

阵列信号处理 (二)

主讲人 宋辉

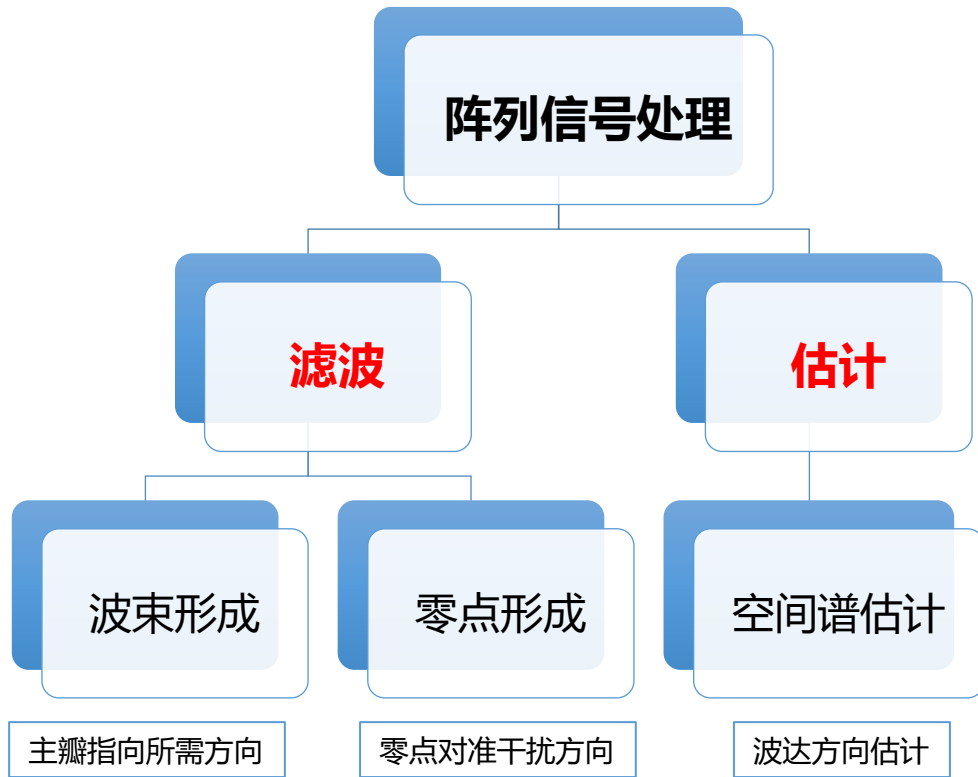
清华大学电子工程系 博士
滴滴AI Labs 语音技术部



7.1 声源定位技术



7.2 波束形成——GSC算法的实际应用





声源定位

确定声音的“方位”

更好的交互体验

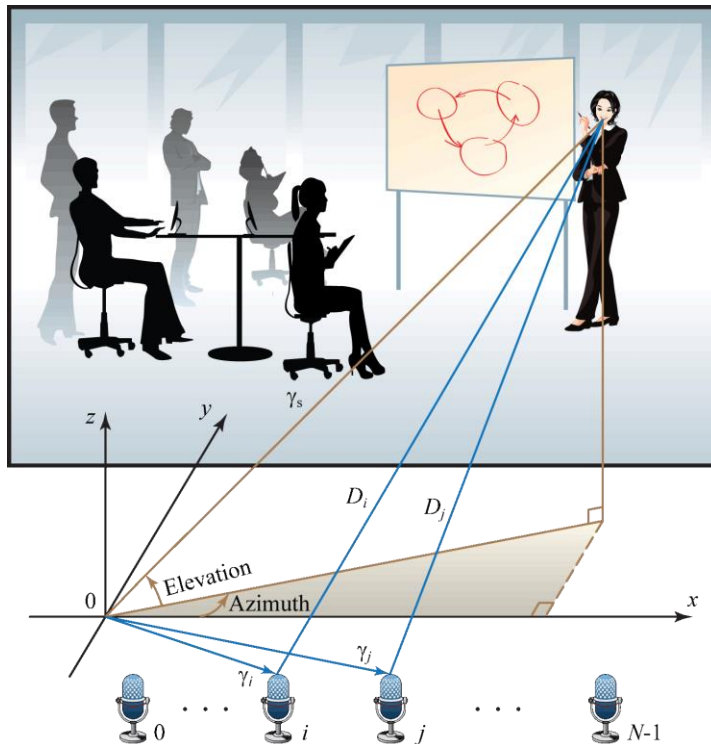
更精准的波束形成

影响声源定位精度的因素

麦克风的数量

阵列构型以及阵列与声源的相对位置关系

背景噪声和混响





7.1 声源定位技术

Maximizing the Steered-response-power (SRP)

High-resolution spectral-estimation

Time difference of arrival (TDOA)

“Brandstein, Ward. Microphone Arrays: Signal Processing Techniques and Applications”



SRP-based Source Localization

“A focused beamformer which steers the array to various locations and searches for a peak in output power.”

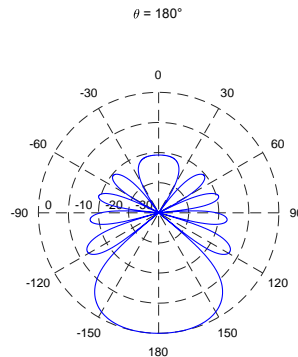
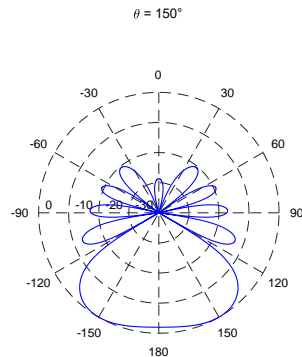
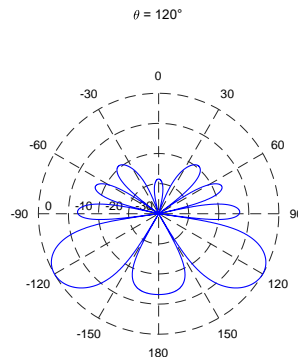
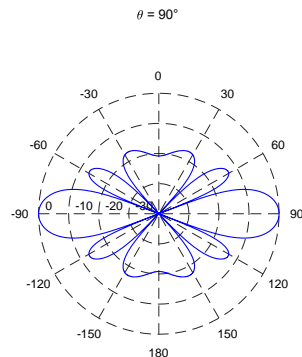
回顾：ULA主波束调向 (beam steering)

调整频率-波数响应函数：

$$Y(\omega, \mathbf{k} - \mathbf{k}_T) \leftarrow Y(\omega, \mathbf{k})$$

修正权向量：

$$\mathbf{w}_T \leftarrow \begin{bmatrix} e^{j\mathbf{k}_T^T \mathbf{p}_0} \\ e^{j\mathbf{k}_T^T \mathbf{p}_1} \\ \vdots \\ e^{j\mathbf{k}_T^T \mathbf{p}_{N-1}} \end{bmatrix} \odot \mathbf{w}$$





SRP-based Source Localization

方法：利用Delay-and-sum波束形成，在每一路输入信号中施加适当的延迟项，补偿声源方向到达麦克风阵列产生的时间差，从而将各个通道中的语音信号成分时间对齐，得到输出信号。

对于阵列输入信号而言：

$$\mathbf{x}(n) = \begin{bmatrix} x_0(n) \\ x_1(n) \\ \vdots \\ x_{N-1}(n) \end{bmatrix} \quad (7.1)$$

经过Delay-and-sum波束形成后的输出信号为：

$$y(n) = \sum_{i=0}^{N-1} x_i(n + \Delta_i) \quad (7.2)$$

其中 $\Delta_i, i = 0, 1, \dots, N - 1$ 称作 *steering delays*



SRP-based Source Localization

当我们把阵列的主波束调向至不同的空间方位时，会产生不同的调向延迟 Δ_i 。显然，Delay-and-sum波束形成的输出信号 $y(n)$ 与主波束的指向相关，故将其记作 $y(n, \mathbf{k}_s)$ ，即：

$$y(n, \mathbf{k}_s) = \sum_{i=0}^{N-1} x_i(n + \Delta_i) \quad (7.3)$$

实际中，由于麦克风阵列的几何布局是已知的，因此，可以选定一个麦克风作为参考麦克风，计算声源到达各个麦克风与参考麦克风之间的到达时间差（TDOA），用 τ 表示，从而将计算绝对延时转化为计算相对延时。

(7.3)式的频域等价形式为：

$$Y(\omega, \mathbf{k}_s) = \sum_{i=0}^{N-1} X_i(\omega) e^{j\omega\Delta_i} \quad (7.4)$$



SRP-based Source Localization

将(7.4)式扩展成更一般的形式:

$$Y(\omega, \mathbf{k}_s) = \sum_{i=0}^{N-1} G_i(\omega) X_i(\omega) e^{j\omega\Delta_i} \quad (7.5)$$

计算SRP:

$$P(\mathbf{k}) = \int_{-\infty}^{\infty} |Y(\omega, \mathbf{k})|^2 d\omega \quad (7.6)$$

Maximizing SRP:

$$\hat{\mathbf{k}}_s = \underset{\mathbf{k}}{\operatorname{argmax}} P(\mathbf{k}) \quad (7.7)$$

“Delay-and-sum SRP” or “Filter-and-sum SRP”



借鉴GCC方法中的PHAT加权思想，定义SRP-PHAT方法：

$$P(\mathbf{k}) = \sum_{i=0}^{N-1} \sum_{l=0}^{N-1} \int_{-\infty}^{\infty} \Psi_{il}(\omega) X_i(\omega) X_l^*(\omega) e^{j\omega\tau_{il}} d\omega \quad (7.8)$$

其中 $\Psi_{il}(\omega)$ 为PHAT加权函数：

$$\Psi_{il}(\omega) = G_i(\omega) G_l^*(\omega) = \frac{1}{|X_i(\omega) X_l^*(\omega)|} \quad (7.9)$$

(7.5)式是SRP-PHAT的频域表达形式。相对应的时域表达形式为：

$$P(\mathbf{k}) = 2\pi \sum_{i=0}^{N-1} \sum_{l=0}^{N-1} r_{il}(\tau_{il}) \quad (7.10)$$

广义互相关函数



SRP-PHAT

SRP-PHAT相当于 N 元麦克风阵列中，对于入射方位 \mathbf{k} 而言，计算所有可能的“麦克风对” (possible pairwise) 的广义互相关函数之和（还包括 N 个零延时的广义自相关函数之和）。

对于只有两个麦克风所组成的阵列，SRP-PHAT等价于广义互相关法（GCC）。随着麦克风数目的增多，GCC方法扩展成为SRP-PHAT方法。

SRP-PHAT运算量偏大，通常需要权衡搜索精度和运算量之间的关系。



High-resolution spectral estimation

Minimum variance (MV) spectral estimation

Eigen-analysis-based techniques

核心：相关矩阵的估计

功率谱密度描述信号功率随频率的分布，是信号的一种频域表示。由于阵列信号处理的主要任务之一是信号的空间参数估计（如DOA估计），所以将功率谱密度的概念在空域加以延伸就显得十分重要。这种广义的功率谱常简称为**空间谱**。空间谱描述了信号的空间参数的分布。



High-resolution spectral estimation

MVDR beamformer:

$$\mathbf{w}_c = \frac{\mathbf{R}_{xx}^{-1}(\omega) \mathbf{v}_k(\mathbf{k}_s)}{\mathbf{v}_k^H(\mathbf{k}_s) \mathbf{R}_{xx}^{-1}(\omega) \mathbf{v}_k(\mathbf{k}_s)} \quad (7.11)$$

MVDR使得目标方向以外的任何干扰贡献的功率最小，同时保持目标方向上的信号功率，可视为一种空间带通滤波器。

MVDR最优波束形成的必要条件：**波达方向 \mathbf{k}_s 的估计。**

最优权重 \mathbf{w}_c 下的阵列输出功率：

$$P(\mathbf{k}_s) = \mathbf{w}_c^H \mathbf{R}_{xx}(\omega) \mathbf{w}_c = \frac{1}{\mathbf{v}_k^H(\mathbf{k}_s) \mathbf{R}_{xx}^{-1}(\omega) \mathbf{v}_k(\mathbf{k}_s)} \quad (7.12)$$



High-resolution spectral estimation

定义空间谱（Capon谱）：

$$P_{Capon}(\mathbf{k}) = \frac{1}{\mathbf{v}_k^H(\mathbf{k}) \mathbf{R}_{xx}^{-1}(\omega) \mathbf{v}_k(\mathbf{k})} \quad (7.12)$$

其峰值对应的入射方向就是声源的入射方向：

$$\hat{\mathbf{k}}_s = \underset{\mathbf{k}}{\operatorname{argmax}} P_{Capon}(\mathbf{k}) \quad (7.13)$$

这就是基于最小方差（MV）的空间谱估计方法。

基于特征值分析的方法：multiple signal classification, MUSIC

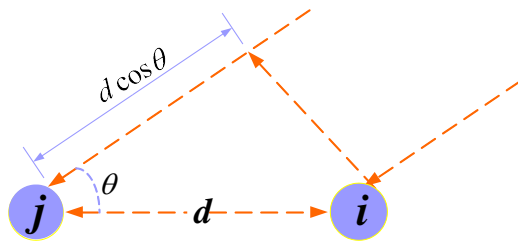
$$P_{MUSIC}(\mathbf{k}) = \frac{1}{\mathbf{v}_k^H(\mathbf{k}) \mathbf{G} \mathbf{G}^H \mathbf{v}_k(\mathbf{k})} \quad (7.14)$$



TDOA-based Source Localization

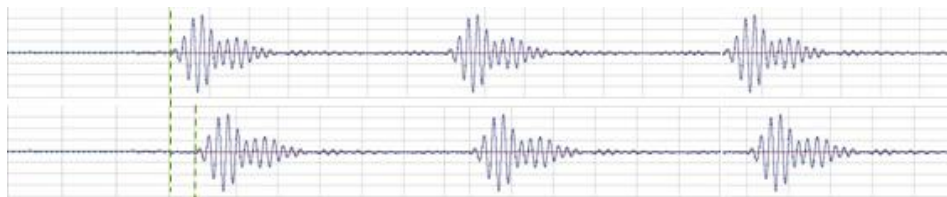
到达时间差 (time difference of arrival, TDOA) :

时延估计 (time delay estimation, TDE) :



麦克风 i

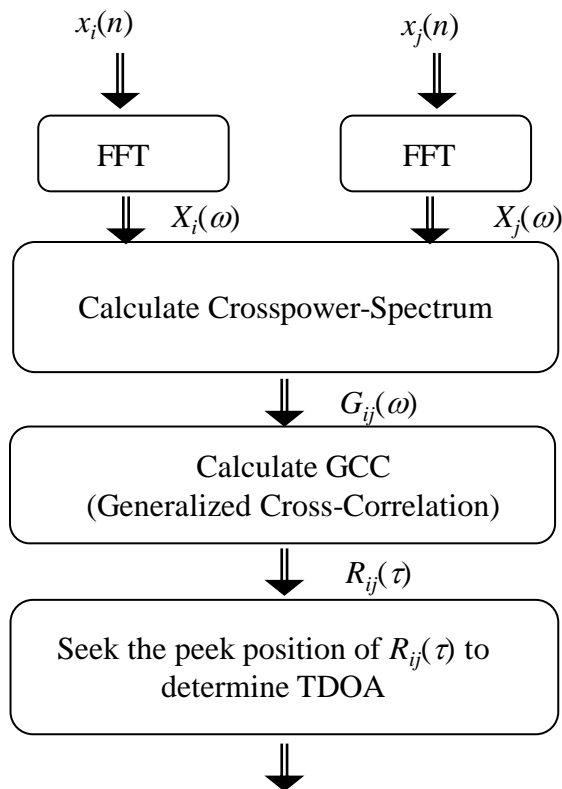
麦克风 j



$$\tau_{ij} = \frac{d \cos \theta}{c}$$



GCC (Generalized cross correlation)



$$R_{ij}(\tau) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \Psi_{ij}(\omega) X_i(\omega) X_j^*(\omega) e^{j\omega\tau} d\omega$$



7.1 声源定位技术

Maximizing the Steered-response-power (SRP)

High-resolution spectral-estimation

Time difference of arrival (TDOA)

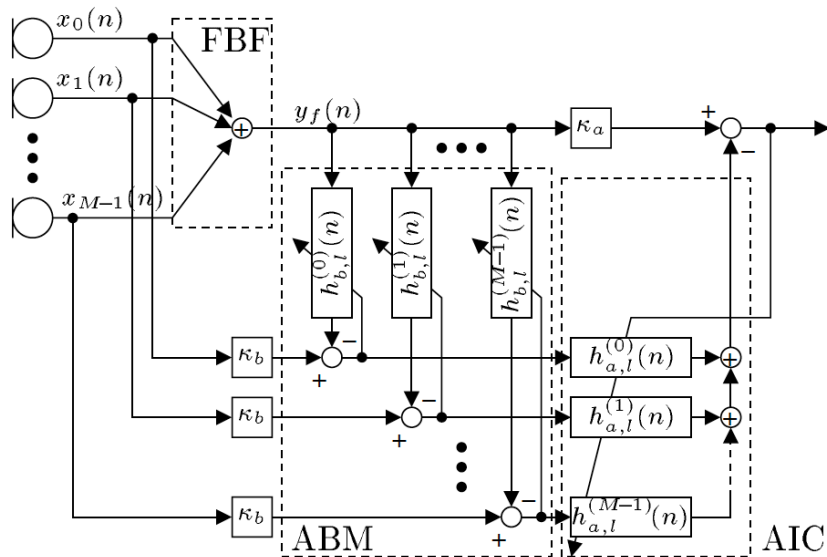


7.2 波束形成——GSC算法的实际应用

Fixed Beamforming

Adaptive Blocking Matrix

Adaptive Interference Canceller



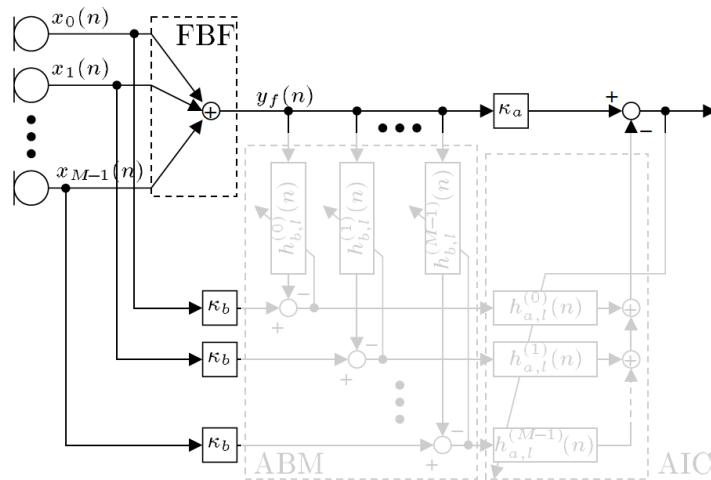
“Efficient Frequency-domain realization of Robust Generalized Sidelobe Cancellers. W. Herbordt, W. Kellermann”



FBF – Fixed Beamforming

Delay-and-sum beamformer

- △ 声源的入射方向已知
- △ 主波束调向 (beam steering) 至声源的入射方向, 对齐每个通道中的语音分量
- △ 求和
- △ 用作ABM的输入信号, 以及AIC模块的期望输出信号



$$y_f(n) = \sum_{m=0}^{M-1} x_m(n) \quad (7.15)$$

steered signals



ABM – Adaptive Blocking Matrix

由 M 条通路构成。

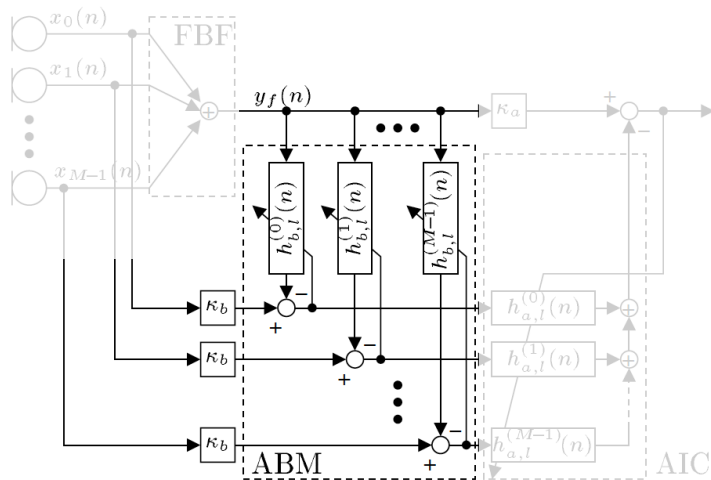
每条通路包含一个自适应滤波器：

△ 自适应消除语音分量

△ 输入： $y_f(n)$

△ 期望： delay of $x_m(n)$

Active during speech presence





ABM – Adaptive Blocking Matrix

以通道 m 为例:

FBF输出 $y_f(n)$ 的频域表示:

$$\mathbf{Y}_f(k) = \text{diag}\{\mathbf{F}(y_f(kL - L), \dots, y_f(kL), \dots, y_f(kL + L - 1))\}^T, \quad (7.16)$$

ABM自适应滤波器系数:

$$\mathbf{H}_b^{(m)}(k) = \mathbf{F}(h_{b,0}^{(m)}(k), \dots, h_{b,L-1}^{(m)}(k), \mathbf{0}_{1 \times L})^T, \quad (7.17)$$

滤波, 频域输出信号为:

$$\mathbf{Y}_b^{(m)}(k) = \mathbf{Y}_f(k) \mathbf{H}_b^{(m)}(k) \quad (7.18)$$

注意到, ABM模块的输出是滤波后的误差信号。



ABM – Adaptive Blocking Matrix

以通道 m 为例：

则时域误差信号：

$$\mathbf{e}_b^{(m)}(k) = \mathbf{x}_m(k - \frac{\kappa_b}{L}) - \mathbf{w} \mathbf{F}^{-1} \mathbf{Y}_f(k) \mathbf{H}_b^{(m)}(k) \quad (7.19)$$

其中：

$$\mathbf{w} = \text{diag}\{(\mathbf{0}_{1 \times L}, \mathbf{1}_{1 \times L})\} \quad (7.20)$$

$$\mathbf{x}_m(k) = (\mathbf{0}_{1 \times L}, x_m(kL), \dots, x_m(kL + L - 1))^T \quad (7.21)$$

则无约束ABM滤波器的更新公式：

$$\mathbf{H}_b^{(m)}(k+1) = \mathbf{H}_b^{(m)}(k) + \mu(k) \mathbf{Y}_f^H(k) \mathbf{E}_b^{(m)}(k) \quad (7.22)$$

Active during speech presence



ABM – Adaptive Blocking Matrix

步长矩阵:

$$\mu(k) = 2\mu \text{diag}\{P_0^{-1}(k) \cdots P_{2L-1}^{-1}(k)\} \quad (7.23)$$

频点功率依旧采用迭代更新:

$$P_l(k) = \lambda P_l(k-1) + (1-\lambda)|Y_{f,l}(k)|^2, \quad l = 0, \dots, 2L-1 \quad (7.24)$$

$Y_{f,l}(k)$ 是 $\mathbf{Y}_f(k)$ 的第 l 个frequency bin。



AIC – Adaptive Interference Canceller

由 M 条通路构成。

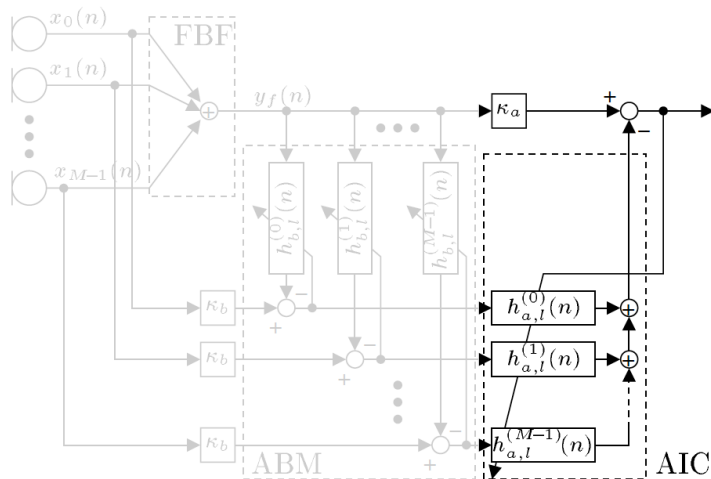
每条通路包含一个自适应滤波器：

△ 自适应滤波器输出结果求和——Interference

△ 输入：ABM的输出信号 $\mathbf{E}_b^{(m)}(k)$

△ 期望：delay of $y_f(n)$

Active during speech absence





AIC – Adaptive Interference Canceller

AIC的输入:

$$\mathbf{X}_a^{(m)}(k) = \text{diag}\{\mathbf{E}_b^{(m)}(k) + \mathbf{J} \mathbf{E}_b^{(m)}(k-1)\}, \quad (7.25)$$

$$\mathbf{J} = \text{diag}\{(1, -1, 1, \dots, -1)_{1 \times 2L}\},$$

则AIC的时域误差信号:

$$\mathbf{e}_a(k) = \mathbf{y}_f(k) - \mathbf{y}_a(k), \quad (7.26)$$

其中:

$$\mathbf{y}_f(k) = (\mathbf{0}_{1 \times L}, y_f(kL - \kappa_a), \dots, y_f(kL + L - 1 - \kappa_a))^T, \quad (7.27)$$

$$\mathbf{y}_a(k) = \mathbf{W} \mathbf{F}^{-1} \sum_{m=0}^{M-1} \mathbf{X}_a^{(m)}(k) \mathbf{H}_a^{(m)}(k). \quad (7.28)$$



AIC – Adaptive Interference Canceller

AIC自适应滤波器更新公式：

$$\mathbf{H}_a^{(m)}(k+1) = \mathbf{H}_a^{(m)}(k) + \boldsymbol{\mu}(k) \mathbf{X}_a^{(m)H}(k) \mathbf{E}_a(k), \quad (7.29)$$

步长矩阵仍然按照(7.23)式的定义，其中频点功率按照如下方式迭代：

$$P_l(k) = \lambda P_l(k-1) + (1-\lambda) \sum_{m=0}^{M-1} |X_{a,l}(k)|^2 \quad (7.30)$$

GSC的最终输出是AIC误差信号 $\mathbf{e}_a(k)$ 的后 L 个点。



本章回顾



7.1 声源定位技术



7.2 波束形成——GSC算法的实际应用

感谢聆听！

Thanks for Listening

