arXiv:2301.07649v1 [eess.AS] 18 Jan 2023

采用具有麦克风相关预测延迟的加权预测误差进行声学传感器网络去混响

Anselm Lohmann[1], *Toon van Waterschoot*[2], *Joerg Bitzer*[3], *Simon Doclo*[1]

[1]奥尔登堡大学医学物理与声学系，德国奥尔登堡[2]鲁汶大学电气工程系（ESAT-STADIUS），比利时鲁汶[3]Fraunhofer IDMT听觉、语音与音频技术项目组，德国奥尔登堡 anselm.lohmann@uni-oldenburg.de

摘要

过去几十年来，人们提出了多种多麦克风语音去混响算法，其中包括加权预测误差（WPE）算法。在WPE算法中，需要设置预测延迟，以降低预测信号与参考麦克风信号中直达分量之间的相关性。对于麦克风间距较近的紧凑阵列，通常采用麦克风无关的预测延迟。而在采用空间分布式麦克风的声学传感器网络中，语音源到达参考麦克风与其他麦克风的时间差（TDOA）通常较大。因此，采用麦克风无关的预测延迟时，参考信号与预测信号之间仍可能存在显著相关性，从而导致去混响输出信号失真。为使信号去相关，本文提出相对于参考麦克风进行到达时间差补偿，从而在WPE算法中采用麦克风相关预测延迟。我们研究了两种情况：在短时傅里叶变换域使用跨频带滤波进行最优到达时间差补偿，以及采用频带间延迟和整数延迟近似。针对不同混响时间，分别使用真实到达时间差和估计到达时间差进行的仿真结果清楚地表明，采用麦克风相关预测延迟具有优势。

索引词— 去混响、加权预测误差、声学传感器网络、预测延迟

1. 引言

在房间内使用麦克风录制语音源时，声学反射产生的混响可能会降低录制语音的质量和可懂度。早期反射可能有益，但晚期混响通常会降低语音可懂度，并影响自动语音识别的性能 [1,2]。因此，许多语音通信应用都需要有效的去混响，例如语音控制系统、助听器和免提电话 [3–14]。一种常用的盲去混响算法是加权预测误差（WPE）算法 [10–14], ，该算法基于多通道线性预测（MCLP）。WPE首先根据延迟的混响麦克风信号，估计参考麦克风中的晚期混响分量，然后从参考麦克风信号中减去该估计值，以实现去混响。该算法已有若干变体，例如，旨在控制时频域中去混响输出信号的稀疏性 [1,2]。

本研究获得了欧盟“地平线2020”研究与创新计划资助，资助依据为玛丽·斯克沃多夫斯卡-居里资助计划协议第 956369 号和欧洲研究委员会整合者资助项目 SONORA（编号 773268）；此外，还获得了德国研究联合会（DFG）在德国卓越战略框架下提供的资助——EXC 2177/1，项目编号 390895286。

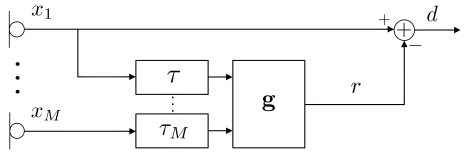


图 1：采用麦克风相关预测延迟的加权预测误差

WPE算法已有若干变体，例如，旨在控制时频域中去混响输出信号的稀疏性 [11,13]。

该延迟称为预测延迟，在加权预测误差中发挥着重要作用。引入它是为了降低预测信号与参考麦克风信号中直达分量之间的相关性，从而力求保留直达分量和早期反射 [10,11]。预测延迟决定了残余混响量与目标分量失真之间的权衡。通常根据语音的自相关来选择，因此往往采用麦克风无关的设置。对于麦克风间距较近的紧凑阵列，这或许是不错的选择；然而，在采用空间分布式麦克风的声学传感器网络中，麦克风间可能相距较远，导致语音源在不同麦克风间的到达时间差（TDOA）较大且各不相同。如果不加以补偿，每个麦克风的有效预测延迟可能并非最优，从而导致目标分量失真或混响过多。

为在短时傅里叶变换（STFT）域中相对于参考麦克风进行到达时间差补偿，本文提出在WPE算法中使用麦克风相关预测延迟。我们采用非整数延迟进行最优到达时间差补偿，或采用整数延迟进行粗略TDOA补偿。在短时傅里叶变换域中，最优到达时间差补偿需要通过跨频带滤波来实现。然而，由于跨频带滤波器会增加计算复杂度，而且使用更多跨频带滤波器并不一定能降低子带中的重建误差 [15], ，我们还提出使用频带间近似。仿真结果表明，对于麦克风空间分布的声学传感器网络，在不同混响时间以及使用估计和理想到达时间差的情况下，所提出的麦克风相关预测延迟优于WPE算法中的麦克风无关预测延迟；采用非整数延迟和跨频带滤波时去混响性能最佳，而计算复杂度较低的频带间近似紧随其后。

2. 信号模型

我们考虑一个语音源在混响房间中由一组 M 麦克风采集的情形。与 [10, 11, 13], 类似，本文不考虑

加性噪声。在短时傅里叶变换域中，令 $s(k,n)$表示干净语音信号，其中 $k \in \{1,...,K\}$ 为子带索引，$n \in \{1,...,N\}$ 为时间帧索引。第 m个麦克风处的混响信号 $x_m(k,n)$可表示为

$$x_m(k,n) = \sum_{l=0}^{L_h-1} h_m(k,l)s(k,n-l) + e_m(k,n), \tag{1}$$

其中，$h_m(k,\ n)$表示语音源与第 m个麦克风之间长度为 L_h 的子带卷积传递函数，$e_m(k,\ n)$表示子带建模误差。不失一般性，我们将第一个麦克风定义为参考麦克风。假设式（1）中的误差项 $e_m(k,\ n)$可以忽略不计，则图1所示的去混响问题可表述为

$$d(k,n) = x_1(k,n) - r(k,n). \tag{2}$$

目标分量 $d(k,\ n) = \sum_{l=0}^{L_d-1} h_1(k,\ l)s(k,\ n-l)$由参考麦克风信号 $x_1(k,\ n)$中的直达路径和早期反射组成，其中 L_d 表示早期反射与晚期反射之间的时间截断点。我们旨在估计的非目标分量 $r(k,\ n) = \sum_{l=L_d}^{L_h-1} h_1(k,\ l)s(k,\ n-l)$是参考麦克风信号 $x_1(k,\ n)$中的晚期混响分量。根据多通道线性预测模型 [10]，非目标晚期混响分量 $r(k,\ n)$可表示为所有混响麦克风信号经过延迟滤波后的总和，即

$$r(k,n) = \sum_{m=1}^{M} \sum_{l=0}^{\tilde{L}_g-1} \tilde{g}_m(k,l)x_m(k,n-\tau-l), \tag{3}$$

其中，$\tilde{g}_m(k,\ n)$ 表示长度为 L_g 的第 m 个预测滤波器，$\tau = L_d$ 表示（麦克风无关的）预测延迟。

预测延迟旨在降低预测信号 $x_m(k,\ n-\tau)$ 与目标分量 $d(k,\ n)$ 之间的相关性。通常根据语音的自相关来选择，因此往往与麦克风无关。本文提出将式（3）中的多通道线性预测模型推广为允许采用麦克风相关预测延迟 τ_m （见图 1），即

$$\boxed{r(k,n) = \sum_{m=1}^{M} \sum_{l=0}^{L_g-1} g_m(k,l)x_m(k,n-\tau_m-l),} \tag{4}$$

其中，$g_m(k,\ l)$ 表示长度为 L_g 的第 m 个预测滤波器。利用式（4），式（2）中的信号模型可用向量表示法重写为

$$\mathbf{d}(k) = \mathbf{x}_1(k) - \mathbf{X}_{\boldsymbol{\tau}}(k)\mathbf{g}(k), \tag{5}$$

with

$$\mathbf{d}(k) = \begin{bmatrix} d(k,1) & \cdots & d(k,N) \end{bmatrix}^T \in \mathbb{C}^N, \tag{6}$$

$$\mathbf{x}_1(k) = \begin{bmatrix} x_1(k,1) & \cdots & x_1(k,N) \end{bmatrix}^T \in \mathbb{C}^N, \tag{7}$$

其中，N 表示时间帧数。式（5）中的多通道延迟卷积矩阵 $\mathbf{X}_{\boldsymbol{\tau}}(k)$ 定义为

$$\mathbf{X}_{\boldsymbol{\tau}}(k) = \begin{bmatrix} \mathbf{X}_{\tau_1}(k) & \cdots & \mathbf{X}_{\tau_M}(k) \end{bmatrix} \in \mathbb{C}^{N \times ML_g}, \tag{8}$$

其中$\mathbf{X}_{\tau_m}(k) \in \mathbb{C}^{N \times L_g}$ 是$\mathbf{x}_m(k)$延迟 τ_m 帧后的卷积矩阵，$\tau_1 = \tau$表示参考麦克风中的预测延迟，而$\mathbf{g}(k) \in \mathbb{C}^{ML_g}$ 是所有预测滤波器系数(k,n)的堆叠向量。语音去混响问题，即估计目标分量$\mathbf{d}(k)$，现在转化为估计用于预测非目标晚期混响的滤波器$\mathbf{g}(k)$。

3. WPE算法

为估计预测滤波器$\mathbf{g}(k)$，文献 [10] 提出使用时变高斯（TVG）模型来表征目标分量 $d(k,n)$。这相当于在每个时频单元中，使用均值为零、方差随时间变化的圆对称高斯分布来表征目标分量 $\lambda(k,n)$，即

$$\mathcal{N}_{\mathbb{C}}(d(k,n);0,\lambda(k,n)) = \frac{1}{\pi\lambda(k,n)} e^{-\frac{|d(k,n)|^2}{\lambda(k,n)}}, \tag{9}$$

其中，方差 $\lambda(k,\ n)$ 是一个未知参数，需要进行估计。由于 TVG 模型不假设频率或时间帧之间存在任何依赖关系，因此似然函数为

$$\mathcal{L}(\mathbf{g}(k),\boldsymbol{\lambda}(k)) = \prod_{n=1}^{N} \mathcal{N}_{\mathbb{C}}(d(k,n);0,\lambda(k,n)), \tag{10}$$

with

$$\boldsymbol{\lambda}(k) = \begin{bmatrix} \lambda(k,1) & \cdots & \lambda(k,N) \end{bmatrix}^{\mathrm{T}} \in \mathbb{R}^N. \tag{11}$$

随后可通过最大化对数似然函数来估计预测滤波器$\mathbf{g}(k)$和方差 $\boldsymbol{\lambda}(k)$，这等价于求解以下优化问题 [10,11]

$$\min_{\boldsymbol{\lambda}(k),\mathbf{g}(k)} \sum_{n=1}^{N} \left(\frac{|d(k,n)|^2}{\lambda(k,n)} + \log\pi\lambda(k,n) \right). \tag{12}$$

由于式 (12) 中的联合优化问题不存在闭式解，文献 [10] 提出采用交替优化过程，以迭代方式求解两个较简单的子问题。具体而言，在每次迭代中，首先假设方差 $\boldsymbol{\lambda}(k)$固定为上一次迭代的取值，并针对预测向量$\mathbf{g}(k)$最小化式 (12) 中的代价函数。使用该预测向量$\mathbf{g}(k)$的取值后，再针对方差 $\boldsymbol{\lambda}(k)$最小化式 (12) 中的代价函数。在给定的迭代 i中，省略索引 k 后，估计的预测滤波器和估计的方差为

$$\hat{\mathbf{g}}^{(i+1)} = \left(\mathbf{X}_{\boldsymbol{\tau}}^H \mathcal{D}_{\hat{\boldsymbol{\lambda}}^{(i)}}^{-1} \mathbf{X}_{\boldsymbol{\tau}} \right)^{-1} \mathbf{X}_{\boldsymbol{\tau}}^H \mathcal{D}_{\hat{\boldsymbol{\lambda}}^{(i)}}^{-1} \mathbf{x}_1, \tag{13}$$

$$\hat{\boldsymbol{\lambda}}^{(i+1)} = |\hat{\mathbf{d}}^{(i+1)}|^2 = |\mathbf{x}_1 - \mathbf{X}_{\boldsymbol{\tau}}\mathbf{g}^{(i+1)}|^2, \tag{14}$$

其中，$\mathcal{D}_{\hat{\boldsymbol{\lambda}}^{(i)}} = \mathrm{diag}(\hat{\boldsymbol{\lambda}}^{(i)})$，且$\hat{\mathbf{d}}^{(i+1)}$ 是使用式 (5) 计算得到的期望分量估计值，其中 $|.|$ 和.^{2} 运算符逐元素应用。

在 [11], 中，通过引入一个带有稀疏促进参数 p 的超先验，对 TVG 模型进行了扩展，以便更好地描述短时傅里叶变换域中语音的稀疏特性。[11] 中表明，式 (14) 中方差的解只需改为

$$\hat{\boldsymbol{\lambda}}^{(i+1)} = |\hat{\mathbf{d}}^{(i+1)}|^{2-p}. \tag{15}$$

4. 麦克风相关的预测延迟

在采用空间分布式麦克风的声学传感器网络中，语音源在各麦克风之间可能存在较大且各异的到达时间差。如果不进行补偿，各麦克风的有效预测延迟可能并非最优，导致 WPE 输出信号出现失真或混响过多（参见第 5 节的评估）。本节提出使用式（4）中的麦克风相关预测延迟进行到达时间差补偿。我们采用最优到达时间差

补偿，即使用非整数的麦克风相关预测延迟；或采用粗略到达时间差补偿，即使用整数的麦克风相关预测延迟。在第 4.1 节中，我们首先在时域中定义这一问题。可以证明，时域中的最优到达时间差补偿滤波器可表示为短时傅里叶变换域中的整数帧延迟与跨频带滤波器的组合（参见第 4.2 节）。为降低计算复杂度，我们还提出采用频带间近似（第 4.3 节）或使用整数帧延迟（第 4.4 节）。

4.1. 时域到达时间差补偿

设 TDOA_m 为麦克风 m 与参考麦克风之间的到达时间差。该到达时间差可分解为

$$\text{TDOA}_m = \delta_m^{\text{frame}} \times L_{\text{shift}} + \delta_m^{\text{samp}} + \delta_m^{\text{frac}}, \tag{16}$$

其中，整数帧延迟 δ_m^{frame}、整数延迟 δ_m^{samp}和分数延迟 δ_m^{frac}定义为

$$\delta_m^{\text{frame}} = \lfloor \frac{\text{TDOA}_m}{L_{\text{shift}}} \rceil, \tag{17}$$

$$\delta_m^{\text{samp}} = \lfloor \text{TDOA}_m - \delta_m^{\text{frame}} \times L_{\text{shift}} \rfloor, \tag{18}$$

$$\delta_m^{\text{frac}} = \text{TDOA}_m - \delta_m^{\text{frame}} \times L_{\text{shift}} - \delta_m^{\text{samp}}, \tag{19}$$

其中，L_{shift} 表示 STFT帧移，$\lfloor . \rceil$ 表示最近整数舍入算子，$\lfloor . \rfloor$ 表示向下取整算子。最优时域TDOA补偿滤波器 $u'_m[t]$可写为

$$\underline{u}'_m[t] = \delta[t - \delta_m^{\text{frame}} \times L_{\text{shift}}] * \delta[t - \delta_m^{\text{samp}}] * \text{sinc}[t - \delta_m^{\text{frac}}], \tag{20}$$

其中，$\delta[t]$ 表示离散时间索引为 t 的单位脉冲，$*$ 表示卷积。需要注意的是，$\delta[t - \delta_m^{\text{frame}} \times L_{\text{shift}}]$ 可在短时傅里叶变换域中通过整数帧延迟实现，而$u_m[t] = \delta[t - \delta_m^{\text{samp}}] * \text{sinc}[t - \delta_m^{\text{frac}}]$ 则需要在短时傅里叶变换域中通过跨频带滤波器实现。

4.2. 使用跨频带滤波实现非整数预测延迟

文献 [15] 表明，任何时域滤波器（例如 $u_m[t]$）都可以通过跨频带滤波器在短时傅里叶变换域中实现

$$u_m(k, k', n) = \{\underline{u}_m[t] * \phi_{k,k'}[t]\}|_{t=nL_{\text{shift}}}, \tag{21}$$

其中，$u_m(k,\ k',\ n)$表示跨频带索引为 k'、位于子带 k中的跨频带滤波器，

$$\phi_{k,k'}[t] = e^{j2\pi k' t} \sum_{m=-K}^{K} \tilde{\psi}[m]\psi[t+m]e^{-j\frac{2\pi}{K}m(k-k')}, \tag{22}$$

其中，$\tilde{\psi}[t]$ 和 $\psi[t]$ 分别表示长度为 K的STFT分析与合成窗。利用跨频带滤波器 $u_m(k,\ k',\ n)$和整数帧延迟 δ_m^{frame}，可以在短时傅里叶变换域中执行最优到达时间差补偿，其效果等同于式（20）中的最优时域TDOA补偿滤波器 $u'_m[t]$ 。

现在，我们对经过预测延迟处理的麦克风信号 $x_m(k,\ n-\tau)$ 应用最优到达时间差补偿，其中 τ是参考麦克风的整数预测延迟。这使得 WPE算法中的麦克风相关预测延迟 τ_m 变为非整数（见图 1），即

$$\boxed{x_m(k, n-\tau_m) = \sum_{k'=1}^{K} \sum_{l=-L_a}^{L_c} x_m(k', n-l-\tau_m^{\text{int}})u_m(k, k', l),} \tag{23}$$

其中，L_a 和 L_c 分别表示$u_m(k,\ k',\ n)$中的非因果抽头数和因果抽头数，而 $\tau_m^{\text{int}} = \tau + \delta_m^{\text{frame}}$表示预测延迟 τ_m的整数分量。

4.3. 频带间近似

由于式（23）中的跨频带滤波器 $u_m(k,\ k',\ n)$计算复杂，而且文献 [15] 已表明，跨频带滤波器不一定能降低 $x_m(k,\ n-\tau_m)$中的重建误差，因此我们提出对式（23）中的最优到达时间差补偿采用频带间近似，即

$$\boxed{x_m(k, n-\tau_m) = \sum_{l=-L_a}^{L_c} x_m(k, n-l-\tau_m^{\text{int}})u_m(k, l),} \tag{24}$$

其中，$u_m(k,n) = u_m(k,k,n)$ 表示式（21）中的频带到频带滤波器。

4.4. 整数预测延迟

为了进一步降低计算复杂度，我们仅考虑式（24）中的整数帧延迟分量，并忽略频带到频带滤波器 $u_m(k,n)$。这相当于对预测延迟麦克风信号 $x_m(k,n-\tau)$ 应用粗略TDOA补偿，从而得到整数麦克风相关预测延迟 $\tau_m = \tau_m^{\text{int}}$，即

$$\boxed{x_m(k, n-\tau_m) = x_m(k, n-\tau_m^{\text{int}}).} \tag{25}$$

5. 实验评估

本节评估麦克风无关预测延迟和所提出的麦克风相关预测延迟在声学传感器网络WPE算法中的性能。该网络采用空间分布式麦克风，并分别使用真实到达时间差和估计到达时间差。第5.1节介绍所考虑的声学设置和性能指标。第5.2节展示第4节提出的不同到达时间差补偿实现方案的仿真结果。

5.1. 声学设置和性能指标

我们考虑一个包含 $M = 9$ 个空间分布式麦克风和一个语音源的声学传感器网络，房间尺寸为 8m×8m×5m。图 2 描绘了麦克风的位置以及房间内所考虑的语音源位置。麦克风按照 3×3 网格布置，间距相等，尺寸为 7.5m×7.5m，高度为 1.5m。参考麦克风选定为左下方的麦克风，并在所有所考虑的语音源位置中保持不变。总共在一个 7×7 网格上考虑了 48 个语音源位置，间距相等，尺寸为 7m×7m，高度为 1.5m；其中移除了一个与选定麦克风位置重叠的语音源位置。所考虑的语音源位置与最近麦克风之间的最小距离为 0.5m。

通过将TIMIT数据库中的一部分无混响语音 [16] 与使用随机镜像法 [17] 模拟的房间脉冲响应（RIR）进行卷积，生成了混响麦克风信号，其中混响时间为 $T_{60} \in \{500, 750, 1000\}$ 毫秒。信号以16 kHz的采样率进行处理，并采用短时傅里叶变换框架，帧长为 $K = 1024$个采样点、帧移为 $L_{\text{shift}} = 256$ 个采样点，同时在（22）中使用平方根Hann分析窗和合成窗。WPE算法在（15）中实现，预测滤波器长度为$L_g \in \{8, 12, 16\}$（与 T_{60}成正比），预测延迟为 $\tau = 2$，稀疏促进参数为 $p = 0.5$ 。

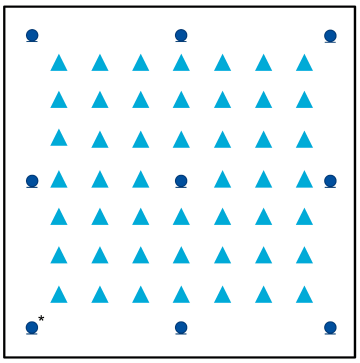


图 2： $M = 9$ 空间分布式麦克风的位置，固定参考麦克风（*）以及语音源的 48 个考察位置

为了估计语音源在各麦克风之间的到达时间差，我们采用了广泛使用的广义互相关相位变换（GCC-PHAT）算法 [18]，其帧长为2048个采样点（即WPE的两倍），帧移为1024个采样点，因为帧长至少需要是最大可能到达时间差的两倍。

使用频率加权分段信噪比（FWSSNR）、语音质量感知评估（PESQ）和倒谱距离（CD）度量 [19]评估去混响性能。这些度量所用的参考信号是参考麦克风中的直达分量。上述度量值取语音源 48 个考察位置的平均值。

5.2. 仿真结果

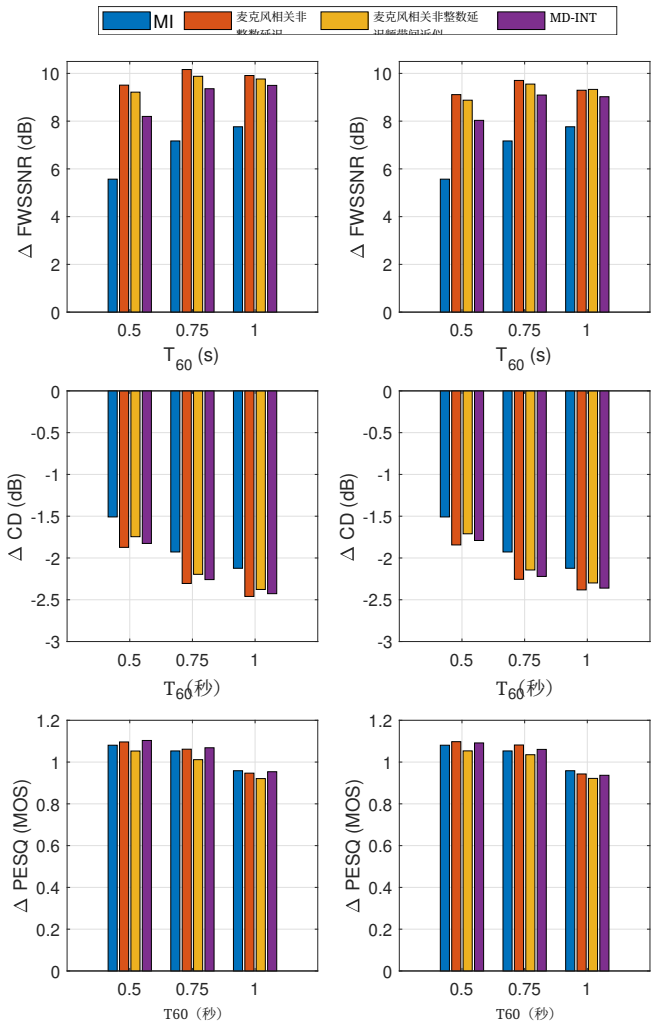
对于 WPE算法，我们考虑以下预测延迟：

- MI：使用麦克风无关预测延迟 τ
- 麦克风相关非整数延迟：使用麦克风相关预测延迟 τ_m，并使用 (23) 中的跨频带滤波进行最优 TDOA 补偿
- MD-NINT-B2B：在式 (24) 中使用麦克风相关预测延迟 τ_m并采用频带间近似
- MD-INT：在式 (25) 中使用麦克风相关整数预测延迟 τ_m

对于所有考察的 WPE 算法，图 3 展示了不同混响时间下的平均性能，包括采用理想到达时间差和估计到达时间差的情况。首先，对于理想到达时间差，可以观察到，在所有混响时间下，就 Δ FWSSNR 和 ΔCD 而言，使用麦克风相关预测延迟的性能显著优于使用麦克风无关预测延迟的性能。就 ΔPESQ 而言，两者的性能相近，不过使用麦克风相关预测延迟和使用麦克风无关预测延迟时，语音质量存在明显的可听差异[1]。比较麦克风相关预测延迟的不同实现方式可以发现，采用跨频带滤波的非整数延迟（MD-NINT）明显优于整数延迟（MD-INT），而采用非整数延迟的频带间近似（MD-NINT-B2B）仅略逊于采用跨频带滤波的方式。其次，使用估计到达时间差进行到达时间差补偿时，性能非常相似

使用理想到达时间差时，所有所考虑的WPE算法均可实现这一点。这表明即使在高度混响环境中，该方法对估计误差也具有很高的鲁棒性。

[1]音频示例见 uol.de/f/6/dept/mediphysik/ag/sigproc/audio/dereverb/mdpd.html



（a）理想到达时间差　　（b）估计到达时间差

图3：对于不同的混响时间 T_{60}，所有所考虑的WPE算法在（a）理想到达时间差和（b）估计到达时间差条件下的频率加权分段信噪比改善、CD改善和PESQ改善方面的平均性能

6. 结论

本文提出了一种采用麦克风相关预测延迟的 WPE 算法，并提出利用这些麦克风相关预测延迟进行到达时间差补偿，这对于麦克风空间分布式的声学传感器网络尤为重要。我们考虑了几种在短时傅里叶变换域中进行到达时间差补偿的方法，即使用跨频带滤波器和非整数延迟实现最优到达时间差补偿、使用频带间滤波器实现近似到达时间差补偿，以及使用整数帧延迟实现粗略到达时间差补偿。实验评估表明，与采用麦克风无关预测延迟相比，采用麦克风相关预测延迟进行到达时间差补偿能够提升性能；其中，使用跨频带滤波器和非整数延迟可获得最佳去混响性能，其次是计算复杂度较低的频带间近似。未来研究可进一步考察所提方案在存在加性噪声、移动声源或多个声源的声学场景中的性能。

7. 参考文献

[1] R. Beutelmann 和 T. Brand，“空间噪声和混响环境下正常听力及听力受损听者的语音可懂度预测”，美国声学学会期刊，第 120 卷，第 1 期，第 331–342 页，2006 年。[2] T. Yoshioka、A. Sehr、M. Delcroix、K. Kinoshita、R. Maas、T. Nakatani 和 W. Kellermann，“让机器理解我们：混响房间中自动语音识别的抗混响能力”，*IEEE* 信号处理杂志，第 29 卷，第 6 期，第 114–126 页，2012 年。[3] E. A. P. Habets 和 P. A. Naylor，“去混响”，载于音频源分离与语音增强，E. Vincent、T. Virtanen 和 S. Gannot 编。Wiley，2018 年。[4] B. Cauchi、I. Kodrasi、R. Rehr、S. Gerlach、A. Juki´c、T. Gerkmann、S. Doclo 和 S. Goetze，“结合 MVDR 波束成形与单通道频谱处理以增强含噪混响语音”，*EURASIP* 信号处理进展期刊，2015 年。[5] S. Braun、A. Kuklasi´nski、O. Schwartz、O. Thiergart、E. A. P. Habets、S. Gannot、S. Doclo 和 J. Jensen，“晚期混响功率谱密度估计器的评估与比较”，*IEEE/ACM* 音频、语音与语言处理汇刊，第 26 卷，第 6 期，第 1056–1071 页，2018 年。[6] T. Dietzen、A. Spriet、W. Tirry、S. Doclo、M. Moonen 和 T. van Waterschoot，“广义旁瓣抵消与多通道线性预测在语音去混响和噪声抑制中的比较分析”，*IEEE/ACM* 音频、语音与语言处理汇刊，第 27 卷，第 3 期，第 544–558 页，2019 年。[7] X. Li、L. Girin、S. Gannot 和 R. Horaud，“基于频谱幅度逆滤波的多通道在线去混响”，*IEEE/ACM* 音频、语音与语言处理汇刊，第 27 卷，第 9 期，第 1365–1377 页，2019 年。[8] D. S. Williamson 和 D. Wang，“用于语音去混响和降噪的复数域时频掩蔽”，*IEEE/ACM* 音频、语音与语言处理汇刊，第 25 卷，第 7 期，第 1492–1501 页，2017 年。[9]J. Lemercier、J. Thiemann、R. Koning 和 T. Gerkmann，“面向助听设备的在线神经网络支持去混响的可定制端到端优化”，载于*IEEE* 国际声学、语音与信号处理会议（*ICASSP*）论文集，新加坡，2022 年，第 171–175 页。

[10] T. Nakatani、T. Yoshioka、K. Kinoshita、M. Miyoshi 和 B. H. Juang，“基于方差归一化延迟线性预测的语音去混响”，*IEEE* 音频、语音与语言处理汇刊，第 18 卷第 7 期，第 1717–1731 页，2010 年。

[11] A. Juki´c、T. van Waterschoot、T. Gerkmann 和 S. Doclo，“基于多通道线性预测和稀疏先验的语音去混响”，*IEEE/ACM* 音频、语音与语言处理汇刊，第 23 卷第 9 期，第 1509–1520 页，2015 年。[12] J. Wung、A. Juki´c、S. Malik、M. Souden、R. Pichevar、J. Atkins、D. Naik 和 A. Acero，“使用加权 Householder 最小二乘格型自适应滤波器进行在线语音去混响的鲁棒多通道线性预测”，*IEEE* 信号处理汇刊，第 68 卷，第 3559–3574 页，2020 年。[13] M. Witkowski 和 K. Kowalczyk，“采用 Split Bregman 方法并强制语音稀疏性的基于线性预测的去混响”，*IEEE* 信号处理快报，第 28 卷，第 942–946 页，2021 年。[14] G. Huang、J. Benesty、I. Cohen 和 J. Chen，“用于自适应加权预测误差语音去混响的克罗内克积多通道线性滤波”，*IEEE/ACM* 音频、语音与语言处理汇刊，第 30 卷，第 1277–1289 页，2022 年。[15] Y. Avargel 和 I. Cohen，“采用跨频带滤波的短时傅里叶变换域系统辨识”，*IEEE* 音频、语音与语言处理汇刊，第 15 卷第 4 期，第 1305–1319 页，2007 年。[16] J. S. Garofolo、L. F. Lamel、W. M. Fisher、J. G. Fiscus、D. S. Pallett 和 N. L. Dahlgren，“TIMIT 声学语音连续语料库”，语言数据联盟，1993 年。[17] E. De Sena、N. Antonello、M. Moonen 和 T. van Waterschoot，“关于房间声学仿真中矩形几何体的建模”，*IEEE/ACM* 音频、语音与语言处理汇刊，第 23 卷第 4 期，第 774–786 页，2015 年。[18] C. Knapp 和 G. Carter，“用于估计时间延迟的广义相关法”，*IEEE* 声学、语音与信号处理汇刊，第 24 卷第 4 期，第 320–327 页，1976 年。[19] K. Kinoshita、M. Delcroix、S. Gannot、E. A. P. Habets、R. Haeb- Umbach、W. Kellermann、V. Leutnant、R. Maas、T. Nakatani、B. Raj、A. Sehr 和 T. Yoshioka，“REVERB 挑战赛综述：混响语音处理研究的最新进展与剩余挑战”，*EURASIP* 信号处理进展期刊，2016 年。