

利用广义相关算法估计远场声源角度

马啸阳 (2018011054), 刘圣禹 (2017010302), 刘坤瓚 (2018011064)

清华大学电子工程系

小组分工—马啸阳: GCC-PHAT; 刘圣禹: SRP-PHAT; 刘坤瓚: GCC-PHAT。

I. 实验原理

A. DOA 估计与经典方法

DOA(Direction Of Arrival) 估计是声学中一个经典的问题, 它的主要内容是通过一组麦克风阵列收到的信号来判断远场声源的角度。在声源定位中, 尤其是在低信噪比、带有回响的场景下, DOA 估计愈加困难。

传统的两级算法 [1] 面临的瓶颈主要在于计算复杂度和低信噪比条件下的时间差估计。通信领域中在信道估计等问题上也有类似的 DOA 估计算法如 MUSIC, 然而由于声音信号可能存在非平稳特性等原因也难以迁移到声源定位中。

B. 广义相关算法 (GCC)

广义相关算法 (Generalized Cross Correlation, GCC)[2] 是一种经典的计算两路信号到达延时的 (Time Difference of Arrival, TDOA) 算法。

假设信号 $s(t)$ 在两个接收机接收时存在时间为 D 的一个很小的延时, 且分别收到带限噪声 $n_1(t), n_2(t)$ 的干扰,

$$x(t) = s(t) + n_1(t), \quad y(t) = s(t - D) + n_2(t)$$

其中观测时间为 $-T/2 \leq t \leq T/2$, 则延时 $D = \hat{D}$ 等于最大化互相关函数,

$$\hat{D} = \arg \max_{\tau} \int_{-T/2}^{T/2} x(t)y(t + \tau)dt.$$

根据本题中的音频信息, 认为满足 GCC 使用条件, 但需要在计算前进行合适的滤波。

C. 相位变换加权的可控响应功率算法 (SRP-PHAT)

已有文章证明了基于可控响应功率 (Steered-Response Power, SRP) 的方法相比两级算法鲁棒性更

好, 事实上基于 SRP 的方法也是利用了成对麦克风阵列收到信号的 TDOA。我们采用相位变换加权的可控响应功率算法 (SRP-PHAT)[3]。

具体来说, 用 $m_i(t)$ 表示在 M 麦克风阵列系统中麦克风 i 收到的信号, 则长度为 T 的信号的 SRP 定义为

$$P_n(\mathbf{x}) = \int_{nT}^{(n+1)T} \left| \sum_{i=1}^M w_i m_i(t - \tau(\mathbf{x}, i)) \right|^2 dt, \quad (1)$$

其中 \mathbf{x} 为声源位置, $\tau(\mathbf{x}, i)$ 为声源 \mathbf{x} 到达麦克风 i 的时间, w_i 为权重。对于本问题中的单声源定位, 我们需要计算

$$\hat{\mathbf{x}}_s^n = \arg \max_{\mathbf{x}} P_n(\mathbf{x}). \quad (2)$$

将 (1) 转换到频域, 利用帕萨瓦尔定理得到

$$P_n(\mathbf{x}) = \sum_{k=1}^M \sum_{l=1}^M \int_{-\infty}^{\infty} \Psi_{kl}(\omega) M_k(\omega) M_l^*(\omega) e^{j\omega(\tau(\mathbf{x}, l) - \tau(\mathbf{x}, k))} d\omega \quad (3)$$

其中 $\Psi_{kl}(\omega) := W_k(\omega) W_l^*(\omega)$ 。

在相位变换 (phase transform, PHAT)[4] 方法中, 取 $\Psi_{kl}(\omega) = 1/M_k(\omega) M_l^*(\omega)$, 问题转换为求 $P'_n(\mathbf{x})$ 最大值, 这可以用随机区域收缩 (Stochastic Region Contraction, SRC)[5] 等办法解决。

II. 实现细节

A. 麦克风方阵列 GCC 的选取组合

相比于由两个麦克风组成的线阵, 此题中由 4 个麦克风构成的方阵将可以计算 6 组两路信号的 GCC 而产生 6 个 TDOA, 如何更好地利用这 6 个 TDOA 计算 DOA 是本节要讨论的问题。

第一种简单的考虑是仅利用两组对角线的 2 个 TDOA, 原因是对角线的两个麦克风距离更远, 产生的 TDOA 更准确。具体地说, 如图 1 所示, 对于 mic(2,4) 构成的阵列, DOA 向法线方向的估计最准确, 若实际

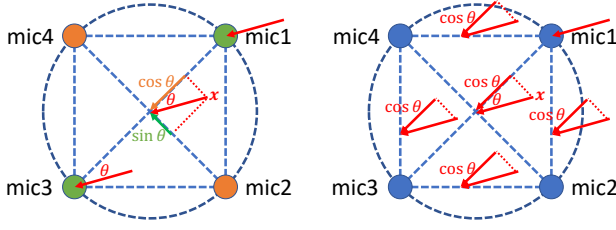


图 1. 两种利用方阵两两 TDOA 的方法。左: 利用两组对角线的 2 个 TDOA。右: 以不同权重利用全部 6 个 TDOA, 以 $\cos \theta$ 方向为例。

DOA 为 θ , 则在法线方向投影大小应为 $\cos \theta$ 。同理, 利用 mic(1,3) 构成的阵列测量另一方向投影大小 $\sin \theta$, 用两个投影来估计 θ 的大小。

第二种考虑是以不同权重利用全部 6 个 TDOA。如图 1 所示, 为了估计第一种方法中的 $\cos \theta$, 同时利用由 mic(1,4) 阵列、mic(4,3) 阵列、mic(3,2) 阵列、mic(2,1) 阵列得到的 TDOA。对角线阵列距离更远, 在计算 TDOA 平均值时取更大的权重, 实现中我们令对角线阵列 TDOA 与水平/竖直阵列 TDOA 权重比例为 2:1。与第一种方法不同的是, 两个水平/竖直阵列在理论上应该得到相同的 TDOA, 所以在计算前应该先确认二者的差是否小于某个可以容许的阈值, 否则取二者中较小的 TDOA 而舍弃较大值。

B. SRP-PHAT 中最大值的求解

对于 SRP-PHAT 中的优化问题(2), 简单的数值求解办法是遍历所有可能声源方向 \mathbf{x} 寻找最大的 $P_n(\mathbf{x})$ 或者随机生成一定数量的声源方向 \mathbf{x} 寻找最大的 $P_n(\mathbf{x})$ 。

然而上述方法中仅考虑了二维平面中一个自由度 θ , 实际实现中表现并不好, 可能的原因是声源到麦克风并不能近似看作一个远场模型。我们改进的办法是假设声源到麦克风是一个近场, 也就是说声源 \mathbf{x} 需要被建模为一个 2-自由度的点源。具体实现中我们随机生成边长为 10m 的方形区域中的 2000000 个点源构成集合 \mathcal{S} , 求解问题

$$\hat{\mathbf{x}}_s^n = \arg \max_{\mathbf{x} \in \mathcal{S}} P_n(\mathbf{x}),$$

以此作为 SRP-PHAT 问题的解。实验表明这种方法比远场模型得到的声源 DOA 更可靠。

III. 结果分析

四种方法在 **train** 数据上的误差如表 III 所示。

- 方法 1: SRP-PHAT
- 方法 2: 利用对角线两组 TDOA
- 方法 3: 利用全部 TDOA
- 方法 4: 综合方法 1, 2 和 3 的结果

方法	1	2	3	4
误差 (°)	2.5335	2.2264	1.7759	2.0945

由于本题中信号的信噪比非常低, 方法 1, 2 和 3 在某些测试中可能出现较大误差, 表现在 DOA 在两个对角线方向投影不满足 $\sin^2 \theta + \cos^2 \theta = 1$ 这一确定的关系。所以我们最终选择综合了方法 1, 2 和 3 的方法 4 作为最终结果, 即选择三种方法中更接近 $\sin^2 \theta + \cos^2 \theta = 1$ 的解。虽然在 **train** 数据上的误差高于方法 3, 但该方法有更好的鲁棒性, 因此会在测试集中取得更好的表现。

对于方法 1, SRP-PHAT 的实现如 II-B 所述需要随机生成大量点源来求解最大值, 其精度依赖于随机实验次数。

IV. 总结

本实验中, 利用 SRP-PHAT 和 GCC-PHAT 多种方法联合估算 DOA。数据集信噪比较低, 利用相位变换加权 (PHAT) 可以有效缓解这一现象, 最终根据多种不同的计算方法, 根据冗余的麦克风信息, 选取有效信息, 最终根据约束条件选择合理的估算, 有效增加了 DOA 估计的鲁棒性。

V. 文件清单

DOA.m	DOA 主脚本
read.m	数据读入函数
srpphat.m	SRP-PHAT
result.txt	输出结果

参考文献

- [1] S. et. al, "Performance of real-time source-location estimators for a large-aperture microphone array," *IEEE Transactions on Speech and Audio Processing*, vol. 13, no. 4, pp. 593–606, 2005.
- [2] K. et. al, "The generalized correlation method for estimation of time delay," *IEEE Transactions on ASSP*, vol. 24, no. 4, pp. 320–327, 1976.
- [3] D. et. al, "A real-time srp-phat source location implementation using stochastic region contraction (src) on a large-aperture microphone array," in *IEEE ICASSP*, vol. 1. IEEE, 2007, pp. I–121.
- [4] S. et. al, "Acoustic source location in a three-dimensional space using crosspower spectrum phase," in *IEEE ICASSP*, vol. 1. IEEE, 1997, pp. 231–234.
- [5] B. et. al, "Microphone array optimization by stochastic region contraction," *IEEE TSP*, vol. 39, no. 11, pp. 2377–2386, 1991.