

一种基于正交非负矩阵分解的多通道线性预测 语音去混响方法

何 冲, 王冬霞, 王旭东, 蒋茂松

(辽宁工业大学电子与信息工程学院, 辽宁锦州 121001)

摘要: 在相对封闭的声学环境中, 由于受到混响的影响, 麦克风阵列采集到的信号清晰度降低、甚至混淆不清。为了解决这一问题, 文章在多通道线性预测(Multi-Channel Linear Prediction, MCLP)语音去混响的基础上, 提出了一种改进的多通道线性预测(Multi-Channel Linear Prediction, MCLP)方法即正交非负矩阵线性预测(Orthogonal Non-negative Matrix Factorization Multi-Channel Linear Prediction, ONMFMCLP)方法。该方法利用纯净语音的短时谱域的稀疏性, 构建了基于正交的非负矩阵分解(Non-negative Matrix Factorization, NMF)的 Kullback-Leibler(KL)问题, 通过对矩阵求迹、利用梯度下降法给出迭代规则, 进而改进了 MCLP 中目标信号矩阵的协方差估计。实验结果表明, 相对于其他方法, ONMFMCLP 方法具有更好的去混响效果。

关键词: 麦克风阵列; 去混响; 多通道线性预测; 非负矩阵分解

中图分类号: TN912.35

文献标识码: A

文章编号: 1000-3630(2018)-05-0468-07

DOI 编码: 10.16300/j.cnki.1000-3630.2018.05.011

Speech dereverbration based on MCLP using orthogonal NMF

HE Chong, WANG Dong-xia, WANG Xu-dong, JIANG Mao-song

(School of Electronic and Information Engineering, Liaoning University of Technology, Jinzhou 121001, Liaoning, China)

Abstract: In a relatively closed acoustic environment, the speech signals can be severely affected by reverberation, which degrades the intelligibility of speech and even results in confusion. In order to solve this problem, this paper presents a new dereverbration algorithm called ONMFMCLP based on the well-known multi-channel linear prediction (MCLP). This algorithm utilizes the sparse nature of clean speech in the short time spectrum domain to construct the KL optimization problem based on the orthogonal NMF. The iterative rules are given through solving the matrix trace and utilizing the gradient descent method, thereby improving the signal covariance matrix in the MCLP algorithm. Experimental results show that the ONMFMCLP algorithm can achieve a better dereverbration performance compared with other algorithms.

Key words: microphone array; dereverbration; Multi-Channel Linear Prediction (MCLP); Non-negative Matrix Factorization (NMF)

0 引言

在一些相对封闭的声学环境中使用免提电话、电话会议、助听器、语音控制系统等通信系统时, 麦克风采集到的信号往往会受到混响的影响导致语音的清晰度、可懂度等降低, 进而影响通信系统的整体功能。因此研究如何去除语音混响即语音去混响也显得愈发重要。

经过多年的发展, 人们提出了诸多去混响方法。主要包括基于多通道均衡的去混响方法^[1]、基

于谱估计和空间滤波器相结合的去混响方法^[2]以及基于概率模型的去混响方法^[3]。I Kodrasi 等^[1]提出的多通道均衡去混响方法在理论上可以实现较好的去混响效果, 但该方法在实际中会受到房间冲激响应估计准确性的影响; Adam 等^[2]根据多通道维纳滤波(Multi-channel Wiener Filter, MWF)理论将 MWF 分解为一个最小方差无失真响应(Minimum Variance Distortionless Response, MVDR)波束形成器和一个单通道维纳滤波器进行去混响处理。该类方法同样也可以实现较好的去混响效果, 但需要估计目标信号方向角等参数; 而 A Jukic 等^[3]提出的多通道线性预测(Multi-Channel Linear Prediction, MCLP)去混响方法既可以实现较好的去混响效果, 又不需要估计任何参数。但该方法会出现目标信号协方差等于零的情况, 当情况严重时会出现输出信号失真。

收稿日期: 2017-06-29; 修回日期: 2017-08-13

基金项目: 辽宁省自然科学基金(201302022)资助项目。

作者简介: 何冲(1990—), 男, 河南南阳人, 硕士研究生, 研究方向为现代信号处理与多媒体技术。

通讯作者: 王冬霞, E-mail: dxwang_lg@126.com

非负矩阵分解(Non-negative Matrix Factorization, NMF)自提出以来便得到学者们的深入研究和广泛的应用^[4-5]。根据 NMF 的分解特性可知, NMF 的分解矩阵具有非负的性质, 而这种非负性使得这种描述数据的方法具有一定的稀疏性。研究表明, 纯净语音的短时傅里叶系数具有一定的稀疏性且大于混响语音信号的短时系数稀疏特性^[6-7]。鉴于此, 考虑将 NMF 应用到 MCLP 去混响中, 这样即可以避免目标信号协方差出现等于零的情况, 又可以进一步提高语音的去混响性能。

本文介绍了混响的声学模型及 MCLP 方法, 提出了一种正交非负矩阵线性预测(Orthogonal Non-negative Matrix Factorization Multi-Channel Linear Prediction, ONMF MCLP) 方法, 最后进行了实验和结果分析。

1 混响信号模型及 MCLP 方法

在一个单声源的封闭声学环境中, 有一个由 M 个全向麦克风组成的均匀线性阵列, 则第 m 个麦克风接收到的语音信号的表达式为^[3]

$$x_m(b) = \sum_{l=0}^{L_r-1} r_m(l) s(b-l) + v_m(b) \quad (1)$$

式中: b 表示离散时间; $s(b)$ 表示时域内的纯净语音; $v_m(b)$ 表示背景噪声; $r_m(l)$ 表示声源与第 m 个麦克风之间的房间冲激响应(Room Impulse Response, RIR), 长度为 L_r 。在时域内 RIR 比较长, 计算复杂, 所以语音去混响通常在频域内执行, 故式(1)的时频域表达式为

$$x_m(k, n) = \sum_{l=0}^{L_r-1} R_m(k, l) s(k, n-l) + e_m(k, n) \quad (2)$$

式(2)中: $n \in \{1, \dots, N\}$ 表示帧指数, $k \in \{1, \dots, K\}$ 表示频带指数; $e_m(k, n)$ 表示模型误差; $R_m(k, l)$ 表示时频域内的 RIR, L_n 为帧数。仅考虑混响对语音信号的影响即假设 $e_m(k, n) = 0$ 。令 $d_m(k, n)$ 表示第 m 个麦克风直达波和早期反射波之和, 则式(2)可以表示为

$$x_m(k, n) = d_m(k, n) + \sum_{l=\tau}^{L_r-1} R_m(k, l) s(k, n-l) \quad (3)$$

$$d_m(k, n) = \sum_{l=0}^{\tau-1} R_m(k, l) s(k, n-l) \quad (4)$$

式(3)和式(4)中的 τ 表示线性延迟。早期反射能够提高语音的清晰度^[8], 而后期混响产生的掩蔽效应会严重降低语音的音质, 故在去混响方法中将 $d_m(k, n)$ 作为期望信号并进行重构。选择一路信号作为参考信号, 如 $m=1$, 则式(3)的 MCLP 表达式为^[3]

$$x_1(k, n) = d_1(k, n) + \sum_{m=1}^M \sum_{l=0}^{L_g-1} g_m(k, l) x_m(k, n-\tau-l) \quad (5)$$

式中, $g_m(k, l)$ 表示自回归系数又称为线性预测系数, 长度为 L_g , 等式右边的第二项为后期混响。令

$$\mathbf{x}_m(k) = [x_m(k, 1), \dots, x_m(k, N)]^T$$

$$\mathbf{d}_1(k) = [d_1(k, 1), \dots, d_1(k, N)]^T$$

$$\mathbf{g}_m(k) = [g_m(k, 0), \dots, g_m(k, L_g-1)]^T$$

$$\mathbf{g}(k) = [\mathbf{g}_1^T(k), \dots, \mathbf{g}_M^T(k)]^T$$

$$\mathbf{X}_\tau(k) = [\mathbf{X}_{1,\tau}(k), \dots, \mathbf{X}_{M,\tau}(k)]$$

$$\mathbf{X}_{m,\tau}(k) = [\bar{\mathbf{X}}_{m,\tau}(k, 1), \dots, \bar{\mathbf{X}}_{m,\tau}(k, N)]^T$$

$$\bar{\mathbf{X}}_{m,\tau}(k, n) = [x_m(k, n-\tau), \dots, x_m(k, n-\tau-L_g+1)]^T$$

则式(5)的向量表达式为

$$\mathbf{x}_1(k) = \mathbf{d}_1(k) + \sum_{m=1}^M \mathbf{X}_{m,\tau}(k) \mathbf{g}_m(k) \quad (6)$$

$$\mathbf{x}_1(k) = \mathbf{d}_1(k) + \mathbf{X}_\tau(k) \mathbf{g}(k) \quad (7)$$

根据式(7)可知, 只要求得自回归系数 $\mathbf{g}(k)$ 便可以得到目标信号的估计值 $\mathbf{d}_1(k)$ 。利用广义高斯先验(Complex Generalized Gaussian, CGG)模型, 将目标信号建模为一个均值为 0、协方差为 $\lambda(k, n)$ 的超高斯先验模型^[3], 则

$$\hat{\mathbf{g}}(k)^{(i+1)} = \left(\mathbf{X}_\tau^H(k) \mathbf{D}_{\hat{\lambda}(k)^{(i)}}^{-1} \mathbf{X}_\tau(k) \right)^{-1} \times \mathbf{X}_\tau^H(k) \mathbf{D}_{\hat{\lambda}(k)^{(i)}}^{-1} \mathbf{x}_1(k) \quad (8)$$

$$\hat{\lambda}^{(i+1)}(k, n) = \left| \hat{\mathbf{d}}_1^{(i+1)}(k, n) \right|^{(2-P)} \quad (9)$$

式(8)中, \mathbf{D} 是由协方差构成的对角矩阵, $\mathbf{X}_\tau^H(k)$ 表示 $\mathbf{X}_\tau(k)$ 的共轭转置, P 表示形状参数。将协方差写成向量形式, 即

$$\hat{\lambda}^{(i+1)}(k) = \left| \hat{\mathbf{d}}_1^{(i+1)}(k) \right|^{\wedge(2-P)}, \quad 0 < P \leq 2 \quad (10)$$

利用参考信号对协方差 $\lambda(k)$ 进行初始化, 然后利用式(8)求出自回归系数 $\mathbf{g}(k)$, 其次再将 $\mathbf{g}(k)$ 代入式(7)求出 $\mathbf{d}_1(k)$, 最后将求得的 $\mathbf{d}_1(k)$ 代入式(9)求出协方差即可。

式(8)~(10)中的 i 表示迭代次数。该类方法同样存在一定的不足之处, 如式(9)中可能会出现 $\lambda(k, n)$ 等于零的情况; 另外, 在执行过程中只是进行了简单的迭代处理而并没有充分利用时频域内语音信号的稀疏特性。针对该问题, 利用纯净语音的短时特性(具有一定的稀疏性)并结合 NMF 的分解特性, 可以通过提高输出信号的稀疏性来产生一个近似于期望信号作为输出信号。

2 基于 NMF 的协方差估计

在原则上, 纯净语音的任何特性都可以加以利用。研究表明, 纯净语音的短时傅里叶变换系数具

有一定的稀疏性^[9], 因此可以利用提高输出信号的稀疏性来产生一个近似于纯净语音的信号, 根据这一特性可以将其应用到去混响中。

根据式(9)的非负性, 首先构造一个矩阵^[10]:

$$\mathbf{M} = \{|d(n, k)|^{(2-p)}\} \in R_{0+}^{N \times K} \quad (11)$$

同理, 再定义一个非负矩阵 \mathbf{V} 也就是协方差矩阵:

$$\mathbf{V} = \{\lambda(n, k)\} \in R_{0+}^{N \times K} \quad (12)$$

式(11)和式(12)中的 N 和 K 分别表示帧总数和频带总数。根据式(10)中协方差与目标信号的关系, 首先用 NMF 对式(11)进行处理。则其低维表示形式为

$$\mathbf{M} \approx \mathbf{W}\mathbf{H}, \quad \mathbf{W} \in R_{0+}^{N \times r}, \quad \mathbf{H} \in R_{0+}^{r \times K} \quad (13)$$

式(13)中的 \mathbf{W} 表示基矩阵, \mathbf{H} 表示系数矩阵, r 表示特征维数。则协方差的估计矩阵为

$$\mathbf{V} = \mathbf{W}\mathbf{H} \quad (14)$$

实际使用过程中, 为了能够对数据进行有效的数据特征提取并进行有效的表述, 需要提高 NMF 分解的稀疏性。

在已知文献中, NMF 直接被应用到语音去混响中, 而没有考虑对 NMF 进行处理, 如提高 NMF 分解的稀疏性。本文将对 NMF 进行深入的研究以达到进一步提高语音去混响的目的。

以 Kullback-Leibler(KL)为代价函数的 NMF 的系数矩阵不具有正交性即变量之间存在冗余信息。为了去除变量中的冗余信息, 可以对其进行正交化处理^[11]。这与稀疏性的思想是契合的, 也就是说这样能够提高 NMF 整体的稀疏性。因此, 基于式(11)以正交的 KL 为代价函数的 NMF 的优化问题可以归纳为

$$\min_{\mathbf{W} \geq 0, \mathbf{H} \geq 0} F_{KL} = \sum_{ij} \left(M_{ij} \lg \frac{M}{(\mathbf{W}\mathbf{H})_{ij}} - M_{ij} + (\mathbf{W}\mathbf{H})_{ij} \right) + \delta \|\mathbf{H}\mathbf{H}^T - \mathbf{I}\|^2 \quad (15)$$

式中, $i=1, \dots, N$; $j=1, \dots, K$ 。 \mathbf{I} 为单位矩阵, δ 为权重系数, 当 $\delta=0$ 时, 对应于标准的 KL 非负矩阵分解^[12]。

$$\min_{\mathbf{W} \geq 0, \mathbf{H} \geq 0} F_{KL} = \sum_{ij} \left(M_{ij} \lg \frac{M}{(\mathbf{W}\mathbf{H})_{ij}} - M_{ij} + (\mathbf{W}\mathbf{H})_{ij} \right) \quad (16)$$

$$\text{Tr}[(\mathbf{H}\mathbf{H}^T - \mathbf{I})(\mathbf{H}\mathbf{H}^T - \mathbf{I})^T] = [\text{Tr}\mathbf{H}\mathbf{H}^T\mathbf{H}\mathbf{H}^T - 2\mathbf{H}\mathbf{H}^T + \mathbf{I}] \quad (17)$$

$$\frac{\partial \text{Tr}[\mathbf{H}\mathbf{H}^T\mathbf{H}\mathbf{H}^T - 2\mathbf{H}\mathbf{H}^T + \mathbf{I}]}{\partial \mathbf{H}} = 4\delta(\mathbf{H}\mathbf{H}^T\mathbf{H}) - 4\delta\mathbf{H} \quad (18)$$

式中, Tr 表示迹算子。根据式(15)、(17)和(18)得到关于 \mathbf{W} 和 \mathbf{H} 的偏导:

$$\frac{\partial F_{KL}}{\partial \mathbf{W}} = -\frac{\mathbf{M}}{\mathbf{W}\mathbf{H}} \mathbf{H}^T + \mathbf{I}_1 \mathbf{H}^T \quad (19)$$

$$\frac{\partial F_{KL}}{\partial \mathbf{H}} = \mathbf{W}^T \mathbf{I}_1 - \mathbf{W}^T \frac{\mathbf{M}}{\mathbf{W}\mathbf{H}} + 4\delta(\mathbf{H}\mathbf{H}^T\mathbf{H}) - 4\delta\mathbf{H} \quad (20)$$

式(17)和式(18)中, \mathbf{I}_1 是一个元素均为 1 的 $N \times K$ 矩阵。根据梯度下降法, 有

$$\mathbf{W} = \mathbf{W} - \eta_w \frac{\partial F}{\partial \mathbf{W}} \quad (21)$$

$$\mathbf{H} = \mathbf{H} - \eta_H \frac{\partial F}{\partial \mathbf{H}} \quad (22)$$

在这里取步长 η_w 和 η_H 为

$$\eta_w = \frac{\mathbf{W}}{\mathbf{I}_1 \mathbf{H}^T} \quad (23)$$

$$\eta_H = \frac{\mathbf{H}}{\mathbf{W}^T \mathbf{I}_1 + 4\delta \mathbf{H}\mathbf{H}^T \mathbf{H}} \quad (24)$$

将式(23)、(24)代入到式(21)、(22)中, 则

$$\mathbf{W} = \mathbf{W} \otimes \frac{\mathbf{M} \mathbf{H}^T}{\mathbf{I}_1 \mathbf{H}^T} \quad (25)$$

$$\mathbf{H} = \mathbf{H} \otimes \left[\frac{\mathbf{W}^T \frac{\mathbf{M}}{\mathbf{W}\mathbf{H}} + 4\delta \mathbf{H}}{\mathbf{W}^T \mathbf{I}_1 + 4\delta \mathbf{H}\mathbf{H}^T \mathbf{H}} \right] \quad (26)$$

式(25)和式(26)中的 \otimes 表示点乘算子。最后, 将式(25)和式(26)代入式(14)中, 则得到协方差的估计矩阵 \mathbf{V} 。

该去混响方法的整体流程框图如图 1 所示。其实现方法只需要对每一次迭代求出的 $\mathbf{d}_1(k)$ 所构成的矩阵 \mathbf{M} 进行 NMF 分解, 然后再利用式(14)求出协方差矩阵 \mathbf{V} 即可。另外, 由图 1 可知, 相对于传统的语音去混响方法, 改进方法的计算复杂度有所增加。

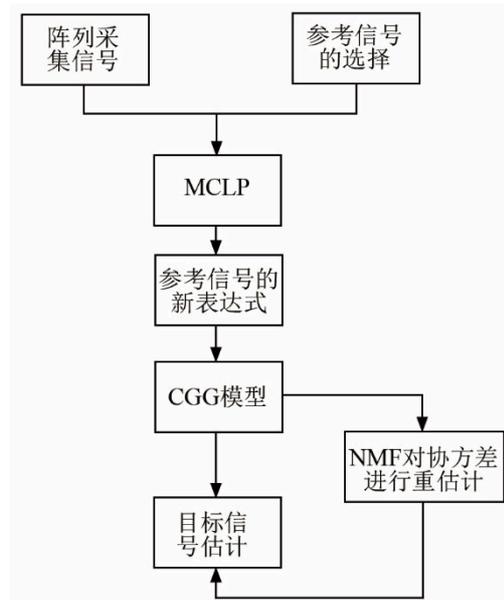


图 1 MCLP 去混响算法框图

Fig. 1 The diagram of speech dereverberation based on MCLP

3 仿真实验与结果分析

3.1 仿真环境

实验中的语音数据通过纯净语音^[13]分别与混响时间 $T_{60}=160\text{ ms}$ (轻度)、 $T_{60}=360\text{ ms}$ (中度)及 $T_{60}=610\text{ ms}$ (重度)情况下的 RIRs^[14]进行卷积得到。语音文本内容为“*She had your dark suit in greasy wash water all year*”。阵列为由 $M=4$ 路全向麦克风组成的均匀线阵。麦克风之间的间距为 8 cm ，声源到阵列中心距离为 2 m ，采样频率为 16 kHz 。短时傅里叶变换的帧长为 64 ms ，帧移为 16 ms ，窗函数选择汉明窗；权重系数 $\delta=0.001$ ， $L_g=8$ ， $\tau=2$ ， $r=60$ ； $P=0.5$ ；语音去混响性能通过语音质量优化和评估参数 PESQ^[15]及频率加权的分段信噪比 FWSegSNR 来评估^[15]。

为分析本文方法 ONMF-MCLP 的有效性，实验选取了经典三种方法即 MCLP^[3]、MCLP+TVG+NMF^[10]和 IRLS^[3]进行了性能比较和分析。

3.2 仿真结果分析

图 2~4 分别是在轻度、中度和重度混响即混响时间 $T_{60}=160\text{ ms}$ 、 $T_{60}=360\text{ ms}$ 和 $T_{60}=610\text{ ms}$ 时，麦克风接收信号和各算法去混响后信号的语音谱。

由图 2~4 可以看出，在各自的混响时间条件下，纯净语音谱的谱分布条理清晰，稀疏性最好。

在轻度混响条件下，混响语音谱的稀疏性仍然比较好；相对于阵列采集信号所对应的混响语音谱，MCLP+TVG+NMF、MCLP、IRLS、NMF-MCLP、ONMF-MCLP 五种方法的去混响效果不明显。

在中度、重度混响条件下，混响语音谱的谱分布布满整个时频域，稀疏性比较差。相对于麦克风阵列采集信号所对应的混响语音谱，MCLP+TVG+NMF、

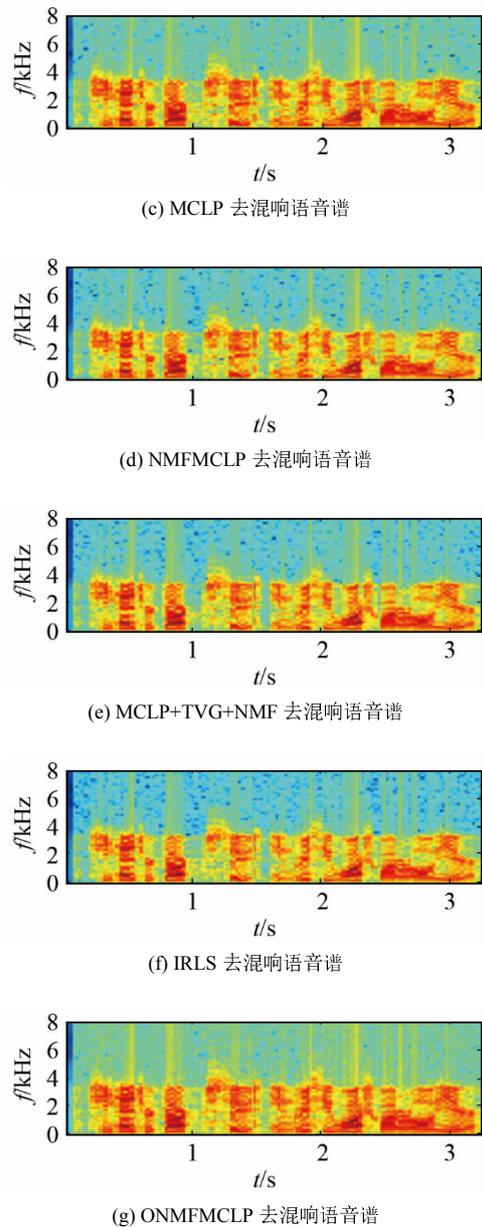
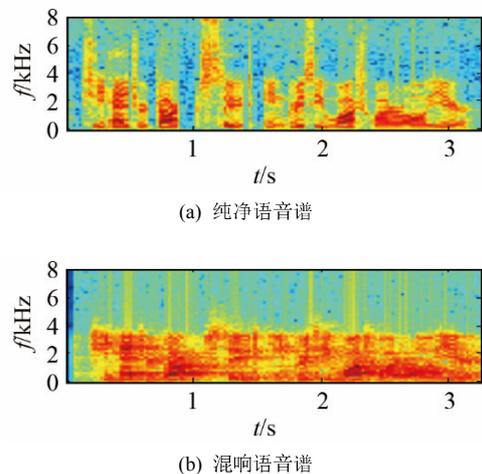
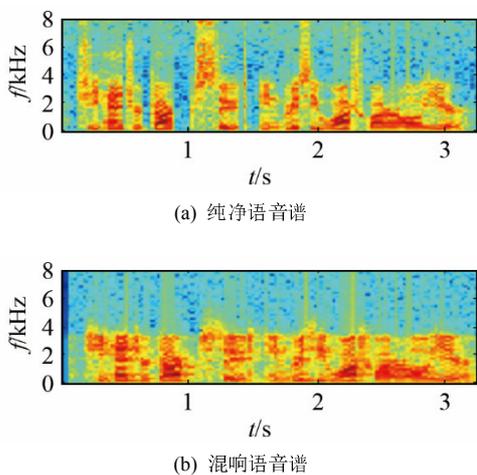
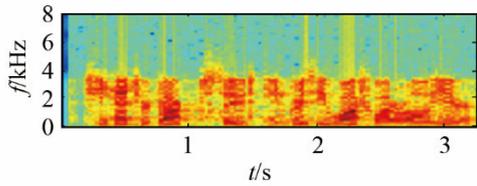
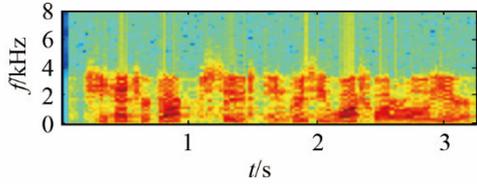


图 2 语音信号谱 $T_{60}=160\text{ ms}$
Fig.2 The spectrograms of the speech signal $T_{60}=160\text{ ms}$

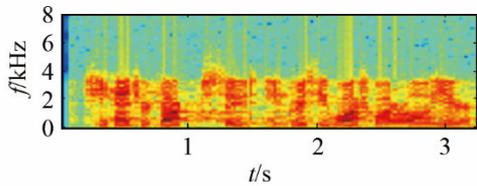




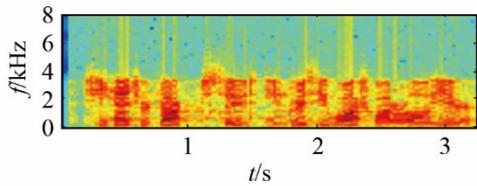
(c) MCLP 去混响语音谱



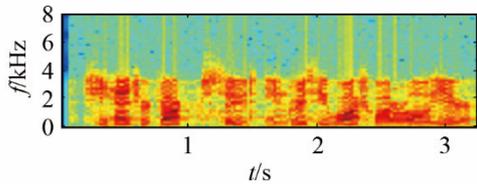
(d) NMF MCLP 去混响语音谱



(e) MCLP+TVG+NMF 去混响语音谱



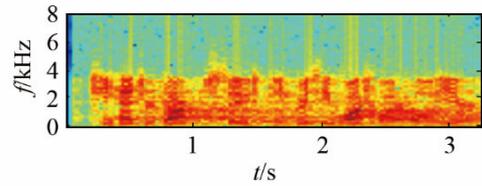
(f) IRLS 去混响语音谱



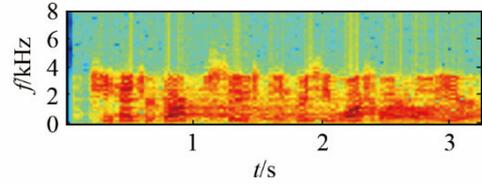
(g) ONMF MCLP 去混响语音谱

图 3 语音信号谱 $T_{60}=360$ ms

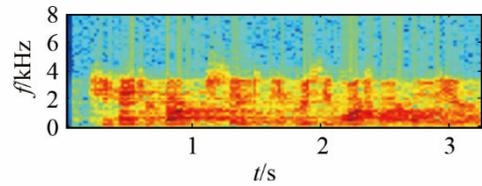
Fig.3 The spectrograms of the speech signals $T_{60}=360$ ms



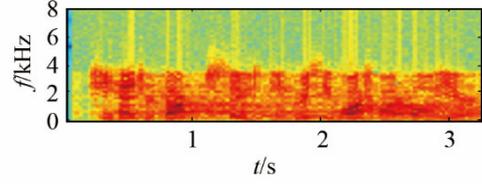
(c) MCLP 去混响语音谱



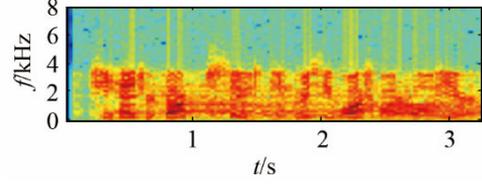
(d) NMF MCLP 去混响语音谱



(e) MCLP+TVG+NMF 去混响语音谱



(f) IRLS 去混响语音谱



(g) ONMF MCLP 去混响语音谱

图 4 语音信号谱 $T_{60}=610$ ms

Fig.4 The spectrograms of the speech signals $T_{60}=610$ ms

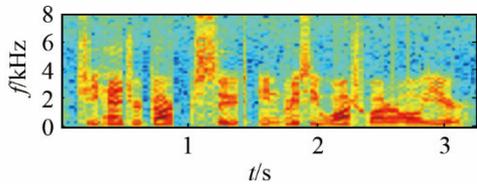
MCLP、IRLS、NMF MCLP、ONMF MCLP 五种方法的去混响语音谱的谱分布不再布满整个时频域内，稀疏性得到了明显的提高。

表 1 是采用主观评测 MOS(Mean Opinion Score)法对去混响语音的评价结果。

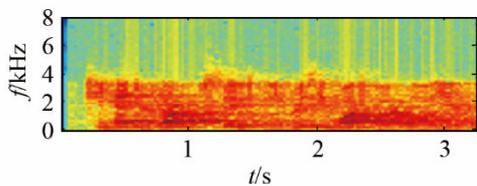
表 1 主观评价结果

Table 1 The subjective evaluation result

去混响方法	$T_{60}=160$	$T_{60}=360$	$T_{60}=610$
混响语音	4	3	2
MCLP+TVG+NMF	4	3.2	2.4
MCLP	4	3.2	2.4
IRLS	4	3	2.4
NMF MCLP	4	3.4	2.6
ONMF MCLP	4	3.2	2.4



(a) 纯净语音谱



(b) 混响语音谱

由表 1 可知，在轻度混响条件下，混响语音及各去混响方法的去混响效果并不是很明显。在中度和重度混响条件下，各去混响方法均起到了一定去混响的效果。

图 5、6 分别为各混响条件下采用相应客观评价指标即 FWSegSNR 和 PESQ 得到的实验结果曲线图。

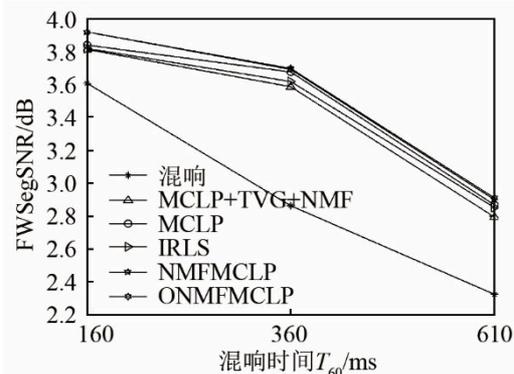


图 5 去混响 FWSegSNR 曲线图
Fig.5 The diagram of FWSegSNR

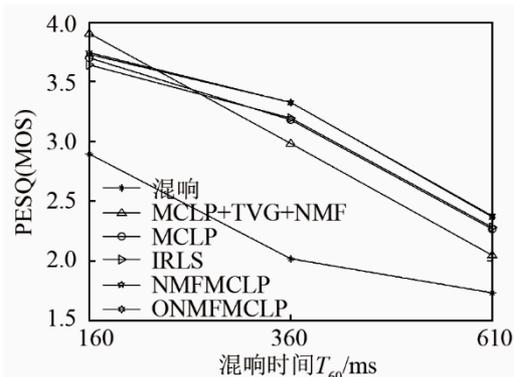


图 6 去混响 PESQ 曲线图
Fig.6 The diagram of PESQ

从图 5、6 中可以看出，在各种混响条件下，相对于混响语音，无论是去混响参数 FWSegSNR 或 PESQ，MCLP+TVG+NMF、MCLP、IRLS、NMF MCLP、ONMF MCLP 五种去混响方法的评估参数都得到了显著的提高。

在轻度混响条件下，MCLP、IRLS、NMF MCLP、ONMF MCLP 四种方法的 FWSegSNR 的提高明显优于 MCLP+TVG+NMF 方法。而 MCLP+TVG+NMF 方法的 PESQ 提高优于其它几种方法。

在中度、重度混响条件下，NMF MCLP，ONMF MCLP 两种方法的性能参数明显高于 MCLP+TVG+NMF、MCLP、IRLS 三种方法且 MCLP+TVG+NMF 方法去混响性能最差；综合以上 FWSegSNR 和 PESQ 两个性能参数，ONMF MCLP 去混响方法几乎等价于 NMF MCLP。

各种语音去混响方法的程序运行时间柱状图

如图 7 所示。

由图 7 可知，MCLP、IRLS 两种方法的程序运行时间一致且最短；NMF MCLP 方法的运行时间次之；MCLP+TVG+NMF、ONMF MCLP 方法的运行时间最长。故改进方法即 NMF MCLP/ONMF MCLP 方法虽然进一步提高了语音的去混响性能，但是其计算复杂度有明显增加。

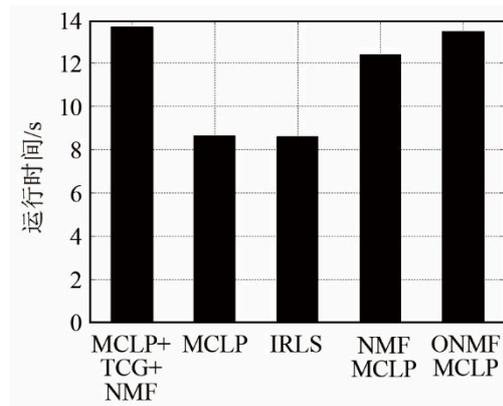


图 7 各去混响方法的运行时间柱状图
Fig.7 The histogram of processing time required for several dereverberation methods

4 结论

本文充分考虑了语音的时频域稀疏性及其 NMF 的特性，对阵列接收目标信号的协方差进行了改进，即通过构建正交 KL 为代价函数的 NMF 优化问题，利用梯度下降法对其进行推导，得到了协方差矩阵的估计值，以达到进一步提高语音的去混响性能。实验结果表明，相对于其他去混响方法，在不需估计混响时间等参数的前提下，NMF MCLP/ONMF MCLP 去混响方法在一定程度上改善了语音去混响性能。但是该方法也有一定的不足之处，例如计算复杂度有明显增加；另外，NMF 的初始化具有随机性。

参 考 文 献

- [1] KODRASI I, JUKIC A, DOČLO S. Robust sparsity promoting acoustic multi-channel equalization for speech dereverberation[C]//IEEE International Conference on Acoustics, 2016: 166-170.
- [2] KUKLASINSKI A, DOČLO S. Maximum likelihood PSD estimation for speech enhancement in reverberation and noise[J]. IEEE/ACM Transactions on Audio Speech & Language Processing, 2016, 24 (9): 1595-1608.
- [3] JUKIC A, WATERSCHOOT T V, GERKMANN T, et al. Multi channel linear prediction-based speech dereverberation with sparse priors[J]. IEEE/ACM Transactions on Audio Speech & Language Processing, 2015, 23 (9): 1509-1520.
- [4] MOHAMMADIHA N, SMARAGDIS P, LEIJON A. Supervised

- and unsupervised speech enhancement using nonnegative matrix factorization[J]. *IEEE Transactions on Audio Speech and Language Processing*, 2013, **21**(10): 2140-2151.
- [5] AKARSH K A, SELVI R S. Speech enhancement using non-negative matrix factorization and enhanced NMF[C]//International Conference on Circuit, 2015: 1-7.
- [6] MOHAMMADIHA N, DOCLO S. Speech dereverberation using non-negative convolutive transfer function and spectro-temporal modeling[J]. *IEEE Transactions on Audio Speech & Language Processing*, 2016, **24**(2): 276-289.
- [7] MOHAMMADIHA N, SMARAGDIS P, DOCLO S. Joint acoustic and spectral modeling for speech dereverberation using non-negative representations[C]//IEEE International Conference on Acoustics, 2015, **8**(4): 4410-4414.
- [8] BRADLEY J S, SATO H, PICARD M. On the importance of early reflections for speech in rooms[J]. *J. Acoust. Soc. Am.*, 2003, **113**(6): 3233-3244.
- [9] TASHEV I, ACERO A. Statistical modeling of the speech signal[C]//International Workshop on Acoustic Echo & Noise Control Tel Aviv Israel, 2010.
- [10] JUKIC A, MOHAMMADIHA N, WATERSCHOOT T V. Multi-channel linear prediction-based on speech dereverberation with low-rank power spectrogram approximation[C]//IEEE International Conference on Acoustics, Speech and Signal Processing, IEEE, 2015, **23**(9): 96-100.
- [11] 崔建. β 散度的非负矩阵分解在基因聚类中的应用研究[D]. 重庆: 重庆大学, 2015.
- CUI Jian. The application of non negative matrix factorization of beta divergence in gene clustering[D]. Chongqing: Chongqing University, 2015.
- [12] FÉVOTTE C, BERTIN N, DURRIEU J L. Nonnegative matrix factorization with the Itakura-Saito divergence: with application to music analysis[J]. *Neural Computation*, 2009, **21**(3): 793.
- [13] GAROFOLO J S, LAMEL L, FISHER W, et al. TIMIT Acoustic-Phonetic Continuous Speech Corpus[J]. *Journal of the Acoustical Society of America*, 1993, **88**(88): 210-221.
- [14] HADAD E, HEESE F, VARY P, et al. Multichannel audio database in various acoustic environments[C]//International Workshop on Acoustic Echo and Noise Control, Antibes, France, 2014: 313-317.
- [15] HU Y, LOIZOU P C. Evaluation of objective quality measures for speech enhancement[J]. *Transactions on Audio Speech and Language Processing*, 2008, **16**(1): 229-238.