reliable estimate for source location. This is not desirable for real-time operation.

3.6 SPHERICAL INTERPOLATION (SI) LOCATOR

In order to overcome the drawback of spherical intersection, a spherical interpolation locator was proposed in [22] which attempts to relax the restriction $\hat{R}_s = \|\hat{\mathbf{r}}_s\|$ by estimating R_s in the least-squares sense.

To begin, substitute the least-squares solution (11.43) into the original spherical equations (11.39) to obtain,

$$R_s\mathbf{P}_{S^\perp}\mathbf{d} = \mathbf{P}_{S^\perp}\mathbf{b}, \tag{11.46}$$

where,

$$\mathbf{P}_{\mathbf{S}^\perp} \triangleq \mathbf{I} - \mathbf{S}(\mathbf{S}^T\mathbf{S})^{-1}\mathbf{S}^T, \tag{11.47}$$

and $\mathbf{I}$ is an $N \times N$ identity matrix. The $\mathbf{P}_{\mathbf{S}^\perp}$ matrix projects a vector onto a space which is orthogonal to the column space of $\mathbf{S}$. It is symmetric ($\mathbf{P}_{\mathbf{S}^\perp} = \mathbf{P}_{\mathbf{S}^\perp}^T$) and idempotent ($\mathbf{P}_{\mathbf{S}^\perp} = \mathbf{P}_{\mathbf{S}^\perp}^2$). Then the least-squares solution to (11.46) is given by,

$$\hat{R}_{\mathrm{s,SI}} = \frac{\mathbf{d}^T\mathbf{P}_{\mathbf{S}^\perp}\mathbf{b}}{\mathbf{d}^T\mathbf{P}_{\mathbf{S}^\perp}\mathbf{d}}. \tag{11.48}$$

Substituting this solution into (11.43) yields the spherical interpolation estimate,

$$\hat{\mathbf{r}}_{\mathrm{s,SI}} = \left(\mathbf{S}^T\mathbf{S}\right)^{-1}\mathbf{S}^T\left[\mathbf{I}_N - \left(\frac{\mathbf{d}\mathbf{d}^T\mathbf{P}_{\mathbf{S}^\perp}}{\mathbf{d}^T\mathbf{P}_{\mathbf{S}^\perp}\mathbf{d}}\right)\right]\mathbf{b}. \tag{11.49}$$

In practice, the SI locator performs better, but is computationally more complex than the SX locator.

3.7 ONE STEP LEAST SQUARES (OSLS) LOCATOR

The SI method tries to solve the spherical equations (11.39) in two separate steps for the source range and its location, both calculated in the least-squares sense. Since the algorithm needs a significant amount of matrix multiplications to be performed, it is not computationally efficient especially when more microphones are applied (N gets larger). In this section, we will simplify the procedure by a one-step least-squares (OSLS) method and show that the OSLS method generates the same results as the SI method while dramatically decreasing the computational complexity, which is desirable for real-time implementations.

The least-squares solution of (11.39) for θ (the source location as well as its range) is given by:

$$\hat{\theta}_{\mathrm{OSLS}} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{b}, \tag{11.50}$$

or written into block form as,

$$\hat{\theta}_{\mathrm{OSLS}} = \left[\begin{array}{c|c} \mathbf{S}^T\mathbf{S} & \mathbf{S}^T\mathbf{d} \\ \hline \mathbf{d}^T\mathbf{S} & \mathbf{d}^T\mathbf{d} \end{array}\right]^{-1}\left[\begin{array}{c} \mathbf{S}_{3\times N}^T \\ \hline \mathbf{d}^T \end{array}\right]\mathbf{b}. \tag{11.51}$$

First, write the matrix that appears in (11.51) as follows:

$$\left[\begin{array}{c|c} \mathbf{S}^T\mathbf{S} & \mathbf{S}^T\mathbf{d} \\ \hline \mathbf{d}^T\mathbf{S} & \mathbf{d}^T\mathbf{d} \end{array}\right]^{-1} = \left[\begin{array}{c|c} \mathbf{Q} & \mathbf{v} \\ \hline \mathbf{v}^T & k \end{array}\right], \tag{11.52}$$

where,

$$\begin{aligned}\mathbf{v} &= -\left(\mathbf{S}^T\mathbf{S} - \frac{\mathbf{S}^T\mathbf{d}\mathbf{d}^T\mathbf{S}}{\mathbf{d}^T\mathbf{d}}\right)^{-1}\frac{\mathbf{S}^T\mathbf{d}}{\mathbf{d}^T\mathbf{d}},\\ \mathbf{Q} &= \left(\mathbf{S}^T\mathbf{S}\right)^{-1}\left[\mathbf{I} - \left(\mathbf{S}^T\mathbf{d}\right)\mathbf{v}^T\right],\\ k &= \frac{1-(\mathbf{d}^T\mathbf{S})\mathbf{v}}{\mathbf{d}^T\mathbf{d}}.\end{aligned}$$

Define the projection matrix onto the **d**-orthogonal space:

$$\mathbf{P}_{\mathbf{d}^\perp} \triangleq \mathbf{I} - \frac{\mathbf{d}\mathbf{d}^T}{\mathbf{d}^T\mathbf{d}}, \tag{11.53}$$

and find,

$$\mathbf{v} = -\left(\mathbf{S}^T\mathbf{P}_{\mathbf{d}^\perp}\mathbf{S}\right)^{-1}\frac{\mathbf{S}^T\mathbf{d}}{\mathbf{d}^T\mathbf{d}}, \tag{11.54}$$

$$\mathbf{Q} = \left(\mathbf{S}^T\mathbf{P}_{\mathbf{d}^\perp}\mathbf{S}\right)^{-1}. \tag{11.55}$$

Substituting (11.52) with (11.54) and (11.55) into (11.51) yields the OSLS estimate,

$$\hat{\mathbf{r}}_{s,\mathrm{OSLS}} = \left(\mathbf{S}^T\mathbf{P}_{\mathbf{d}^\perp}\mathbf{S}\right)^{-1}\mathbf{S}^T\mathbf{P}_{\mathbf{d}^\perp}\mathbf{b}, \tag{11.56}$$

which is the minimizer of

$$J_{\mathrm{OSLS}}(\mathbf{r}_s) = \|\mathbf{P}_{\mathbf{d}^\perp}\mathbf{b} - \mathbf{P}_{\mathbf{d}^\perp}\mathbf{S}\mathbf{r}_s\|, \tag{11.57}$$

or the least-squares solution to the linear equations,

$$\mathbf{P}_{\mathbf{d}^\perp}\mathbf{S}\mathbf{r}_s = \mathbf{P}_{\mathbf{d}^\perp}\mathbf{b}. \tag{11.58}$$

In fact, the OSLS algorithm tries to approximate the projection of the observation vector **b** with the projections of the column vectors of the microphone location matrix **S** onto the **d**-orthogonal space. The source location estimate is the coefficient vector associated with the *best* approximation. Clearly from (11.58), the OSLS algorithm is the generalization of the linear intersection method proposed in [23].

By using the Sherman-Morrison formula,

$$\left(\mathbf{A} + \mathbf{x}\mathbf{y}^T\right)^{-1} = \mathbf{A}^{-1} - \frac{\mathbf{A}^{-1}\mathbf{x}\mathbf{y}^T\mathbf{A}^{-1}}{1 + \mathbf{y}^T\mathbf{A}^{-1}\mathbf{x}}, \tag{11.59}$$

the item $\left(\mathbf{S}^T\mathbf{P}_{\mathbf{d}^\perp}\mathbf{S}\right)^{-1} = \left(\mathbf{S}^T\mathbf{S} - (\mathbf{S}^T\mathbf{d})(\mathbf{S}^T\mathbf{d})^T/(\mathbf{d}^T\mathbf{d})\right)^{-1}$ in (11.56) is expanded and finally it can be shown that the OSLS solution (11.56) is equivalent to the SI estimate (11.49), i.e. $\hat{\mathbf{r}}_{s,\mathrm{OSLS}} \equiv \hat{\mathbf{r}}_{s,\mathrm{SI}}$.

Now, we consider the computational complexity for the SI and OSLS methods. To calculate the inverse of an $M \times M$ matrix by using the Gauss-Jordan method without *pivoting*, the numbers of necessary scalar multiplications and additions are given by,

$$\text{Mul}_{\text{mi}} = \frac{4}{3}M^3 - \frac{2}{3}M, \quad \text{Add}_{\text{mi}} = \frac{4}{3}M^3 - \frac{3}{2}M^2 - \frac{1}{6}M. \tag{11.60}$$

To multiply matrix $\mathbf{X}_{p\times k}$ with matrix $\mathbf{Y}_{k\times q}$, the numbers of scalar multiplications and additions are,

$$\text{Mul}_{\text{mm}} = pqk, \quad \text{Add}_{\text{mm}} = pq(k-1). \tag{11.61}$$

For the SI locator (11.49), many matrix multiplications and one 3×3 matrix inverse need to be performed. Estimating the source location requires,

$$\text{Mul}_{\text{SI}} = N^3 + 6N^2 + 22N + 35 \sim \text{O}(N^3), \tag{11.62}$$

$$\text{Add}_{\text{SI}} = N^3 + 4N^2 + 15N + 11 \sim \text{O}(N^3), \tag{11.63}$$

multiplications and additions, respectively. Both are on the order of $\text{O}(N^3)$. However, for the OSLS locator (11.50), only three matrix multiplications and one 4×4 matrix inverse are required. The numbers of scalar multiplications and additions are found as,

$$\text{Mul}_{\text{OSLS}} = 36N + 84 \sim \text{O}(N), \tag{11.64}$$

$$\text{Add}_{\text{OSLS}} = 32N + 42 \sim \text{O}(N), \tag{11.65}$$

which are on the order of only $\text{O}(N)$. When $N \geq 4$ (5 microphones), the OSLS method is more computationally efficient than the SI method.

4. SYSTEM IMPLEMENTATION

A real-time passive acoustic source localization system for video camera steering has been developed recently by the authors at Bell Labs, Lucent Technologies. This system consists of a front-end 6-element microphone array with pre-amps, a Sonorus AUDI/O™ AD/24 8-channel A/D converter, a video camera, and a Pentium III™ 500 MHz PC equipped with Sonorus STUDI/O™ 16-channel digital audio interface board and video capture card. A block diagram of the system infrastructure is shown in Fig. 11.4.

In this system, the microphone array, shown in Fig. 11.5, was designed according to our simulation results. This array uses six Lucent Speech Tracker Directional™ hypercardioid microphones. The reference microphone 0 is located at the origin of the coordinate. Microphones 1, 2, and 4 are placed with equal distance of 50 centimeters from the origin on the $x - y$ (horizontal) plane. For symmetric performance to the horizontal plane, two more microphones, 3

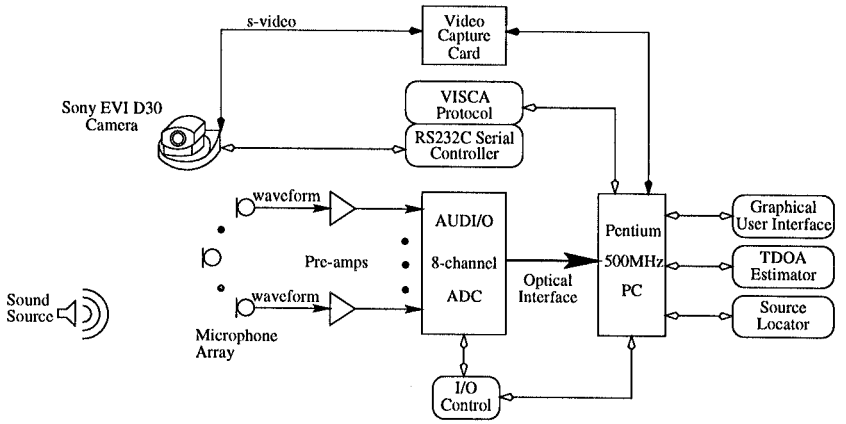

Figure 11.4 Schematic block diagram of the real-time system infrastructure.

and 5, are installed on z and $-z$ axis, both with 40 centimeters from the origin. In order to achieve desired portability, the microphone array is integrated with a Sony EVI-D30 video camera by using a frame. This camera, with pan, tilt, and zoom controls, allows for optimum coverage of a normal conference room. It has two high speed motors for pan and tilt. Pan, tilt, and zoom operation can be performed at the same time. Therefore, the camera is able to capture the full interaction of video-conference participants at remote locations.

The software for this system is running on Microsoft Windows98™. It consists of five major modules: acoustic signal I/O control, time delay estimator, source locator, graphical user interface, and video camera communications protocols. The acoustic signal is sampled at 8.82 kHz and quantized with 16 bits per sample. Five TDOAs were then estimated by the AEDA algorithm. The AEDA estimator uses a fixed step size which is $\mu = 0.01$ and the length of the adaptation vector $\mathbf{u}(n)$ is $L = 2M = 1024$. These TDOAs are supplied to the OSLS source locator. Benefiting from the efficient time delay estimator and source locator, the overall system has a relatively small computational requirement that can be easily met by the general-purpose Pentium III™ processor without any additional specific digital signal processor. The source location estimate is updated at 8 times/sec and used to drive the video camera. The PC communicates with the camera through two layers of protocol, namely RS232C serial control and the Video System Control Architecture (VISCA™) protocol. Finally, the video output is fed into the PC through video capture card and displayed on the system monitor.

This system has been tested by many people in two different rooms. One is a normal laboratory and the other is a conference room at Bell Labs. The laboratory has some computers and is much noisier than the conference room.

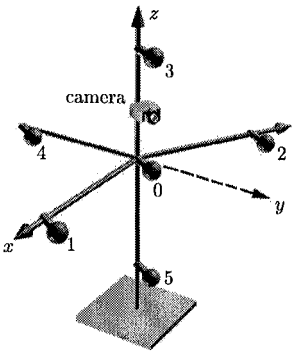

Figure 11.5 Three-dimensional microphone array for passive acoustic source localization.

The system can provide a quality video sequence in both environments and demonstrates robustness to noise. The speaker could be tracked with ease if he or she kept talking while moving. If the sound source is fixed at one arbitrary position, the system can drive the camera to focus on that position with high accuracy.

5. SUMMARY

In this chapter, we have looked at passive acoustic source localization techniques for video camera steering using a microphone array. The first part of the chapter was concerned with time delay estimation between distinct microphone sensor pairs. Beginning with the traditional generalized cross-correlation method, we found that this approach was of limited use since it did not take into account room reverberation and, as a result, performs poorly in a reverberant environment. We then derived the adaptive eigenvalue decomposition algorithm, which tried to directly estimate the speaker-microphone channel impulse response. This is a more realistic signal propagation channel model and the results demonstrated more immunity to noise and reverberation.

In the second part of the chapter, we considered the problem of determining the location of the sound source given the time delays of arrival and array geometry. A number of different approaches, including maximum likelihood estimation and several closed-form solutions, were derived. It was shown that the spherical equations could be used to localize the source. The least-squares solutions to the spherical equations were more robust to noise in the TDOAs than that to the hyperbolic equations. Even though all of the developed methods are somewhat limited, the one-step least-squares algorithm is favored over the others in terms of robustness, accuracy, and computational complexity.

In the last section of this chapter, we looked briefly at a real-time passive acoustic source localization system for video camera steering. The system ap-

plied the adaptive eigenvalue decomposition algorithm for time delay estimation and the one-step least-squares algorithm for source localization.

References

[1] M. S. Brandstein and H. F. Silverman, "A practical methodology for speech source localization with microphone arrays," *Comput., Speech, Language*, vol. 2, pp. 91-126, Nov. 1997.

[2] C. H. Knapp and G. C. Carter, "The generalized correlation method for estimation of time delay," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-24, pp. 320-327, Aug. 1976.

[3] M. S. Brandstein, "A pitch-based approach to time-delay estimation of reverberant speech," in *Proc. IEEE ASSP Workshop Appls. Signal Processing Audio Acoustics*, 1997.

[4] M. Bodden, "Modeling human sound-source localization and the cocktail-party-effect," *Acta Acoustica* 1, pp. 43-55, 1993.

[5] B. Champagne, S. Bédard, and A. Stéphenne, "Performance of time-delay estimation in the presence of room reverberation," *IEEE Trans. Speech Audio Processing*, vol. 4, pp. 148-152, Mar. 1996.

[6] Y. Huang, J. Benesty, and G. W. Elko, "Adaptive eigenvalue decomposition algorithm for realtime acoustic source localization system," in *Proc. IEEE ICASSP*, 1999, vol. 2, pp. 937-940.

[7] J. Benesty, "Adaptive eigenvalue decomposition algorithm for passive acoustic source localization," *J. Acoust. Soc. Am.*, vol. 107, pp. 384-391, Jan. 2000.

[8] M. Wax and T. Kailath, "Optimum localization of multiple sources by passive arrays," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-31, pp. 1210-1218, Oct. 1983.

[9] H. Wang and P. Chu, "Voice source localization for automatic camera pointing system in videoconferencing," in *Proc. IEEE ASSP Workshop Appls. Signal Processing Audio Acoustics*, 1997.

[10] H. C. Schau and A. Z. Robinson, "Passive source localization employing intersecting spherical surfaces from time-of-arrival differences," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-35, pp. 1223-1225, Aug. 1987.

[11] J. O. Smith and J. S. Abel, "Closed-form least-squares source location estimation from range-difference measurements," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-35, pp. 1661-1669, Dec. 1987.

[12] Y. Huang, J. Benesty, and G. W. Elko, "Passive acoustic source localization for video camera steering," Lucent Bell Laboratories technical memorandum, 1999.

[13] D. V. Rabinkin, R. J. Ranomeron, J. C. French, and J. L. Flanagan, "A DSP Implementation of Source Location Using Microphone Arrays," in *Proc. SPIE* 2846, 1996, pp. 88-99.

[14] C. Wang and M. S. Brandstein, "A hybrid real-time face tracking system," in *Proc. IEEE ICASSP*, 1998.

[15] M. Omologo and P. Svaizer, "Acoustic source location in noisy and reverberant environment using CSP analysis," in *Proc. IEEE ICASSP*, 1996, pp. 921-924.

[16] J. Benesty, F. Amand, A. Gilloire, and Y. Grenier, "Adaptive filtering algorithms for stereophonic acoustic echo cancellation," in *Proc. IEEE ICASSP*, 1995, pp. 3099-3102.

[17] G. Xu, H. Liu, L. Tong, and T. Kailath, "A least-squares approach to blind channel identification," *IEEE Trans. Signal Processing*, vol. 43, pp. 2982-2993, Dec. 1995.

[18] O. L. Frost, III, "An algorithm for linearly constrained adaptive array processing," *Proc. of the IEEE*, vol. 60, pp. 926-935, Aug. 1972.

[19] D. Mansour and A. H. Gray, JR., "Unconstrained frequency-domain adaptive filter," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-30, pp. 726-734, Oct. 1982.

[20] D. H. Youn, N. Ahmed, and G. C. Carter, "On using the LMS algorithm for time delay estimation," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-30, pp. 798-801, Oct. 1982.

[21] M. S. Brandstein, J. E. Adcock, and H. F. Silverman, "A closed-form method for finding source locations from microphone-array time-delay estimates," in *Proc. IEEE ICASSP*, 1995, pp. 3019-3022.

[22] J. S. Abel and J. O. Smith, "The spherical interpolation method for closed-form passive source localization using range difference measurements," in *Proc. IEEE ICASSP*, 1987, pp. 471-474.

[23] R. O. Schmidt, "A new approach to geometry of range difference location," *IEEE Trans. Aerosp. Electron.*, vol. AES-8, pp. 821-835, Nov. 1972.