

利用相位时频掩蔽的麦克风阵列噪声消除方法

何 礼 周 翊 刘宏清

(重庆邮电大学通信与信息工程学院, 重庆 400065)

摘 要: 本文提出了一种在干扰声源和背景噪声存在条件下麦克风阵列噪声消除的方法。麦克风阵列通过波束形成增强由导向矢量所指定方向的目标声源来抑制背景噪声。然而, 现有的波束形成算法在干扰声源存在的情况下, 无法进行准确的导向矢量估计。为此, 本文提出一种基于音频信号互相关功率谱相位的麦克风阵列噪声消除方法。首先通过音频信号的相位时频掩码估计导向矢量, 并对其进行波束形成, 从而有效抑制干扰声源和背景噪声; 然后利用语音存在概率, 采用最大似然的方法估计波束形成后信号中残留的干扰噪声功率谱密度, 对其进行后处理, 进一步抑制残留干扰和噪声。实验结果表明在干扰声源和背景噪声存在的条件下, 所提方法有效地实现了麦克风阵列噪声消除, 且各种性能指标优于基线方法。

关键词: 时频掩蔽; 相位; 麦克风阵列; 语音存在概率; 后置滤波

中图分类号: TN912.3

文献标识码: A

DOI: 10.16798/j.issn.1003-0530.2018.12.010

Microphone Array Noise Cancellation Method Using Phase Time-frequency Masking

HE Li ZHOU Yi LIU Hong-qing

(School of Communication and Information Engineering, Chongqing University of Posts and Telecommunications, Chongqing 400065, China)

Abstract: A noise cancellation method is studied for microphone array under the interfering sound source and background noise. The microphone array suppresses background noise by reinforcing the target sound source in the direction specified by the steering vector by beamforming. However, the present beamforming algorithms cannot accurately estimate the steering vector when interfering sound source presents. Therefore, this paper presents a noise cancellation method using the phase of audio signal cross-correlation power spectrum based on the microphone array. Firstly, the proposed method estimates the steering vector by the phase time-frequency mask of the audio signal and carries out the beamforming, so as to effectively suppress the interference sound source and background noise. Then the speech presence probability is calculated to estimate residual interference and noise power spectral density using maximum likelihood method and employed on the post-beamforming signal to further suppress residual interference and noise. The experimental results demonstrate that under the interfering sound source and background noise, the proposed microphone array method can effectively suppress noise, which is superior to the baseline methods.

Key words: time-frequency masking; phase; microphone array; speech presence probability; post-filtering

1 引言

近年来, 自动语音识别技术的飞速发展催生

了一系列语音应用, 例如: 手机、智能音箱、助听器和电话会议系统等。为了保证上述设备在复杂声学环境下运行的鲁棒性, 这些设备通常配有噪声

消除算法。这些算法分为基于单个麦克风的单通道噪声消除算法^[1]和基于麦克风阵列的多通道噪声消除算法。很多单通道噪声消除算法依赖于语音活动检测,只适合平稳噪声,对非平稳噪声鲁棒性不好。麦克风阵列基于多个麦克风,在时频域的基础上进一步获得空间域信息。因此,与只用时频信息进行噪声消除的单麦克风方案相比,麦克风阵列通常具有更好的性能,因而得到了广泛应用。

波束形成器是麦克风阵列噪声消除算法的核心。它可设计成固定波束形成器,如延迟求和波束形成器、超指向性波束形成器;也可设计成基于某个优化准则的自适应波束形成器,如线性约束最小方差、最小方差无失真响应(Minimum Variance Distortionless Response, MVDR)、最小均方误差和最大信噪比^[2]。这些现有的技术解决方案通常依赖于扩散场声音相干性的模型,从而区分观测信号中由目标语音主导的时频点和由背景噪声主导的时频点。然而这些解决方案并不适用于噪声环境中干扰声源存在的情况。另外,单独的波束形成器并不总是能够达到理论上的效果,波束形成处理后的输出信号中仍然保留了期望信号方向及波束旁瓣上的残留背景噪声,因此通常会在波束形成后采用单通道后置滤波技术对残留噪声进一步消除。麦克风阵列噪声消除算法通常使用带有单通道维纳后置滤波的 MVDR 波束形成器来实现^[2]。文献[3-4]中推导了多种不同的后置滤波器,它们的主要区别在于对残留噪声功率谱密度的估计,而共同点是都未考虑干扰声源存在的情况,无法对波束形成输出信号中残留干扰的功率谱密度进行估计。

针对上述问题,本文提出了基于相位时频掩蔽的麦克风阵列噪声消除算法。首先通过互相关功率谱的相位信息估计出目标声源和干扰声源的时频掩码。为了克服线性麦克风阵列前后指向性模糊和端射方向敏感等问题,本文采用均匀环形麦克风阵列。然后通过时频掩码估计出导向矢量和噪声协方差矩阵,从而对目标声源进行波束形成,克服了传统导向矢量估计算法受干扰声源影响的缺点。最后,通过参考麦克风的语音存在概率修正相位时频掩码,更加准确估计残留干扰噪声功率谱密

度,经过维纳滤波器生成出最终的增强语音。以下是本文剩余部分的结构安排:第 2 节将介绍麦克风阵列噪声消除的基本原理。第 3 节将重点介绍所提算法。第 4 节进行实验结果描述和分析。第 5 节是结论。

2 麦克风阵列噪声消除基本原理

设在同一个房间内有 M 个麦克风,第 $m=1,2,\dots,M$ 个麦克风接收到的音频信号 $x_m(t)$ 可建模为^[5]:

$$x_m(t) = h_m * s(t) + \sum_{r=1}^R h_{m,r} * i_r(t) + \eta_m(t) \quad (1)$$

其中, h_m 是目标声源与第 m 个麦克风之间的房间冲激响应, $s(t)$ 是目标声源, $h_{m,r}$ 是第 r 个干扰声源与第 m 个麦克风之间的房间冲激响应, $i_r(t)$ 是第 r 个干扰声源, $r=1,2,\dots,R$, $\eta_m(t)$ 是第 m 个麦克风处的背景噪声, ‘*’ 表示卷积运算。

式(1)可简化为:

$$x_m(t) = h_m * s(t) + v_m(t) \quad (2)$$

其中, $v_m(t) = \sum_{r=1}^R h_{m,r} * i_r(t) + \eta_m(t)$ 。

为了克服语音信号非平稳性带来的信号估计问题,通常将语音信号通过短时傅里叶变换(Short Time Fourier Transformer, STFT)转换到时频域进行处理,假设房间冲激响应是时不变的,对式(2)应用 STFT,式(2)在频域中表示为:

$$\mathbf{X}(t,k) = \mathbf{H}(k)S(k) + \mathbf{V}(t,k) \quad (3)$$

其中, $\mathbf{X}(t,k)=[X_1(t,k),\dots,X_M(t,k)]^T$, $\mathbf{H}(k)=[H_1(k),\dots,H_M(k)]^T$, $\mathbf{V}(t,k)=[V_1(t,k),\dots,V_M(t,k)]^T$, 且 $X_m(t,k)$ 、 $H_m(k)$ 和 $V_m(t,k)$ 分别是 $x_m(t)$ 、 h_m 和 $v_m(t)$ 的 STFT, t 表示间索引, k 表示频率索引。 $H_m(k)$ 可建模为延迟衰减模型^[6]:

$$H_m(k) = D_m(k) = \alpha_m e^{-j\omega_k \tau_m} \quad (4)$$

其中, τ_m 是第 m 个麦克风与参考麦克风 $m=1$ 之间的到达时间差, ω_k 是频点 k 角频率的弧度表示。 $\mathbf{D}(k)=[D_1(k),\dots,D_M(k)]^T$ 是麦克风阵列的导向矢量^[2]。文献[7]证明最优多通道最小均方误差语音增强滤波器 W_{opt} 可分解为:

$$W_{\text{opt}} = \frac{\Phi_{vv}^{-1} \mathbf{D}}{\mathbf{D}^H \Phi_{vv}^{-1} \mathbf{D}} \frac{\phi_{ss}}{\phi_{ss} + \phi_{nn}} \quad (5)$$

其中, $W_{\text{MVDR}} = \frac{\Phi_{vv}^{-1} \mathbf{D}}{\mathbf{D}^H \Phi_{vv}^{-1} \mathbf{D}}$ 为最小方差无失真响应波束形成器^[8], Φ_{vv} 为噪声协方差矩阵, 单通道维纳滤波器 $\mu(t, k)$ 为

$$\mu(t, k) = \frac{\phi_{ss}}{\phi_{ss} + \phi_{nn}} \quad (6)$$

其中, ϕ_{ss} 和 ϕ_{nn} 分别为波束形成输出结果中目标信号和噪声信号的功率谱密度。最后, 目标信号估计为:

$$\mathbf{Y}(t, k) = W_{\text{opt}}^H(t, k) \mathbf{X}(t, k) \quad (7)$$

3 基于相位时频掩蔽的麦克风阵列噪声消除方法

3.1 算法整体框架

算法的整体框架如图1所示。

算法包含四部分: 时频掩码估计、波束形成、干扰噪声功率谱密度估计和单通道维纳滤波。在时频掩码估计部分, 结合多麦克风音频信号估计的相位时频掩码和参考麦克风估计的语音存在概率, 得到最终的时频掩码。在波束形成部分, 通过时频掩码估计出导向矢量和噪声协方差矩阵, 基于最小方差无失真响应对目标信号波束形成, 对于干扰信号和背景噪声进行消除。在干扰噪声功率谱密度估计部分, 通过参考麦克风语音存在概率修正的相位时频掩码最大似然估计出上一阶段波束形成后残留的干扰噪声功率谱密度。在单通道维纳滤波部分, 通过上一阶段估计的干扰噪声功率谱密度进一步消除波束形成输出结果中残留的干扰噪声, 生成增强语音。

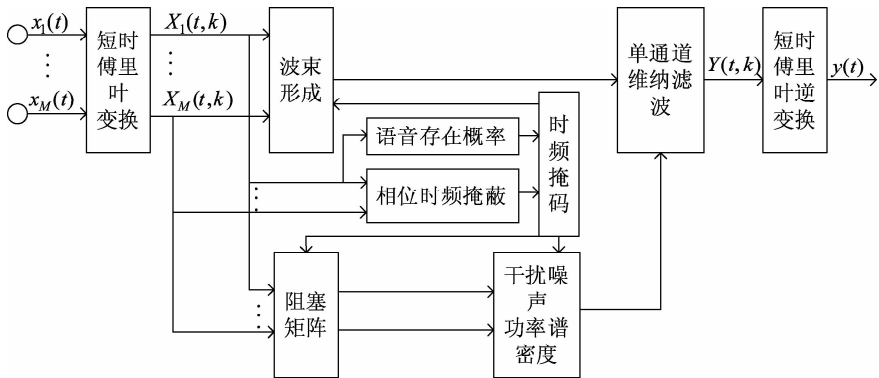


图1 算法噪声消除流程

Fig. 1 The flowchart of noise cancellation

3.2 基于均匀环形麦克风阵列的相位时频掩蔽

为了克服线性麦克风阵列前后指向性模糊和端射方向敏感的问题, 本文采用均匀环形麦克风阵列, 如图2所示。设式(1)是基于均匀环形麦克风阵列建立的模型, 包含 M 个麦克风。对于一个给定方向的声源, 相邻麦克风对表示为 $\{m_i, m_{i+1}\}$, 最后一对麦克风对表示为 $\{m_M, m_1\}$ 。

将环形阵列所在平面分为 M 个均匀扇区, 每个扇区的序号表示为 $g, g=1, 2, \dots, M$, 每个扇区的预设角度为扇区中心线和 x 轴正向的夹角 ϕ_g , 则每个扇区预设角度的预设相位 $G_{m_i \rightarrow m_1}^{(t, k)}(\phi_g)$ 可由下式计算得出^[9]:

$$G_{m_i \rightarrow m_1}^{(t, k)}(\phi_g) \triangleq e^{-j\omega \tau_{m_i \rightarrow m_1}(\phi_g)} \quad (8)$$

其中, $\tau_{m_i \rightarrow m_1}(\phi_g) = \tau_{m_1 m_2}(\phi_g) - \tau_{m_i m_{i+1}}(\phi_g)$ 为 $\{m_1, m_2\}$ 麦克风对与 $\{m_i, m_{i+1}\}$ 麦克风对的相对时延^[9]。

麦克风对音频信号互相关功率谱的相位在单源区域内估计为:

$$G_{m_i m_{i+1}}(t, k) = \frac{X_i(t, k) X_{i+1}^*(t, k)}{|X_i(t, k) X_{i+1}^*(t, k)|} \quad (9)$$

其中, $*$ 表示复共轭。

由式(8)和式(9)分别已知预设角度相位和音频信号相位, 时频掩码索引 $I_{\text{circular}}(t, k)$ 可通过下式估计为:

$$I_{\text{circular}}(t, k) \triangleq \arg \min_{i=1}^m \|G_{m_i m_{i+1}}(t, k) - G_{m_i \rightarrow m_1}^{(t, k)}(\phi_g)\|_2 \quad (10)$$

其中, $I_{\text{circular}}(t, k)$ 表示时频掩码索引, $\|\cdot\|_2$ 表示欧氏范数。

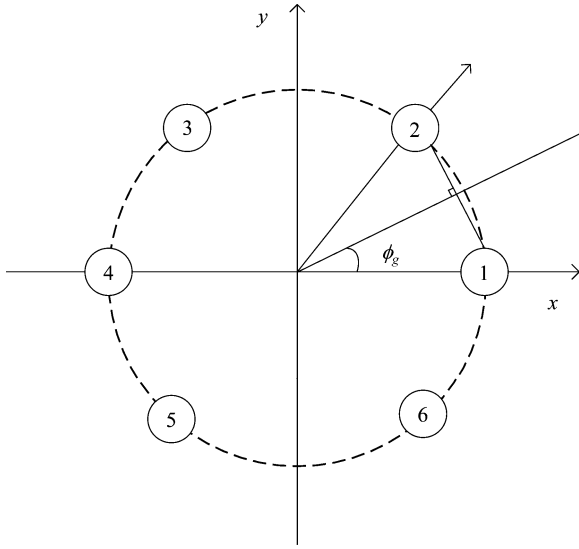


图 2 均匀环形麦克风阵列

Fig. 2 Uniform circular microphone array

针对线性麦克风阵列端射方向敏感的问题,考虑到提出的环形麦克风阵列基于相位时频掩蔽的算法特点,本文在式(10)的基础上增加基于预设角度的权重因子,以弱化误差较大的麦克风对对相位的影响,强化误差较小的麦克风对对相位的影响。因此,式(10)修改为:

$$I_{\text{circular}}(t, k) \triangleq$$

$$\arg \min \sum_{i=1}^m \|G_{m_i m_{i+1}}(t, k) W_{\text{weight}}(\phi_g) - G_{m_i \rightarrow m_1}^{(t, k)}(\phi_g)\|_2 \quad (11)$$

其中,权重因子 $W_{\text{weight}}(\phi_g)$ 定义为:

$$W_{\text{weight}}(\phi_g) = \alpha_{\text{weight}} \cos(\phi_g) + \beta_{\text{weight}} \quad (12)$$

其中, α_{weight} 和 β_{weight} 为常数,设目标声源来源于第 g 个扇区,则基于相位估计的时频掩码 $M_{\text{circular}}(t, k)$ 为:

$$M_{\text{circular}}(t, k) = \begin{cases} 1, & I_{\text{circular}}(t, k) = g \\ 0, & \text{其他} \end{cases} \quad (13)$$

式(13)估计的时频掩码是依赖于不同扇区声源的相位差异,故在目标声源和干扰声源都不存在的情况下是不准确的。因此,式(13)中的时频掩码需要依赖于参考麦克风的语音存在概率进行修正得到最终的时频掩码 $M(t, k)$,如式(14)所示:

$$M(t, k) = \begin{cases} M_{\text{circular}}(t, k), & \text{spp} > \text{threshold} \\ 0, & \text{其他} \end{cases} \quad (14)$$

其中, threshold 为常数, spp 为基于文献[10]的方法计算的参考麦克风的语音存在概率,如式(15)所示:

$$\text{spp}(t, k) = \left\{ 1 + \frac{q(t, k)}{1 - q(t, k)} (1 + \xi(t, k)) \times \exp\left(-\frac{\gamma(t, k)\xi(t, k)}{1 + \xi(t, k)}\right) \right\}^{-1} \quad (15)$$

其中, $q(t, k)$ 是先验语音不存在概率, $\xi(t, k)$ 是先验信噪比, $\gamma(t, k)$ 是后验信噪比。当语音存在时,时频掩码可以分离出目标信号和干扰信号;当语音不存在时,时频掩码可以分离出语音信号和噪声信号。

设 $M^v(t, k)$ 表示只包含干扰噪声信号的时频点, $M^{s+v}(t, k)$ 表示带噪信号的时频点,则只包含干扰噪声信号的协方差矩阵 $R^v(t, k)$ 和带噪信号的协方差矩阵 $R^{s+v}(t, k)$ 分别为:

$$R^v(t, k) = \frac{1}{\sum_t M^v(t, k)} \sum_t M^v(t, k) X(t, k) X^H(t, k) \quad (16)$$

$$R^{s+v}(t, k) = \frac{1}{\sum_t M^{s+v}(t, k)} \sum_t M^{s+v}(t, k) X(t, k) X^H(t, k) \quad (17)$$

因此,目标语音的协方差矩阵表示为:

$$R^s(t, k) = R^{s+v}(t, k) - R^v(t, k) \quad (18)$$

导向矢量可由目标语音协方差矩阵广义特征值分解的最大特征值对应的特征向量估计^[8]。

3.3 基于 spp 的最大似然干扰噪声功率谱密度估计

文献[11]提出在类广义旁瓣消除结构下用最大似然的方法估计波束形成后残留的噪声功率谱密度,取得了比单通道残留噪声功率谱密度估计^[9]更好的效果。但是文献[11]并未考虑干扰的影响,因此,本文通过上文基于参考麦克风语音存在概率修正的相位时频掩码得到目标语音存在概率,准确估计出波束形成后残留的干扰噪声功率谱密度。

当语音不存在时,干扰噪声功率谱密度 $\hat{\Phi}_v(t_0,$

k_0)通过最小值递归平均^[10]方法估计为:

$$\hat{\Phi}_v(t_0, k_0) = \alpha(t_0, k_0) \hat{\Phi}_v(t_0 - 1, k_0) + (1 - \alpha(t_0, k_0)) \mathbf{X}(t_0, k_0) \mathbf{X}(t_0, k_0)^H \quad (19)$$

其中, (t_0, k_0) 表示目标语音不存在时的时频点索引, $\alpha(t_0, k_0)$ 表示目标语音存在概率, 定义为:

$$\alpha(t_0, k_0) = \tilde{\alpha} + (1 - \tilde{\alpha}) M(t_0, k_0) \quad (20)$$

其中, $\tilde{\alpha}$ 是平滑因子, 为常数。由式(20)可知, 相比文献[11]只考虑目标声源和背景噪声的差异估计目标语音存在概率, 本文通过上文得到的时频掩码估计目标语音存在概率, 可以克服干扰的影响, 因此, 能够准确估计出干扰噪声功率谱密度。

由式(19)可知 $\hat{\Phi}_v(t_0, k_0)$ 为目标语音不存在时的干扰噪声功率谱密度。本文假设目标语音存在时的干扰噪声功率谱密度 $\hat{\Phi}_v(t, k)$ 满足下列模型^[11]:

$$\hat{\Phi}_v(t, k) = C^2(t) \hat{\Phi}_v(t_0, k_0), t > t_0 \quad (21)$$

其中, $C^2(t)$ 为一个时变的比例因子。

通过最大似然^[11]的方法估计出 $C_{ML}^2(t)$, 通过式(5)的 MVDR 波束形成器可推导出波束形成后残留的干扰噪声功率谱密度 $\hat{\Phi}_{mn}(t, k)$ 为:

$$\hat{\Phi}_{mn}(t, k) = (\mathbf{W}_{MVDR})^H C_{ML}^2(t) \hat{\Phi}_v(t_0, k_0) \mathbf{W}_{MVDR} \quad (22)$$

由式(5)可得出波束形成后信号的功率谱密度为:

$$\hat{\Phi}_{ss}(t, k) = |(\mathbf{W}_{MVDR})^H \mathbf{X}(t, k)|^2 \quad (23)$$

将式(22)、(23)代入式(6)的维纳后置滤波中, 可得到增强语音 $\mathbf{Y}(t, k)$, 如式(24)所示:

$$\mathbf{Y}(t, k) = \left[\frac{\hat{\Phi}_{ss}(t, k)}{\hat{\Phi}_{ss}(t, k) + \hat{\Phi}_{mn}(t, k)} \right] [(\mathbf{W}_{MVDR})^H \mathbf{X}(t, k)] \quad (24)$$

4 仿真实验与结果分析

本节中, 将提出的算法在 MATLAB 环境下进行仿真实验。实验中所用的 30 条纯净语音和噪声数据均来自 CHIME4 语音库^[12], 采样率为 16 kHz, 量化精度为 16 bit。噪声数据包括 BUS 噪声、CAF 噪声、PED 噪声和 STR 噪声。为消除混响对算法的影响, 选择低混响房间, 混响时间 T_{60} 在 200 ms 以内。

环形六麦阵列原点位于房间中央, 阵列半径为 5 cm, 目标声源位于麦克风阵列的 0° 方向, 干扰声源位于麦克风阵列的 180° 方向, 两者距离麦克风阵列原点都为 1 m。在背景噪声下, 目标声源发声, 4 s 之后干扰声源发声, 12 s 之后目标声源结束, 4 s 之后干扰声源结束, 4 s 之后背景噪声结束。通过 MATLAB 对目标声源、干扰声源和背景噪声进行混合, 信干比为 0 dB 的条件下信噪比分别为 -5 dB、0 dB、5 dB 和 10 dB。增强算法 STFT 使用的帧长为 512 个采样点, 帧移为 128 个采样点。

为验证本文所提算法的性能, 本文采用经典的基于广义互相关相位 (Generalized Cross-Correlation Phase, GCC-Phat) 的延迟求和波束形成器和最优修正对数幅度谱 (Optimal Modified Log Spectral Amplitude, OMLSA) 后置滤波器的麦克风阵列噪声消除算法为基线方法。为能客观评估算法性能, 本文采用感知语音质量 (Perceptual Evaluation of Speech Quality, PESQ)、频域分段信噪比 (Frequency Weighted Segment Signal Noise Ratio, fwsegSNR)、信号干扰比 (Signal to Interference Ratio, SIR)、信号失真比 (Signal to Distortion Ratio, SDR) 这一系列指标分别对原始语音 (original)、基线方法 (baseline)、本文所提算法 (proposed) 与加入多通道后置滤波的本文所提算法 (proposed + multipost) 的性能进行全面衡量。

4.1 PESQ 评价法

PESQ 是评估语音主观试听感受的客观计算方法^[13]。PESQ 分值越高说明其语音感知质量越高。

图 3 为在 BUS 噪声、CAF 噪声、PED 噪声和 STR 噪声环境下各算法在各信噪比条件下的平均 PESQ 得分结果。可见本文所提算法在低信噪比 (-5 dB) 条件下, 相比基线方法具有显著提升的增强效果, 同时随着信噪比提高 (0 dB、5 dB、10 dB), 通过加入多通道后置滤波, 在高信噪比 (5 dB、10 dB) 条件下, 语音感知质量进一步提高。4 种噪声条件下, STR 噪声语音感知质量提升幅度相对有限, 说明所提算法在背景噪声平稳时的优势有限。

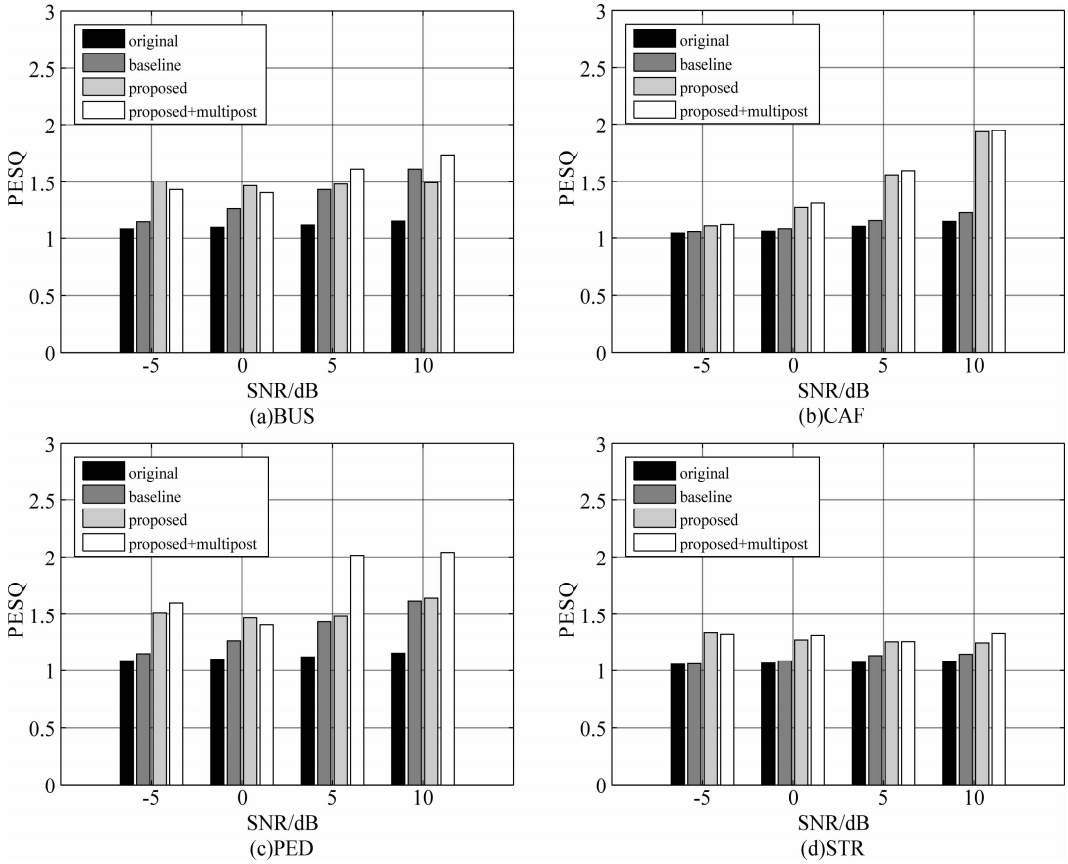


图 3 4 种背景噪声环境下,各算法在各信噪比条件下的 PESQ 均值

Fig. 3 The average PESQ of each algorithm under various SNR conditions and four noise types

4.2 fwsegSNR 评价法

频域分段信噪比是衡量增强语音中所含噪声能量的常规指标,其定义式如式(25)所示:

$$\text{fwsegSNR} = \frac{10 \sum_{m=1}^{M-1} \sum_{j=1}^k B_j \log_{10} [F^2(m, j) / (F(m, j) - \hat{F}(m, j))^2]}{\sum_{j=1}^k B_j} \quad (25)$$

其中, B_j 为位于第 j 个频带的权重, k 为划分的频带总个数, M 为信号的总帧数。 $F(m, j)$ 为第 m 帧语音信号第 j 个频带的幅度, $\hat{F}(m, j)$ 为增强后的信号在第 m 帧第 j 个频带的幅度。 fwsegSNR 值越大说明噪声消除得越好。

图 4 为在 BUS 噪声、CAF 噪声、PED 噪声和 STR 噪声环境下,各算法在各信噪比条件下的平均 fwsegSNR 值。总体上看,本文所提算法在各背景噪

声环境下表现的噪声消除能力最强,在低信噪比 (-5 dB) 条件下加入后置滤波后, fwsegSNR 值有所下降,说明在噪声消除的同时去除了较多语音部分。在 CAF 噪声、PED 噪声环境下,本文所提算法相比基线方法, fwsegSNR 值提升最大,说明所提算法在背景噪声非平稳时效果最好。

4.3 SIR 评价法

信干比是衡量算法抑制干扰声源性能的常规指标^[14],其定义式如式(26)所示:

$$\text{SIR} = 10 \log_{10} \frac{|s_{\text{target}}|^2}{|e_{\text{interf}}|^2} \quad (26)$$

其中, s_{target} 为增强语音中的目标语音, e_{interf} 为增强语音中的干扰语音。 SIR 值越大说明干扰消除得越好。

图 5 为在 BUS 噪声、CAF 噪声、PED 噪声和 STR 噪声环境下,各算法在各信噪比条件下的平均 SIR 值。可见,基线方法对干扰几乎没有消除作用,

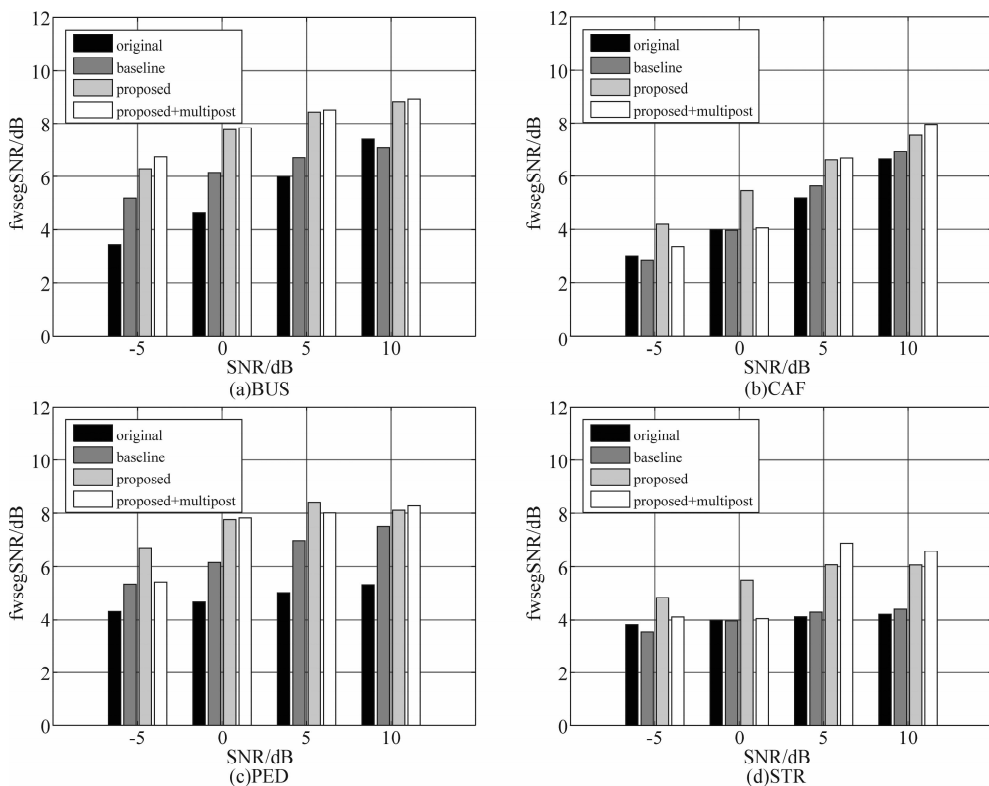


图4 4种背景噪声环境下,各算法在各信噪比条件下的fwsegSNR均值

Fig. 4 The average fwsegSNR of each algorithm under various SNR conditions and four noise types

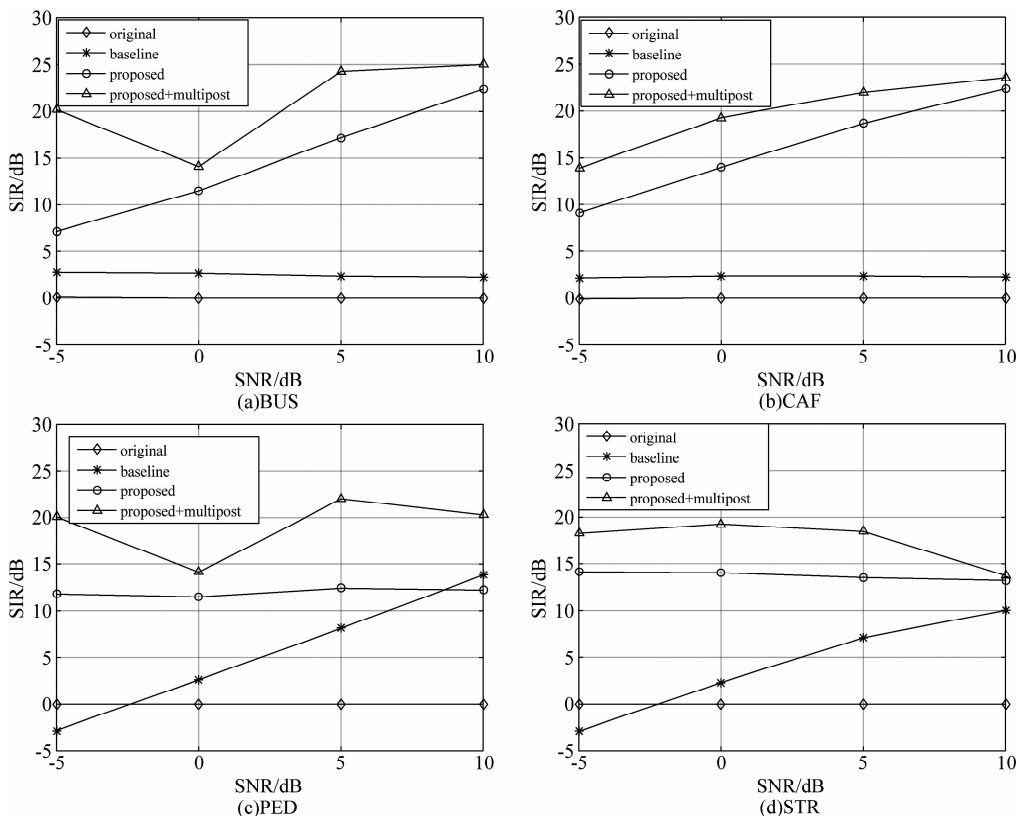


图5 4种背景噪声环境下,各算法在各信噪比条件下的SIR均值

Fig. 5 The average SIR of each algorithm under various SNR conditions and four noise types

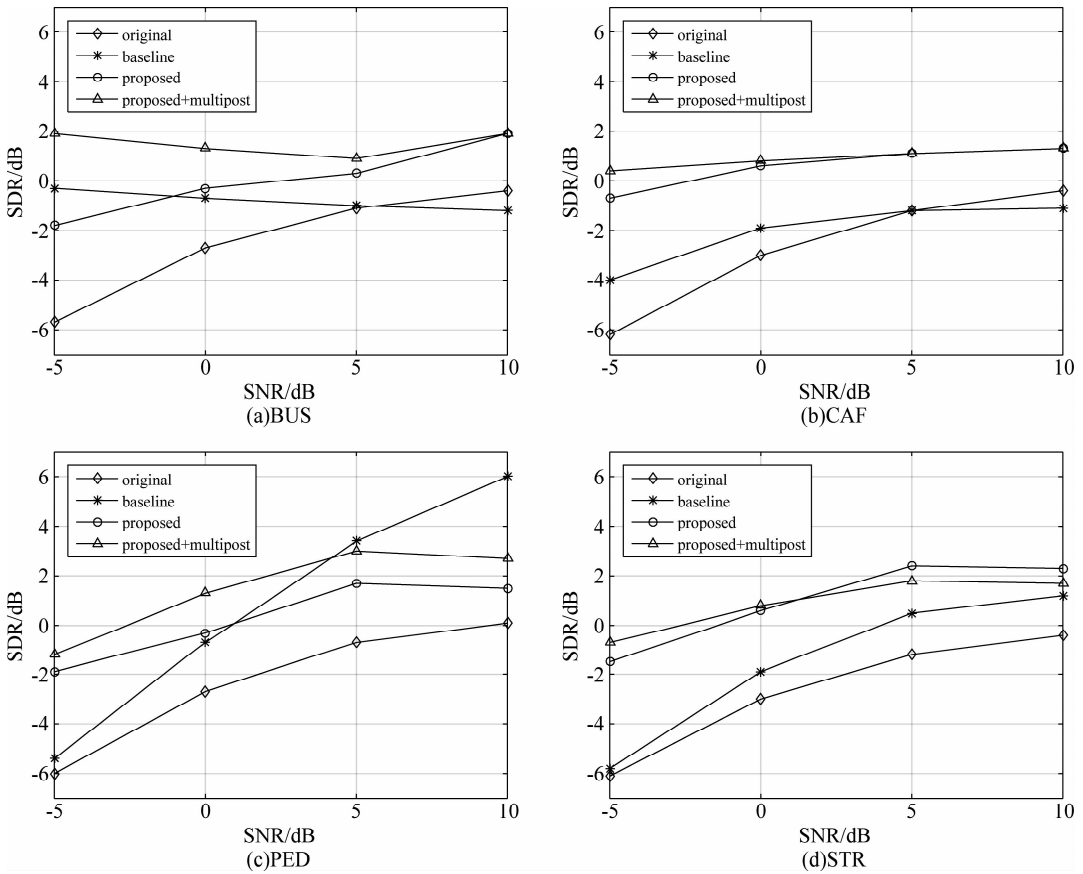


图 6 4 种背景噪声环境下,各算法在各信噪比条件下的 SDR 均值

Fig. 6 The average SDR of each algorithm under various SNR conditions and four noise types

而本文所提算法具有显著的抑制干扰效果。在信干比为 0 dB 的条件下,本文所提算法在低信噪比(-5 dB)时,抑制干扰 10 dB 左右;在高信噪比(10 dB)时,抑制干扰 20 dB 左右。同时,加入后置滤波后,SDR 值有所提升,说明本文提出的基于 spp 的后置滤波具有抑制干扰的效果。

4.4 SDR 评价法

信号失真比是衡量算法抑制干扰过程中语音失真程度的常规指标^[14],其定义式如(27)所示:

$$SDR = 10 \log_{10} \frac{|s_{target}|^2}{|e_{interf} + e_{noise} + e_{artif}|^2} \quad (27)$$

其中, s_{target} 为增强语音中的目标语音, e_{interf} 为增强语音中的干扰语音, e_{noise} 为增强语音中的噪声分量, e_{artif} 为增强语音中的人为噪声。SDR 值越小说明增强语音过程中语音失真越严重。

图 6 为在 BUS 噪声、CAF 噪声、PED 噪声和 STR 噪声环境下,各算法在各信噪比条件下的平均 SDR 值。在低信噪比条件下(-5 dB),本文所提算

法与基线方法相比,造成的语音失真更小,但随着信噪比提高(0 dB、5 dB、10 dB),其优势逐渐减小。后置滤波加入后,在高信噪比(10 dB)条件下,SDR 值有所下降,说明后置滤波在抑制干扰的同时会使语音有一定程度的失真。

5 结论

本文针对在干扰声源和背景噪声的环境下传统波束形成和后置滤波算法难以有效实现噪声消除的问题,提出一种基于相位时频掩蔽的麦克风阵列噪声消除算法。该算法利用不同声源的相位差异和语音存在概率估计时频掩码对目标语音波束形成,对干扰声源和背景噪声有效抑制,然后通过后置滤波进一步抑制残留干扰和噪声。在 4 种不同噪声以及 4 种不同高低信噪比环境下进行了实验,证明所提算法能有效提高语音的质量。

在后续研究中,将考虑所提算法的实时性,同时提高现有算法的计算效率,减小后置滤波引起的

语音失真,此外针对所提算法在混响环境下的适用性问题展开研究。

参考文献

- [1] 胡永刚,张雄伟,邹霞,等.改进的非负矩阵分解语音增强算法[J].信号处理,2015,31(9):1117-1123.
Hu Yonggang, Zhang Xiongwei, Zou Xia, et al. Improved Nonnegative Matrix Factorization based Speech Enhancement Algorithm[J]. Journal of Signal Processing, 2015, 31(9): 1117-1123. (in Chinese)
- [2] Gannot S, Vincent E, Golan S M, et al. A consolidated perspective on multimicrophone speech enhancement and source separation[J]. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2017, 25(4): 692-730.
- [3] McCowan I A, Bourlard H. Microphone array post-filter based on noise field coherence[J]. IEEE Transactions on Speech and Audio Processing, 2003, 11(6): 709-716.
- [4] Lefkimmiatis S, Maragos P. A generalized estimation approach for linear and nonlinear microphone array post-filters[J]. Speech Communication, 2007, 49(7): 657-666.
- [5] Brutti A, Tsiami A, Katsamanis A, et al. A Phase-Based Time Frequency masking for multi-channel speech enhancement in domestic environments[C]//Proceedings of Interspeech, 2016: 2875-2879.
- [6] Doclo S, Moonen M. Design of far-field and near-field broadband beamformers using eigenfilters[J]. Signal Processing, 2003, 83(12): 2641-2673.
- [7] Simmer K U, Bitzer J, Marro C. Microphone Arrays: Signal Processing Techniques and Applications[M]. Berlin, Heidelberg: Springer Berlin Heidelberg, 2001, ch. Post-Filtering Techniques.
- [8] Higuchi T, Ito N, Yoshioka T, et al. Robust mvdr beamforming using time-frequency masks for online/offline asr in noise[C]//Shanghai: IEEE Int. Conf. Acoustic, Speech and Signal Process (ICASSP), 2016: 5210-5214.
- [9] Karbasi A, Sugiyama A. A new DOA estimation method using a circular microphone array[C]//in Proceedings of the European Signal Processing Conference (EUSIPCO), 2007: 778-782.
- [10] Cohen I, Berdugo B. Speech enhancement for non-stationary noise environments[J]. Signal Process, 2001, 81(11): 2403-2418.
- [11] Kjems U, Jensen J. Maximum likelihood based noise covariance matrix estimation for multi-microphone speech enhancement[C]//in Proceedings of the European Signal Processing Conference (EUSIPCO), 2012: 295-299.
- [12] Vincent E, Watanabe S, Nugraha A A, et al. An analysis of environment, microphone and data simulation mismatches in robust speech recognition[J]. Computer Speech and Language, 2017, 46(8): 535-557.
- [13] Hu Y, Loizou P C. Evaluation of objective quality measures for speech enhancement[J]. IEEE Transactions on Audio, Speech, and Language Processing, 2008, 16(1): 229-238.
- [14] Emiya V, Vincent E, Harlander N, et al. Subjective and objective quality assessment of audio source separation[J]. IEEE Transactions Audio, Speech, Language Process, 2011, 19(7): 2046-2057.

作者简介



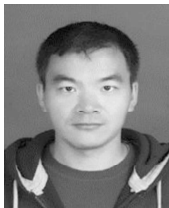
何礼男,1994年生,四川人。重庆邮电大学硕士研究生。研究方向为语音信号处理、麦克风阵列语音增强和去混响。

E-mail: s160101042@stu.cqupt.edu.cn



周翊男,1974年生,四川人。重庆邮电大学通信与信息工程学院教授,硕士生导师。研究方向为语音信号处理、智能听觉分析和机器听觉。

E-mail: zhouj@cqupt.edu.cn



刘宏清男,1980年生,黑龙江人。重庆邮电大学信息与通信工程学院教授,硕士生导师。研究方向为稀疏信号处理、语音信号处理和图像信号处理。

E-mail: liuhq@cqupt.edu.cn