第七讲： 声源定位

李军锋

中国科学院声学研究所

提纲

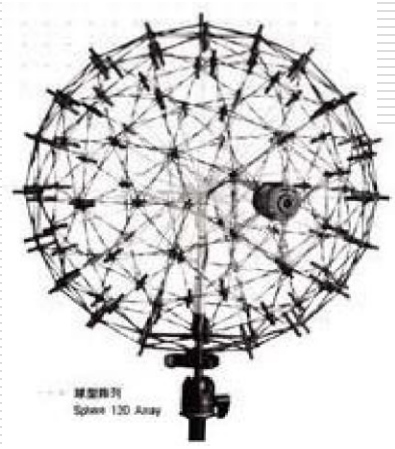
- 简介
- 时间延迟方法
  - 单声源线性阵列
  - 单声源平面阵列
- 多源信号分类(MUSIC)
  - 主分量分析(PCA)
  - MUSIC的原理与方法
- 波束形成的定位方法
- 语音稀疏性与多声源定位
- 基于深度学习的方法

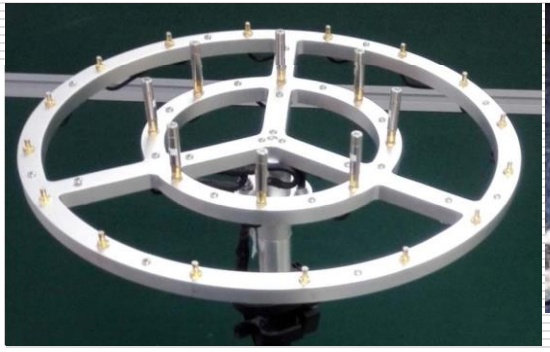
提纲

- 简介
- 时间延迟方法
  - 单声源线性阵列
  - 单声源平面阵列
- 多源信号分类(MUSIC)
  - 主分量分析(PCA)
  - MUSIC的原理与方法
- 波束形成的定位方法
- 语音稀疏性与多声源定位
- 基于深度学习的方法

麦克风阵列硬件系统

- 什么是麦克风阵列？
  - 麦克风按照某种几何拓扑结构排列，空间位置被精确固定；
  - 不同麦克风具有相似的时频响应；
  - 在模拟信号向数字信号转化的过程中，各路信号的编码器具有同步的时钟。



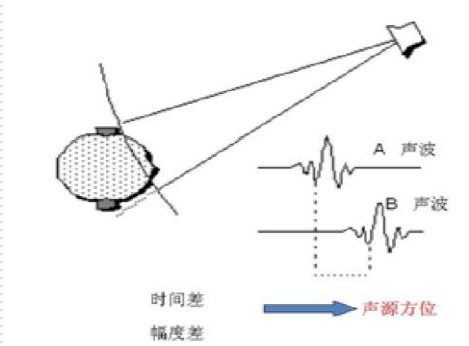


麦克风阵列应用

- 远讲语音增强与识别
- 狙击手定位
- 话者追踪

麦克风阵列的作用

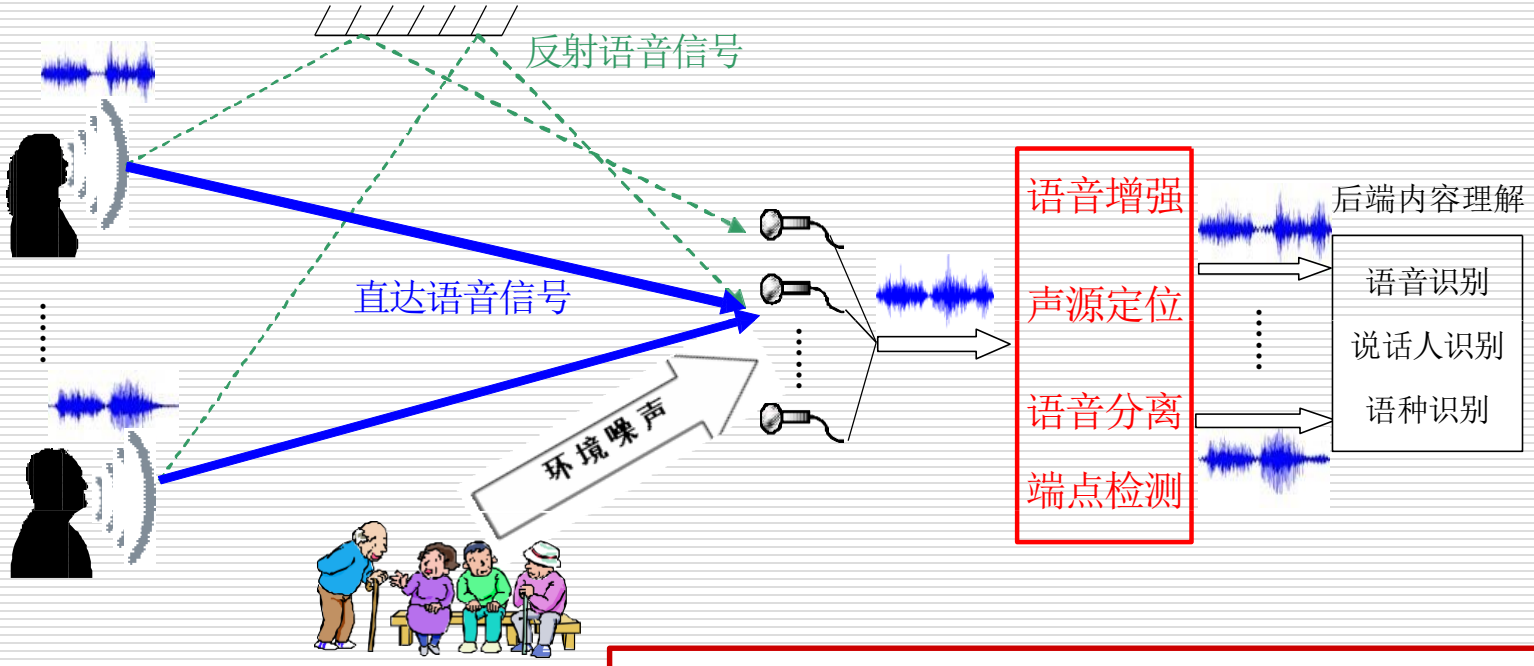
- 捕捉声源的空间位置信息
  - 声源信号到达不同麦克的时间差信息
  - 参考麦克风阵列的的几何拓谱结构



- 利用空间信息对目标源进行定向声聚焦
  - 抑制环境噪声对于目标信号的干扰
  - 抑制房间混响对声源的干扰

基于麦克风阵列的语音信号处理

- 经典多源语音场景



麦克风阵列使得场景中的各路信号具备空间可区分性，为前端信号处理提供空间信息的线索。

多源语音场景:难点与挑战

- 混响的未知性
  - 环境很难预知，如房间大小、形状、固体表面反射系数、声源与麦克几何位置等不可预知；
  - 多路混响传递函数的盲估计，增加不确定性。
- 环境噪声
  - 环境噪声对部分语音频谱的掩蔽作用；
  - 环境噪声的非稳定性，统计特征难以确定。
- 多路语音源信号的叠加
  - 语音源信号
  - 声源的运动导致位置的连续变化

声源定位在阵列处理中的作用

- 在语音增强中，从空间信息区分目标信号和噪声干扰。可以处理非稳定噪声，使得语音增强的效果得到大幅改 善。

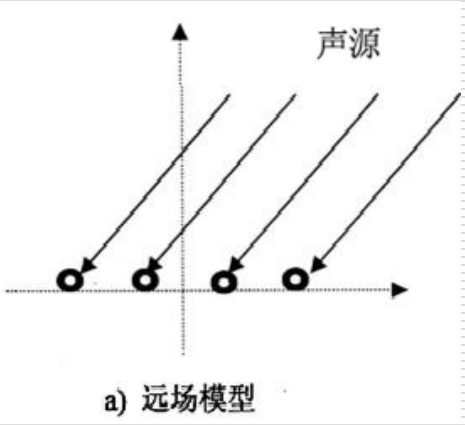
- 在语音分离中，可以从空间信息区别各语音源，甚至把叠加的几路信号还原，这是单通道语音分离方法难以完成的任务。

- 在解混响应用中，可以从空间信息拾取直达声源信号，能够使阵列阵列抑制混响，而不需要预知混响时间。

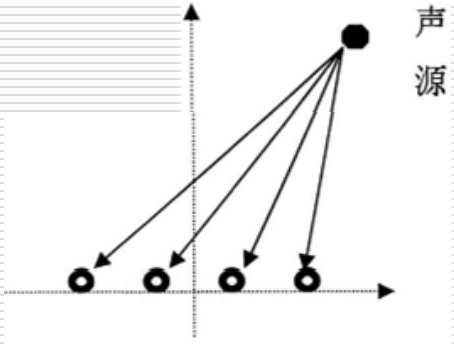
这一切的前提：求取声源的空间位置。

近场与远场

- 远场
  - 声源到麦克的距离足够远，以致于声源到各麦克的直达声传播路径是平行的。在远场假设条件下，难以推测声源的绝对空间位置（三维坐标）。只能估计声源入射的方向。
- 近场
  - 声源到麦克风的距离足够近，声源到达各麦克的传播距离和方向有显著差异，声源传播路径是辐射状。在近场条件下，可以定位声源的三维空间坐标位置。
- 远近场区分标准：$r = 2L^2/\lambda$
  - 其中λ表示波长，L表示阵列长度。声源到阵列中心的距离大于r表示远场，反之为近场。



a) 远场模型



b) 近场模型

提纲

- 简介
- 时间延迟方法
  - 单声源线性阵列
  - 单声源平面阵列
- 多源信号分类(MUSIC)
  - 主分量分析(PCA)
  - MUSIC的原理与方法
- 波束形成的定位方法
- 语音稀疏性与多声源定位
- 基于深度学习的方法

基于时间延迟的声源定位(远场)

- 时域模型

$$x_k(t) = s(t - \psi_k) + n_k(t)$$

  - k表示麦克风序号，ψ_k表示声波传播时间，$n_k(t)$表示加性噪声，那么声源到达两个麦克的时间差表示为：

$$\tau = \psi_2 - \psi_1$$

  - 相应地，入射方向可以表示为：

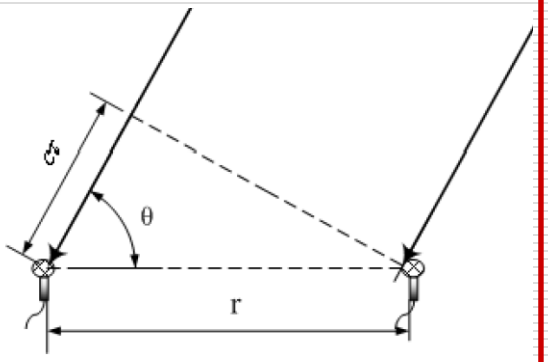
$$\theta = \arg\cos(c\tau/r)$$

- 频域模型

$$X_k(\omega_i) = S(\omega_i)e^{-j\omega_i\psi_k} + N_k(\omega_i)$$

  - ω_i表示角频率，j表示虚数单位。相应地，入射方向的求解值为：

$$\widehat{\theta_i} = \arg\cos(c\widehat{\tau_i}/r) \quad \widehat{\tau_i} = \frac{\angle X_2(\omega_i) - \angle X_2(\omega_i)}{\omega_i}$$



远场模型，两个麦克间距为r，声源入射方向为θ，c为声速，τ为声源在两个麦克间传播的时间差。

混叠问题

- 相位的周期性

*p*有无数种取值可能，怎么办？

$$X_k(\omega_i) = S(\omega_i)e^{-j\omega_i\psi_k} + N_k(\omega_i)$$
$$= S(\omega_i)e^{-j\omega_i\psi_k + 2p\pi} + N_k(\omega_i)$$

$$\hat{\tau}_i = \frac{\angle X_2(\omega_i) - \angle X_2(\omega_i) + 2p\pi}{\omega_i}$$

- 麦克间距约束

$$-r/c \le \tau \le r/c$$

$$\frac{-r\omega_i/c - \angle X_2(\omega_i) + \angle X_1(\omega_i)}{2\pi} \le p \le \frac{r\omega_i/c - \angle X_2(\omega_i) + \angle X_1(\omega_i)}{2\pi}$$

- p没有合适的整数值，求解无效；
- p有唯一整数取值，理想；
- p有多个整数取值产生空间混叠需要解混叠。

时间延迟的快计——交叉相关法

- 交叉相关的定义：

$$c_{12}(\tau) \equiv \int_{-\infty} x_1(t)x_2(t+\tau)dt$$

- 交叉相关的傅里叶变换：

$$C_{12}(\omega) = \int_{-\infty} c_{12}(\tau)e^{-j\omega\tau}d\tau$$

- 根据傅里叶变化的性质，时域卷积等于频域的点乘：

$$C_{12}(\omega) = X_1(\omega)X_2'(\omega)$$

- 对C_{12}进行反傅里叶变换得到：

$$c_{12}(\tau) = \frac{1}{2\pi}\int_{-\infty}^{\infty} X_1(\omega)X_2'(\omega)e^{j\omega\tau}d\omega$$

时间延迟的快计——交叉相关法

- 时间延迟的求解：

$$\tau_{12} = \arg\max_{\tau} c_{12}(\tau)$$

$$-\frac{r}{c} \le \tau \le \frac{r}{c}$$

- 理论解释——交叉相关
  - 假设：

$$X_1(\omega_i) = X_2(\omega_i) \times e^{-j\omega_i \tau_{12}}$$

  - 那么：

$$c_{12}(\tau) = \int_{-\infty}^{\infty} X_1(\omega) X_2^H(\omega) e^{j\omega\tau} d\omega$$

$$= \int_{-\infty}^{\infty} |X_1(\omega)|^2 e^{j\omega(\tau-\tau_{12})} d\omega = \sum_{\omega} |X_1(\omega)|^2 \left[\cos(\omega(\tau-\tau_{12})) + i * \sin(\omega(\tau-\tau_{12}))\right]$$

$$= \sum_{\omega} |X_1(\omega)|^2 \cos\left[\omega(\tau-\tau_{12})\right]$$

- 交叉相关与相位差方法的优劣
  - 交叉相关法,不受混叠影响,但受限于离散傅里叶变换的时间解析度
  - 相位差法时间分辨率高,容易产生时间延迟

提纲

- 简介
- **时间延迟方法**
  - 单声源线性阵列
  - **单声源平面阵列**
  - 语音稀疏性与多声源
- 多源信号分类(MUSIC)
  - 主分量分析(PCA)
  - MUSIC的原理与方法
- 波束形成的定位方法
- 基于深度学习的方法

多麦克平面阵列的定位(远场)

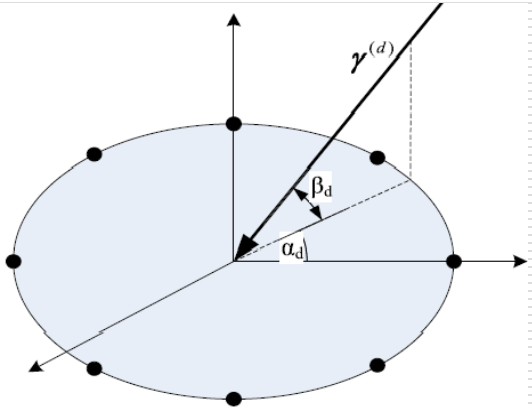
- 空间几何关系
  - 平面阵列相对于线性阵列具有更强的空间区分性;
  - 对于平面阵列，入射方向需要使用方位角α和仰角β表示。也可以使用单位向量γ表示入射方向。

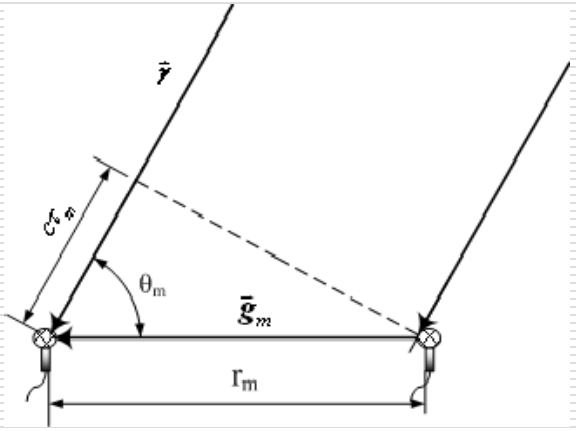
$$\gamma = [\cos\alpha\cos\beta \quad \sin\alpha\cos\beta \quad \sin\beta]^T$$

  - 假设平面阵列由K个麦克构成，这些麦克组合成$K(K-1)/2$个麦克风对。考虑其中第m对麦克，入射方向与麦克连线的夹角为θ_m，它的余弦表示为:

$$\cos\hat{\theta}'_{m,i} = c\hat{\tau}_{m,i}/r_m$$

$$\cos\hat{\theta}_m = \hat{\gamma}_i^T \mathbf{g}_m$$





多麦克平面阵列的定位(远场)

- 代价函数的定义
  - 在理想条件下，没有任何干扰，两个余弦互等

$$\cos\hat{\theta}'_{m,i} = c\hat{\tau}_{m,i}/r_m \quad \Longleftrightarrow \quad \cos\hat{\theta}_m = \hat{\gamma}_i^T \mathbf{g}_m$$

  - 在一般条件下，产生干扰，两者间存在误差，据此定义误差代价函数：

$$f_i(\hat{\gamma}_i) = \sum_{m=1}^{M} \left[\cos\hat{\theta}'_{m,i} - \cos\hat{\theta}_{m,i}\right]^2$$
$$= \sum_{m=1}^{M} \left[c\hat{\tau}_{m,i}/r_m - \hat{\gamma}_i^T \mathbf{g}_m\right]^2$$

  - 最终求解得入射方向

$$\begin{bmatrix} \hat{\gamma}_{1,i} \\ \hat{\gamma}_{2,i} \end{bmatrix} = \left[\sum_{m=1}^{M} \mathbf{g}'_m \mathbf{g}'^T_m\right]^{-1} \sum_{m=1}^{M} c\hat{\tau}_{m,i}\mathbf{g}'_m/r_m$$
$$\hat{\gamma}_{3,i} = \sqrt{1 - \hat{\gamma}_{1,i}^2 - \hat{\gamma}_{2,i}^2}.$$

多麦克平面阵列的定位(远场)

- 入射方向求解
  - 约束最小化求解

$$\hat{\boldsymbol{\gamma}}_i = \min_{\boldsymbol{\gamma}} f_i(\boldsymbol{\gamma})$$
$$\text{subjected to: } \boldsymbol{\gamma}^T\boldsymbol{\gamma} = 1.$$

  - Kuhn-Tucker必要条件

$$L(\boldsymbol{\gamma}, \mu) = f_i(\boldsymbol{\gamma}) + \mu(\boldsymbol{\gamma}^T\boldsymbol{\gamma} - 1)$$

  - 最终入射方向求解

$$\begin{bmatrix} \hat{\gamma}_{1,i} \\ \hat{\gamma}_{2,i} \end{bmatrix} = \left[\sum_{m=1}^{M} \mathbf{g}'_m \mathbf{g}'^T_m\right]^{-1} \sum_{m=1}^{M} c\hat{\tau}_{m,i}\mathbf{g}'_m / r_m$$
$$\hat{\gamma}_{3,i} = \sqrt{1 - \hat{\gamma}^2_{1,i} - \hat{\gamma}^2_{2,i}}.$$

  - 注意：这里的空间混叠问题被忽略

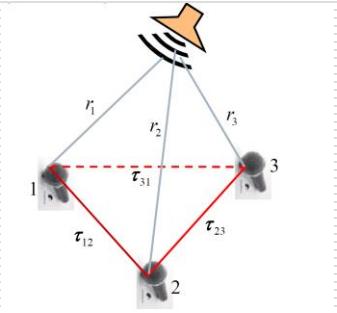
多麦克平面阵列的定位(远场)

- 理想的三角平面阵列定位示例
  - r_k表示声源到达第k麦克的距离，声源的三维空间位置由(r_1, r_2, r_3)三元组确定。
  - 三个时间延迟确定三个方程，求解方程获得空间位置

$r_1 - r_2 = c\tau_{12}$

$r_2 - r_3 = c\tau_{23}$

$r_3 - r_1 = c\tau_{31}$



- 一般平面阵列
  - 噪声、混响、测量误差导致球面不可能汇聚于一点；
  - 多个方程对应绝对坐标(x y z) 形成超定方程；
  - 定义代价函数，求解最优。

$$f(x,y,z)=\sum_{k_1}\sum_{k_2}\left[\sqrt{(x-g_{k1,1})^2+(y-g_{k1,2})^2+(z-g_{k1,3})^2}-\sqrt{(x-g_{k2,1})^2+(y-g_{k2,2})^2+(z-g_{k2,3})^2}-c\tau_{k_1k2}\right]^2$$

提纲

- 简介
- 时间延迟方法
  - 单声源线性阵列
  - 单声源平面阵列
- 多源信号分类（MUSIC）
  - 主分量分析（PCA）
  - MUSIC的原理与方法
- 波束形成的定位方法
- 语音稀疏性与多声源定位
- 基于深度学习的方法

一般直角坐标系表示

一般直角坐标系表示

- 两个坐标轴由$[0\ 1]^T$和$[1\ 0]^T$两个单位方向矢量构成，这一般角坐标系表示两个矢量正交。对于空间中任何一个给定的点(x,y)，它的坐标值在两个坐标轴上的投影值即，该点可由两个坐标轴向量表示：

$$\begin{bmatrix} x \\ y \end{bmatrix} = x \times \begin{bmatrix} 1 \\ 0 \end{bmatrix} + y \times \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

- 给定坐标轴向量的数据描述方法不能反映数据分布特征，对于给定某种特殊数据分布，需要新的坐标系反映数据特征。

二维特征空间示例

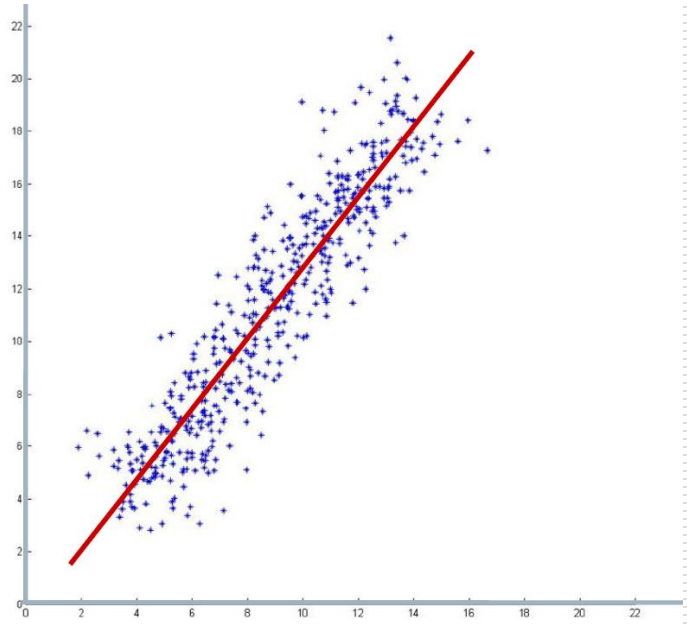
- 二维数据生成

$$y = rand(t) \times \begin{bmatrix} 2 \\ 3 \end{bmatrix}$$

$$\mathbf{x}_t = y + \begin{bmatrix} w_{t1} \\ w_{t2} \end{bmatrix} + 4$$

$$w_{t1} \sim N(t; 0, 1)$$

$$w_{t2} \sim N(t; 0, 1)$$



- x_t 的两个维度产生关联，w_{t_1} 和 w_{t_2} 是遵循零均值高斯分布的随机数。
- 图中点的分布明显表现出一定的取向性，原始的直角坐标系不能反映这种分布特征。我们需要采用新的坐标系，表现数据分布的特征，使得新的坐标系表现出数据的取向性，以及数据的相关性。

二维特征空间的生成

- 解数据的协方差矩阵

$$\mathbf{R} = E\left[\mathbf{x}_t \mathbf{x}_t^T\right]$$

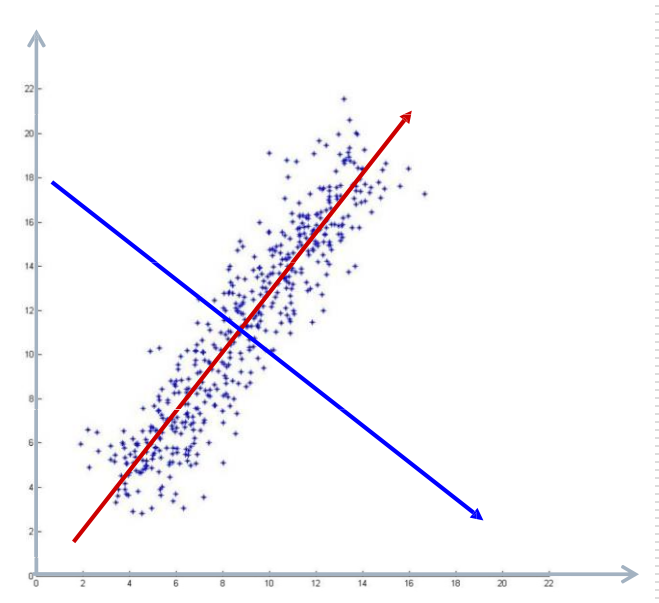
- 特征值分解

$$\mathbf{R} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{U}^H$$

$$\boldsymbol{\Sigma} = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ & \cdots & \cdots & \\ 0 & 0 & \cdots & \lambda_K \end{bmatrix}$$

$$\mathbf{U} = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \cdots & \mathbf{a}_K \end{bmatrix}$$

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_K$$



$$\boldsymbol{\Sigma} = \begin{bmatrix} 0.54 & -0.84 \\ 0.84 & 0.54 \end{bmatrix} \qquad \lambda_1 = 13610 \quad \lambda_2 = 477$$

在新坐标系中，数据投影在主分量方向上的方差最大，在该方向上数据 获得最好的区分度；而另一方向方差较小，数据区分度很小，它表示了 所含信息较少的特征分量

特征征空间与声源定位

- 双麦克单源定位中的频域观察值示例

$$X_k(\omega_i) = S(\omega_i)e^{-j\omega_i\psi_k} + N_k(\omega_i)$$

$$\begin{bmatrix} X_1(\omega_i) \\ X_2(\omega_i) \end{bmatrix} = S(\omega_i) \times \begin{bmatrix} e^{-j\omega_i\psi_1} \\ e^{-j\omega_i\psi_2} \end{bmatrix} + \begin{bmatrix} N_1(\omega_i) \\ N_2(\omega_i) \end{bmatrix}$$

观察值　　声源看做增益系数　　导向矢量由传播时间决定　　噪声可以看作扰动

- 分析表明：对于若干个随时间变化的观察值，观察值信号 主要沿着导向矢量方向分布，白噪声使得观察值偏离导向 矢量。如果我们采用特征空间分析，那么导向矢量代表了 特征空间信息。换句话说：特征空间分析提取了导向矢量。

提纲

- 简介
- 时间延迟方法
  - 单声源线性阵列
  - 单声源平面阵列
- 多源信号分类(MUSIC)
  - 特征空间分析
  - MUSIC的原理与方法
- 波束形成的定位方法
- 语音稀疏性与多声源定位
- 基于深度学习的方法

多源信号叠加模型

- 在时间域

$$x_k(t) = \sum_{d=1}^{D} s_d(t - \psi_{k,d}) + n_k(t)$$

  - D个声源叠加在一起，d为声源序号，$n_k(t)$为白噪声

- 在频率域，i表示频率索引

$$X_k(\omega_i) = \sum_{d=1}^{D} S_d(\omega_i) e^{-j\omega_i \psi_{k,d}} + N_k(\omega_i)$$

- 空间导向矢量

$$\begin{aligned}\mathbf{a}_i(\boldsymbol{\gamma}_d) &= \left[e^{-j\omega_i\psi_{1,d}}, e^{-j\omega_i\psi_{2,d}}, \cdots, e^{-j\omega_i\psi_{K,d}}\right]^T \\ &= e^{-j\omega_i\psi_{1,d}} \times \left[1, e^{-j\omega_i\tau_{1,2,d}}, \cdots, e^{-j\omega_i\tau_{1,K,d}}\right]^T\end{aligned}$$

多源信号叠加模型

- 基于向量的描述方式

$$\mathbf{x}(\omega_i) = \sum_{d=1}^{D} \mathbf{a}_i(\gamma_d) S_d(\omega_i) + \mathbf{n}(\omega_i)$$

$$\mathbf{x}(\omega_i) = [X_1(\omega_i), \cdots, X_K(\omega_i)]^T$$
$$\mathbf{n}(\omega_i) = [N_1(\omega_i), \cdots, N_K(\omega_i)]^T$$

- 自相关描述

$$\mathbf{R}(\omega_i) = E\left[\mathbf{x}(\omega_i)\mathbf{x}^H(\omega_i)\right] = \mathbf{A}_i \mathbf{R}_i^{(s)} \mathbf{A}_i^H + \sigma^2 \mathbf{I}$$

$$\mathbf{R}_i^{(s)} = E\left[\mathbf{s}(\omega_i)\mathbf{s}^H(\omega_i)\right]$$
$$\mathbf{A}_i = \left[\mathbf{a}_i(\gamma_1), \mathbf{a}_i(\gamma_2), \cdots, \mathbf{a}_i(\gamma_D)\right]$$

  - I为单位矩阵，$(\cdot)^H$表示共轭转置，矩阵$R_i^{(s)}$和$A_i R_i^{(s)} A_i^H$的秩由声源数目决定。

自相关矩阵特征分解

- 特征分解

$$\mathbf{R}_i = \mathbf{U}_i \mathbf{\Sigma}_i \mathbf{U}_i^H = [\mathbf{G}_i, \mathbf{V}_i] \begin{bmatrix} \mathbf{\Sigma}_i' & \mathbf{O} \\ \mathbf{O} & \sigma^2 \mathbf{I}_{K-D} \end{bmatrix} \begin{bmatrix} \mathbf{G}_i^H \\ \mathbf{V}_i^H \end{bmatrix}$$

  - 信号子空间： $\mathbf{G}_i = [\mathbf{u}_{i,1}, \cdots, \mathbf{u}_{i,D}]$
  - 噪声子空间： $\mathbf{V}_i = [\mathbf{u}_{i,D+1}, \cdots, \mathbf{u}_{i,K}]$

- 特征属性

$$\mathbf{R}_i \mathbf{V}_i = [\mathbf{G}_i, \mathbf{V}_i] \mathbf{\Sigma}_i \begin{bmatrix} \mathbf{G}_i^H \\ \mathbf{V}_i^H \end{bmatrix} \mathbf{V}_i = [\mathbf{G}_i, \mathbf{V}_i] \mathbf{\Sigma}_i \begin{bmatrix} \mathbf{O} \\ \mathbf{I} \end{bmatrix} = \sigma^2 \mathbf{V}_i$$

$$\mathbf{R}_i \mathbf{V}_i = \sigma^2 \mathbf{V}_i$$

自相关矩阵特征分解

- 特征属性

$$\mathbf{R}_i = \mathbf{A}_i \mathbf{R}_i^{(s)} \mathbf{A}_i^H + \sigma^2 \mathbf{I}$$

$$\mathbf{R}_i \mathbf{V}_i = \mathbf{A}_i \mathbf{R}_i^{(s)} \mathbf{A}_i^H \mathbf{V}_i + \sigma^2 \mathbf{V}_i$$

$$\mathbf{R}_i \mathbf{V}_i = \sigma^2 \mathbf{V}_i$$

$$\mathbf{A}_i \mathbf{R}_i^{(s)} \mathbf{A}_i^H \mathbf{V}_i = 0$$

$$\mathbf{V}_i^H \mathbf{A}_i \mathbf{R}_i^{(s)} \mathbf{A}_i^H \mathbf{V}_i = 0$$

$$\mathbf{A}_i^H \mathbf{V}_i = 0 \quad \Rightarrow \quad \mathbf{a}_i^H(\gamma_d) \mathbf{V}_i = 0$$

正交性的物理意义：

声源信号可以看做导向矢量与一个标量的乘积。本质上，信号空间可以看做是导向矢量生成的空间，因此，任何一个声源的导向矢量都可以由信号空间中的特征向量线性表示。而这些特征向量与噪声子空间正交，因而导向矢量也与特征空间正交。

MUSIC代价函数设计

- 全频带代价函数

$$f_{MUSIC}(\gamma) = \frac{1}{\sum_{i=1}^{F} \mathbf{a}_i^H(\gamma)\mathbf{V}_i\mathbf{V}_i^H\mathbf{a}_i(\gamma)}$$

- 方向求解

$$\hat{\gamma} = \arg\max_{\gamma} f_{MUSIC}(\gamma)$$

  - 由于代价函数无法对方向矢量直接求导，因而不得不采用空间 遍历的方法，对于方位角α~$[0°, 360°]$和仰角β~$[0°, 90°]$，在二维格点上逐一搜索，$[\alpha, \beta] \rightarrow \gamma$，然后将$\gamma$ 代入代价函数，求取最 小值。

自相关矩阵特征分解

- 特征分解

$$\mathbf{R}_i = \mathbf{U}_i \mathbf{\Sigma}_i \mathbf{U}_i^H = [\mathbf{G}_i, \mathbf{V}_i] \begin{bmatrix} \mathbf{\Sigma}_i' & \mathbf{O} \\ \mathbf{O} & \sigma^2 \mathbf{I}_{K-D} \end{bmatrix} \begin{bmatrix} \mathbf{G}_i^H \\ \mathbf{V}_i^H \end{bmatrix}$$

  - 信号子空间： $\mathbf{G}_i = [\mathbf{u}_{i,1}, \cdots, \mathbf{u}_{i,D}]$
  - 噪声子空间： $\mathbf{V}_i = [\mathbf{u}_{i,D+1}, \cdots, \mathbf{u}_{i,K}]$

- 特征属性

$$\mathbf{R}_i \mathbf{V}_i = [\mathbf{G}_i, \mathbf{V}_i] \mathbf{\Sigma}_i \begin{bmatrix} \mathbf{G}_i^H \\ \mathbf{V}_i^H \end{bmatrix} \mathbf{V}_i = [\mathbf{G}_i, \mathbf{V}_i] \mathbf{\Sigma}_i \begin{bmatrix} \mathbf{O} \\ \mathbf{I} \end{bmatrix} = \sigma^2 \mathbf{V}_i$$

$$\mathbf{R}_i \mathbf{V}_i = \sigma^2 \mathbf{V}_i$$

提纲

- 简介
- 时间延迟方法
  - 单声源线性阵列
  - 单声源平面阵列
- 多源信号分类(MUSIC)
  - 主分量分析(PCA)
  - MUSIC的原理与方法
- 波束形成定位方法
- 语音稀疏性与多声源定位
- 基于深度学习的方法

空间能量

- SRP算法
  - 空间搜索，选取搜索空间中能量最大的方向。
  - 对每个搜索方向：
    - 线阵：半平面搜索，搜索范围-90°~90°
    - 平面阵：半球面搜索，搜索范围 -180°~180°，0°~90°
    - 立体阵：球面搜索，搜索范围 -180°~180°，-90°~90°
  - 选取搜索方向：与阵列结构相关
    - 在该方向上做beamforming：delay and sum
    - 选取能量最大的方向作为DOA方向

SRP算法

- 对每个搜索方向q:
  - 其到达麦克风对的时间差为m_q
  - 该方向上的SRP值为:

$$P^{SRP}(q) = \sum_n \left| x_1(n) + x_2\left(n - m_q\right) \right|^2$$

在频域可以表示为:

$$P^{SRP}(q) = \frac{1}{2\pi} \sum_{k=0}^{N-1} X_1(k) \cdot X_2^*(k) \cdot \exp(j \frac{2\pi k m_q}{N})$$

$$= \frac{1}{2\pi} \sum_{k=0}^{N-1} X_{12}(k) \cdot \exp(j \frac{2\pi k m_q}{N})$$

SRP算法

- 对每个搜索方向q:
  - 其到达麦克风对的时间差为m_q
  - 该方向上的SRP值为:

$$P^{SRP}(q) = \sum_n \left| x_1(n) + x_2\left(n - m_q\right) \right|^2$$

在频域可以表示为:

$$P^{SRP}(q) = \frac{1}{2\pi} \sum_{k=0}^{N-1} X_1(k) \cdot X_2^*(k) \cdot \exp(j \frac{2\pi k m_q}{N})$$

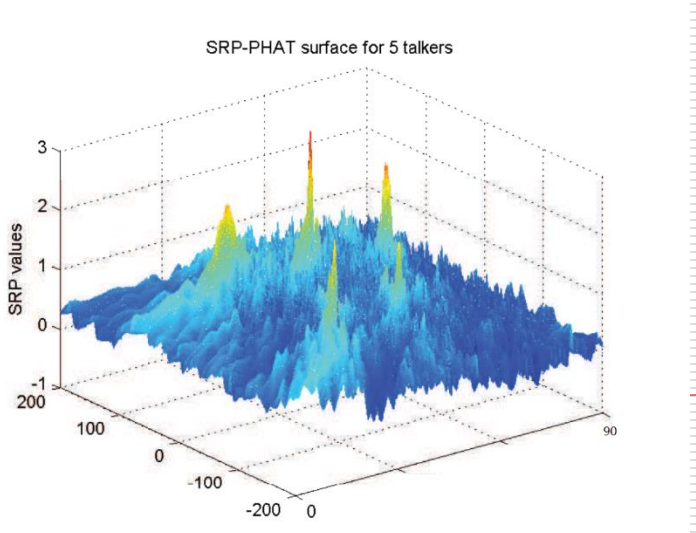
$$= \frac{1}{2\pi} \sum_{k=0}^{N-1} X_{12}(k) \cdot \exp(j \frac{2\pi k m_q}{N})$$

SRP-PHAT算法

- 对每个搜索方向q:
  - 只利用相位信息：SRP-PHAT算法

$$P_{PHAT}^{SRP}(q)=\frac{1}{2\pi}\sum_{k=0}^{N-1}\psi_{PHAT}(k)\cdot X_{12}(k)\cdot\exp(j\frac{2\pi km_q}{N})$$

- 输出
  - 对每个搜索方位，beamforming在该位置的能量
  - 五个声源
  - 搜索空间：上半单位球面



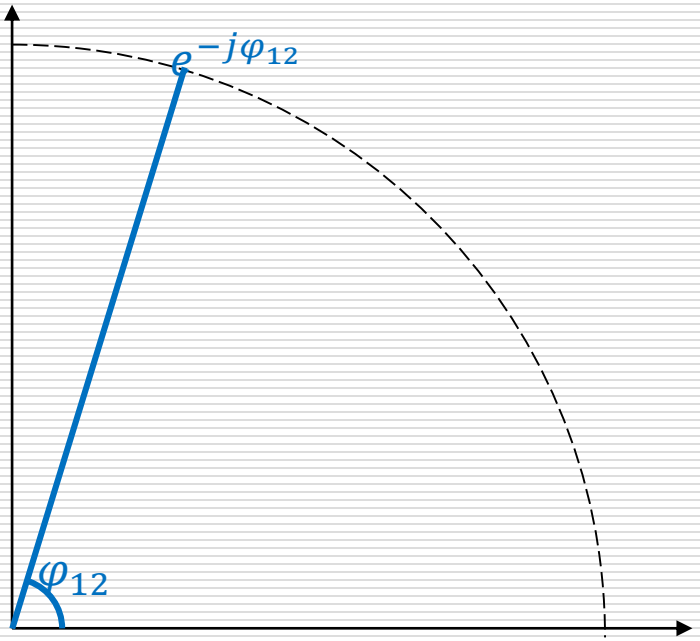
SRP-PHAT算法

- 在每个频率处：

$$S_p(k) = \mathcal{R}\left\{\psi_{PHAT}(k)\cdot X_{12}(k)\exp\left(j\frac{2\pi k m_q}{N}\right)\right\}$$

- Mic12之间拾取数据的相位差：

$$e^{-j\varphi_{12}} = \psi_{PHAT}(k)\cdot X_{12}(k)$$



SRP-PHAT算法

- 在每个频率处：

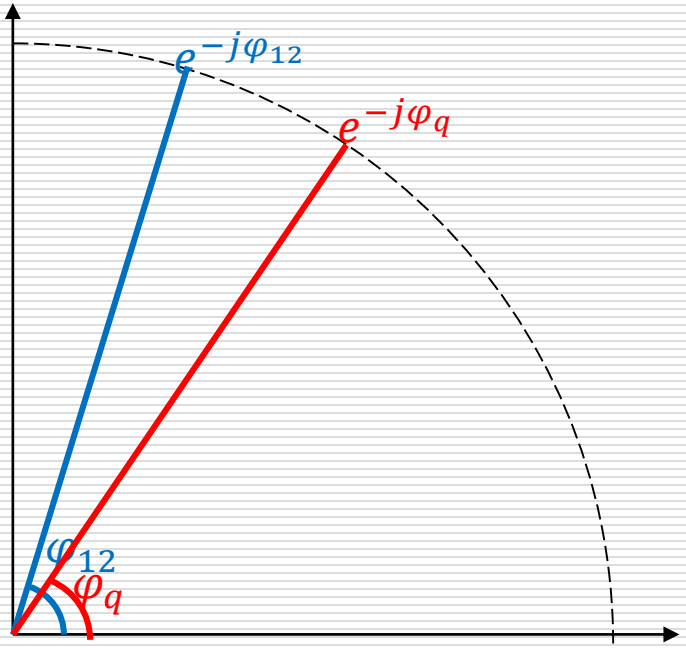
$$S_p(k) = \mathcal{R}\left\{\psi_{PHAT}(k)\cdot X_{12}(k)\cdot \exp\left(j\frac{2\pi k m_q}{N}\right)\right\}$$

- Mic12之间拾取数据的相位差：

$$e^{-j\varphi_{12}} = \psi_{PHAT}(k)\cdot X_{12}(k)$$

- 搜索位置q到mic12的相位差：

$$e^{-j\varphi_q} = \exp\left(-j\frac{2\pi k m_q}{N}\right)$$



SRP-PHAT算法

- 在每个频率处：

$$S_p(k) = \mathcal{R}\left\{\psi_{PHAT}(k)\cdot X_{12}(k)\cdot \exp\left(j\frac{2\pi k m_q}{N}\right)\right\}$$

- Mic12之间拾取数据的相位差：

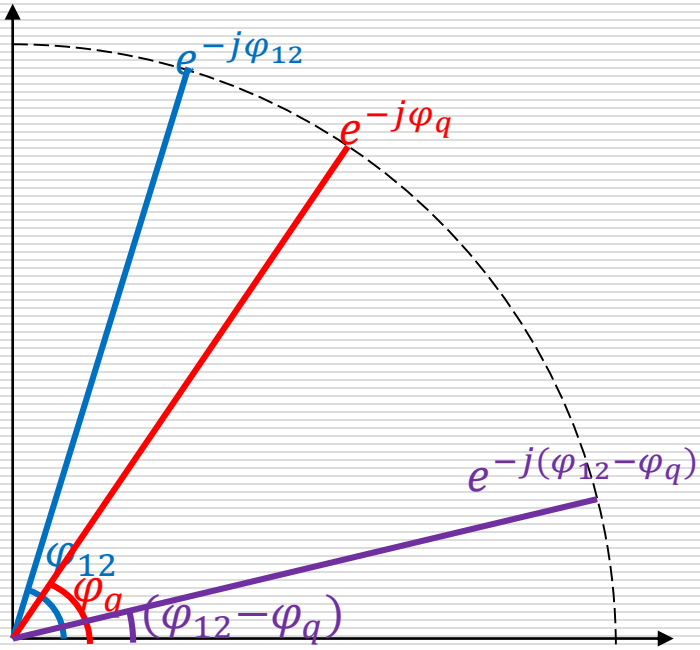
$$e^{-j\varphi_{12}} = \psi_{PHAT}(k)\cdot X_{12}(k)$$

- 搜索位置q到mic12的相位差：

$$e^{-j\varphi_q} = \exp\left(-j\frac{2\pi k m_q}{N}\right)$$

- 两个相位差的差

$$e^{-j(\varphi_{12}-\varphi_q)} = \psi_{PHAT}(k)\cdot X_{12}(k)\cdot \exp\left(j\frac{2\pi k m_q}{N}\right)$$



SRP-PHAT算法

- 在每个频率处：

$$S_p(k) = \mathcal{R}\left\{\psi_{PHAT}(k)\cdot X_{12}(k)\cdot \exp\left(j\frac{2\pi k m_q}{N}\right)\right\}$$

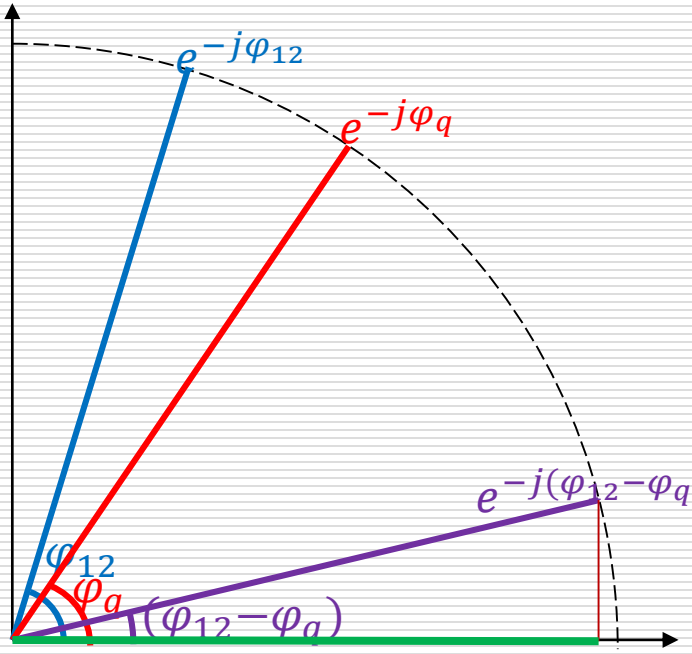
- 两个相位差的差

$e^{-j(\varphi_{12}-\varphi_q)}$

$$= \psi_{PHAT}(k)\cdot X_{12}(k)\cdot \exp\left(j\frac{2\pi k m_q}{N}\right)$$

- 两个相位差之间的相似度

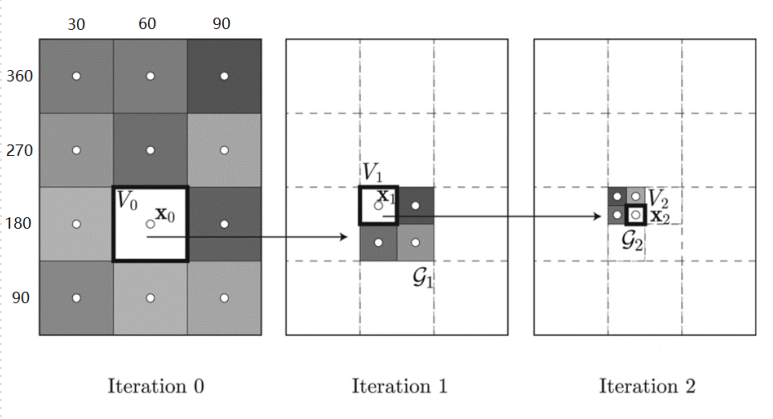
$$\mathcal{R}\{e^{-j(\varphi_{12}-\varphi_q)}\}$$



即：SRP-PHAT是在每个频率处检测搜索位置到麦克风的相位差与真实拾取到数据的相位差的相似度

SRP-PHAT算法

- 简化计算
  - 层次搜索



  - SVD分解

$$X_{i,j}[k] = \frac{X_i[k]X_j[k]^*}{|X_i[k]||X_j[k]|}$$

$$W_{q,i,j}[k] = \exp\left(2\pi\sqrt{-1}k\tau_{q,i,j}/N\right)$$

$$Y_q = \Re\left\{\sum_{i=1}^{M}\sum_{j=(i+1)}^{M}\sum_{k=0}^{N/2} W_{q,i,j}[k]X_{i,j}[k]\right\}$$

矩阵计算

$$\mathbf{X} = \begin{bmatrix} X_{1,2}[0] & X_{1,2}[1] & \cdots & X_{M-1,M}[N/2] \end{bmatrix}^T$$

$$\mathbf{W} = \begin{bmatrix} W_{1,1,2}[0] & W_{1,1,2}[1] & \cdots & W_{1,M-1,M}[N/2] \\ \vdots & \vdots & \ddots & \vdots \\ W_{Q,1,2}[0] & W_{Q,1,2}[1] & \cdots & W_{Q,M-1,M}[N/2] \end{bmatrix}$$

$$\mathbf{Y} = \begin{bmatrix} Y_1 & \dots & Y_Q \end{bmatrix}^T = \Re\{\mathbf{WX}\}$$

利用SVD优化矩阵乘法计算

空间信噪比

- 对每个搜索方向：
  - 求解该方向上声源信号的能量$E_s(\tau)$

$$
\begin{aligned}
E_\tau^{(\mathrm{DS})} &= \frac{\mathbf{d}_\tau{}^H \widehat{\mathbf{\Phi}}_{\mathbf{xx}} \mathbf{d}_\tau}{4} \\
E_\tau^{(\mathrm{MVDR})} &= (\mathbf{d}_\tau{}^H \widehat{\mathbf{\Phi}}_{\mathbf{xx}}^{-1} \mathbf{d}_\tau)^{-1}
\end{aligned}
$$

  - 求解噪声能量$E_N(\tau)$

$$
\begin{aligned}
\mathrm{SNR}_{\mathrm{DS}} &= \frac{\mathbf{d}_\tau{}^H \widehat{\mathbf{\Phi}}_{\mathbf{xx}} \mathbf{d}_\tau}{2\mathrm{tr}(\widehat{\mathbf{\Phi}}_{\mathbf{xx}}) - \mathbf{d}_\tau{}^H \widehat{\mathbf{\Phi}}_{\mathbf{xx}} \mathbf{d}_\tau} \\
\mathrm{SNR}_{\mathrm{MVDR}} &= \frac{(\mathbf{d}_\tau{}^H \widehat{\mathbf{\Phi}}_{\mathbf{xx}}^{-1} \mathbf{d}_\tau)^{-1}}{\frac{1}{2}\mathrm{tr}(\widehat{\mathbf{\Phi}}_{\mathbf{xx}}) - (\mathbf{d}_\tau{}^H \widehat{\mathbf{\Phi}}_{\mathbf{xx}}^{-1} \mathbf{d}_\tau)^{-1}}
\end{aligned}
$$

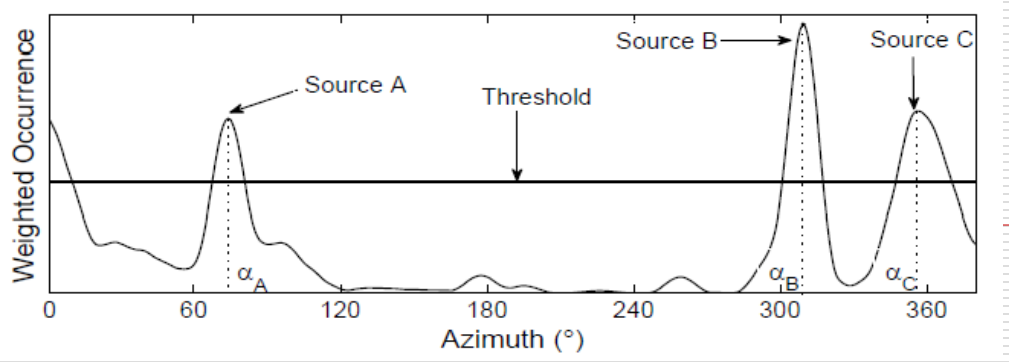
  - 信噪比$SNR(\tau) = E_s(\tau)/E_N(\tau)$
  - 选取空间中信噪比最大的方向作为声源方位

提纲

- 简介
- 时间延迟方法
  - 单声源线性阵列
  - 单声源平面阵列
- 多源信号分类(MUSIC)
  - 主分量分析(PCA)
  - MUSIC的原理与方法
- 波束形成的定位方法
- 语音稀疏性与多声源定位
- 基于深度学习的方法

语信号的稀疏性分布与源定位策略

- 语音稀疏分布特征
  - 语音信号的主要能量分布并非覆盖整个频域，它的主要能量集中在诸如谐波结构等少数频点上。
  - 在多个声源分布的情况下，通常只有一个频点处于支配性的地位，它在该频点上的能量值远远高于其他频点的值，因为其他频点可以忽略。
- 多源定位策略
  - 在每个频点上进行单声源定位，然后确定入射方向
  - 总结各频点的定位方向，采用直方图或聚类方法进行汇总，判断声源数目和入射方向。下例所示为直方图汇总示意图。

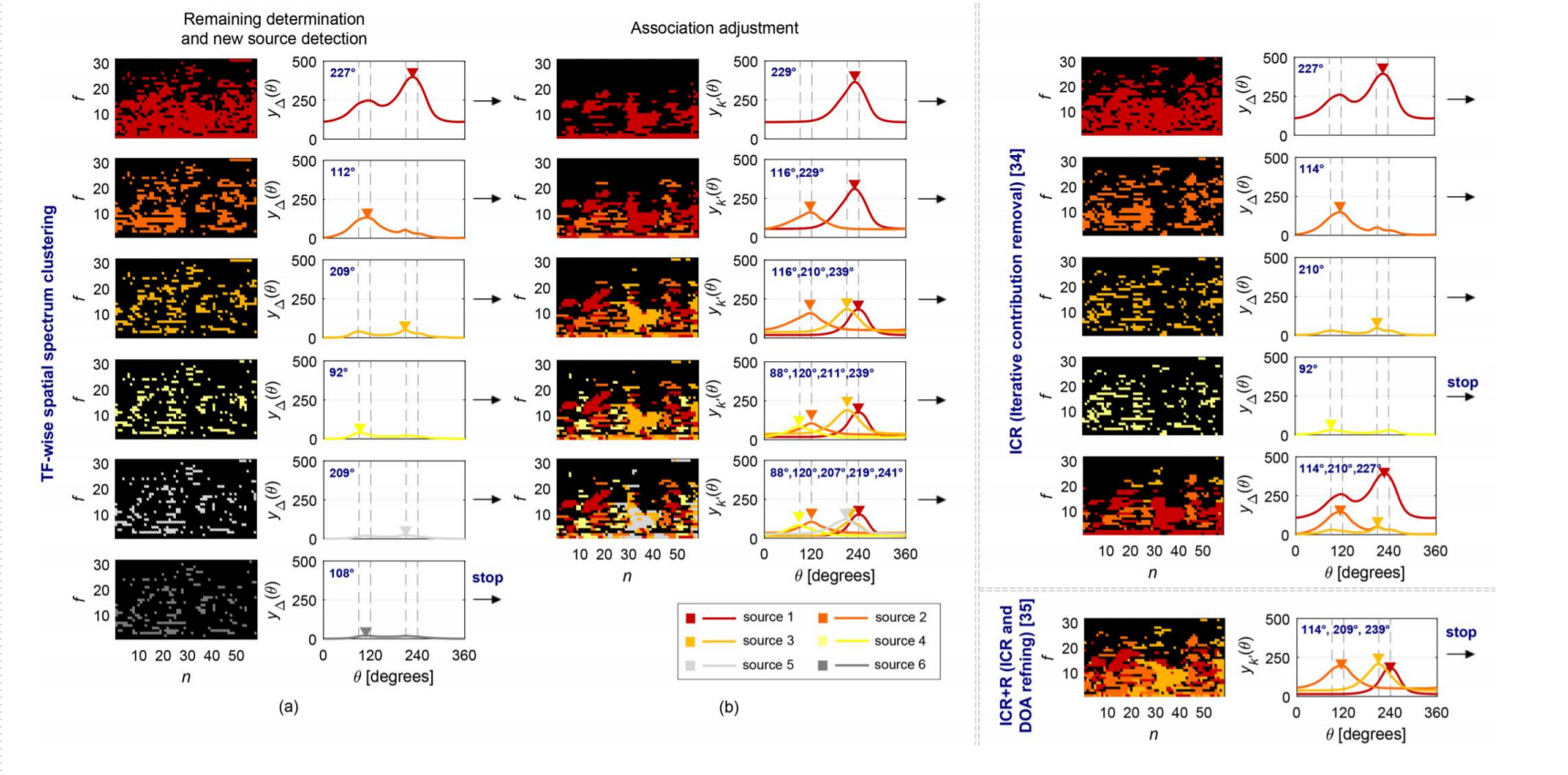


时频点选择

- 基于GMM的定位算法
  - 初步定位得到声源方位$d_1, d_2, \ldots, d_P$
  - 利用GMM算法，根据声源方位d_p反选该方位对应的频点，得到频点集合G_p
  - 使用每个声源对应的时频点集合G_p重新计算该声源的方位d_p
  - 重复2，3直至声源方位d_p收敛
  - 缺点：
    - 一般情况下需要知道声源个数
    - 多声源时初始方位确定不准确

时频点选择

- 逐步递归定位声源

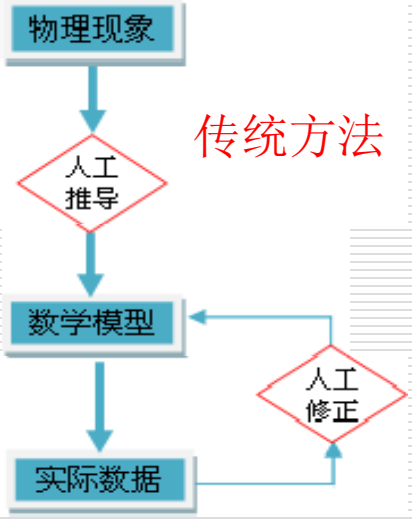


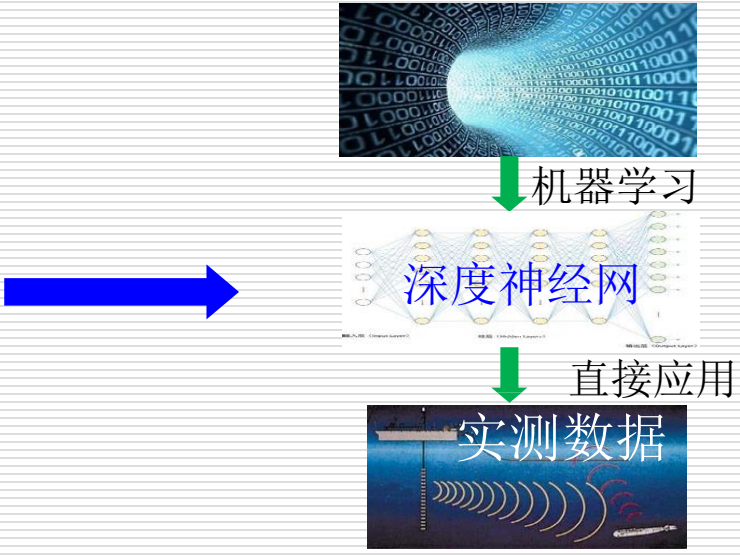
提纲

- 简介
- 时间延迟方法
  - 单声源线性阵列
  - 单声源平面阵列
- 多源信号分类(MUSIC)
  - 主分量分析(PCA)
  - MUSIC的原理与方法
- 波束形成的定位方法
- 语音稀疏性与多声源定位
- 基于深度学习的方法

深度学习Vs.物理建模

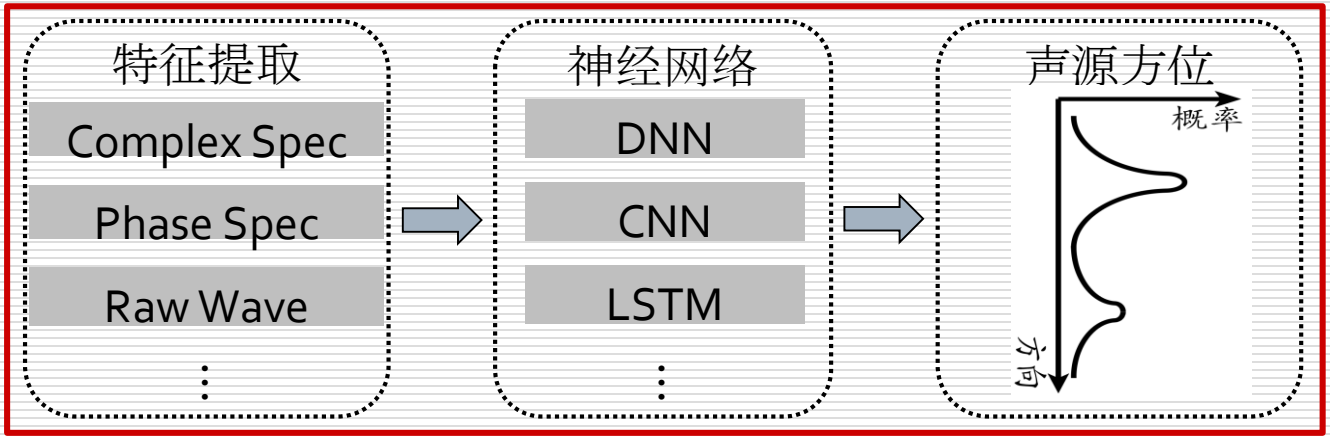
- 物理建模
  - 观察物理现象，构建数学模型，实验估计和修正 参数。
- 深度学习
  - 从数据中学习物理现象所需的最佳数学公式和参 数，无需人工推导和估计。





基于深度学习的方法

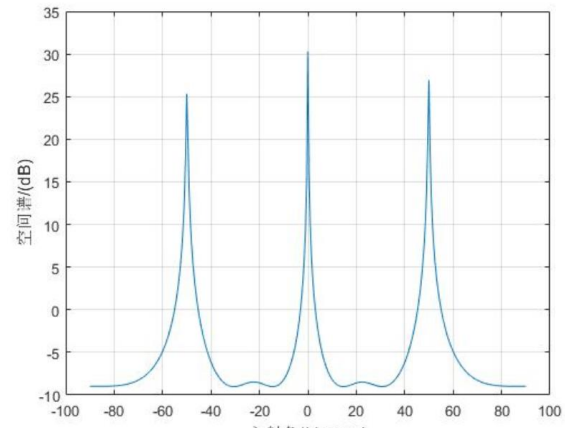
- 端到端方法

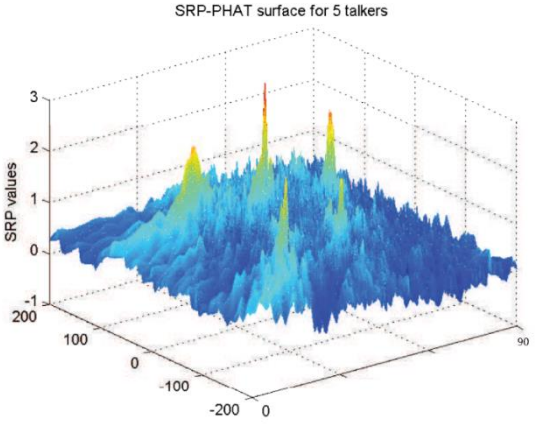


- 利用神经网络估计传统定位算法中某些参数
  - 利用神经网络估计语音存在概率
  - 利用神经网络增强相位差
  - ……

端到端声源定位方法

- 网络的输出
  - DOA的类别
    - 按照一定精度把空间划分为不同的区域
  - 空间谱
    - 利用直达声获得空间谱
  - 声源的坐标
    - 直接估计声源的x, y, z坐标
  - DOA的角度值
    - 直接估计声源的水平角与俯仰角
- 训练的损失函数
  - 对于分类问题，一般采用交叉熵
  - 对于回归问题，一般采用MSE





空间谱

深度学习结合传统信号处理算法

- 选取直达声时频点
  - 估计目标
    - 直达声占主导的概率：
      - IRM：只从能量信息判断

$$\mathrm{IRM}_p(t,f) = \sqrt{\frac{|c_p(f)\,s(t,f)|^2}{|c_p(f)\,s(t,f)|^2 + |h_p(t,f) + n_p(t,f)|^2}}$$

      - PSM：在IRM的基础上加入相位的信息

$$\mathrm{PSM}_p(t,f) = \max\{0, \mathrm{IRM}_p(t,f)\cos(\angle y_p(t,f) - \angle(c_p(f)\,s(t,f)))\}$$

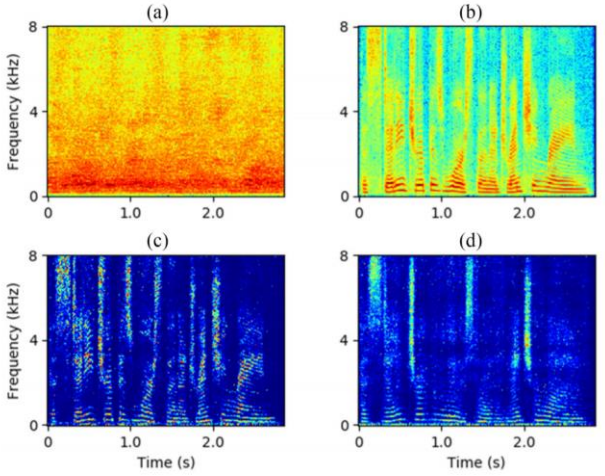


Fig. 3. Illustration of an estimated IRM for a mixture with babble noise in the two-microphone setup (SNR = −6 dB and T60 = 0.9 s). (a) Mixture log power spectrogram; (b) clean log power spectrogram; (c) IRM; (d) estimated IRM.

深度学习结合传统信号处理算法

- 选取直达声时频点
  - 与传统算法的结合
    - 利用估计出来的Mask对算法进行加权
    - Mask加权GCC算法

$$GCC_{p,q}(t,f,k) = \mathcal{R}e\left\{\frac{y_p(t,f)y_q(t,f)^H}{|y_p(t,f)||y_q(t,f)^H|}e^{-j2\pi\frac{f}{N}f_s\tau_{p,q}(k)}\right\}$$

$$= \cos\left(\angle y_p(t,f) - \angle y_q(t,f) - 2\pi\frac{f}{N}f_s\tau_{p,q}(k)\right)$$

    - Mask加权MUSIC算法

$$\hat{\Phi}_{p,q}^{(s)}(f) = \frac{\sum_t M_{p,q}^{(s)}(t,f)\,\boldsymbol{y}_{p,q}(t,f)\,\boldsymbol{y}_{p,q}(t,f)^H}{\sum_t M_{p,q}^{(s)}(t,f)}$$

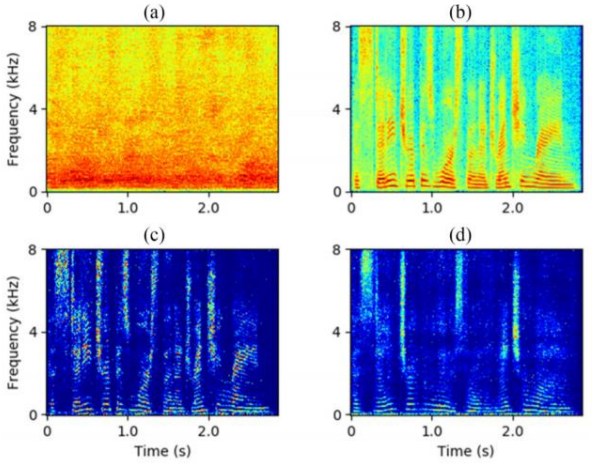


Fig. 3. Illustration of an estimated IRM for a mixture with babble noise in the two-microphone setup (SNR = −6 dB and T60 = 0.9 s). (a) Mixture log power spectrogram; (b) clean log power spectrogram; (c) IRM; (d) estimated IRM.

深度学习结合传统信号处理算法

- 利用神经网络恢复方向信息
  - 估计目标
    - 两通道之间的相位差的正弦余弦值
    - 因为相位差具有周期性(2π)，因此不直接估计

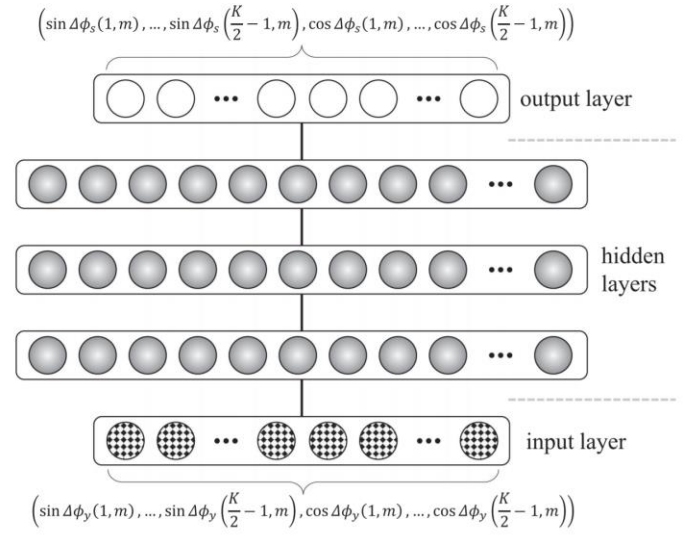


Fig. 1. DNN architecture for the "clean and dry" phase difference estimation.

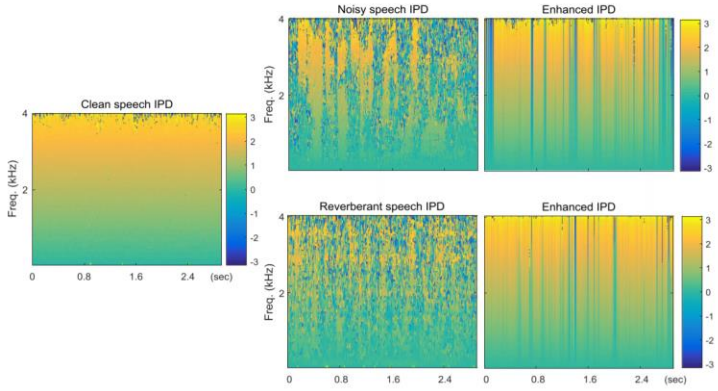


Fig. 3. Interchannel phase differences for clean speech (left), noisy speech corrupted by *Babble* noise at 5 dB SNR (top center), reverberant speech with $RT_{60} = 0.4$ sec (bottom center), and the estimated phase differences from them (right), when a single source is located at $\theta = 40°$.

深度学习结合传统信号处理算法

- 利用神经网络恢复方向信息
  - 与传统算法结合
    - 利用增强后的正弦余弦值计算增强后的相位差：

$$\widehat{\Delta\phi_s}(k,m) = \arctan 2(\widehat{\sin\Delta\phi_s}(k,m), \widehat{\cos\Delta\phi_s}(k,m))$$

    - 结合阵列形状计算声源方位角

$$\theta(k,m) = \arcsin\left(\frac{c\cdot\Delta\phi(k,m)}{2\pi f d}\right)$$

    - 统计
      - 利用直方图统计所有时频点的方位角
      - 利用k-means算法对所有时频点的方位角进行聚类

谢谢

训练数据生成

- 数据生成
  - 麦克风阵列结构多样，声源位置多样，噪声混响多样
  - 难以获取真实场景下的大量带有声源位置标注的数据
  - 利用仿真数据进行神经网络的训练
- 混响仿真
  - 房间冲激响应(RIR)仿真：镜像源法
- 噪声仿真
  - 扩散场噪声仿真工具：ANF-Generator
  - 点噪声仿真：利用RIR仿真
- 声源位置
  - 利用RIR生成不同位置的声源

单通道语音信号处理的弊端

- 前端信号处理包括语音增强、语音分离、解混响等三个环节，现有基于单麦克风的方法在理论上遭遇瓶颈，它们几乎全部存在很强的假设性，而现实需求往往难以满足这些假设。
- 在语音增强中，要求噪声信号稳定，而现实中的噪声大多是不稳定的；
- 在语音分离中，要求语音满足瞬时叠加的要求，不考虑混响，而现实中的语音叠加几乎都存在混响的影响；
- 解混响通常要求预知混响时间，而一般应用中难以预知混响时间。
- 即便是目前流行的深度学习方法，也只是在某些方面减弱了对这些假设的依赖。通过机器学习的方法，也难以满足实用化的需求。