Chapter 4

Performance of Combined Localization and Tracking

4.1 Introduction

Locating acoustic sources using a sensor array is an important task in hands-free telephony and video conferencing. Room environments, however, pose a major challenge to source localization algorithms, since they corrupt sound signals with reverberation. In this chapter we determine the extent of improvement that can be obtained by fusing the localization algorithm with a source tracking algorithm.

Localization algorithms can be categorized as one of three different types: (i) steered beamformer techniques that function by maximizing the steered response power of a beamformer [94], (ii) time-delay estimation (TDE) based methods and (iii) subspace-based methods.

Unfortunately, the performance of all three algorithms fall significantly in moderately reverberant environments. Subspace-based methods are virtually unusable. Time-delay estimation methods are widely used in acoustic applications but still suffer considerably from reverberation. The maximum likelihood generalized cross correlation time-delay estimation technique performs poorly [10]. Some improvement has been made by using more robust frequency weighting functions such as the phase transform [55], least squares fitting of phase [11] and cepstral prefiltering [93]. More promising is the adaptive eigenvalue decomposition algorithm [6], which computes the time delay from estimates of the acoustic impulse response. Steered beamforming methods are also often used in source localization [15,94] and suffer similar problems to reverberation [108].

Recently, significant improvement to source localization algorithms has been seen by fusing the location estimator with a source tracking algorithm. By incorporating a dynamic model of source motion, source tracking provides a way of rejecting outliers. Sturim et al. [95] achieved some outlier rejection by Kalman filtering. Ward et al. [108] saw significant improvement by source tracking with a
particle filtering algorithm. Unlike Kalman filtering, which presumes the distribution of location estimates is uni-modal, the success of particle filter lie in recognizing that the distribution is multi-modal.

Other improvements to localization have been shown by incorporating a signal model, such as the harmonic structure of voiced speech into the localization algorithm [10], by using a hybrid of steered beamforming and time-delay estimation [14] and by using clustering of location estimates for outlier rejection [17].

In this chapter, we quantify the improvement in reverberant rooms that is possible by combining source localization with tracking. We study this improvement achievable using steered beamforming localization, as a function of room acoustic parameters and the beamformer directivity. We show that increased beamformer directivity, though improving localization estimates, also retards the performance of the tracking algorithm.

Previewing the chapter contents, in Section 4.2 we overview the analysis approach. In Section 4.3 we describe the soundfield and signal models. In Section 4.4 we define a generic algorithm for steered beamforming. In Section 4.5 we describe a general way of analyzing source tracking of bearing estimates. In Section 4.6 we perform the analysis, deriving an expression for the probability of anomalous position estimation. In Section 4.7 we explore a principle for extending range of applicability of the analysis results. Finally, in Section 4.8 we apply the results to steered beamforming examples.

### 4.2 Overview of Analysis Approach

The top level operation of a system for combined localization and tracking is illustrated in Figure 4.1. A sound source moves in the farfield of a sensor array, possessing a bearing angle $\phi_s(t)$ at each time $t$. Sound from the source, comprising of direct sound and reverberant reflections, is captured by a microphone array and passed to a steered beamforming source localization algorithm. The output of this algorithm is the angular spectrum $A(\theta; \omega, t)$ that quantifies the sound energy coming from each direction $\theta$ of angular frequency $\omega$ at time $t$. The source tracking algorithm processes the angular spectra using a model of sound source motion, to produce an estimate of the source bearing $\hat{\phi}_s(\omega, t)$.

It is clear from the literature that tracking can improve the accuracy of source localization [17, 95, 108]. The central issue addressed in this chapter is in establishing limits on how well tracking can improve it. Because the sound pressure between closely spaced points in a soundfield is correlated, the angular spectra estimated from closely-spaced source locations will be similar. This diminishes the utility of source tracking.

Depicted in Figure 4.1 is localization of a moving sound source. Ideally we would
Figure 4.1: Combined localization and tracking to a sound source in the farfield with a microphone array.

consider this problem, but due to the challenges in modelling the reverberant field created by a moving source, we analyze the parallel problem of localizing a fixed source with a moving sensor array. Extension to the moving source case is discussed in Section 4.7.

We perform the analysis in this chapter using several simplifications: (i) steered beamforming is used for location tracking, (ii) the algorithms presented are narrow-band, allowing study of dependence on frequency\(^1\) and (iii) Analysis is restricted to the two dimensional case with the sound source lying in the farfield\(^2\). The steered beamformer is chosen for analysis because of its linear behavior.

We commence analysis in the next section by deriving the model for the beamformer output signal.

### 4.3 Signal Model

Consider an array of \(M\) omnidirectional microphones with each positioned at \(x_1(t), x_2(t), \ldots, x_L(t)\) and moving with time \(t\). An acoustic source located in

\(^1\)It is beneficial for localization algorithms to exploit the wide-band nature of acoustic signals. However to avoid broadband design issues, we leave the performance a function of frequency.

\(^2\)The analysis however has obvious extension to the three dimensional case.
the farfield transmits the signal \( s(t) \). The signal captured at each sensor \( m_\ell(t) \) is:

\[
m_\ell(t) = h(\mathbf{x}_\ell(t), t) \ast s(t) + \eta_\ell(t),
\]

where \( \ast \) is the time varying convolution operator defined by:

\[
h(\mathbf{x}_\ell(t), t) \ast s(t) = \int_{-\infty}^{\infty} h(\mathbf{x}_\ell(t - \tau), \tau) s(t - \tau) d\tau,
\]

\( h(\mathbf{x}_\ell(t), \tau) \) is the acoustic impulse response at time \( \tau \) between the source and an omnidirectional sensor \( \ell \) positioned at \( \mathbf{x}_\ell(t) \), and \( \eta_\ell(t) \) is additive noise. In contrast to classical work, the focus of this chapter is on the reverberant term of \((4.1)\). We assume the additive noise term \( \eta_\ell(t) \) is negligible.

For each sensor signal \( m_\ell(t) \), create \( P \) frames of signal output of duration \( 2t_F \) sampled at each time \( t_p \), \( m_\ell(t) : t \in [t_p - t_F, t_p + t_F] \). To simplify the analysis, we reduce the system \( h(\mathbf{x}_\ell(t), t) \) to a time invariant system over each frame by assuming each sensor position is fixed (i) over the duration of the frame and (ii) for a duration of one reverberation time preceding. In that case, only one acoustic impulse response \( h(\mathbf{x}_\ell(t_p), \tau) \) contributes to the frame and \((4.3)\) reduces to:

\[
m_\ell(t) = \int_{-\infty}^{\infty} h(\mathbf{x}_\ell(t_p), \tau) s(t - \tau) d\tau, \quad t \in [t_p - t_F, t_p + t_F].
\]

Because the sensors move, we consider a time-frequency analysis of captured sound signals, defining the short time Fourier transform (STFT):

\[
X(\omega, t) = \int_{-\infty}^{\infty} g(\tau - t) x(\tau) e^{-i\omega \tau} d\tau,
\]

where \( g(t) \) is a window function and \( \omega \) is angular frequency. Computing the STFT of both sides of \((4.2)\) for \( t \in [t_p - t_F, t_p + t_F] \), yields:

\[
M_\ell(\omega, t) = H_\ell(\mathbf{x}(t_p); \omega, t) S(\omega, t).
\]

The microphones signals are passed to a steered beamformer to localize the source. Here, each microphone signal is processed by the frequency-dependent filter weights \( \{W_\ell(\theta; \omega)\}_{\ell=1}^{L} \) to yield the beamformer output signal \( B(\theta; \omega, t) \):

\[
B(\theta; \omega, t) = \sum_{\ell=1}^{L} M_\ell(\omega, t) W_\ell(\theta; \omega).
\]

The parameter \( \theta \) is called the steering angle. It describes the beampattern of the beamformer by defining the direction in which the beamformer is pointing. The beamformer output signal is hence a function of this parameter.

Following Chapter 3, we model the soundfield as the sum of a direct field compo-
nent and a diffuse reverberant field component. Due to the linearity of the STFT, 
\( B(\theta; \omega, t) \) can be separated into direct and reverberant components:

\[
B(\theta; \omega, t) = B_d(\theta; \omega, t) + B_r(\theta; \omega, t). \tag{4.4}
\]

To facilitate writing an expression for the output signal of the beamformer, we use the farfield beampattern \( D_\theta(\phi; \omega) \). The beampattern allows abstraction of the beamformer properties from the details of the beamformer design. Weighting the direct path by the beampattern function in the source direction, the direct components is given by:

\[
B_d(\theta; \omega, t_p) = D_\theta(\phi_s(t_p); \omega)\xi_d(\omega)e^{-ikx(t_p) \cdot \hat{\phi}_s(t_p)}.
\tag{4.5}
\]

where \( \xi_d(\omega) \) is the direct path amplitude and \( \hat{\phi}_s(t_p) \triangleq [\cos \phi_s(t_p), \sin \phi_s(t_p)]^T \). In a diffuse field, the reverberant field component is a result of plane waves coming from all directions. If in each direction \( \phi \), the plane wave has amplitude \( \xi_r(\phi; \omega) \), the reverberant component is given by:

\[
B_r(\theta; \omega, t_p) = \int_0^{2\pi} D_\theta(\phi; \omega)\xi_r(\phi; \omega)e^{-ikx(t_p) \cdot \hat{\phi}_s(t_p)}d\phi,
\tag{4.6}
\]

and \( \hat{\phi} \triangleq [\cos \phi, \sin \phi]^T \). As described in Section 2.4, in a diffuse soundfield the variance of the plane wave amplitude \( \sigma^2_{\xi}(\phi; \omega) = E\{|\xi_r(\phi; \omega)|^2\} \) is constant in \( \phi \).

From Chapter 3, the beamformer output signal is a circularly complex Gaussian random variable with a mean of \( B_d(\theta; \omega, t) \) and variance of \( \sigma^2_B(\theta; \omega) \triangleq E\{|B_t(\theta; \omega, t)|^2\} \) that is equal to:

\[
\sigma^2_B(\theta; \omega) = \sigma^2_{\xi}(\omega)\int_0^{2\pi} |D_\theta(\phi; \omega)|^2d\phi.
\tag{4.7}
\]

The variance \( \sigma^2_B(\theta; \omega) \) is independent of beamformer position \( x(t_p) \) and hence frame time \( t_p \).

An important characteristic is the covariance between beamformer output signals at different times. From Theorem 3.3.1:

\[
E\{B^*_t(\theta; \omega, t_p)B_t(\theta; \omega, t_q)\} = \sigma^2_{\xi}(\omega)\int_0^{2\pi} |D_\theta(\phi; \omega)|^2e^{ik[x(t_p)-x(t_q)] \cdot \hat{\phi}_s(t_p)}d\phi.
\]

Collect the reverberant beamformer component for each steering angle \( \theta \) into the vector

\[
b_t = [B_t(\theta; \omega, t_1), B_t(\theta; \omega, t_2), \ldots, B_t(\theta; \omega, t_P)]^T.
\]

This vector of correlated complex Gaussian variable can be written as a linear
transformation of $P$ uncorrelated complex Gaussian variables:

$$b_t = Tz,$$

where $T$ is a transformation matrix and $z$ is a vector of $P$ zero mean independent Gaussian random variables of unity variance. Matrix $T$ is related to the covariance matrix of $b_t$:

$$E\{b_t b_t^H\} = TTE\{zz^H\}^H = TT^H,$$

where we have applied the property $E\{zz^H\} = I$ for uncorrelated Gaussian random variables and $I$ is the identity matrix. Each element of $E\{b_t b_t^H\}$ is given by:

$$[E\{b_t b_t^H\}]_{qp} = E\{B_r^*(\theta; \omega, t_p)B_r^*(\theta; \omega, t_q)\},$$

where $[\cdot]_{qp}$ denotes the element in the $q$th row and $p$th column of a matrix.

A useful measure of the similarity between beamformer output signals is the correlation coefficient between different beamformer signals:

$$\rho_B(\theta, \vartheta; \omega, t_p, t_q) = \frac{E\{B_r^*(\theta; \omega, t_p)B_r^*(\theta; \omega, t_q)\}}{\sqrt{E\{|B_r(\theta; \omega, t_p)|^2\}E\{|B_r(\theta; \omega, t_q)|^2\}}}.$$ (4.10)

The relative magnitude of the direct and reverberant parts is measured with the beamformed direct-to-reverberant ratio (DRR), which shall be defined:

$$\gamma_B(\theta; \omega, t_p) \triangleq \frac{|B_d(\theta; \omega, t_p)|^2}{E\{|B_r(\theta; \omega, t_p)|^2\}}.$$  

The beamformed DRR $\gamma_B(\theta; \omega, t_p)$ is in general a function of frame time $t_p$ and the steering angle $\theta$.

### 4.4 Generic Algorithm for Steered Beamforming

In this chapter we explore the performance of combined source localization and tracking utilizing a steered beamforming. In Section 4.4.1 we describe a generic algorithm for steered beamforming. In Section 4.4.2, we define the beamformer design specifications that are required for accurate bearing estimation in general and for the analysis in particular.

#### 4.4.1 Description of Steered Beamformer Algorithm

In the traditional steered beamforming approach to source localization, a beamformer is steered through all directions, and the source bearing is estimated as the
beamformer steering direction that yields the largest beamformer output power.

In search for the source, at each time $t_p$ a beamformer is steered through all steering angles $\theta$. We define the STFT of the output signal of the beamformer as $B(\theta_n; \omega, t_p)$. The angular spectrum function is then the beamformer output signal energy density:

$$A(\theta; \omega, t_p) \triangleq |B(\theta; \omega, t_p)|^2.$$  \hspace{1cm} (4.11)

The true source bearing $\phi_s(t_p)$ is estimated as the angle $\hat{\phi}_s(\omega, t_p)$ that maximizes the angular spectrum function:

$$\hat{\phi}_s(\omega, t_p) = \arg \max_{\theta} A(\theta; \omega, t_p).$$

In practice, the search for the angle $\hat{\phi}_s(\omega, t_p)$ is discretized. The beamformer output signal is sampled for $N$ steering directions $\theta_1, \theta_2, \ldots, \theta_N$ and the amplitudes compared:

$$\hat{\phi}_s'(\omega, t_p) = \arg \max\{A(\theta; \omega, t_p) : \theta = \theta_1, \theta_2, \ldots, \theta_N\}.$$  

Denoting the steering angle closest to true source bearing\(^3\) as $\theta_{n_s}(t_p)$, we see the algorithm functions correctly when $\hat{\phi}_s'(\omega, t_p) = \theta_{n_s}(t_p)$.

Next, the choice of beamformer design for steered beamforming is described.

### 4.4.2 Beamformer Specifications

One is required to select a different beampattern for each steering direction. In this section, we discuss how to do this.

We typically want to choose each beampattern $D_{\theta_n}(\phi; \omega)$ so the amplitude is concentrated about the steering direction. In Figure 4.2, we illustrate the following

\(^3\)That is $\theta_{n_s}(t_p) \triangleq \arg \min_{\phi} \{ |\theta(t_p) - \phi| : \phi = \theta_1, \theta_2, \ldots, \theta_N \}$.  

---

Figure 4.2: Beamformer specification for a steered beamformer focussed in a steering direction $\theta_n$. The beampattern is constrained to each of the hatched areas.
(i) a unity magnitude about \( \theta_n \): 
\[
1 - \epsilon_2 < |D_{\theta_n}(\phi; \omega)| < 1 \text{ for } |\phi - \theta_n| < \Delta \theta,
\]

(ii) a thin transition interval: 
\[
\epsilon_1 < |D_{\theta_n}(\phi; \omega)| < 1 - \epsilon_2 \text{ for } \Delta \theta < |\phi - \theta_n| < \Delta \theta + \delta \theta,
\]

(iii) a small magnitude at angles much different from \( \theta_n \): 
\[
|D_{\theta_n}(\phi; \omega)| < \epsilon_1 \text{ for } |\phi - \theta_n| > \Delta \theta + \delta \theta.
\]

For good spatial selectivity, we require \( \epsilon_1 + \epsilon_2 \ll 1 \). For good resolution in localization, the beamwidth \( \Delta \theta \) must be small.

Like in Chapter 3, we define the beamformer gain in direction of interest \( \phi \) [47]:
\[
\varepsilon_{\theta_n}(\phi; \omega) \triangleq \frac{|D_{\theta_n}(\phi; \omega)|^2}{\frac{1}{2\pi} \int_0^{2\pi} |D_{\theta_n}(\phi; \omega)|^2 d\phi}.
\] (4.12)

For accurate estimation of bearing angle, we design the beamformers with a large gain for \( |\phi - \theta_n| < \Delta \theta \) and a small gain for \( |\phi - \theta_n| > \Delta \theta + \delta \theta \) and ensure the transition intervals are small. Now for \( \epsilon_1, \epsilon_2 \) and \( \delta \theta \) small, the denominator of (4.12) is approximately equal to \( \Delta \theta / \pi \). From the above specifications, the gain is approximately lower bounded by \( \varepsilon_{\theta_n}(\phi) > \pi (1 - \epsilon_2)^2 / \Delta \theta \) for \( |\phi - \theta_n| < \Delta \theta \) and approximately upper bounded by \( \varepsilon_{\theta_n}(\phi) < \pi \epsilon_2^2 / \Delta \theta \) for \( |\phi - \theta_n| > \Delta \theta + \delta \theta \). We hence constrain parameters so that \( \epsilon_2 \ll 1, \Delta \theta / \pi \ll 1, \epsilon_1^2 \ll \Delta \theta / \pi, \) and \( \delta \theta \ll \Delta \theta \).

In the limit \( \epsilon_1, \epsilon_2, \delta \theta \to 0 \), each beampattern possesses a constant amplitude over \( [\theta_n - \Delta \theta, \theta_n + \Delta \theta] \) and is zero everywhere else (see Figure 4.3(b)). In Section 4.8, we explore the performance of source localization in this limit.

It is reasonable to assume that in situations of interest the direct-to-reverberant ratio of sound pressure \( \gamma_P(\omega) \) at the microphone array is of moderate magnitude. At extremely low DRRs (\( \gamma_P(\omega) \ll 1 \)), localization tends to be a fruitless task. At high DRRs (\( \gamma_P(\omega) \gg 1 \)), reverberation is irrelevant. We hence consider the case \( \gamma_P(\omega) \sim 1 \). Recalling that beamformed DRR is the product of beamformer gain and the DRR of pressure, the above beamformer gain properties then imply the following:

(i) \( E\{|B_d(\theta_{n_s}; \omega, t_p)|^2\} \ll |B_d(\theta_{n_s}; \omega, t_p)|^2 \). Beamformed output is dominated by its direct component, and can be assumed deterministic.

(ii) \( E\{|B_t(\theta_n; \omega, t_p)|^2\} \gg |B_d(\theta_n; \omega, t_p)|^2, n \neq n_s \). Beamformed output is dominated by its reverberant component, and can thus be assumed zero mean.

These properties assist analysis of the estimation error in Section 4.6.

Also to aid the analysis, we space the steering angles at least \( 2(\Delta \theta + \delta \theta) \) apart,
so that beampatterns possess the following negligible overlap property:

\[
\int_{0}^{2\pi} |D_{\theta_n}(\phi; \omega)D_{\theta_m}(\phi; \omega)|d\phi \ll \sqrt{\int_{0}^{2\pi} |D_{\theta_n}(\phi; \omega)|^2d\phi \int_{0}^{2\pi} |D_{\theta_m}(\phi; \omega)|^2d\phi, m \neq n.
\]

This property is illustrated in Figure 4.3(a), where the overlap are shown as hatched areas. From it we can show that beamformer output signals sampled at different steering directions are independent:

**Theorem 4.4.1** In a diffuse field, beamformers with approximately non-overlapping beampatterns \(D_{\theta_n}(\phi; \omega)\) and \(D_{\theta_m}(\phi; \omega)\) satisfying (4.13) have output signals \(B_{r}(\theta_n; \omega, t_p)\) and \(B_{r}(\theta_m; \omega, t_q)\) which are approximately statistically independent.

The proof of this theorem is presented in the appendix at the end of the chapter.

### 4.5 Algorithm Independent Description of Source Tracking

In this section, we describe a means of analyzing source tracking in a way that is relatively independent of the details of the tracking algorithm and source dynamics.

In the source tracking algorithm, the aim is to estimate the bearing angles \(\phi_s(t_1), \phi_s(t_2), \ldots, \phi_s(t_P)\) from the angular spectra \(\{A(\theta; \omega, t_p)\}_{t=1}^{P}\). In the context of source localization, a source tracking algorithm consists of a dynamic model of the target coupled with a way of transforming physical measurements (the angular spectra) into bearing estimates.

The dynamic model can be as complicated as a high dimensional model of the source kinematics, or as simple as modelling position coordinates as a random processes. Several dynamic models used to track the time-varying location of a person are presented in [45]. For our analysis we desire an approach that is free from the details of the such model. We simply assume that the source lies so much
in the farfield that a small change in the position of the sensor array or source will not change the bearing $\phi_s(t)$. Then for tracking purposes, the sound source appears stationary, i.e. $\phi_s(t) \approx \phi_s$.

Imperfections in the dynamic model in a tracking algorithm are typically accounted for with a random process. For example, speaker motion modelled with a Langevin model in [108] and constant velocity Kalman filter model in [95], where they let target velocity be a first order Markov process. The accuracy of the model is reflected by the variance of this random process. We make the idealization that given the tracking model is given no new state information, it can perfectly track over a number of iterations $P$, after which the algorithm diverges.

The basic idea of tracking is then to exploit the fact that the peak in the angular spectrum due to the true source location will be in the same position in successive frames, while the spurious peaks caused by reverberation will vary in position. Reverberant peaks can then be damped by averaging the spectra, motivating definition of the tracked angular spectrum:

$$\bar{A}(\theta; \omega) = \sum_{p=1}^{P} A(\theta; \omega, t_p), \quad (4.14)$$

and the tracked bearing estimate:

$$\hat{\phi}_s(\omega) = \arg \max \{ \bar{A}(\theta; \omega) : \theta = \theta_1, \theta_2, \ldots, \theta_N \}.$$

Using the signal model of Section 4.3, we derive an expression for the mean of the tracked angular spectrum $\bar{A}(\theta_n; \omega)$. Taking the expectation of (4.14) and noting that $A(\theta_n; \omega, t_p) = |B(\theta_n; \omega, t_p)|^2$:

$$E\{ \bar{A}(\theta_n; \omega) \} = \sum_{p=1}^{P} E\{ |B(\theta_n; \omega, t_p)|^2 \}. \quad (4.15)$$

Separating $B(\theta_n; \omega, t_p)$ into direct and reverberant parts using (4.4), and applying the zero mean property of the reverberant part:

$$E\{ |B(\theta_n; \omega, t_p)|^2 \} = |B_d(\theta_n; \omega, t_p)|^2 + E\{ |B_r(\theta_n; \omega, t_p)|^2 \}. \quad (4.16)$$

Now if the source is stationary, the direct component $|B_d(\theta_n; \omega, t_p)|^2$ will be the same in every frame. Also, as is seen from (4.7), the reverberant component $\sigma_B^2(\theta_n; \omega) = E\{ |B_r(\theta_n; \omega, t_p)|^2 \}$ is also constant with frame time. Consequently the beamformed DRR $\gamma_B(\theta_n; \omega, t_p)$ is constant across all frame times. Factoring $\sigma_B^2(\theta_n; \omega, t_p)$ out of (4.16) and identifying the beamformed DRR:

$$E\{ |B(\theta_n; \omega, t_p)|^2 \} = \sigma_B^2(\theta_n; \omega) [\gamma_B(\theta_n; \omega) + 1]. \quad (4.17)$$
Substituting (4.17) into (4.15), the expectation of the tracked angular spectrum in (4.14) is calculated as:

\[
E\{\bar{A}(\theta_n; \omega, t_p)\} = \sum_{p=1}^{P} \sigma_B^2(\theta_n; \omega)[\gamma_B(\theta_n; \omega) + 1] \\
= P\sigma_B^2(\theta_n; \omega)[\gamma_B(\theta_n; \omega) + 1]. \tag{4.18}
\]

Next we derive expressions for the probability of wrongly estimating the source bearing direction.

### 4.6 Probability of Estimation Error

In this section we develop an expression for the probability of incorrectly estimating the source direction \(\phi_s\), specifically the probability that \(\hat{\phi}_s(\omega) \neq \theta_{ns}\). This is the probability that the sampled angular spectrum \(\{\bar{A}(\theta_n; \omega)\}_{n=1}^{N}\) is not at its maximum at angle \(\theta_{ns}\):

\[
Pr(\text{Error}) = 1 - Pr\left[\max_{n \neq ns} \bar{A}(\theta_n; \omega) < \bar{A}(\theta_{ns}; \omega)\right].
\]

From Theorem 4.4.1, using the beamformer design specifications from Section 4.4.2, since each \(\bar{A}(\theta_n; \omega)\) is a sum of squared beamformer output signals \(\{B(\theta_n; \omega, t_p)\}_{p=1}^{P}\), \(\bar{A}(\theta_n; \omega)\) and \(\bar{A}(\theta_m; \omega)\) are approximately independent for \(n \neq m\). The probability of estimation error can hence be rewritten:

\[
Pr(\text{Error}) = 1 - \prod_{n=1, n \neq ns}^{N} Pr[\bar{A}(\theta_n; \omega) < \bar{A}(\theta_{ns}; \omega)].
\]

The \(E\{|B_r(\theta_n; \omega, t_p)|^2\} \ll |B_d(\theta_{ns}; \omega, t_p)|^2\) property of Section 4.4.2 ensures that the angular spectrum in direction \(\theta_{ns}\) is dominated by its mean \(E\{\bar{A}(\theta_{ns}; \omega)\}\). Consequently we can replace this random variable by its expectation:

\[
Pr(\text{Error}) = 1 - \prod_{n=1, n \neq ns}^{N} F_{\bar{A}(\theta_n; \omega)}\left[E\{\bar{A}(\theta_{ns}; \omega)\}\right] \tag{4.19}
\]

where \(F_{\bar{A}(\theta_n; \omega)}(a) \triangleq Pr[\bar{A}(\theta_n; \omega) < a]\) is the cumulative density function (CDF) of \(\bar{A}(\theta_n; \omega)\). This estimation error is dependent solely on \(F_{\bar{A}(\theta_n; \omega)}(a)\). We want each \(F_{\bar{A}(\theta_n; \omega)}(a)\) large at \(a = E\{\bar{A}(\theta_{ns}; \omega)\}\) so that estimation error is small.

Further analysis will depend whether the beamformer output signals are correlated. We derive the CDF of \(\bar{A}(\theta_n; \omega)\), first for the case of uncorrelated beamformer output signals (Section 4.6.1) and then for the more practical case of correlated...
beamformer output signals (Section 4.6.2). We then make some notes on the results (Section 4.6.3).

4.6.1 Upper Performance Limit

In the case that beamformer output signals are uncorrelated, tracking performance achieves its upper limit. We derive the CDF of $\mathcal{A}(\theta_n; \omega)$ for this case.

From Section 4.4.2, provided $n \neq n_s$, $B(\theta_n; \omega, t_p)$ is a circularly complex Gaussian random variable of zero mean. Circularity of complex Gaussian random variables implies the real and imaginary parts are independent so that $A(\theta_n; \omega) = |B(\theta_n; \omega, t_p)|^2$ is chi-squared with 2 degrees of freedom. $\mathcal{A}(\theta_n; \omega)$ is thus the sum of $P$ independent chi-squared random variables $|B(\theta_n; \omega, t_p)|^2 \sim \sigma_B^2 \chi^2(2), p = 1, 2, \ldots, P$, which is also chi-squared. The tracked angular spectrum is hence

$$
\mathcal{A}(\theta_n; \omega) \sim \sum_{p=1}^{P} \sigma_B^2 \chi^2(2) = \sigma_B^2 \chi^2(2P), n \neq n_s,
$$

where $\chi^2(m)$ is the chi-squared distribution with $m$ degrees of freedom, and the probability density function (PDF) is

$$
f_{\chi^2(m)}(x) = \frac{1}{2^{m/2} \Gamma(m/2)} x^{m/2 - 1} e^{-x/2},
$$

and $\Gamma(m)$ is the gamma function. Integrating, the CDF is given by:

$$
F_{\chi^2(m)}(x) = \int_0^x f_{\chi^2(m)}(u) du = G\left(\frac{x}{2}, \frac{m}{2}\right),
$$

(4.20)

where $G(x, m)$ is the regularized incomplete Gamma function [3, pp. 260 - 263]:

$$
G(x, m) = \frac{1}{\Gamma(m)} \int_0^x u^{m-1} e^{-u} du.
$$

In the case $m$ is an integer,

$$
G(x, m) = 1 - e^{-x} \sum_{n=0}^{m-1} \frac{x^n}{n!}.
$$

From (4.20) and the CDF scaling property $F_{\sigma X}(x) = F_X(x/\sigma)$:

$$
F_{\mathcal{A}(\theta_n; \omega)}(a) = G\left(\frac{a}{2\sigma_B}, P\right).
$$

In case $P = 1$ of localization without tracking, the CDF reduces to:

$$
F_{\mathcal{A}(\theta_n; \omega)}(a) = F_{\chi^2(2)}\left(\frac{a}{\sigma_B^2}\right) = 1 - e^{-a/2\sigma_B^2}, n \neq n_s.
$$
and the probability of error to:

$$\text{Pr}(\text{Error}) = 1 - \exp \left[ -\frac{E\{\overline{A}(\theta_{n_s}, \omega)\}}{2E\{|B_i(\theta_n, \omega)|^2\}} \right].$$  \hfill (4.21)

### 4.6.2 Practical Performance

In practice, the beamformer output signals are correlated. We derive the CDF of $\overline{A}(\theta_n, \omega)$ in such a case.

From Section 4.4.2, provided $n \neq n_s$, $\overline{A}(\theta_n; \omega)$ is the sum of the square magnitude of $P$ correlated circularly complex Gaussian random variables of zero mean. In matrix-vector form, $\overline{A}(\theta_n; \omega) = b^H(\theta_n; \omega)b(\theta_n; \omega)$ where $b(\theta_n; \omega)$ is the vector of beamformer output signals defined in Section 4.3. Utilizing (4.8) we see that

$$\overline{A}(\theta_n; \omega) = z^H Q z, \quad n \neq n_s$$  \hfill (4.22)

where $z$ is a vector of uncorrelated complex Gaussian random variables and $Q = T^HT$ is the matrix of the quadratic form. For the rest of this subsection, we suppress dependence of variables on $\theta_n$ and $\omega$ which are held constant over this derivation. From [16], the bilateral Laplace Transform of the PDF of a quadratic form of circularly complex Gaussian random variables is the well-known result:

$$\tilde{f}_A(s) = \frac{1}{\det(I + 2sQ)},$$

where $\mathcal{L}\{\cdot\}(s)$ is the Laplace transform operator and $\tilde{f}_A(s) \triangleq \mathcal{L}\{f_A(a)\}(s)$. Since in the $s$ domain, integration is accomplished by dividing by $s$, the Laplace transform of the CDF of $\overline{A}$, $\tilde{F}_A(s) \triangleq \mathcal{L}\{F_A(a)\}(s)$ is given by:

$$\tilde{F}_A(s) = \frac{1}{s \det(I + 2sQ)}. \hfill (4.23)$$

$\tilde{F}_A(s)$ is now expressed as a function of the $P$ eigenvalues $\nu_1, \nu_2, \ldots, \nu_P$ of the covariance matrix $E\{b_i b_i^H\} = TT^H$ of Section 4.4.2. Since $TT^H$ is a correlation matrix, it is positive semidefinite and eigenvalues are all either positive or zero. One can show then from the singular value decomposition of $TT^H$ and $T^HT$, that the singular values and hence the eigenvalues of $TT^H$ and $T^HT$ are the same [33]. Thus the eigenvalues of $Q$ are $\nu_1, \nu_2, \ldots, \nu_P$. From the characteristic polynomial of $Q$, one can write:

$$\det(I + 2sQ) = \prod_{p=1}^P (1 + 2s\nu_p), \hfill (4.24)$$

Substituting (4.24) into (4.23), we can then use the method of residues to evaluate the inverse Laplace transform. The poles of (4.23) are $s_0 = 0$ and $s_p = -1/2\nu_p, p = 1, 2, \ldots, P$. In the case that all eigenvalues are distinct, the CDF of $\overline{A}$ can be
Table 4.1: The probability of estimation error for combined localization and tracking over $P$ frames of signal output. Expressions for $E\{|B_t(\theta_n;\omega)|^2\}$ and $E\{A(\theta_n;\omega)\}$ are presented in (4.7) and (4.15) respectively.

written as

$$F_{\overline{A}}(a) = \frac{1}{2\pi i} \oint e^{as} \overline{F}_A(s) ds = \sum_{p=0}^{P} \varsigma_p e^{-\frac{a}{\nu_p}},$$

where the contour integral runs clockwise and encloses all the poles of $\overline{F}_A(s)$ and the $\varsigma_p$'s are given by:

$$\varsigma_p = \prod_{q=0}^{P} \frac{s_p}{s_p - s_q}, \quad p = 0, 1, \ldots, P.$$ 

Consequently the CDF of $\overline{A}$ is summarized as:

$$F_{\overline{A}(\theta_n,\omega)}(a) = 1 - \sum_{p=1}^{P} \left( \prod_{q=1}^{P} \frac{\nu_p}{\nu_p - \nu_q} \right) e^{-\frac{a}{2\nu_p}},$$

The complete expressions for the probability of estimation error are summarized in Table 4.1.

4.6.3 Notes on Analysis

Correlation of Beamformer Signals

We see from the above analysis a dependence of performance on the correlation of beamformer output signals. Section 4.6.2 reveals that practical performance is a function of the eigenvalues $\nu_1, \nu_2, \ldots, \nu_P$ of the correlation matrix $E\{b_t(\theta_n;\omega)b_t^H(\theta_n;\omega)\}$. This matrix is basically a matrix of correlation coefficients,
scaled by factor $\sigma_{b}^{2}(\theta_{n};\omega)$:

$$[E\{b_{r}(\theta_{n};\omega)b_{r}^{H}(\theta_{n};\omega)\}]_{qp} = \sigma_{B}^{2}(\theta_{n};\omega)\rho_{B}(\theta_{n},\theta_{n};\omega, t_{p}, t_{q}).$$

As will be seen in Section 4.7, higher correlation degrades the performance.

From the robustness of equalization analysis in Chapter 3, strong correlation in beamformer output signals sampled at two points $\mathbf{x}(t_{p})$ and $\mathbf{x}(t_{q})$ arises from:

(i) sampling the soundfield at points that are closely spaced and

(ii) capturing sound signals using a beamformer with a high directivity.

We shall see in the example of Section 4.8 that, for the same tracking improvement, beamformers with higher directivity require a large distance of separation between successive measurements of angular spectra.

**Beampatterns with Circular Symmetry**

Consider the case where each steered beampattern has the same shape, i.e. $D_{\theta_{n}}(\phi;\omega) = D(\phi-\theta_{n};\omega)$ for some generating function $D(\phi;\omega)$. Such beampatterns can be achieved with a uniform circular array (i) exactly when the number of sensors is an integral multiple of the number of steering directions $(M|N)$, and (ii) approximately for large numbers of sensors.

For circular symmetry, the term $E\{|\mathbf{A}(\theta_{ns};\omega)|/E\{|B_{r}(\theta_{ns};\omega, t_{p})|^{2}\}$ present in Table 4.1 reduces to a simple functional form. From (4.7) the beamformer directivity constraint ensures that the variance of the beamformer output signal $E\{|B_{r}(\theta_{ns};\omega, t_{p})|^{2}\}$ is the same for each $\theta_{ns}$:

$$E\{|\mathbf{A}(\theta_{ns};\omega)|/E\{|B_{r}(\theta_{ns};\omega, t_{p})|^{2}\} = E\{|\mathbf{A}(\theta_{ns};\omega)|/E\{|B_{r}(\theta_{ns};\omega, t_{p})|^{2}\}.$$

Substituting (4.18) for $E\{|\mathbf{A}(\theta_{ns};\omega)|$, we see that:

$$E\{|\mathbf{A}(\theta_{ns};\omega)|/E\{|B_{r}(\theta_{ns};\omega, t_{p})|^{2}\} = P[\gamma_{B}(\theta_{ns};\omega) + 1].$$

(4.25)

Substituting (4.25) into (4.21), the probability of error in the upper performance limit is shown to possess an explicit dependence on beamformed DRR $\gamma_{B}(\theta_{ns};\omega)$. We recall from (3.15) that beamformed DRR is simply the product of beamformer gain and the DRR of sound pressure. We have hence shown the intuitive fact that localization accuracy is improved by increasing beamformer gain and the DRR of sound pressure.
Figure 4.4: Performance expected by combined source tracking and localization at 2kHz when the tracking algorithm is accurate over $P$ positions, at direct-to-reverberant ratios $\gamma_P = (-6\text{dB}, -3\text{dB}, 0\text{dB}, 3\text{dB})$. The dashed curves represent the limit of perfect tracking and a fast moving array. The solid curves are for an array moving at 1 m/s and sampling frames at 10Hz.

### 4.7 Analysis for Moving Sound Source

The analysis thus far has been for a moving microphone array and fixed sound source. This scenario certainly has applications. However, we desire to apply the analysis to the case of a fixed array and moving source.

Looking over the analysis, one can see that if an expression for covariance matrix $E\{b_r(\theta_n; \omega)b^H_r(\theta_n; \omega)\}$ can be found for movement of the sound source, the above analysis can be applied to the moving sound source case. Unfortunately the direct plus diffuse soundfield model does not predict the covariance $E\{B_r^c(\theta_n; \omega, t_p)B_r(\theta_n; \omega, t_q)\}$ for such a case.

In Chapter 3, we also required an expression for this covariance. Section 3.4 proposed a reciprocity principle which asserted a statistical equivalence of the beamformer output signal between cases of a moving source and fixed beamformer, and a fixed source and moving beamformer. However we leave investigation into such a principle to future research.

### 4.8 Examples

We now explore the improvements obtainable by source tracking with the following example. Consider a sensor array moving linearly at constant speed $v$ m/s. The localization algorithm samples a frame of sound data every $t_h$. The steered beamforming algorithm consists of $N$ steered beamformers, each possessing an idealized
beampattern described by:

\[ D_{\theta_n}(\phi; \omega) = \begin{cases} 
1, & -\pi/N \leq |\phi - \theta_n| \leq \pi/N \\
0, & \text{otherwise},
\end{cases} \]

which is depicted in Figure 4.3(b).\(^4\) Simulation is performed at a frequency of 2kHz.

The accuracy of the tracking algorithm is quantified by the *number* of frames \( P \) over which it can track without further state knowledge and without diverging. In Figure 4.4 we investigate of probability of error as a function of \( P \). Source bearing is estimated with six steering angles \((N = 6)\). The dashed curves denote the probability of error in the limit of perfect tracking and the array moving fast enough so that\(^5\) \( vt_s \gg 1 \). This curve shows the maximum obtainable improvement by combined localization and tracking. The solid curves represent the performance improvements achievable for a source moving at \( v = 1 \) m/s and with a real tracking algorithm and \( t_s = 0.1s \). Due to the correlation between the beamformer output signals, performance is diminished from the ideal case.

The spatial correlation of beamformer output signals suggests there is an upper theoretical limit to the improvements obtainable by source tracking. In Figure 4.5, we explore this limit by performing localization estimates *continuously* along the path of movement. As we can see in Figure 4.5(a), improvements only start becoming significant at 2kHz if tracking is performed over a distance of 0.5m or more.

Further Figure 4.5(b) shows that if beamformers of narrower beamwidth are used, tracking must be accurately performed over a larger distance. Narrowing the beamwidths results in higher correlation between beamformer output signals. Though this improves the accuracy of localization (shown by a more rapid drop in error probability than in Figure 4.5(a)), it also retards the performance boost obtained by tracking (shown by the tighter concentration of probability curves).

In each of the simulation cases above, we note that the direction of movement of the sensor array was unimportant. No matter whether the beamformer moves radially toward the source, or tangentially to it, the probability of error curves remained the same. This result contrasts with that in robustness of equalization, where the direction is important.

### 4.9 Summary and Contribution

In this chapter, we have quantified the upper limits of improvement to source localization in reverberant rooms achievable by combined it with a tracking algorithm.

\(^4\)This beampattern has a beamformer gain of \( N \).

\(^5\)This condition ensures beamformer output signals are uncorrelated.
Figure 4.5: Maximum performances expected by combined source tracking and localization in the limit that localization estimates are made continuously over the distances (1/16m, 1/8m, 1/4m, 1/2m, 1m), at 2kHz. (a) Using 6 steering angles with a beamformer of beamwidth 30°. (b) Using 12 steering angles with a beamformer of beamwidth 15°. The dotted line represents no tracking.

The major contribution made in this chapter are:

(i) Exploration of an upper limit on tracking when localizing the source with a steered beamformer algorithm, showing how the performance is affected by the correlation between the beamformer output signals at different beamformer locations.

(ii) Quantifying the improvement to steered beamformer localization as a function of the accuracy of the tracking algorithm and beamwidth of the steered beamformer.

These results pertain to tracking a fixed source with a moving sensor array. We have proposed investigation of a theorized reciprocity principle of Chapter 3 that will allow us to extend analysis to movement of the sound source.

4.10 Appendix

4.10.1 Proof of Theorem 4.4.1

We recall form Theorem 3.3.1 that in the diffuse field, beamformer output signals are complex Gaussian. To prove $B_t(\theta_n; \omega, t_p)$ and $B_t(\theta_m; \omega, t_q)$ are independent, we simply demonstrate that they are approximately uncorrelated, by showing that the correlation coefficient $\rho_B(\theta_n, \theta_m; \omega, t_p, t_q)$ in (4.7) is negligible. Calculating the
numerator of (4.10), using the covariance expression (4.6):

\[
E\{B_t^* (\theta_n; \omega, t_p) B_t (\theta_m; \omega, t_q)\} = \int_0^{2\pi} \int_0^{2\pi} D_{\theta_n}^* (\phi_1; \omega) D_{\theta_m} (\phi_2; \omega) E\{\xi_t (\phi_1; \omega) \xi_t (\phi_2; \omega)\}
\times e^{ikx(t_p) \cdot \hat{\phi}_1} e^{-ikx(t_q) \cdot \hat{\phi}_2} d\phi_1 d\phi_2,
\]

\[
= \sigma_\xi^2 (\omega) \int_0^{2\pi} D_{\theta_n}^* (\phi; \omega) D_{\theta_m} (\phi; \omega) e^{ik[x(t_p) - x(t_q)] \cdot \hat{\phi}} d\phi,
\]

where (2.20) was applied in the second step. Next taking the magnitude of this expectation and moving it inside the integral:

\[
|E\{B_t^* (\theta_n; \omega, t_p) B_t (\theta_m; \omega, t_q)\}| \leq \sigma_\xi^2 (\omega) \int_0^{2\pi} |D_{\theta_n}^* (\phi; \omega) D_{\theta_m} (\phi; \omega) e^{ik[x(t_p) - x(t_q)] \cdot \hat{\phi}}| d\phi
\]

\[
\leq \sigma_\xi^2 (\omega) \int_0^{2\pi} |D_{\theta_n} (\phi; \omega) D_{\theta_m} (\phi; \omega)| d\phi. \quad (4.26)
\]

The denominator of (4.10), the beamformer output signal \(E\{|B_t (\theta_n; \omega, t_p)|^2\}\), is given by (4.7). Substituting (4.7) and (4.26) into (4.10),

\[
|\rho_B (\theta_n, \theta_m; \omega, t_p, t_q)| \leq \frac{\int_0^{2\pi} |D_{\theta_n} (\phi; \omega) D_{\theta_m} (\phi; \omega)| d\phi}{\sqrt{\int_0^{2\pi} |D_{\theta_n} (\phi; \omega)|^2 d\phi \int_0^{2\pi} |D_{\theta_m} (\phi; \omega)|^2 d\phi}}.
\]

From non-overlap property (4.13) it then follows that \(|\rho_B (\theta_n, \theta_m; \omega, t_p, t_q)| \ll 1\). The beamformer signals are hence effectively uncorrelated, and by their Gaussianity property are approximately independent. \(\square\)