
Linear and Parametric Microphone Array Processing

Part 4 - Parametric Spatial Processing

Emanuël A. P. Habets¹ and Sharon Gannot²

- ¹ International Audio Laboratories Erlangen, Germany
A joint institution of the University of Erlangen-Nuremberg and Fraunhofer IIS
- ² Faculty of Engineering, Bar-Ilan University, Israel



Overview

- 1 Overview
- 2 Sound Field and Signal Models
- 3 Parameter Estimation
- 4 Estimation of the Sound Pressures
- 5 Applications and Examples
- 6 Summary

1. Overview

Parametric-based spatial audio processing makes use of an efficient parametric representation of the sound-field. A major advantage compared to classical spatial processing is the limited number of parameters.

Examples:

- Directional Audio Coding (DirAC)*
- High Angular Resolution Planewave Expansion (HARPEX)*
- Computational Auditory Scene Analysis (CASA)
- Some well-known dereverberation techniques that make use of the reverberation time and signal-to-reverberation ratio.

Applications:

- Spatial audio recording, coding and reproduction
- Source separation / source extraction / noise reduction / dereverberation
- Acoustic scene analysis / localization

*Patented technologies.

1. Overview

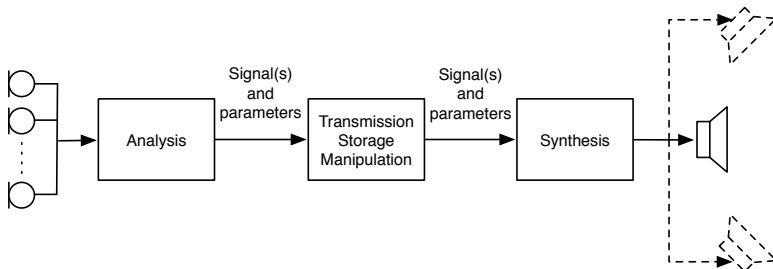


Figure: Block diagram of a parametric-based spatial audio processing system.

1. Overview

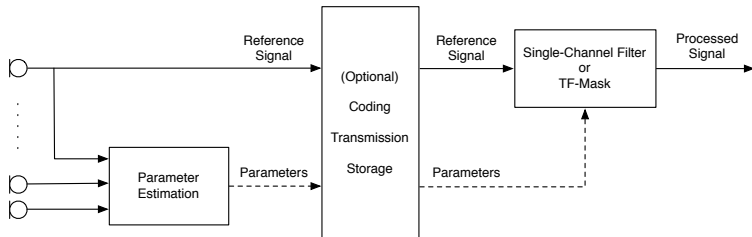


Figure: An example of the analysis stage (left) and processing stage (right) using a single reference signal.

Overview

- 1 Overview
- 2 Sound Field and Signal Models**
- 3 Parameter Estimation
- 4 Estimation of the Sound Pressures
- 5 Applications and Examples
- 6 Summary

2. Sound Field and Signal Models

The sound pressure at a position \mathbf{p}_i , time frame m and discrete frequency k , can be decomposed into a direct sound component and a diffuse sound component:

$$S_i(k, m) \triangleq S(k, m, \mathbf{p}_i) = S_{\text{dir}}(k, m, \mathbf{p}_i) + S_{\text{diff}}(k, m, \mathbf{p}_i). \quad (1)$$

While other models have been presented in the literature, we will focus on the model in (1) in this tutorial.

We assume that

- The direct components are sparse in the time-frequency domain.
- The directional sound components are produced by i) a plane wave or ii) an isotropic point-like source (IPLS).
- The direct and diffuse components are uncorrelated, i.e.,
 $E \{S_{\text{dir}}(k, m, \mathbf{p}_i) S_{\text{diff}}^*(k, m, \mathbf{p}_i)\} = 0$.
- $E \{|S_{\text{diff}}(k, m, \mathbf{p}_i)|^2\} = P_{\text{diff}}(k, m)$ for all positions i .
- Optional: $E \{|S_{\text{dir}}(k, m, \mathbf{p}_i)|^2\} = E \{|S_{\text{dir}}(k, m, \mathbf{p}_j)|^2\}$ when $\|\mathbf{p}_i - \mathbf{p}_j\|$ is small.

2. Sound Field and Signal Models

The directional sound component $S_{\text{dir}}(k, m, \mathbf{p}_i)$ can be expressed as:

1) An isotropic point-like source:

$$S_{\text{dir}}(k, m, \mathbf{p}_i) = H_{\text{dir}}(k, \mathbf{p}_i, \mathbf{p}_s) \sqrt{P_s(k, m, \mathbf{p}_s)} e^{j\phi_0(k, m, \mathbf{p}_s)},$$

where $H_{\text{dir}}(k, \mathbf{p}_i, \mathbf{p}_s)$ describes the transfer function from the source position $\mathbf{p}_s(k, m)$ to position \mathbf{p}_i and $P_s(k, m, \mathbf{p}_s)$ is the power of the source.

2) A plane wave:

$$S_{\text{dir}}(k, m, \mathbf{p}_i) = e^{-j\mu(k, \phi_s)} \sqrt{P_s(k, m)} e^{j\phi_0(k, m)},$$

where $\mu(k, \phi_s)$ denotes the spatial frequency that depends on the direction-of-arrival (DOA) $\phi_s(k, m)$.

The i -th microphone signal can be expressed as

$$X_i(k, m) \triangleq X(k, m, \mathbf{p}_i) = S(k, m, \mathbf{p}_i) + V(k, m, \mathbf{p}_i),$$

where V denotes the noise received at position \mathbf{p}_i and S and V are assumed to be uncorrelated, i.e., $E \{S(k, m, \mathbf{p}_i) V^*(k, m, \mathbf{p}_j)\} = 0$ for all i, j, k and m .

Overview

- 1 Overview
- 2 Sound Field and Signal Models
- 3 **Parameter Estimation**
 - Direction of Arrival Estimation
 - Position Estimation
 - Signal-to-Diffuse Ratio
 - Diffuseness Estimation
- 4 Estimation of the Sound Pressures
- 5 Applications and Examples
- 6 Summary

3. Parameter Estimation

- Commonly performed in the time-frequency domain (e.g., short-time Fourier transform (STFT) domain or perceptually motivated transform domain).
- It is assumed that the sound sources do not overlap in the time-frequency domain. See also [Rickard and Yilmaz, 2002] in which the W-disjoint orthogonality of speech is studied.
- We desire high temporal, spatial and spectral resolutions.
- Parameters:
 - ▶ Direction of arrival
 - ▶ Position
 - ▶ Signal-to-diffuse ratio (SDR)
 - ▶ Diffuseness
- In the context of CASA, the following parameters are used:
 - ▶ Interaural Level Differences
 - ▶ Interaural Time Differences / Interaural Phase Differences
 - ▶ Interaural Coherence

3.1 Direction of Arrival Estimation

- The active sound intensity vector describes the direction and magnitude of the net flow of energy characterizing the sound field at the measurement position.
- The instantaneous active sound intensity vector is given by

$$\mathbf{i}_a(k, m, \mathbf{p}_i) \propto \text{Re} \{ X(k, m, \mathbf{p}_i) \mathbf{u}^*(k, m, \mathbf{p}_i) \},$$

where $X(k, m, \mathbf{p}_i)$ is the sound pressure and $\mathbf{u}(k, m, \mathbf{p}_i)$ is a vector containing the particle velocities in the x, y and z dimensions.

- Signals proportional to the sound pressure and particle velocity (a.k.a. B-format signals) can be measured using an:
 - ▶ Acoustic vector sensor
 - ▶ Sound field microphone
 - ▶ Circular microphone array
 - ▶ Spherical microphone array

3.1 Direction of Arrival Estimation

- The instantaneous DOA can be expressed by means of the unit vector \mathbf{e}_{DOA} , which points towards to the direction where the sound is coming from, namely

$$\mathbf{e}_i(k, m) = - \frac{\mathbf{i}_a(k, m, \mathbf{p}_i)}{\|\mathbf{i}_a(k, m, \mathbf{p}_i)\|}.$$

- It is well known that the response of differential arrays deviates from the desired figure-of-eight response. The effect of such deviations in the context of DirAC parameter estimation was studied in [Kallinger et al., 2008]. In the absence of noise, it was shown that the relation between the observed azimuth $\hat{\phi}$ and the corresponding true azimuth ϕ can be formulated analytically:

$$\hat{\phi} = \arctan \left(\frac{\sin \left(\frac{2\pi kr}{Kc} \sin \phi \right)}{\sin \left(\frac{2\pi kr}{Kc} \cos \phi \right)} \right),$$

where r denotes the radius of the array, c denotes the speed of sound, and K denotes the number of subbands.

3.1 Direction of Arrival Estimation

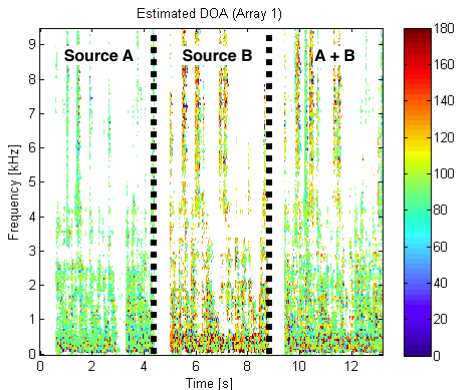


Figure: Instantaneous DOA estimates obtained using a circular array ($r = 2.65$ cm and $M = 4$). Source A was located at 90° and source B was located at 120° .

3.1 Direction of Arrival Estimation

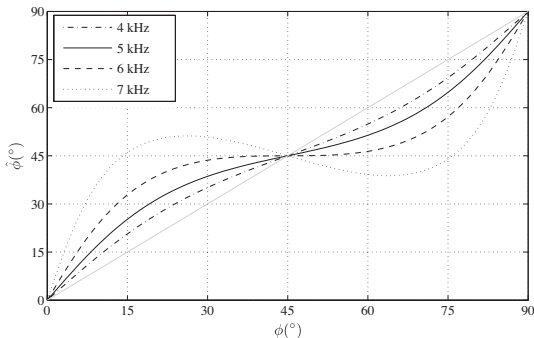


Figure: Illustration of bias for an azimuth estimation based on active sound intensity vectors (cylindrical array: $r = 2.65$ cm and $M = 4$). The solid gray line corresponds to an unbiased angle estimation.

3.1 Direction of Arrival Estimation

- We can estimate the instantaneous DOA using, for example, the ESPRIT (Estimation of Signal Parameters by Rotational Invariance Techniques) algorithm.

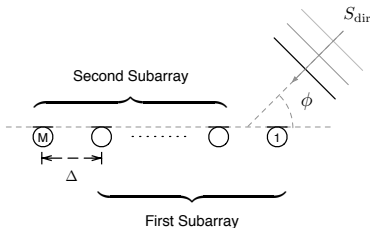


Figure: Two subarrays of a uniform linear array.

- A unit vector pointing towards the direct sound source is given by

$$\mathbf{e}_i(k, m) = \begin{bmatrix} \cos[\phi(k, m)] \\ \sin[\phi(k, m)] \end{bmatrix}.$$

3.1 Direction of Arrival Estimation

- Assuming a sound at a fixed position, we can express the direct components received by M microphones as:

$$\mathbf{s}_{\text{dir}}(k, m) = \mathbf{d}(k, m) \sqrt{P_s(k, m)} e^{j\phi_0(k, m)},$$

where $\mathbf{d}(k, m) = \left[1 e^{-j\mu(k, m)} \dots e^{-j(M-1)\mu(k, m)} \right]^T$ with

$$\mu(k, m) = \frac{2\pi k}{Kc} \Delta \cos[\phi(k, m)]$$

and K denotes the number of subbands.

- The power spectral density (PSD) matrix (in the absence of noise) is given by

$$\begin{aligned} \Phi_{\mathbf{s}_{\text{dir}}} &= \mathbb{E} \left\{ \mathbf{s}_{\text{dir}} \mathbf{s}_{\text{dir}}^H \right\} \\ &= P_s \begin{bmatrix} 1 & e^{j\mu} & \dots & e^{j(M-1)\mu} \\ e^{-j\mu} & 1 & \dots & e^{-j(M-2)\mu} \\ \vdots & \dots & \ddots & \vdots \\ e^{-j(M-1)\mu} & e^{-j(M-2)\mu} & \dots & 1 \end{bmatrix} = P_s \begin{bmatrix} \mathbf{a}_1 \\ \mathbf{A} \\ \mathbf{a}_2 \end{bmatrix}. \end{aligned}$$

3.1 Direction of Arrival Estimation

- Using the two subarrays, we can now form the following expression

$$\begin{bmatrix} \mathbf{a}_1 \\ \mathbf{A} \end{bmatrix} \rho = \begin{bmatrix} \mathbf{A} \\ \mathbf{a}_2 \end{bmatrix}.$$

where ρ is a complex scalar.

- We can for example solve ρ in the least-squares sense such that the estimated DOA is given by

$$\phi(k, m) = \arccos \left(\frac{cK \angle \rho_{\text{LS}}(k, m)}{2\pi k \Delta} \right).$$

- Alternative estimators (some of which require the computation of the eigenvalue decomposition of the PSD matrix) are available [van Trees, 2002].

3.1 Direction of Arrival Estimation

- The narrowband DOA estimators provide incorrect results above the spatial aliasing frequency, which is given by:

$$f_a = \sin\left(\frac{\pi}{M}\right) \frac{c}{r} \text{ Hz}$$

for a uniform circular microphone array (r denotes the radius of the array and c denotes the speed of sound) and

$$f_a = \frac{1}{2} \frac{c}{\Delta} \text{ Hz}$$

for a uniform linear microphone array (Δ denotes the inter-microphone distance).

- To overcome this problem, an envelope-based DOA estimator was recently presented in [Kratschmer et al., 2012]. The estimator can be used to reliably estimate the DOA above the spatial aliasing frequency.
- In particular, the time delay between the Hilbert envelope of two subband microphone signals was used to estimate the DOA.

3.1 Direction of Arrival Estimation

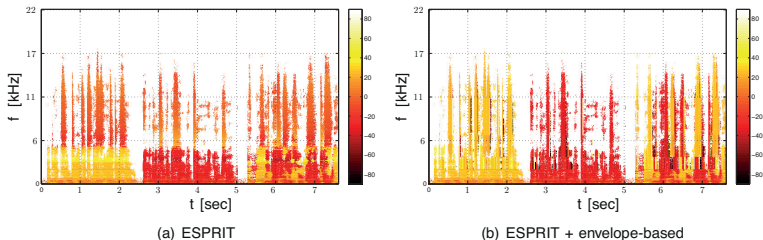


Figure: Estimated DOA for (a) ESPRIT and (b) ESPRIT combined with envelope-based DOA estimator above spatial aliasing frequency for a recorded reverberant double-talk scenario [Kratschmer et al., 2012]. STFT analysis with 1024-point STFT and 50% overlap. Sampling frequency $f_s = 44.1$ kHz.

3.2 Position Estimation

- We can use two or more microphone arrays with known relative position and orientation.
- **Step 1:** At each array the direction of arrival (DOA) is determined for each time-frequency instance (k, m) .
- **Step 2:** The position (w.r.t. a reference coordinate system) of each IPLS is found via triangulation. In this example, using the fact that $\mathbf{p}_1 + \|\mathbf{d}_1\| \mathbf{e}_1 = \mathbf{p}_2 + \|\mathbf{d}_2\| \mathbf{e}_2$.

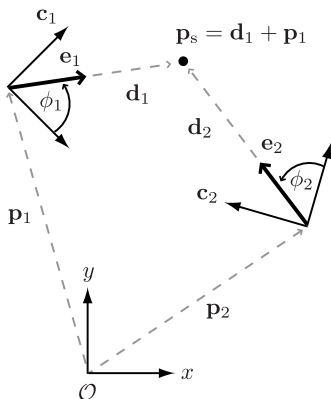


Figure: Position estimation using the DOAs estimated using two arrays.

3.3 Signal-to-Diffuse Ratio

- The signal-to-diffuse ratio (SDR) is defined as

$$\Gamma(k, m, \mathbf{p}_i) = \frac{P_{\text{dir}}(k, m, \mathbf{p}_i)}{P_{\text{diff}}(k, m, \mathbf{p}_i)},$$

where P_{dir} is the power of the direct component and P_{diff} is the power of the diffuse component.

- Different estimation methods have been proposed using:
 - ▶ Omni-directional microphones and assuming the DOA=90° (real part of the complex coherence) [Jeub et al., 2011]
 - ▶ Omni-directional microphones (complex coherence) [Thiergart et al., 2012b]
 - ▶ Using virtual first-order microphones (complex coherence) [Thiergart et al., 2011]
 - ▶ First-order directional microphones (complex coherence) [Thiergart et al., 2012a]

3.3 Signal-to-Diffuse Ratio

- For two omni-directional microphones and $P_{\text{diff}} = P_{\text{diff}}(\mathbf{p}_i) \forall \mathbf{p}_i$, an estimate of the SDR is given by [Thiergart et al., 2012b]

$$\hat{\Gamma}(k, m) = \text{Re} \left\{ \frac{\gamma_{\text{diff}}(k) - \hat{\gamma}_{\text{sig}}(k, m)}{\hat{\gamma}_{\text{sig}}(k, m) - e^{j\hat{\mu}(k, m)}} \right\},$$

where $\gamma_{\text{diff}}(k)$ is the theoretical (or measured) complex spatial coherence between the two sensors in a purely diffuse field, $\gamma_{\text{sig}}(k, m)$ is the complex spatial coherence between S_i and S_j , and $\hat{\mu}(k, m)$ is the estimated spatial frequency of the direct signal components.

- In [Thiergart et al., 2012b], we proposed the following estimator for $\mu(k, m)$:

$$\hat{\mu}(k, m) = \angle \psi_{ij}(k, m),$$

where $\psi_{ij}(k, m) = E \{ X_i(k, m) X_j^*(k, m) \}$ is the cross-PSD.

3.3 Signal-to-Diffuse Ratio

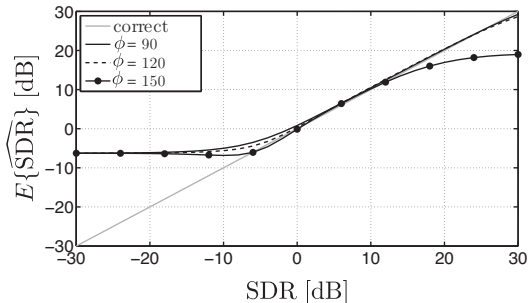


Figure: SDR estimated from the complex spatial coherence as a function of the true SDR at $f = 2.48$ kHz with 10 time frames (≈ 110 ms).

3.4 Diffuseness Estimation

- The diffuseness $\Psi(k, m) \in [0, 1]$ can be expressed in terms of the SDR:

$$\Psi(k, m) = \frac{1}{1 + \Gamma(k, m)} = \frac{P_{\text{diff}}(k, m)}{P_{\text{dir}}(k, m) + P_{\text{diff}}(k, m)}.$$

- The diffuseness can be approximated by the square-root of one minus the coefficient of variation of the active intensity vector $\mathbf{i}_a(k, m)$ [Ahonen and Pulkki, 2009]:

$$\Psi(k, m) = \sqrt{1 - \frac{\|E\{\mathbf{i}_a(k, m)\}\|}{E\{\|\mathbf{i}_a(k, m)\|\}}}$$

- This estimator determines the diffuseness of sound field by comparing the length of the average intensity vector (numerator) with the average length of the intensity vector (denominator).

3.4 Diffuseness Estimation

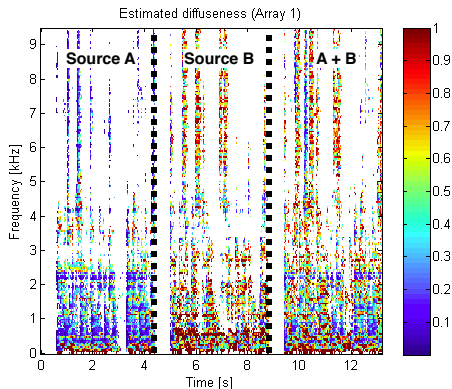


Figure: Diffuseness estimates obtained using a circular array with $M = 4$ microphones using the coefficient of variation method. Source A was located at 90° at a distance of 1 m and source B was located at 120° at a distance of 2 m.

Overview

- 1 Overview
- 2 Sound Field and Signal Models
- 3 Parameter Estimation
- 4 Estimation of the Sound Pressures**
- 5 Applications and Examples
- 6 Summary

4. Estimation of the Sound Pressures

- To manipulate or synthesize the recorded reference signal, we estimate S_{dir} and S_{diff} based on the SDR or diffuseness.
- In principle, more recorded signals can be transmitted and used.
- In the absence of noise, we can for example estimate S_{dir} using the square-root Wiener filter using a reference signal $R(k, m)$:

$$\begin{aligned}\widehat{S}_{\text{dir}}(k, m) &= \sqrt{\frac{P_{\text{dir}}(k, m)}{P_{\text{dir}}(k, m) + P_{\text{diff}}(k, m)}} R(k, m) \\ &= \sqrt{\frac{\Gamma(k, m)}{1 + \Gamma(k, m)}} R(k, m) \\ &= \sqrt{1 - \Psi(k, m)} R(k, m)\end{aligned}$$

such that $E \left\{ |\widehat{S}_{\text{dir}}(k, m)|^2 \right\} = P_{\text{dir}}(k, m)$.

- Note that the square-root Wiener filter has received lots of attention in the context of noise reduction for hearing aid devices.

4. Estimation of the Sound Pressures

- In a similar way, we can estimate the diffuse sound component S_{diff} :

$$\begin{aligned}\widehat{S}_{\text{diff}}(k, m) &= \sqrt{\frac{P_{\text{diff}}(k, m)}{P_{\text{dir}}(k, m) + P_{\text{diff}}(k, m)}} R(k, m) \\ &= \sqrt{\frac{1}{1 + \Gamma(k, m)}} R(k, m) \\ &= \sqrt{\Psi(k, m)} R(k, m).\end{aligned}$$

- Note that $E \left\{ |\widehat{S}_{\text{dir}}(k, m)|^2 \right\} + E \left\{ |\widehat{S}_{\text{diff}}(k, m)|^2 \right\} = E \left\{ |R(k, m)|^2 \right\}$.
- Obviously, many existing estimators can be used as well as techniques to reduce musical noise and other artifacts.
- Finally, we can take into account the ambient noise $V(k, m)$.

Overview

- 1 Overview
- 2 Sound Field and Signal Models
- 3 Parameter Estimation
- 4 Estimation of the Sound Pressures
- 5 Applications and Examples**
 - Spatial Reproduction
 - Directional Filtering
 - Dereverberation
 - Acoustic Zoom
 - Virtual Microphone
- 6 Summary

5.1 Spatial Reproduction

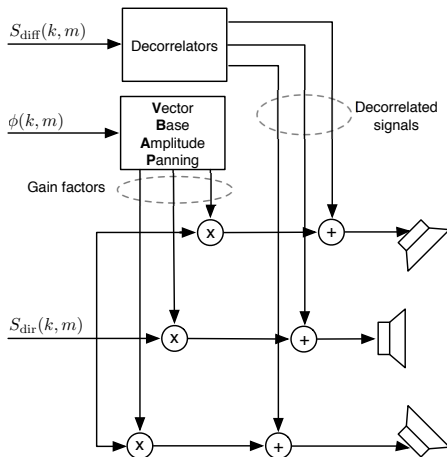


Figure: DirAC spatial reproduction [Pulkki, 2007] - here presented without the inverse STFT.

5.2 Directional Filtering

- In [Kallinger et al., 2009] a directional filter was proposed in the parametric domain by modifying the reference signal $R(k, m)$, the diffuseness $\Psi(k, m)$, and the DOA $\phi(k, m)$ (i.e., the signal and parametric side information used in DirAC).
- In this tutorial, we apply two gain functions directly to the directional and diffuse components:

$$\tilde{S}_{\text{dir}}(k, m) = G_{\text{dir}}[\phi(k, m)] \hat{S}_{\text{dir}}(k, m)$$

$$\tilde{S}_{\text{diff}}(k, m) = G_{\text{diff}} \hat{S}_{\text{diff}}(k, m)$$

- The modified reference signal is now given by:

$$\tilde{R}(k, m) = \tilde{S}_{\text{dir}}(k, m) + \tilde{S}_{\text{diff}}(k, m).$$

- The directional response of a first-order directional microphone is given by

$$D[\phi(k, m)] = \alpha + (1 - \alpha) \cos[\phi(k, m) - \phi_d],$$

where α is the shape parameter of a first-order directivity response and ϕ_d is the desired look-direction.

5.2 Directional Filtering

- The power of a diffuse component at the output of such a first-order directional microphone is given by

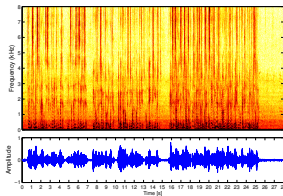
$$Q = \frac{1}{2\pi} \int_0^{2\pi} D^2[\phi] d\phi = \frac{3}{2}\alpha^2 - \alpha + \frac{1}{2}.$$

- The gain functions are therefore given by

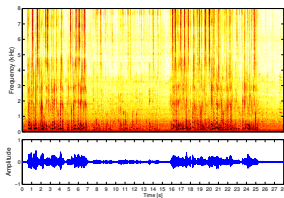
$$G_{\text{dir}}[\phi(k, m)] = D[\phi(k, m)]$$
$$G_{\text{diff}} = \sqrt{Q}.$$

- Obviously, other directivity patterns can be used. Because of errors in the DOA estimates, we cannot use very sharp responses. In practice, a beam-width of 60° can be used without introducing severe audible artifacts (assuming the direct components are sparse in the time-frequency domain).

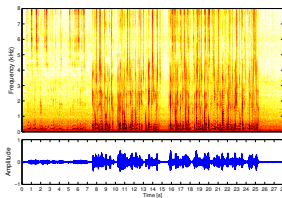
5.2 Directional Filtering



(a) Reference microphone.



(b) Focus on female speaker.



(c) Focus on male speaker.

Figure: Directional filtering example with two speakers. Thanks to Markus Kallinger for providing the audio examples.

5.3 Dereverberation

- A signal which contains less reverberation compared to the reference signal $R(k, m)$ is given by [Kallinger et al., 2011]

$$\tilde{R}(k, m) = S_{\text{dir}}(k, m) + \beta S_{\text{diff}}(k, m)$$

where β ($0 \leq \beta \leq 1$) is the reverberation reduction factor.

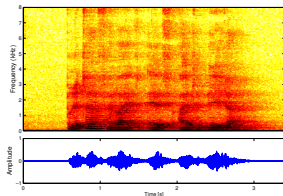
- We seek a (real-valued) gain function that can be applied directly to the reference signal in order to estimate $\tilde{R}(k, m)$, i.e.,

$$\tilde{R}(k, m) = G(k, m) R(k, m).$$

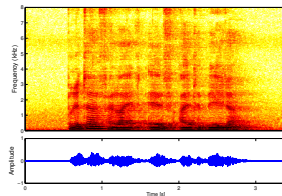
- The gain function that minimizes the error in the minimum mean square sense is given by

$$\begin{aligned} G(k, m) &= \underset{G(k, m)}{\operatorname{argmin}} E \left\{ \left| \tilde{R}(k, m) - G(k, m) R(k, m) \right|^2 \right\} \\ &= 1 - (1 - \beta) \Psi(k, m) = \frac{\Gamma(k, m) + \beta}{\Gamma(k, m) + 1}. \end{aligned}$$

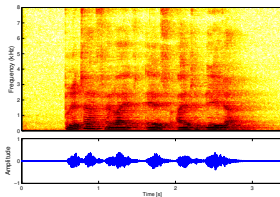
5.3 Dereverberation



(a) Reference microphone.



(b) MVDR beamformer (super-directive).



(c) Parametric -6 dB suppression.

Figure: Dereverberation example - Thanks to Markus Kallinger for providing the audio examples.

5.4 Acoustic Zoom

- In [Schultz-Amling et al., 2010], a technique was proposed for an acoustic zoom, which allows us to virtually change the recording position.
- To change the recording position, we need to:
 1. Change the DOAs of the directional sound sources.
 2. Change the signal-to-diffuse ratio and the levels of the direct sound components.

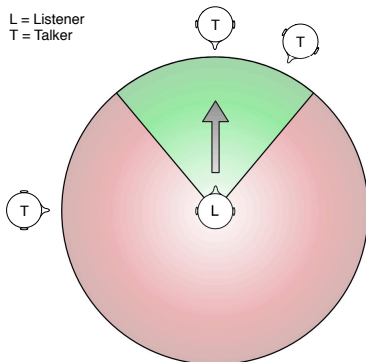


Figure: Acoustic zoom.

5.4 Acoustic Zoom

- It was proposed to remap the DOAs such that they correspond to the new listening position.
- The region of interest increases from 2ϕ to $2\phi'$ when the listener moves d meters closer.
- The following mapping function was derived:

$$\phi' = \arccos \left(\frac{r^2 \cos(\phi) + d^2 - r d [1 + \cos(\phi)]}{(r - d) \sqrt{d^2 + r^2 - 2 r d \cos(\phi)}} \right).$$

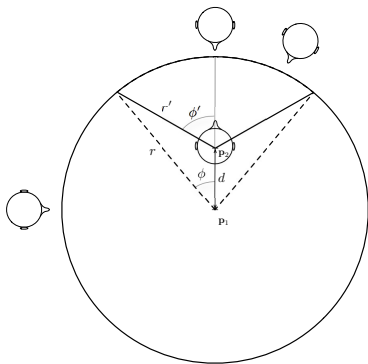


Figure: Details of the geometric setup.

5.4 Acoustic Zoom

- Three assumptions were made for a zoomed-in audio scene:
 1. A sound source becomes louder while approaching it.
 2. Sound coming from the side and back should be attenuated as it moves out of focus.
 3. A sound source moving closer become less diffuse and sound sources moving to the background becomes more diffuse.
- This can be accomplished by extending the aforementioned directional filtering technique which will now depend on the DOA, the radius r and the distance d . The direct and diffuse signals can be modified as follows:

$$\begin{aligned}\tilde{S}_{\text{dir}}(k, m) &= G_{\text{dir}}[\phi(k, m), d, r] \hat{S}_{\text{dir}}(k, m) \\ \tilde{S}_{\text{diff}}(k, m) &= G_{\text{diff}}[\phi(k, m), d, r] \hat{S}_{\text{diff}}(k, m)\end{aligned}$$

- More details can be found in [Schultz-Amling et al., 2010].

5.4 Acoustic Zoom

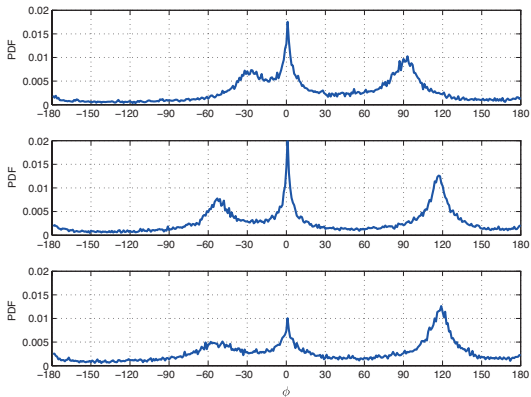


Figure: PDF of the azimuth for the given scenario of three simultaneous talkers [Schultz-Amling et al., 2010]. Top: Microphone at position \mathbf{p}_1 . Middle: Microphone at position \mathbf{p}_2 . Bottom: Microphone at position \mathbf{p}_1 and virtually moved to \mathbf{p}_2 with the acoustic zoom processing.

5.5 Virtual Microphone

- In [Del Galdo et al., 2011], a technique was proposed to generate virtual microphone signals.
- The virtual microphone signal is computed using the **position of the IPLS** as denoted by \mathbf{p}_s . In the following, we assume that $S_{\text{diff}}(k, m) = 0$.
- The **position of the virtual microphone** is defined by the user and is denoted by \mathbf{p}_v .

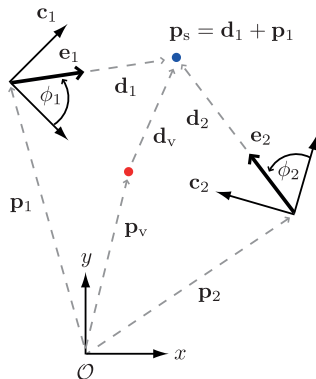


Figure: Geometric illustration of the problem.

5.5 Virtual Microphone

- In the following we use $X(k, m, \mathbf{p}_1)$ as a reference signal. We could also use any other microphone signal or a combination of the microphone signals.

- According to the model and in the absence of noise we have

$$X(k, m, \mathbf{p}_1) = H_{\text{dir}}(\mathbf{p}_1, \mathbf{p}_s) S(k, m, \mathbf{p}_s).$$

- Our objective is to compute a signal that sounds perceptually similar to a signal recorded using a microphone placed at position \mathbf{p}_v :

$$\begin{aligned} X(k, m, \mathbf{p}_v) &= H_{\text{dir}}(\mathbf{p}_v, \mathbf{p}_s) S(k, m, \mathbf{p}_s) \\ &= H_{\text{dir}}(\mathbf{p}_v, \mathbf{p}_s) H_{\text{dir}}^{-1}(\mathbf{p}_1, \mathbf{p}_s) X(k, m, \mathbf{p}_1). \end{aligned}$$

- As we do not know H_{dir} , we propose to use a simple model in which we only model the attenuation of the sound pressure:

$$H_{\text{dir}}[\mathbf{p}_1, \mathbf{p}_s(k, m)] = \frac{1}{\|\mathbf{p}_s(k, m) - \mathbf{p}_1\|} = \frac{1}{\|\mathbf{d}_1(k, m)\|}.$$

5.5 Virtual Microphone

- Using the same model, we can now predict the attenuation from the IPLS to the position of the virtual microphone, i.e.,

$$H_{\text{dir}}[\mathbf{p}_v, \mathbf{p}_s(k, m)] = \frac{1}{\|\mathbf{p}_s(k, m) - \mathbf{p}_v\|} = \frac{1}{\|\mathbf{d}_v(k, m)\|}.$$

- Therefore, the virtual microphone signal is given by

$$X(k, m, \mathbf{p}_v) = \frac{\|\mathbf{d}_1(k, m)\|}{\|\mathbf{d}_v(k, m)\|} X(k, m, \mathbf{p}_1).$$

5.5 Virtual Microphone

- We can simulate any arbitrary directional response by defining the angle $\phi_v(k, m)$ that represents the DOA of the IPLS from the perspective of the virtual microphone:

$$\phi_v(k, m) = \arccos \left(\frac{\mathbf{d}_v(k, m) \cdot \mathbf{c}_v}{\|\mathbf{d}_v(k, m)\|} \right),$$

where \mathbf{c}_v is a unit vector describing the orientation of the virtual microphone.

- Finally, the virtual microphone signal is now given by

$$X(k, m, \mathbf{p}_v) = D[\phi_v(k, m)] \frac{\|\mathbf{d}_1(k, m)\|}{\|\mathbf{d}_v(k, m)\|} X(k, m, \mathbf{p}_1).$$

- We can for instance use

$$D[\phi_v(k, m)] = \frac{1}{2} + \frac{1}{2} \cos[\phi_v(k, m)]$$

to simulate a virtual microphone with cardioid directivity.

5.5 Virtual Microphone

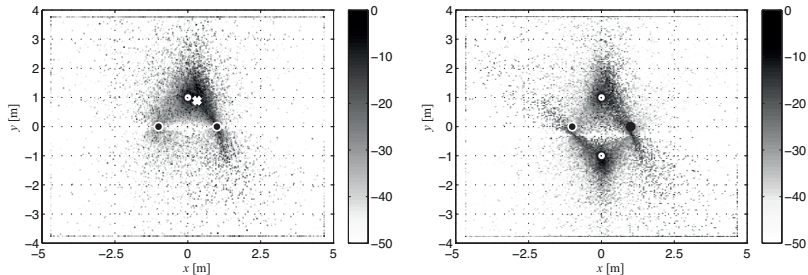


Figure: Spatial power density obtained using two circular arrays ($M = 4$ and $r = 1.6$ cm) for a one talker (left) and two talkers (right).

5.5 Virtual Microphone

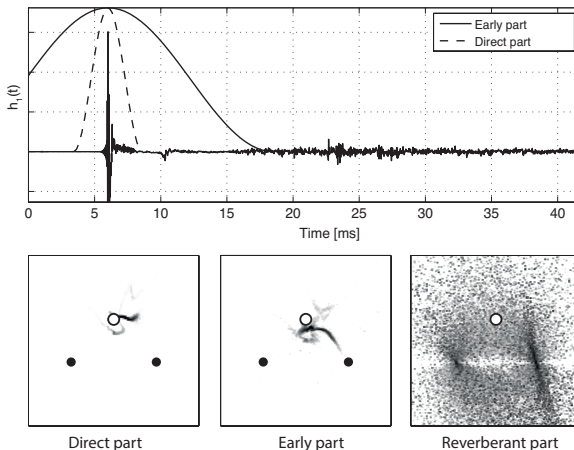


Figure: Spatial power densities (in dB) of the localized positions p_s for a single sound source.

5.5 Virtual Microphone

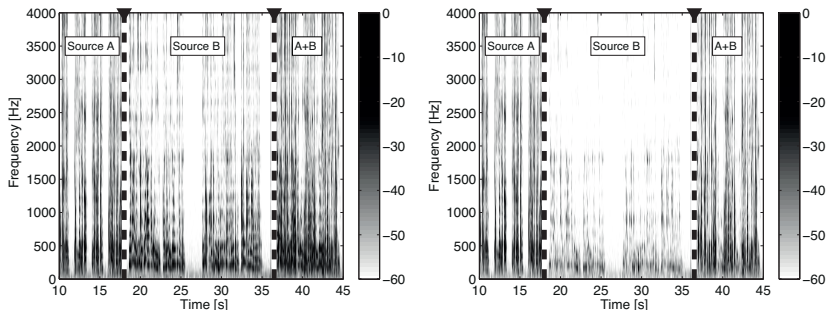


Figure: Spectrogram of a virtual omnidirectional microphone signal (left) and a virtual cardioid microphone pointing to Source A (right).

Overview

- 1 Overview
- 2 Sound Field and Signal Models
- 3 Parameter Estimation
- 4 Estimation of the Sound Pressures
- 5 Applications and Examples
- 6 Summary**

6. Summary

- Parametric-based spatial audio processing relies on a simple yet powerful description of the sound-field.
- Accurate estimation of the time and frequency dependent parameters is paramount.
- Possible model violations can introduce artifacts.
- Several applications have been developed over the last five years.

References I



Ahonen, J. and Pulkki, V. (2009).

Diffuseness estimation using temporal variation of intensity vectors.

In *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, pages 285–288, San Francisco, USA.



Del Galdo, G., Thiergart, O., Weller, T., and Habets, E. A. P. (2011).

Generating virtual microphone signals using geometrical information gathered by distributed arrays.

In *Proc. Hands-Free Speech Communication and Microphone Arrays (HSCMA)*, Edinburgh, United Kingdom.



Jeub, M., Nelke, C. M., Beaugeant, C., and Vary, P. (2011).

Blind estimation of the coherent-to-diffuse energy ratio from noisy speech signals.

In *Proc. European Signal Processing Conf. (EUSIPCO)*, pages 1347–1351.



Kallinger, M., Del Galdo, G., Kuech, F., and Thiergart, O. (2011).

Dereverberation in the spatial audio coding domain.

In *130th AES Convention*, Paper 8429, London, UK.



Kallinger, M., Kuech, F., Schultz-Amling, R., Del Galdo, G., Ahonen, J., and Pulkki, V. (2008).

Enhanced direction estimation using microphone arrays for directional audio coding.

In *Proc. Hands-Free Speech Communication and Microphone Arrays (HSCMA)*, pages 45–48.

References II



Kallinger, M., Ochsenfeld, H., Del Galdo, G., Kuech, F., Mahne, D., Schultz-Amling, R., and Thiergart, O. (2009).

A spatial filtering approach for directional audio coding.

In *126th AES Convention*, Paper 7653, Munich, Germany.



Kratschmer, M., Thiergart, O., and Pulkki, V. (2012).

Envelope-based spatial parameter estimation in directional audio coding.

In *Audio Engineering Society Convention 133*, San Francisco, USA.



Pulkki, V. (2007).

Spatial sound reproduction with directional audio coding.

J. Audio Eng. Soc., 55(6):503–516.



Rickard, S. and Yilmaz, Z. (2002).

On the approximate W-disjoint orthogonality of speech.

In *Proc. IEEE Intl. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*.



Schultz-Amling, R., Kuech, F., Thiergart, O., and Kallinger, M. (2010).

Acoustical zooming based on a parametric sound field representation.

In *Audio Engineering Society Convention 128*, London UK.

References III



Thiergart, O., Del Galdo, G., and Habets, E. A. P. (2011).

Diffuseness estimation with high temporal resolution via spatial coherence between virtual first-order microphones.

In *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, pages 217–220.



Thiergart, O., Del Galdo, G., and Habets, E. A. P. (2012a).

On the spatial coherence in mixed sound fields and its application to signal-to-diffuse ratio estimation.

J. Acoust. Soc. Am., 132(4):2337–2346.



Thiergart, O., Del Galdo, G., and Habets, E. A. P. (2012b).

Signal-to-reverberant ratio estimation based on the complex spatial coherence between omnidirectional microphones.

In *Proc. IEEE Intl. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*.



van Trees, H. L. (2002).

Detection, Estimation, and Modulation Theory, volume IV, Optimum Array Processing. Wiley, New York, USA.