高技术通讯 2024 年 第 34 卷 第 12 期:1307-1317

doi:10.3772/j.issn.1002-0470.2024.12.007

基于 GSMDS 超指向性波束形成的室内定向语音增强方法①

雷世伟②* 刘祖斌③* **

(* 浙江工业大学特种装备制造与先进加工技术教育部重点实验室 杭州 310014)

(** 浙江工业大学台州研究院 台州 318014)

摘 要 针对室内定向语音增强时常出现的混响干扰、语音失真问题,本文提出一种基于 Gram-Schmidt 模态波束分解与综合(GSMDS)的超指向性波束形成室内定向语音增强方法。首先进行基于加权预测误差的去混响前处理,然后采用 GSMDS 超指向性波束形成进行定向目标语音增强,再利用归一化最小均方误差自动增益控制算法对增强语音信号进行增益调节来控制响度的稳定,从而完成对目标语音的定向增强。以 4 阵元端射线阵进行数值仿真,结果表明,该方法相比于常规波束形成(CBF)方法增益提高了 10.1 dB,最高平均增益达到了 22.7 dB,同时端射方向的语音质量感知语音质量评价(PESQ)值相比于原接收混响语音信号 PESQ 值提高了 0.2。在经过装修的房间内进行 4 阵元端射线阵实验,实验结果表明,该方法能获得较好的定向增强效果。

关键词 室内定向语音增强; Gram-Schmidt 模态波束分解与综合(GSMDS); 超指向性; 自动增益控制; 端射线阵

近年来,基于麦克风阵列的波束形成方法[1]被广泛应用于语音增强,比如音视频会议系统、智能机器人、智能音箱等领域。对于声源方向固定的语音增强,可采用麦克风线阵对目标方向语音信号进行定向增强,比如枪型麦克风用于直播、电视广播等录音场景。但在对目标说话人语音进行增强时,容易受到干扰噪声的影响,尤其室内条件下,混响会造成增强语音信号语音质量下降和定向准确性下降,甚至导致语音失真。

经过多年发展,出现了许多经典的波束形成语音增强算法。常规波束形成(conventional beamforming, CBF)方法依据各个阵元接收信号的延时导致相位上的不同,设置一个加权向量补偿延时使得期望方向语音信号对齐,再进行求和输出,完成波束形成。CBF 算法处理简单、稳健性高,但在语音主要频谱内的空间指向的主瓣宽度大、分辨率低[2]。Capon[3]通过约束期望方向波束响应,同时保持总输出功率最小,提出最小方差无失真响应(minimum variance distortionless response, MVDR)波束形成方法;但 MVDR 算法稳健性低,对宽带语音信号处理的频率一致性差。Frost[4]提出线性约束最小方差(linearly constrained minimum variance, LCMV)波束形成方法,它是 MVDR 算法的推广,将 MVDR 约束条件用线性方程组表示,能设置多个约束条件对干扰方向进行零陷抑制,对期望方向保持输出,但是 LCMV 算法在弱相干噪声场和混响环境中处理效果差,性能严重降低。随后 Griffith 和 Jim[5]提出广义旁瓣抵消(generalized sidelobe canceller, GSC)算法将 LCMV 中的约束最优化问题转化为无约束最优化问题,优化求解过程,但是 GSC 算法的阻塞矩阵输出中含有语音成分,会造成语音信号相消,算法性能下降。Gannot 等人[6]通过计算期望信号传递函数比率,构

① 国家自然科学基金(61601407)和浙江省自然科学基金探索(LY20A040007)资助项目。

② 男,1997 年生,硕士生;研究方向:语音增强与声源定位;E-mail: svlab_leishiwei@163.com。

③ 通信作者,E-mail: liuzb@zjut.edu.cn。

(收稿日期:2023-03-15)

建 GSC 中的阻塞矩阵,称为 TF-GSC 算法,此算法方便求解阻塞矩阵,但在非平稳噪声情形下算法性能差[7]。

相比于上述经典波束形成方法,对于定向语音增强而言,在特定方向能获得更高的指向性尤为重要。超指向性波束形成方法[8]通过利用声场特性,推导得到具有高指向性因子(directivity factor, DF)的波束。文献[9-11]提出的基于特征波束分解与综合(eigen-beam decomposition and synthesis, EBDS)的圆环阵超指向性波束形成方法,利用圆环阵协方差矩阵的循环特性推导出最优加权向量,合成得到的最优波束可表示成有限阶子分量的叠加和形式。当合成阶数越高,指向性指数越大,但对误差越敏感,稳健性变差;当合成阶数越低,指向性指数越小,指向性减弱,但稳健性提高。EBDS 超指向性波束形成方法能提供超指向性,但阵列的形状限定为圆环阵,不适用于端射线阵。除此之外,厉剑等人[12]提出一种球谐域类正则化宽带超指向性波束形成算法,该算法结合了球谐域超指向性波束形成器和延迟求和波束形成器,能有效抑制定向噪声和弥漫性噪声,提高语音质量水平。Wang 等人[13]提出一种双耳超指向性波束形成算法,此算法令波束形成器输出的目标信号是同相位的,而输出的白噪声成分是随机相位的,因此可以更好地将目标声音信号从白噪声中分离出来。Wang 等人[14]提出了 Gram-Schmidt 模态-波束分解与合成(Gram-Schmidt mode-beam decomposition and synthesis, GSMDS)波束形成算法,该算法利用 Gram-Schmidt 正交化方法计算噪声协方差矩阵,并用于构建加权向量,可将合成波束进行分阶[15],对比 EBDS 方法,此方法可用于任意阵型的阵列,尤其适用于端射线阵。同时 GSMDS 算法合成最优波束的指向性指数更大,稳健性相比 EBDS 方法也更好。

综上所述,相比于 CBF 等经典波束形成方法,本文选择 GSMDS 超指向性波束形成进行室内定向语音增强,可以获得更高的语音增益水平和更稳健的定向效果,但是 GSMDS 增强后存在响度不稳定问题,需要进行响度控制。为此,首先采用加权预测误差算法对多通道信号进行去混响处理,再进行基于端射线阵的 GSMDS 超指向性波束形成计算,然后对输出信号进行基于归一化最小均方误差的自动增益控制,最后得到高增益、清晰的语音信号。

1　理论部分

1.1　GSMDS 超指向性波束形成

GSMDS 超指向性波束形成应用于端射线阵,能在较小阵元间距下实现宽频带的高指向性。图 1 所示为一个具有 M 个阵元的均匀线阵,相邻阵元间距为 d, 假设远场有一目标语音信号的入射角度为 θ, 参考阵元为阵元 1,则阵列的方向向量 $\boldsymbol{a}(\theta)$ 为

$$\begin{aligned}\boldsymbol{a}(\theta) &= \exp\left(-j\cdot\frac{2\pi}{\lambda}\cdot\boldsymbol{q}\cos\theta\right)\\ &= [a_0(\theta),a_1(\theta),\cdots,a_{M-1}(\theta)]^{\mathrm{T}}\end{aligned}\tag{1}$$

其中,$j=\sqrt{-1}$,λ 为信号波长,$\boldsymbol{q}=[0,d,\cdots,(M-1)d]^{\mathrm{T}}$ 为阵元的位置向量。由此可得波束响应 $B(\theta_0)$ 为

$$B(\theta_0)=\boldsymbol{w}^{\mathrm{H}}(\theta_0)\boldsymbol{a}(\theta)\tag{2}$$

其中,θ_0 为预设导向角度,$\boldsymbol{w}$ 为加权向量,上标 H 表示共轭转置。

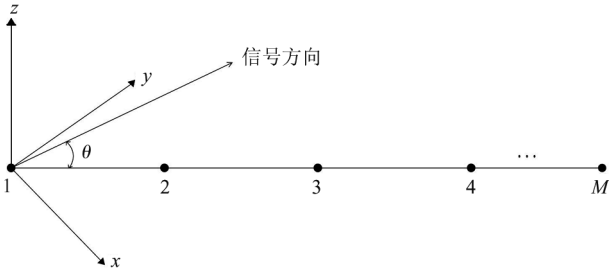


图 1　M 阵元均匀线阵示意图

GSMDS 超指向性波束形成算法利用 Gram-Schmidt 正交化方法消除干扰声,可表示为

$$\boldsymbol{V}=\boldsymbol{C}\boldsymbol{U}\tag{3}$$

其中,$\boldsymbol{C}$ 为正交化矩阵,$\boldsymbol{U}$ 为干扰声向量,$\boldsymbol{V}$ 为预白化后输出的干扰声矩阵,$\boldsymbol{V}$ 中第 k 项 $\boldsymbol{V}_k=\boldsymbol{U}_k-\sum_{j=0}^{k-1}\frac{\langle\boldsymbol{U}_k,\boldsymbol{V}_j\rangle}{\langle\boldsymbol{V}_j,\boldsymbol{V}_j\rangle}\boldsymbol{V}_j$, 其中$\langle\cdot\rangle$表示内积运算,代入式(3),化简可得最优加权向量 $\boldsymbol{w}_{\mathrm{opt}}$为

$$\boldsymbol{w}_{\mathrm{opt}}=\boldsymbol{C}^{\mathrm{T}}\boldsymbol{D}^{-2}\boldsymbol{C}\cdot\boldsymbol{a}(\theta_0)\tag{4}$$

其中,$\boldsymbol{D}=\mathrm{diag}\{|V_0|,|V_1|,\cdots,|V_{M-1}|\}$,$\mathrm{diag}\{\cdot\}$

表示对角矩阵,|·|表示取模运算,利用 $\boldsymbol{w}_{\text{opt}}$ 进行波束形成,得到输出信号 z:

$$z = \boldsymbol{w}_{\text{opt}}\boldsymbol{x} \tag{5}$$

将 $\boldsymbol{w}_{\text{opt}}$ 代入式(2)可得波束响应为

$$\begin{aligned} B(\theta_0) &= \boldsymbol{w}_{\text{opt}}^{\text{H}}\boldsymbol{a}(\theta) = [\boldsymbol{C}\cdot\boldsymbol{a}(\theta_0)]^{\text{H}}\boldsymbol{D}^{-2}[\boldsymbol{C}\cdot\boldsymbol{a}(\theta)] \\ &= \sum_{k=0}^{M-1}\frac{[\boldsymbol{C}_k^{\text{T}}\boldsymbol{a}(\theta_0)]^{*}\cdot[\boldsymbol{C}_k^{\text{T}}\boldsymbol{a}(\theta)]}{\boldsymbol{C}_k^{\text{T}}(\boldsymbol{C}^{\text{T}}\boldsymbol{D}^{-2}\boldsymbol{C})\boldsymbol{C}_k} \\ &= \frac{E_k^{*}(\theta_0)\cdot E_k(\theta)}{\eta_k} = \sum_{k=0}^{M-1}b_k(\theta,\phi) \end{aligned} \tag{6}$$

其中 $E_k(\theta) = \boldsymbol{C}_k^{\text{T}}\boldsymbol{a}(\theta)$,$\eta_k = \boldsymbol{C}_k^{\text{T}}(\boldsymbol{C}^{\text{T}}\boldsymbol{D}^{-2}\boldsymbol{C})^{-1}\boldsymbol{C}_k$,分解的 b_k 为第 $k(k=0,1,\cdots,M-1)$ 阶 GS 模态波束。

由上述可得指向性因子 DF 为

$$\begin{aligned} DF &= [\boldsymbol{C}\boldsymbol{a}(\theta_0)]^{\text{H}}\boldsymbol{D}^{-2}[\boldsymbol{C}\boldsymbol{a}(\theta_0)] \\ &= \sum_{k=0}^{M-1}\frac{|E_k(\theta_0)|^2}{\eta_k} = \sum_{k=0}^{M-1}Q_k \end{aligned} \tag{7}$$

其中 $Q_k = \dfrac{|E_k(\theta_0)|^2}{\eta_k}$ 为第 k 阶模态波束的指向性因子。各阶模态波束指向性因子越大,波束的指向性越好。图 2 所示为一个 4 阵元端射线阵的各阶波束指向性因子与 d/λ(阵元间距与波长之比)值的变化关系图。从图中可看出当 $d/\lambda \leqslant 0.5$ 时,各阶波束的指向性因子 DF 均大于等于 1,且频率越低,DF 越大,指向性因子水平越高,定向增强水平就越高;当 $d/\lambda > 0.5$ 时,各阶波束指向性因子 DF 在常量 1 附近波动,定向增强水平就较低。

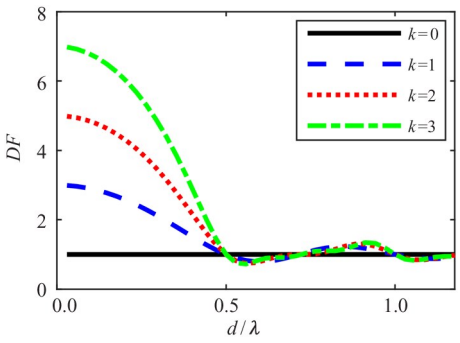


图 2　GSMDS 波束形成端射方向各阶波束指向性因子 DF 随 d/λ 变化关系图($M=4$)

综上所述,GSMDS 超指向性波束形成算法利用 Gram-Schmidt 正交化方法消除干扰噪声,增强目标声信号,在目标方向获得更高的指向性因子,实现超指向性。与其他波束形成相比,GSMDS 波束形成方法可以通过设计阵列获得语音主要频率范围内大于 1 的指向性因子水平,从而获得更好的定向增强水平。

1.2　基于 GSMDS 波束形成的室内定向语音增强方法

在常见的室内环境中,利用端射线阵 GSMDS 超指向性波束形成算法进行定向语音增强时,一方面混响(REVERB)干扰会导致增强得到的语音信号的语音质量下降;另一方面 GSMDS 波束形成的指向性因子随着频率发生改变,导致增强语音信号在不同时刻的增益水平产生明显差异,造成响度不稳定问题,语音质量下降。针对上述问题,本文提出使用基于加权预测误差(weighted prediction error, WPE)去混响和结合响度控制的归一化最小均方误差方法[16](normalized least mean square, NLMS)自动增益调节算法的 GSMDS 波束形成室内定向语音增强方法,前者用于消除混响干扰,后者用于控制 GSMDS 波束形成增强语音的响度稳定。下面对本室内定向语音增强方法的 WPE 去混响算法和 NLMS 自动增益控制算法进行介绍。

首先在 GSMDS 超指向性波束形成增强前加入去混响处理,消除混响的干扰。WPE 的去混响算法[17-19]利用多通道线性预测思想,通过已接收信号构造一个自回归系数矩阵估计晚期混响,再从接收信号中除去晚期混响,即可得到所需的直达声(含早期混响)。

WPE 去混响算法的处理流程图如图 3 所示。

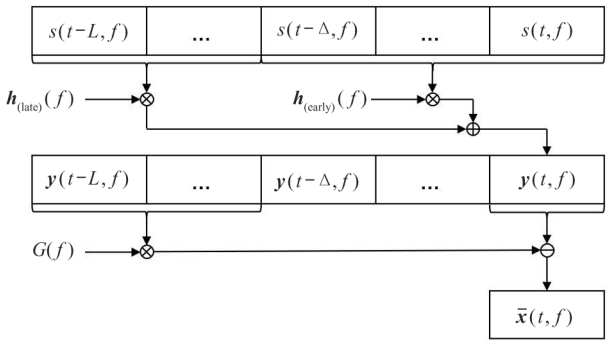


图 3　WPE 算法处理示意图

其中,$s(t-L,f),\cdots,s(t-\Delta,f),\cdots,s(t,f)$ 为纯净语音信号,t 为时间索引,f 为频率索引,L 为房间脉冲响应长度,Δ 为预测延迟,用于分割直达声(包含

早期混响)与晚期混响。$\boldsymbol{y}(t-L,f),\cdots,\boldsymbol{y}(t-\Delta,f),\cdots,\boldsymbol{y}(t,f)$ 为麦克风接收信号的时频域表示，$\boldsymbol{y}(t,f)$ 为 $M\times 1$ 列向量，M 为麦克风数量。WPE 算法通过构造一个混响预测滤波器 $G(f)$（长度为 $J=L-\Delta$），将其与麦克风接收信号 $\boldsymbol{y}(t-L,f),\cdots,\boldsymbol{y}(t-L+1,f),\cdots,\boldsymbol{y}(t-\Delta-1,f)$ 卷积来估计晚期混响部分，$y(t,f)$ 中减去此部分即可得到直达声部分。

WPE 去混响算法首先由接收语音信号 $\boldsymbol{y}(t,f)$ 初始化早期混响和直达声信号 $\bar{\boldsymbol{x}}(t,f)$：

$$\bar{\boldsymbol{x}}(t,f)=\boldsymbol{y}(t,f) \tag{8}$$

然后由 $\bar{\boldsymbol{x}}(t,f)$ 计算初始方差 $\hat{\sigma}_{t,f}^2=\max\{|\bar{\boldsymbol{x}}(t,f)|^2,\varepsilon_f\}$（$\varepsilon_f$ 是设置的方差的下界），根据初始方差 $\hat{\sigma}_{t,f}^2$ 计算混响预测滤波器 $G(f)$，$\hat{G}(f)$ 计算式为

$$\hat{G}(f)=\left(\sum_t \frac{\bar{\boldsymbol{x}}_{t-\Delta,f}\bar{\boldsymbol{x}}_{t-\Delta,f}^{\mathrm{H}}}{\hat{\sigma}_{t,f}^2}\right)^{+}\sum_t \frac{\bar{\boldsymbol{x}}_{t-\Delta,f}(\boldsymbol{x}_{t,f}^{(1)})^{*}}{\hat{\sigma}_{t,f}^2} \tag{9}$$

再由混响预测滤波器 $\hat{G}(f)$ 计算直达声信号 $\hat{\boldsymbol{x}}^{(e)}(t,f)$：

$$\bar{\boldsymbol{x}}^{(e)}(t,f)=\bar{\boldsymbol{x}}_{t,f}^{(1)}-(\hat{G}(f))^{\mathrm{H}}\bar{\boldsymbol{x}}_{t-\Delta,f} \tag{10}$$

接着更新方差 $\hat{\sigma}_{t,f}^2$，重复计算式(9)～(10)，直到算法收敛，输出直达声信号 $\hat{\boldsymbol{x}}^{(e)}(t,f)$。

将 WPE 去混响语音信号 $\hat{\boldsymbol{x}}^{(e)}(t,f)$ 进行 GSMDS 超指向性波束形成语音增强，增强语音信号的能量明显提高，但是在听感上的某些时刻会出现爆音，响度过大，而在某些语音过渡段声音能量相比之前降低，过渡不自然，即出现语音信号前后响度不稳定问题，因此需要对增强语音信号进行响度控制。本文利用自动增益控制算法实现响度控制，此算法采用 NLMS，并通过在 NLMS 算法中引入一个响度参考量来有效控制语音信号的响度，输出平稳的语音信号。

结合响度控制的 NLMS 自动增益控制算法，首先将输入语音信号进行分组，计算每组输入语音信号 $x(n)$（n 为采样点）的功率参数 $P_x(n)$，利用线性加权方法计算：

$$P_x(n)=\alpha P_x(n-1)+(1-\alpha)x^2(n) \tag{11}$$

其中，α 称为平滑因子。α 越大，前一时刻的影响越大，功率变化越平缓；α 越小，功率变化越大。为了保持信号的原有波动特性而不至于剧烈抖动，一般选择 $0<\alpha<1$。

然后判断 $P_x(n)$ 是否过小，若小于下界值 $P_{\min}$（如 $P_{\min}=0.001$），此时为静音点，令 $g(n)=g(n-1)$；若大于下界值 $P_{\min}$，增益 $g(n)$ 计算式为

$$g(n)=g(n-1)(1+\mu(n)P_x(n)(P_V-P_y(n-1))) \tag{12}$$

其中 $\mu(n)$ 称为归一化步长因子，$\mu(n)$ 的计算式为

$$\mu(n)=\frac{\mu_0}{\xi+P_x(n)} \tag{13}$$

其中 μ_0 对自动增益控制算法的收敛速率和稳定性有影响，μ_0 越大，算法收敛加快，但是算法收敛的稳定性变差；μ_0 越小，算法收敛减慢，而算法收敛的稳定性变好。μ_0 值一般取为 0.0～0.1 之间的常数。ξ 的大小也为 0.0～0.1 之间的常数，用来防止分母过小。P_V 为响度参考量，结合文献[20]和实验，将 $\frac{1}{N}\sum_{n=1}^{N}x^2(n)$（$N$ 为语音段采样总点数）这一参数作为响度参考量引入 NLMS 算法中，来稳定整体信号响度。$P_V-P_y(n-1)$ 作为 NLMS 算法中误差 $e(n)$，$g(n-1)\mu(n)P_x(n)$ 作为误差 $e(n)$ 的迭代步长。

利用求得的增益系数，计算输出信号 $y(n)$，即：

$$y(n)=g(n)x(n) \tag{14}$$

再根据输出信号计算输出信号的功率参数 $P_y(n)$，$P_y(n)$ 的计算式与 $P_x(n)$ 的计算式(11)相同。

对每组语音信号进行上述步骤处理，再将各组语音信号拼接得到整段语音信号，最后输出增益调节后的语音信号，即室内定向语音增强处理后输出语音信号。

综上所述，基于 GSMDS 超指向性波束形成的室内定向语音增强方法分为 3 部分：WPE 去混响、GSMDS 超指向性波束形成和 NLMS 自动增益控制，如图 4 所示。

图 4　基于 GSMDS 超指向性波束形成的室内定向语音增强方法流程图

2 仿真分析

为了分析基于 GSMDS 超指向性波束形成的室内定向语音增强方法的适用性,本文采用数值仿真方法进行计算。设计一个 4 元均匀线阵,阵元间距 d 根据语音信号的主要频率范围 100 ~ 4 000 Hz 和指向性因子大于 1 的要求设计为 0.04 m。空气声速 c 为 343 m · s^{-1},入射信号选取为一单频信号,期望入射角度为 0 °(GSMDS 算法为端射方向,CBF、MVDR 算法为垂射方向)。为比较 CBF、MVDR、GSMDS 波束形成算法在不同频率范围对不同角度的响应能力差异,在 100 ~ 4 000 Hz 频率范围改变单频信号的频率,计算 4 元线阵 3 种波束形成算法的波束响应。图 5 所示为 CBF、MVDR 和 GSMDS 算法的波束响应三维图。

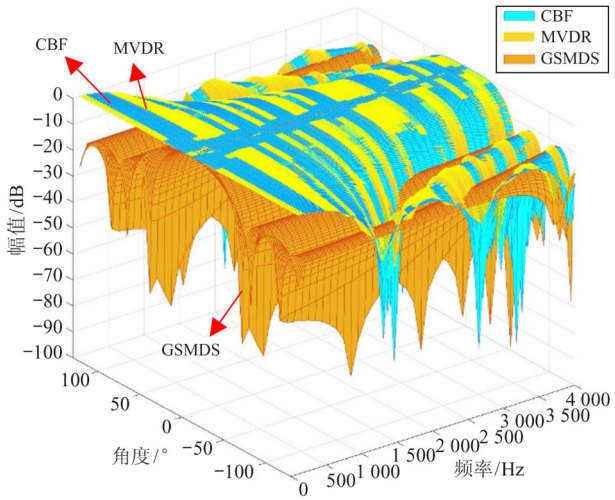


图 5 CBF、MVDR、GSMDS 波束响应三维图

由图 5 可知,CBF 和 MVDR 算法在 100 ~ 4 000 Hz 频率范围内在期望方向的主瓣宽度过大,随着频率的降低波束主瓣宽度更大。而 GSMDS 波束形成算法相比于经典的 CBF、MVDR 波束形成算法,在该频率范围内的期望方向得到的波束主瓣宽度显著更小,且旁瓣水平低,对期望方向信号的分辨力高,响应能力强。并且在各频率范围内主瓣宽度和旁瓣水平能维持不变,表明 GSMDS 波束形成算法在各频率范围内均能在端射方向获得高响应能力,具有频率不变性。综上所述,GSMDS 波束形成算法能在不同频率范围内的端射方向获得更强的波束响应能力,实现超指向性。

随后将该端射线阵应用到室内定向语音增强中。假设有一长方体房间尺寸为 8.6 m × 7.8 m × 4.0 m,端射线阵的 4 个麦克风位置坐标为(4.3 m, 3.9 m, h_i),i = 1 ~ 4, 其中 h_i = 3.61 m + 0.04 m × i,依次命名为麦克风 01 ~ 04,室内混响时间设置为 0.5 s。选取一段时长 5 s 的纯净中文女声音频[21],在 9 个声源位置点 $\boldsymbol{r}_s$ = (4.3 m, y_j, 1.65 m) (j = 1 ~ 9) 轮流播放,其中 y_j = 1.4 m + 0.5 m × j,依次命名为声源位置点 1 ~ 9,将声源位置点与麦克风 01 连线与线阵所在直线方向的夹角记作 θ($\theta \in [-45°, 45°]$)。信号采样频率为 16 kHz,信号分帧帧长为 0.032 s,窗函数为矩形窗。

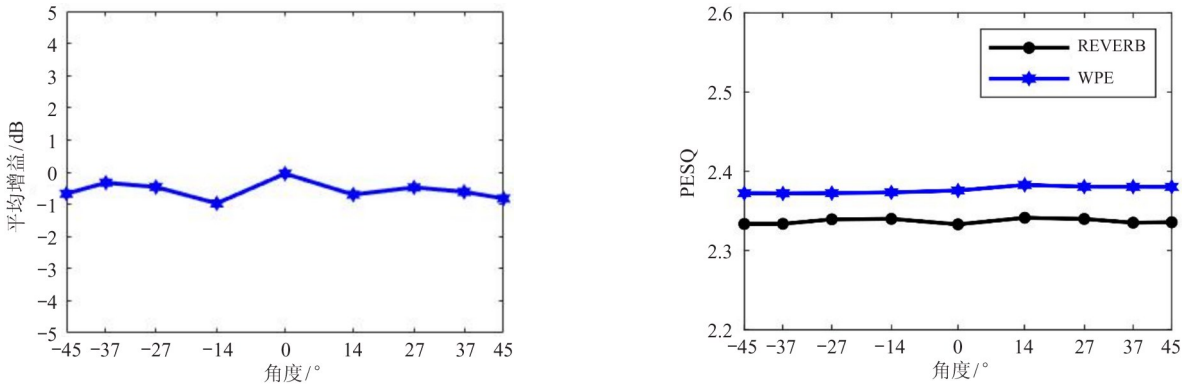
首先采用 Image 法[22]进行混响环境仿真,获得接收的混响语音信号;再对接收语音信号进行 WPE 去混响处理。语音信号经过每一步骤的处理都会发生变化,其获得的平均增益的计算式如式(15)所示。

$$G = 20\log_{10}\left(\frac{\bar{A}_{\text{out}}}{\bar{A}_{\text{in}}}\right) \tag{15}$$

其中,$\bar{A}_{\text{out}}$ 为输出语音信号的平均幅值;$\bar{A}_{\text{in}}$ 为输入语音信号的平均幅值;G 为平均增益,单位为 dB。由图 6(a) WPE 去混响平均增益值折线图可知,WPE 去混响处理得到的平均增益水平在 -1 ~ 0 之间,表明 WPE 去混响后语音信号的能量略减小,这部分能量减小是消除混响造成的。

同时,语音信号的语音质量以感知语音质量评价(perceptual evaluation of speech quality, PESQ)值[23-24]作为衡量指标,PESQ 值越大代表语音质量越高。由图 6(b)所示 WPE 去混响前后的语音信号 PESQ 值折线图可知,WPE 去混响后语音信号的 PESQ 值相比于原接收混响语音信号在各个方向均有所提升,表明 WPE 去混响后语音信号的语音质量得到提升。

在 WPE 去混响后,进行 GSMDS 波束形成增强计算。GSMDS 波束形成增强前后语音信号的时域波形对比图如图 7 所示。从中可知,GSMDS 波束形成增强后语音信号的幅值显著增大,幅值增大约10



(a)平均增益值折线图　　(b)PESQ 值折线图

图 6　不同方向各声源点 WPE 去混响语音信号的平均增益值折线图和 PESQ 值折线图

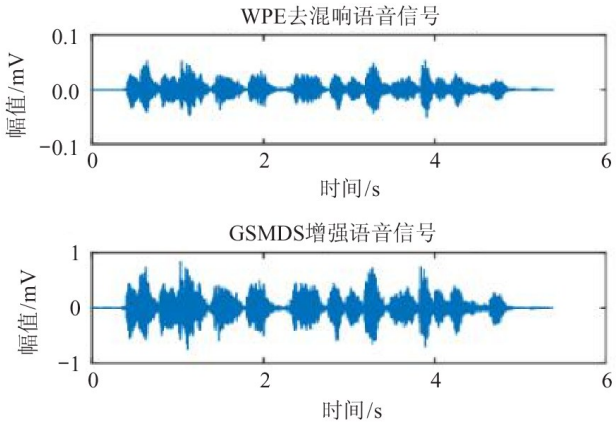
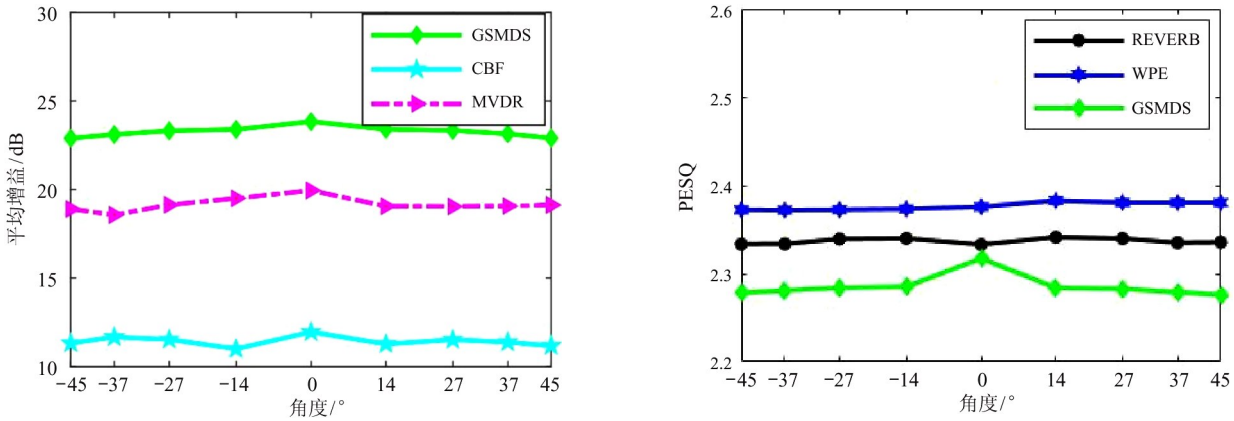


图 7　GSMDS 波束形成增强前后语音信号的时域波形对比图

倍。不同方向各声源点经 GSMDS 波束形成增强后的语音信号的平均增益值折线图如图 8(a)所示。从中可知,增强后语音信号的平均增益水平为 23.20 dB,相比于增强前的语音信号提高了 23.80 dB,同时端射方向的平均增益最大,最大平均增益为 23.83 dB。CBF 波束形成增强得到平均增益水平为 11.40 dB,只有 GSMDS 波束形成增强算法的一半左右,MVDR 波束形成增强得到平均增益水平为 19.14 dB,比 GSMDS 增强的增益水平低 4.06 dB。由此可知 GSMDS 相比于 CBF、MVDR 算法能获得更高平均增益水平,且对端射方向的语音信号响应能力更强,即定向增强的效果更显著。

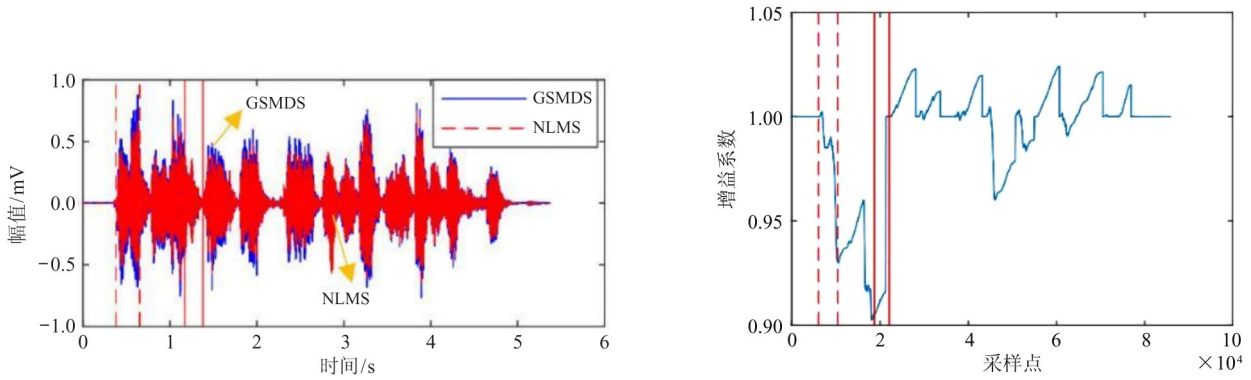


(a)平均增益值折线图　　(b)PESQ 值折线图

图 8　不同方向各声源点 GSMDS 波束形成增强语音信号平均增益值折线图和 PESQ 值折线图

图 8(b)所示为不同方向各声源点 GSMDS 波束形成增强前后语音信号的 PESQ 值折线图。然而,GSMDS 波束形成增强后的语音信号的 PESQ 值显著降低,且相比于原接收混响语音信号的 PESQ 值更低,其原因可归于 GSMDS 波束形成算法的指向性因子随频率发生改变的性质。结合图 2 分析可

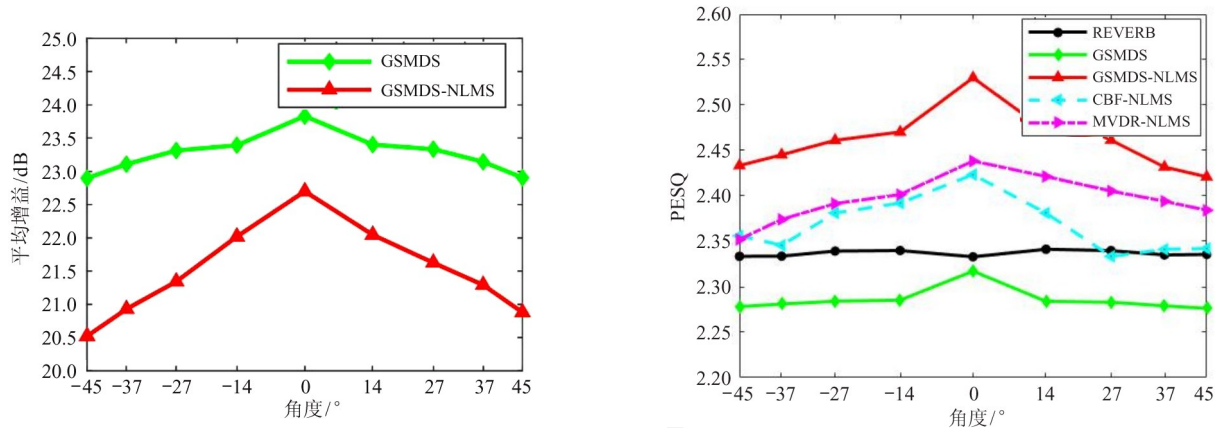
知,GSMDS 波束形成的指向性因子在低频段 100 ~ 500 Hz 远大于高频段 1 500 ~ 3 400 Hz,而语音信号为非平稳信号,不同时刻频率不断变化,导致 GSMDS 波束形成聚集得到语音信号在不同时刻的能量产生了显著差异,致使增强后语音信号在不同时段响度存在差异,出现语音信号过大或过小、响度不稳定问题。针对这一问题,本文提出用 NLMS 自动增益控制算法维持语音信号的响度稳定。端射方向 NLMS 自动增益调节语音信号时域波形图和不同时刻采样点的增益系数变化图如图 9 所示。对于语音信号中静音点,如语音开头段和语音结尾段的静音段不进行增益调节,增益系数设置为 1。在语音信号幅值突然变大时,采样点 6 080 处(虚线之间部分),NLMS 增益调节算法能有效设置较小的增益系数,减小幅值,防止幅值过大引起的响度过大;在语音信号幅值突然变小时,在采样点 18 720 处(实线之间部分),NLMS 增益调节算法提高增益,增强响度,防止幅值突然减小引起的响度过小,由此解决 GSMDS 波束形成输出的语音信号在不同时段的响度不稳定问题,从而提升语音质量。



(a)语音信号时域波形图　　(b)增益系数变化图

图 9　NLMS 自动增益调节语音信号时域波形图和增益系数变化图

图 10(a)为不同方向各声源点 NLMS 自动增益调节语音信号的平均增益值折线图。从中可知,NLMS 增益调节后各方向语音信号的平均增益均略有降低,平均增益水平为 21.50 dB,相比于 CBF 波束形成增强方法提高 10.10 dB,且端射方向平均增益最大,随方向角度增大呈对称性减小趋势,其中最大平均增益为 22.70 dB。图 10(b)为不同方向各声源点 NLMS 自动增益调节后语音信号的 PESQ 值。从中可知,NLMS 增益调节后的语音信号的 PESQ 值相比于GSMDS增强语音信号的PESQ值得到明显



(a)平均增益值折线图　　(b)PESQ 值折线图

图 10　不同方向各声源点 NLMS 自动增益调节语音信号的平均增益值折线图和 PESQ 值折线图

提高,相比于增益调节前提高了平均 0.22,且端射方向的 PESQ 值最大,为 2.53,相比于原接收混响语音信号提高了 0.20。同时,与 CBF 和 MVDR 算法进行对比,本方法(GSMDS-NLMS)的平均 PESQ 水平更高,比 CBF 算法平均高 0.11,比 MVDR 平均高 0.07。结果表明 NLMS 自动增益控制能有效稳定响度,解决 GSMDS 波束形成增强后语音信号的响度不平衡问题,从而提高语音质量水平。

因此,结合 WPE 去混响和 NLMS 自动增益控制算法的 GSMDS 超指向性波束形成方法能完成室内定向语音的增强,既大幅提高增益水平,又能提高语音质量水平。

3　实验验证

为验证基于 GSMDS 波束形成的室内定向语音增强方法在实际房间中的语音增强效果,选择在一个经过装修的真实房间内进行定向语音增强实验。房间的平面尺寸图如图 11 所示,其中位置点 1 ~9 为选定声源位置点,麦克风 01 为均匀线阵靠近声源位置的第一个麦克风,将麦克风 01 与各声源点连线的夹角记作 θ,$\theta \in [-22°, 22°]$。实验房间内设备实物图如图 12 所示,其中麦克风选择为 B&K 4966 型传声器,数采卡型号为 3 160-A-042,喇叭为佳讯 MTD-350 中音喇叭。利用气球爆破和脉冲响应积分法[25]测得此房间的混响时间 RT60 为 0.514 s。麦克风阵列为4阵元均匀线阵,阵元间距为0.040m,

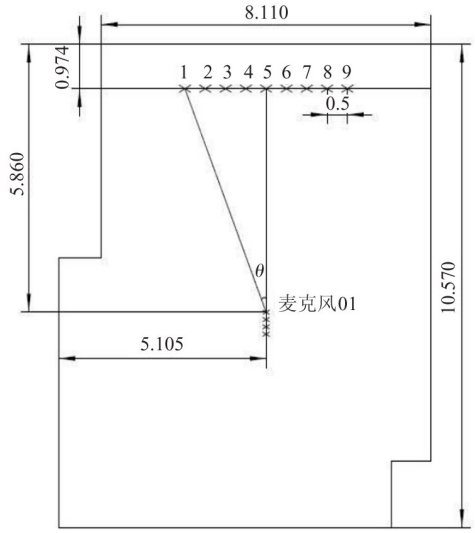


图 11　房间的平面尺寸图(m)

图 12　实验房间及设备实物图

利用水平尺和激光水平仪调整阵列位置使其在水平面上且朝向位置 5 方向。逐一选取位置 1 ~9(位置 5 位于端射方向)为喇叭声源点,喇叭离地面高度不变,为 1.652 m,重复播放同一段中文女声音频(与仿真分析中音频相同)。然后使用麦克风采集信号,采样频率设置为 131 072 Hz,信号采样时长设置为 10.000 s。采集后的各阵元信号根据有效部分截断为 6.000 s,信号处理帧长设置为 0.032 s,后续为方便计算 PESQ 值,采集信号降采样处理到采集频率16 000 Hz。

根据第 2 节仿真分析介绍的基于 GSMDS 波束形成的室内定向语音增强方法,首先对接收信号进行 WPE 去混响处理。由仿真分析可知去混响处理能得到可预期的较小的负增益和 PESQ 值提升,所以此处不做详细介绍。

然后进行 GSMDS 波束形成增强计算。不同方向各声源点 GSMDS 波束形成增强语音信号的平均增益值折线图如图 13 所示。从图中可知,各方向平

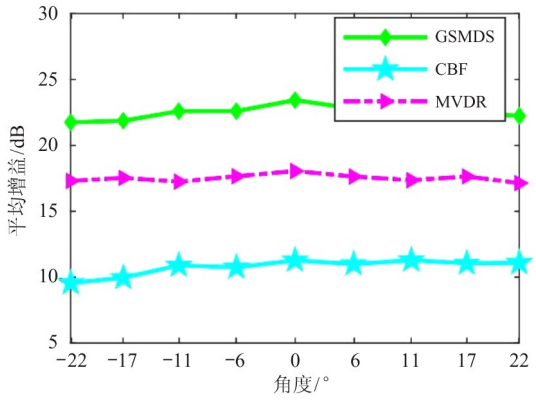
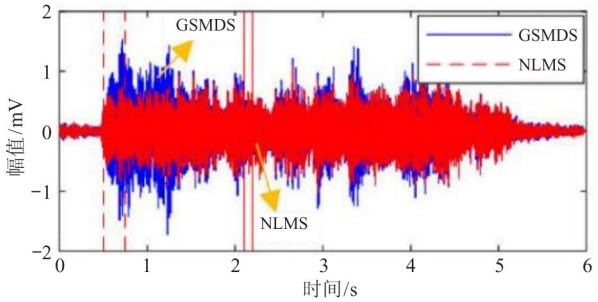


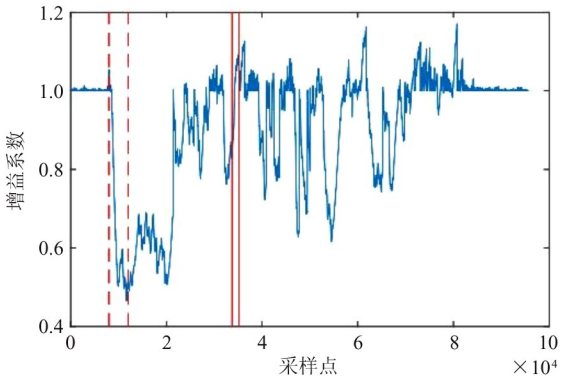
图 13　不同方向各声源点 GSMDS 波束形成增强语音信号的平均增益值折线图

均增益水平均大于 21.74 dB,且端射方向的平均增益最大,为 23.42 dB,相比于 CBF 波束形成增强平均增益水平 10.75 dB 提高了 10.99 dB,超过了 102.23%,同时比 MVDR 增强平均增益水平 17.50 dB 高出 4.98 dB。

随后选取端射方向位置 5 的 GSMDS 波束形成增强语音信号进行 NLMS 自动增益调节。图 14(a)和(b)为端射方向 NLMS 自动增益调节语音信号时域波形图和增益系数变化图。从中可知,在语音信号幅值突然增大处,如在采样点 8 000 处开始,对应时刻为 0.5 s(虚线开始处),该算法能迅速设置较小的增益系数,减小响度;而在语音信号幅值突然变小处,如采样点 33 600 处,对应时刻 2.1 s(实线开始处),该算法能迅速增大增益系数、增大响度。



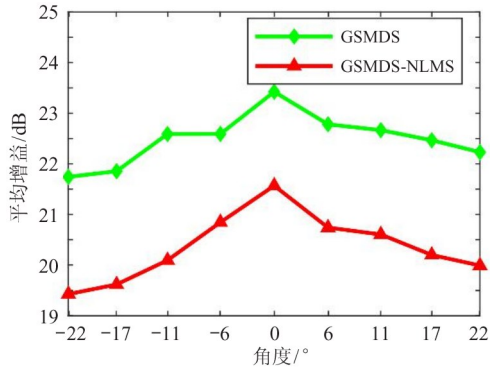
(a)时域波形图



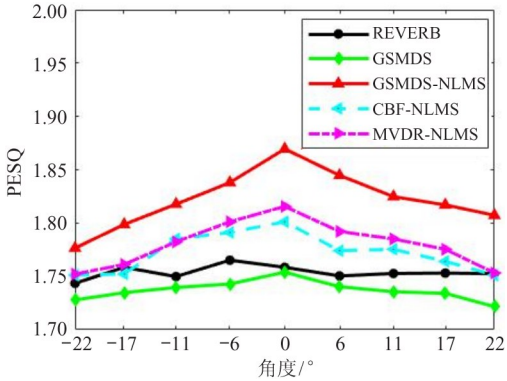
(b)增益系数变化图

图 14　NLMS 自动增益调节语音信号时域波形图和增益系数变化图

图 15(a)为 NLMS 自动增益调节前后语音信号的平均增益水平图,从中对比发现,NLMS 自动增益调节降低了增益水平,平均下降了 2.14 dB,但仍相比于 CBF 增强语音信号的增益提高了 9.59 dB,且端射方向的语音信号平均增益仍最大,最大平均增益为 21.57 dB。图 15(b)为 NLMS 自动增益调节前后语音信号的 PESQ 值折线图,从中对比发现,NLMS 自动增益调节后的语音信号的 PESQ 值水平提高,平均提高了 0.08,NLMS 自动增益调节方法能有效提高语音质量,解决响度不稳定造成的语音质量下降问题。其中,在端射方向进行室内语音增强得到语音信号的 PESQ 值最大,PESQ 值为 1.87,相比于原接收混响语音信号的 PESQ 值提高 0.11。同时,与 CBF 和 MVDR 算法进行对比,本方法(GSMDS-NLMS)的平均 PESQ 水平更高,比 CBF 算法平均高 0.06,比 MVDR 平均高 0.03。由此可知,基于 GSMDS 波束形成的室内定向语音增强方法能有效提升不同方向语音信号的语音质量水平,且在端射方向语音质量最高,此方法能在端射方向获得高增益、高语音质量的语音信号。



(a)平均增益值折线图



(b)PESQ 值折线图

图 15　不同方向各声源点 NLMS 自动增益调节语音信号的平均增益值折线图和 PESQ 值折线图

4 结 论

本文提出一种结合 WPE 去混响前处理与 NLMS 自动增益控制的 GSMDS 超指向性波束形成室内定向语音增强方法。室内定向语音增强方法首先通过 WPE 去混响算法有效减弱混响对 GSMDS 波束形成的影响,然后进行 GSMDS 波束形成进行定向增强,最后通过结合响度控制的 NLMS 自动增益控制算法进行实时增益调节,得到较高增益和较优语音质量的语音信号。该算法的语音增强效果通过 4 阵元端射线阵的数值仿真和实地实验得到了验证,实现了对室内目标说话人的定向语音增强,相比于 CBF 和 MVDR 波束形成语音增强具有更好的性能。但对其他方向存在噪声源的问题仍缺少研究,下一步将对此室内定向语音增强方法在干扰噪声环境下的语音增强进行研究,进一步拓展该方法的适用性和实用性。

参考文献

[1] BENESTY J, CHEN J, HUANG Y. Microphone array signal processing[M]. Berlin: Springer-Verlag, 2008: 11-15.

[2] 季浩然, 马晓川, 张舒皓, 等. 功率受限的常规波束形成后处理拟合方法[J]. 声学学报, 2020,45(1):1-14.

[3] 杨洋, 杨震. 基于图信号处理的 MVDR 波束形成多通道语音增强[J]. 南京邮电大学学报(自然科学版), 2021,41(6):35-40.

[4] FROST O L. An algorithm for linearly constrained adaptive array processing[J]. Proceedings of the IEEE, 1972,60(8):926-935.

[5] GRIFFITHS L, JIM C. An alternative approach to linearly constrained adaptive beamforming[J]. IEEE Transactions on Antennas and Propagation, 1982,30(1):27-34.

[6] GANNOT S, BURSHTEIN D, WEINSTEIN E. Signal enhancement using beamforming and nonstationarity with applications to speech[J]. IEEE Transactions on Signal Processing, 2001,49(8):1614-1626.

[7] 马子骥, 倪忠, 余旭. 基于 TF-GSC 的多通道后置滤波语音增强算法[J]. 传感器与微系统, 2018,37(5): 105-107, 111.

[8] 张英豪, 汪勇, 杨益新. 障板表面圆环阵频率不变响应波束形成方法研究[J]. 哈尔滨工程大学学报, 2021,42(9):1249-1255.

[9] MA Y, YANG Y, HE Z, et al. Theoretical and practical solutions for high-order superdirectivity of circular sensor arrays[J]. IEEE Transactions on Industrial Electronics, 2013,60(1):203-209.

[10] 汪勇, 杨益新. 圆环形传感器阵列超指向性频率不变响应波束优化方法[J]. 声学学报. 2018,43(4):620-632.

[11] WANG Y, YANG Y, MA Y. Arbitrary-order superdirectivity of circular sensor arrays[J]. The Journal of the Acoustical Society of America, 2020,148(3):1425.

[12] 厉剑, 柯雨璇, 彭任华, 等. 球谐域类正则化宽带超指向性波束形成算法[J]. 声学学报. 2020,45(2): 145-160.

[13] WANG Y, CHEN J, BENESTY J, et al. Binaural heterophasic superdirective beamforming[J]. Sensors, 2020,21(1):74.

[14] WANG Y, YANG Y, HE Z, et al. A general superdirectivity model for arbitrary sensor arrays[J]. EURASIP Journal on Advances in Signal Processing, 2015(1):1-16.

[15] WANG Y, YANG Y. Superdirectivity solutions of circular arrays with acoustic particle velocity sensors[J]. Digital Signal Processing, 2020,106:102811.

[16] 许芳芳, 何培宇, 潘帆, 等. 用于语音处理的变步长 LMS 自适应增益控制新方案[J]. 信号处理, 2021,37(3):456-462.

[17] NAKATANI T, YOSHIOKA T, KINOSHITA K, et al. Speech dereverberation based on variance-normalized delayed linear prediction[J]. IEEE Transactions on Audio, Speech, and Language Processing, 2010,18(7):1717-1731.

[18] YOSHIOKA T, NAKATANI T. Generalization of multichannel linear prediction methods for blind mimo impulse response shortening[J]. IEEE Transactions on Audio, Speech, and Language Processing, 2012,20(10):2707-2720.

[19] 张国昌, 吴鸣, 杨军. 基于加权预测误差算法的声源定位方法[J]. 网络新媒体技术, 2021,10(1):12-19.

[20] 吴振宙. 音频信号 AGC 及响度控制算法的研究[D].

南京: 南京大学, 2014.

[21] SHI Y, BU H, XU X, et al. AISHELL-3: a multi-speaker mandarin TTS corpus and the baselines[EB/OL]. (2021-04-22)[2023-03-15]. https://doi.taixueshu.com/foreign/arXiv201011567.

[22] ALLEN J B, BERKLEY D A. Image method for efficiently simulating small-room acoustics[J]. The Journal of the Acoustical Society of America, 1979, 65(4): 943-950.

[23] 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.

[24] 王霞, 马俊晖, 王光艳, 等. 面罩语音质量评价算法适用性研究[J]. 计算机工程与应用, 2017, 53(19): 114-117, 156.

[25] 郑凯桐, 郑成诗, 张玉龙, 等. 房间脉冲响应模拟法及混响时间盲估计应用[J]. 应用声学, 2023, 42(2): 260-268.

An indoor directional speech enhancement method based on GSMDS superdirectivity beamforming

LEI Shiwei*, LIU Zubin*,**

(*Key Laboratory of Special Equipment Manufacturing and Advanced Processing Technology of Ministry of Education, Zhejiang University of Technology, Hangzhou 310014)

(**Taizhou Research Institute, Zhejiang University of Technology, Taizhou 318014)

Abstract

Aiming at the problems of reverberation interference and poor directivity of the main lobe during indoor directional speech enhancement, an indoor directional speech enhancement method for superdirectivity beamforming based on Gram-Schmidt mode-beam decomposition and synthesis (GSMDS) is proposed. Firstly, the weighted prediction error is used for dereverberation, then GSMDS superdirectivity beamforming is used for directional target speech enhancement, and then the normalized minimum mean square error automatic gain control algorithm is used for gain adjustment of the enhanced speech signal to control the loudness stability, so as to complete directional enhancement of the target speech. Among them, the gain regulation step also compares the frequency band equalization control algorithm. The numerical simulation results show that compared with conventional beamforming(CBF) method, the gain of the proposed method is improved by 10.1 dB and the maximum average gain is up to 22.7 dB. Meanwhile, the perceptual evaluation of speech quality(PESQ) value of the speech quality in the endfire direction is improved by 0.2, compared with the original received reverb speech signal. The four-element endfire array experiment is carried out in a decorated room. The experimental results show that this method can obtain good directional enhancement effect in a decorated room.

Key words: indoor directional speech enhancement, Gram-Schmidt mode-beam decomposition and synthesis (GSMDS), superdirectivity, automatic gain control, endfire array