ISSN 1004-9037,CODEN SCYCE4
Journal of Data Acquisition and Processing Vol. 39,No. 3,May2024,pp. 649－658
DOI:10. 16337/j. 1004-9037. 2024. 03. 012

http://sjcj. nuaa. edu. cn
E-mail:sjcj@nuaa. edu. cn
Tel/Fax:＋86-025-84892742

基于SDW-MMSE的广义特征值稳健波束形成方法

李海龙[1,2]，杨　飞[1,2]，杨诗童[1,2]，路晓庆[1,2]

(1. 武汉大学电气与自动化学院,武汉 430072;2. 武汉大学综合能源电力装备及系统安全湖北省重点实验室,武汉 430072)

摘　要：最大输出信噪比(Signal-to-noise ratio，SNR)准则下，广义特征值(Generalized eigenvalue，GEV)波束形成存在复系数难以控制的问题,在复杂的声学环境中容易导致输出信号严重失真。针对复系数估计问题,本文提出一种基于最小均方误差(Minimum mean square error，MMSE)的复系数估计方法,并通过引入语音失真权重因子(Speech distortion weight，SDW),调节降噪效果和语音失真之间的权重关系,进而提出了基于SDW-MMSE的广义特征值稳健波束形成方法。通过最大似然法估计目标信号和噪音信号的功率谱,进而求解主广义特征向量。进一步基于SDW-MMSE估计复系数,将复系数与主广义特征向量相结合,从而得到基于SDW-MMSE的广义特征值稳健波束形成滤波向量。仿真实验结果表明,本文提出的波束形成方法可有效消除相干噪声和非相干噪声,具有输出信噪比高、语音失真少等稳健性能。

关键词：语音增强;广义特征值波束形成;最小均方误差;语音失真权重;最大似然参数估计

中图分类号：TN912.35　　**文献标志码**:A

Generalized Eigenvalue Robust Beamforming Based on SDW-MMSE

LI Hailong[1,2]，YANG Fei[1,2]，YANG Shitong[1,2]，LU Xiaoqing[1,2]

(1. School of Electrical Engineering and Automation，Wuhan University，Wuhan 430072，China;2. Hubei Key Laboratory of Power Equipment & System Security for Integrated Energy，Wuhan University，Wuhan 430072，China)

Abstract：Under the criterion of maximum output signal-to-noise ratio (SNR)，the problem of difficult control of complex-valued coefficients in generalized eigenvalue (GEV) beamforming is encountered，and severe distortion of the output signal can be caused in complex acoustic environments. To address the issue of complex-valued coefficient estimation，a complex-valued coefficient estimation method based on minimum mean square error (MMSE) is proposed in this paper. By introducing a speech distortion weight factor (SDW)，the weight relationship between noise reduction and speech distortion is adjusted，thereby proposing a method for generalized eigenvalue robust beamforming based on SDW-MMSE. The power spectra of the target and noise signals are estimated using maximum likelihood method，and the main generalized eigenvectors are then determined. Furthermore，the complex-valued coefficients are estimated ，and the complex coefficients are combined with the principal generalized eigenvector to obtain the generalized eigenvalue robust beamforming filter vector based on SDW-MMSE. Through simulation experiments，it is demonstrated that the proposed beamforming method effectively eliminates coherent and

基金项目:国家自然科学基金(52377155)。
收稿日期:2023-06-20;**修订日期**:2023-11-28

incoherent noise, and exhibits robust performance with high output SNR and low speech distortion.

Key words: speech enhancement; generalized eigenvalue beamforming; minimum mean square error; speech distortion weight; maximum likelihood parameter estimation

引　言

语音增强技术是语音研究领域的重要分支,一直是该领域研究的热点问题。谱减法[1]、维纳滤波[2]等单通道算法往往只利用了时间信息,而基于阵列信号处理的多通道自适应波束形成方法能同时利用时间和空间信息,在时频域增强特定的信号和实现特定的需求,去除噪声、干扰和混响,从而提高语音信号的信噪比和可懂度[3]。多通道的语音增强方法大体分为盲源分离[4]和波束形成。盲源分离旨在分离语音中的所有相关源,而波束形成只增强特定的语音源,将其他语音源认定为需要消除的干扰。典型的波束形成方法有最小方差无失真响应波束形成[5]、线性约束最小方差波束形成[6]和广义旁瓣相消波束形成[7]等。其中,广义旁瓣相消波束形成具有结构简单且鲁棒性强的特点,因此被广泛使用。Park等[8]提出了一种基于行列式的广义旁瓣相消器,该方法首先使用基于行列式的维纳滤波器对带噪信号进行滤波,然后通过基于行列式的自适应模式控制器来消除残余噪声,在多噪声和混响环境下达到了良好效果。

基于最大输出信噪比(Signal-to-noise ratio, SNR)准则的广义特征值(Generalized eigenvalue, GEV)波束形成是一种经典的统计最优波束形成方法。Asano等[9]首次提出在多通道语音增强中应用最大输出信噪比准则下的广义特征值波束形成方法,广义特征值波束形成方法可最大化输出信噪比,极易导致严重的语音失真,但该语音失真可以通过调节波束形成滤波向量中的复系数得到改善。Warsitz等[10]提出了基于盲解析归一化(Blind analytical normalization, BAN)算法,通过归一化主广义特征向量来估计复系数,使得波束形成滤波向量在空间响应上对相对传递函数具有单位增益,从而控制语音失真程度。Krueger等[11]将广义特征值分解方法应用于广义旁瓣相消波束形成结构中阻塞矩阵的构建,通过求解每个频域的广义特征值,实现自适应特征向量跟踪,从而加快了算法收敛速度。Huang等[12]研究和分析了单通道和多通道情况下基于最大输出信噪比准则的广义特征值波束形成方法,并对其二者在时域和短时傅里叶变换域下的降噪效果进行了比较。Pfeifenberger等[13]提出了基于相位感知归一化(Phase aware normalization, PAN)复系数估计的广义特征值波束形成方法,进行目标信号协方差矩阵的特征分解,求得其主特征向量以代替相对传递函数,并考虑相应的相位信息估计出复系数,从而得到滤波向量。上述提出的基于盲解析归一化和基于相位感知归一化复系数估计的广义特征值波束形成方法分别考虑了幅值和相位信息,改进了广义特征值波束形成方法,提高了输出语音信号的信噪比和减少了语音失真。但仍存在一些问题,如在多噪声和混响的环境下降噪效果欠佳、降噪效果和语音失真之间的权重关系不明确等。

本文提出了一种基于最小均方误差(Minimum mean square error, MMSE)的广义特征值波束形成方法,该方法通过最小化输出语音信号与目标信号之间的均方误差来确定最优复系数,从而提高降噪效果,降低语音失真程度,同时在复杂的声学环境下还能具有良好的稳健性。此外,本文引入了语音失真权重因子(Speech distortion weight, SDW)[14],将SDW与MMSE相结合得到基于SDW-MMSE的广义特征值波束形成方法,进一步调节了降噪效果和语音失真之间的权重关系。仿真实验结果表明,本文提出的方法极大提高了输出语音信号的信噪比和语音质量,同时显著降低了语音失真程度。本文对降噪效果和语音失真的权衡进行综合考虑,从而为语音信号的增强提供了一种有效的波束形成方法。

1 信号模型

本文中麦克风阵列由M个麦克风按一定的拓扑结构排列而成,在存在相干噪声和非相干噪声的嘈杂环境下进行语音信号的采集,则麦克风阵列接收到的信号在时域中可表示为[15]

$$y_m(t)=a_m(t)*s(t)+n_m(t)\quad m=1,2,\cdots,M \tag{1}$$

式中:$y_m(t)$表示第m个麦克风接收到的信号;$a_m(t)$表示纯净语音信号到第m个麦克风的冲激响应函数;$*$表示卷积;$s(t)$表示纯净语音信号;$n_m(t)$表示第m个麦克风接收到的噪声信号,其表示相干噪声和非相干噪声经空间卷积后的结果。

对式(1)进行短时傅里叶变换,并表示为如下向量形式[16]

$$\boldsymbol{y}(k,l)=\boldsymbol{a}(k)S(k,l)+\boldsymbol{n}(k,l) \tag{2}$$

$$\boldsymbol{a}(k)=\begin{bmatrix}a_1(k) & a_2(k) & \cdots & a_m(k) & \cdots & a_M(k)\end{bmatrix}^{\mathrm{T}} \tag{3}$$

式中:k和l分别为短时傅里叶变换后的频率索引和帧索引;$\boldsymbol{y}(k,l)$、$\boldsymbol{n}(k,l)$分别表示$y_m(t)$、$n_m(t)$的短时傅里叶变换矢量;$\boldsymbol{a}(k)$表示纯净语音信号到麦克风阵列的传递函数矢量;$S(k,l)$表示$s(t)$的短时傅里叶变换;$a_m(k)$表示第m个麦克风的传递函数。

相对传递函数(Relative transfer function, RTF)[17]以一个麦克风为参考,将其余麦克风的传递函数转换为相对于参考麦克风的相对传递函数。因此,相较于传递函数,它所需的先验知识更少,在实际情况中更符合声音的实际传播模型且更容易获得。本文采用相对传递函数模型来描述语音传播模型,则接收的信号在短时傅里叶变换域中可表示为

$$\boldsymbol{y}(k,l)=\boldsymbol{h}(k)a_1(k)S(k,l)+\boldsymbol{n}(k,l) \tag{4}$$

$$\boldsymbol{h}(k)=\frac{\boldsymbol{a}(k)}{a_1(k)}=\begin{bmatrix}1 & \dfrac{a_2(k)}{a_1(k)} & \cdots & \dfrac{a_M(k)}{a_1(k)}\end{bmatrix}^{\mathrm{T}} \tag{5}$$

式中:$\boldsymbol{h}(k)$表示麦克风阵列相对于参考麦克风的相对传递函数矢量,目标信号是参考麦克风接收到的纯净语音信号$a_1(k)S(k,l)$。

2 广义特征值波束形成

以第1个麦克风为参考麦克风,则麦克风阵列接收到的信号又可表示为

$$\boldsymbol{y}(k,l)=\boldsymbol{x}(k,l)+\boldsymbol{n}(k,l)=\boldsymbol{h}(k)S_1(k,l)+\boldsymbol{n}(k,l) \tag{6}$$

$$\boldsymbol{x}(k,l)=\boldsymbol{h}(k)S_1(k,l) \tag{7}$$

$$S_1(k,l)=a_1(k,l)S(k,l) \tag{8}$$

式中:$\boldsymbol{x}(k,l)$表示麦克风阵列接收的纯净语音信号;$\boldsymbol{n}(k,l)$表示麦克风阵列接收的噪声信号;$S_1(k,l)$表示参考麦克风接收的目标语音信号。为简洁起见,后文中一律省略频率索引k和帧索引l。假设目标信号与噪声信号不相关,则麦克风阵列接收的语音信号协方差矩阵可以表示为

$$\boldsymbol{\Phi}_{yy}=E\left\{\boldsymbol{y}\boldsymbol{y}^{\mathrm{H}}\right\}=\boldsymbol{\Phi}_{xx}+\boldsymbol{\Phi}_{nn} \tag{9}$$

$$\boldsymbol{\Phi}_{xx}=E\left\{\boldsymbol{x}\boldsymbol{x}^{\mathrm{H}}\right\}=\boldsymbol{h}\phi_{\mathrm{S}_1}\boldsymbol{h}^{\mathrm{H}} \tag{10}$$

$$\boldsymbol{\Phi}_{nn}=E\left\{\boldsymbol{n}\boldsymbol{n}^{\mathrm{H}}\right\}=\phi_{\mathrm{n}}\boldsymbol{\Gamma} \tag{11}$$

式中:$E\{\cdot\}$和$(\cdot)^{\mathrm{H}}$分别表示期望和共轭转置;$\boldsymbol{\Phi}_{xx}$和$\boldsymbol{\Phi}_{nn}$分别表示纯净语音信号的协方差矩阵和噪声信号的协方差矩阵;ϕ_{S_1}和ϕ_{n}分别表示目标语音信号的功率谱和噪声信号的功率谱;$\boldsymbol{\Gamma}$表示噪声场的空间相干矩阵。

为恢复参考麦克风接收到的目标信号$S_1(k,l)$和抑制一系列相干和不相干噪声,使用波束形成的滤波向量对麦克风阵列接收信号进行加权求和,得到的估计语音信号可表示为

$$Z=\boldsymbol{w}^{\mathrm{H}}\boldsymbol{y}=\boldsymbol{w}^{\mathrm{H}}\boldsymbol{x}+\boldsymbol{w}^{\mathrm{H}}\boldsymbol{n} \tag{12}$$

$$\boldsymbol{w}=[w_1 \quad w_2 \quad \cdots \quad w_M]^{\mathrm{T}} \tag{13}$$

式中$\boldsymbol{w}$表示波束形成滤波向量。估计语音信号的功率谱可以表示为

$$\phi_Z=\boldsymbol{w}^{\mathrm{H}}\boldsymbol{\Phi}_{yy}\boldsymbol{w}=\boldsymbol{w}^{\mathrm{H}}\boldsymbol{\Phi}_{xx}\boldsymbol{w}+\boldsymbol{w}^{\mathrm{H}}\boldsymbol{\Phi}_{nn}\boldsymbol{w} \tag{14}$$

该波束形成方法的窄带输出信号信噪比可以表示为

$$\mathrm{SNR}_{\mathrm{O}}=\frac{\boldsymbol{w}^{\mathrm{H}}\boldsymbol{\Phi}_{xx}\boldsymbol{w}}{\boldsymbol{w}^{\mathrm{H}}\boldsymbol{\Phi}_{nn}\boldsymbol{w}}=\frac{\boldsymbol{w}^{\mathrm{H}}\boldsymbol{\Phi}_{yy}\boldsymbol{w}}{\boldsymbol{w}^{\mathrm{H}}\boldsymbol{\Phi}_{nn}\boldsymbol{w}}-1 \tag{15}$$

输出信噪比的大小代表着波束形成方法滤波性能的高低。显然,由式(15)可知,最大化输出信噪比问题可等效为一个广义瑞利商问题[10],从而可将最大输出信噪比准则下的滤波向量求解问题转化为一个广义特征值分解问题,即

$$\boldsymbol{\Phi}_{yy}\boldsymbol{w}=\lambda\boldsymbol{\Phi}_{nn}\boldsymbol{w} \tag{16}$$

式中$\boldsymbol{w}$表示广义特征值λ对应的广义特征向量。特别令$\lambda_{\max}$为最大广义特征值,其对应的广义特征向量用$\boldsymbol{w}_{\max}$表示,$\boldsymbol{w}_{\max}$又称为主广义特征向量。有广义瑞利商小于等于最大广义特征值,即

$$\frac{\boldsymbol{w}^{\mathrm{H}}\boldsymbol{\Phi}_{yy}\boldsymbol{w}}{\boldsymbol{w}^{\mathrm{H}}\boldsymbol{\Phi}_{nn}\boldsymbol{w}}\leqslant\lambda_{\max} \tag{17}$$

由式(17)可知,当把主广义特征向量作为滤波向量时,最大输出信噪比准则下广义特征值波束形成的输出信噪比可取得最大值。此时,该波束形成方法的输出信噪比为

$$\mathrm{SNR}_{\mathrm{O}}=\lambda_{\max}-1 \tag{18}$$

直接将主广义特征向量$\boldsymbol{w}_{\max}$作为波束形成的滤波向量,这使得该波束形成方法具有最大窄带输出信噪比,但会导致目标语音信号严重失真。为解决语音信号失真问题,通常在主广义特征向量前乘一个复系数,用该复系数降低目标语音信号失真程度并保持最大输出信噪比,则引入复系数的最大输出信噪比广义特征值波束形成滤波向量可表示为

$$\boldsymbol{w}_{\mathrm{GEV}}=G\boldsymbol{w}_{\max} \tag{19}$$

式中G为一个不为零的复系数。将式(19)代入式(15)中,可以发现,窄带输出信噪比不变且广义瑞丽商仍然取得最大值。

综上,在通过广义特征值分解求得主广义特征向量的前提下,广义特征值波束形成滤波向量的求解问题可转为复系数G的求解问题。

3 复系数估计方法

3.1 基于MMSE的复系数估计算法

复系数的估计重点在于降低输出信号的语音失真程度,使输出信号尽可能地等于或接近目标语音信号,而最小均方误差准则的思想是得到一个期望误差最小的估计值。通过将最小均方误差准则引入到复系数的估计中,在$\boldsymbol{w}_{\max}$已知的前提下,求得最小均方误差准则下最优的复系数G。即最小化输出语音信号与目标语音信号的均方误差,波束形成输出语音信号和目标语音信号的均方误差可以表示为

$$E\left\{\left|\boldsymbol{w}_{\mathrm{GEV}}^{\mathrm{H}}\boldsymbol{y}-S_1\right|^2\right\} \tag{20}$$

将式(19)代入式(20),对应的代价函数可表示为

$$J(G)=E\left\{\left|G^{*}\boldsymbol{w}_{\max}^{\mathrm{H}}\boldsymbol{y}-S_1\right|^{2}\right\} \tag{21}$$

式中$(\cdot)^{*}$表示复数的共轭,化简式(21)可以得到

$$J(G)=|G|^{2}\boldsymbol{w}_{\max}^{\mathrm{H}}\boldsymbol{\Phi}_{yy}\boldsymbol{w}_{\max}-G^{*}\boldsymbol{w}_{\max}^{\mathrm{H}}\boldsymbol{y}S_1^{*}-G\boldsymbol{y}^{\mathrm{H}}\boldsymbol{w}_{\max}S_1+\phi_{S_1} \tag{22}$$

当代价函数$J(G)$取最小值时的复系数,就是在最小均方误差准则下的最优复系数。即

$$G_{\mathrm{MMSE}}=\frac{\boldsymbol{w}_{\max}^{\mathrm{H}}\boldsymbol{h}\phi_{S_1}}{\boldsymbol{w}_{\max}^{\mathrm{H}}\boldsymbol{\Phi}_{yy}\boldsymbol{w}_{\max}} \tag{23}$$

由式(23)可得,基于MMSE的广义特征值波束形成滤波向量可表示为

$$\boldsymbol{w}_{\text{GEV-MMSE}}=G_{\mathrm{MMSE}}\boldsymbol{w}_{\max} \tag{24}$$

可以看出,基于MMSE复系数估计算法的公式推导过程简洁明了,能够最小化输出语音信号与目标语音信号的均方误差。

3.2 基于SDW-MMSE的复系数估计算法

基于MMSE的广义特征值波束形成方法能够降低噪声和改善语音失真,但降噪效果和语音失真之间的权重关系仍不明朗。为了更加明确清晰地调节降噪效果和语音失真之间的权重关系,同时受到文献[14]的启发,本文引入语音失真权重因子,提出了基于SDW-MMSE的复系数估计算法。最小化语音失真能量和残余噪声能量的加权和,则相应均方误差可表示为

$$E\left\{\left|\boldsymbol{w}_{\mathrm{GEV}}^{\mathrm{H}}\boldsymbol{x}-S_1\right|^{2}\right\}+\mu E\left\{\left|\boldsymbol{w}_{\mathrm{GEV}}^{\mathrm{H}}\boldsymbol{n}\right|^{2}\right\} \tag{25}$$

式中μ为语音失真权重因子,为均方误差中残余噪声能量的系数。当$\mu=1$时,式(25)则为基于MMSE的复系数估计。如果$\mu<1$,则以降低输出信噪比为代价来改善语音失真;如果$\mu>1$,则以增加语音失真程度为代价来提高输出信噪比。将式(19)代入到式(25)可得代价函数

$$J(G)=E\left\{\left|G^{*}\boldsymbol{w}_{\max}^{\mathrm{H}}\boldsymbol{x}-S_1\right|^{2}\right\}+\mu E\left\{\left|G^{*}\boldsymbol{w}_{\max}^{\mathrm{H}}\boldsymbol{n}\right|^{2}\right\} \tag{26}$$

化简代价函数式(26),可得

$$J(G)=|G|^{2}\boldsymbol{w}_{\max}^{\mathrm{H}}\boldsymbol{\Phi}_{xx}\boldsymbol{w}_{\max}-G^{*}\boldsymbol{w}_{\max}^{\mathrm{H}}\boldsymbol{x}S_1^{*}-G\boldsymbol{x}^{\mathrm{H}}\boldsymbol{w}_{\max}S_1+\phi_{S_1}+\mu|G|^{2}\boldsymbol{w}_{\max}^{\mathrm{H}}\boldsymbol{\Phi}_{nn}\boldsymbol{w}_{\max} \tag{27}$$

同理,当代价函数最小时,可得基于SDW-MMSE准则的复系数为

$$G_{\text{SDW-MMSE}}=\frac{\boldsymbol{w}_{\max}^{\mathrm{H}}\boldsymbol{h}\phi_{S_1}}{\boldsymbol{w}_{\max}^{\mathrm{H}}(\boldsymbol{\Phi}_{xx}+\mu\boldsymbol{\Phi}_{nn})\boldsymbol{w}_{\max}} \tag{28}$$

则基于SDW-MMSE的广义特征值波束形成滤波向量可表示为

$$\boldsymbol{w}_{\text{GEV-SDW-MMSE}}=G_{\text{SDW-MMSE}}\boldsymbol{w}_{\max} \tag{29}$$

基于SDW-MMSE复系数估计与基于MMSE复系数估计的区别主要在于语音失真权重因子的引入。语音失真权重因子的取值影响到降噪效果和语音失真之间的权重关系,由式(25)和式(28)可知,μ越大,表示波束成形方法将更多注意力放在降低残余噪声上,从而该波束形成方法降噪效果更强,但语音失真程度也在变大。

4 仿真结果

本研究仿真实验依托MATLAB平台和Campbell等[18]提出的Roomsim工具箱进行。通过Roomsim工具箱模拟麦克风阵列,并考虑不同混响时间所对应的不同房间墙壁吸收系数,生成具有不同混响时间和不同声源位置的冲激响应信号,与单通道语音信号相卷积得到麦克风阵列接收信号。房间大小设置为长3 m、宽4 m、高3 m。设置5个麦克风组成均匀线性阵列,每个麦克风之间的间隔为5 cm。将最左侧麦克风作为参考麦克风,目标语音信号与麦克风水平方向的夹角为45°,噪声干扰信号与水平

方向的夹角为135°。目标语音信号是选自TIMIT语音库中时长为3 s的纯净女声，采样频率为8 kHz。噪声信号包括从NOISEX-92噪声数据库中随机选取的babble等噪声，用来模拟相干噪声，同时选择高斯白噪声来模拟非相干噪声。短时傅里叶变换的帧长为256，帧移为64，使用Hamming窗函数。

本文采用输出信噪比、语音失真指数(Speech distortion index, SDI)和语音质量感知评估(Perceptual evaluation of speech quality, PESQ)[19]作为语音质量评价指标，其中输出信噪比表示输出信号在频域中纯净语音与噪声语音的总功率之比，以分贝表示，数值越高表示纯净语音所占比重越大，语音越清晰。语音失真指数定义为目标信号与其估计值之间的均方误差，并通过目标信号的功率进行归一化，以分贝表示，数值越高表示目标信号与其估计值之间的差异越大，失真程度更严重。PESQ是一种用于评估语音质量的方法，其评分值通常在－0.5和4.5之间，评分值越高表示语音质量越高。

实验在不同实际情况、不同混响时间、不同噪声类型和不同输入信噪比等条件下进行，通过最大似然(Maximum likelihood, ML)参数估计法[20]估计目标信号和噪声信号的功率谱。首先在不同实际情况下，如在只有干扰和有干扰又有混响等情况下，通过改变语音失真权重因子的值，得到不同实际情况下的语音评价指标，以验证所提波束形成方法的有效性和选取合适的语音失真权重因子值。然后设置不同的混响时间，不同的噪声类型和不同的信噪比，通过相应语音评价指标比较各种波束形成方法的性能好坏，最后分析和对比不同波束形成方法增强后的波形图和语谱图。本文将提出的基于MMSE的广义特征值波束形成方法(GEV-MMSE)和基于SDW-MMSE的广义特征值波束形成方法(GEV-SDW-MMSE)与文献[10]提出的基于盲解析归一化的广义特征值波束形成方法(GEV-BAN)、文献[13]提出的基于相位感知归一化的广义特征值波束形成方法(GEV-PAN)进行对比，在多噪声和混响等复杂环境下，相较于前文有所提升，能够进一步提升语音质量和减少语音失真，印证了文本工作。

4.1 语音失真权重因子取值分析

语音失真权重因子将更多的注意力放在了降低噪声上面，但也导致了更多的语音失真。现仿真分析语音失真权重因子μ的变化引起的输出信噪比、语音失真指数和PESQ的变化情况。在只有干扰、只有混响、有干扰和混响、有干扰和白噪声、既有干扰又有白噪声和混响的5种情况下，研究基于SDW-MMSE的广义特征值波束形成方法相应评价指标的变化情况。其中，干扰和干扰加白噪声情况下输入信噪比为0 dB，有混响的情况下输入信噪比为－15 dB、目标语音信号和干扰语音信号的混响时间均为200 ms，结果如图1所示。

图1显示了语音失真权重因子在不同情况下从0到10变化过程中，输出信噪比、语音失真指数和PESQ的变化情况。由图1可知，输出信噪比与语音失真权重因子μ正相关，在不存在混响的情况下，输出信噪比随μ的增大而增大，且提高明显。在存在混响的情况下，输出信噪比提高较少且随μ的增大而变化不大。PESQ在不存在混响的情况下较大，在存在混响的情况下较小。语音失真指数与μ正相关，随着μ的增大而增大。语音失真指数在存在混响的情况下较大，在只存在干扰情况下最小。输出信噪比、语音失真指数和PESQ在不存在混响的情况下表现较好，这说明不存在混响时滤波效果良好。在存在混响的情况下，输出信噪比等较未滤波之前有所提高。

从图1可以看出，随着μ的增大，输出信噪比和语音失真指数增大，PESQ变化较为缓慢，这说明μ的增大更有利于降低噪声，但会导致更大的语音失真。因此对于滤波效果而言，选择合适的μ值至关重要。由图1可知，μ在5左右时输出信噪比增大趋近平缓，语音失真指数和PESQ下降程度也可接受，语音失真较少，因此在后续的仿真实验中μ取5。

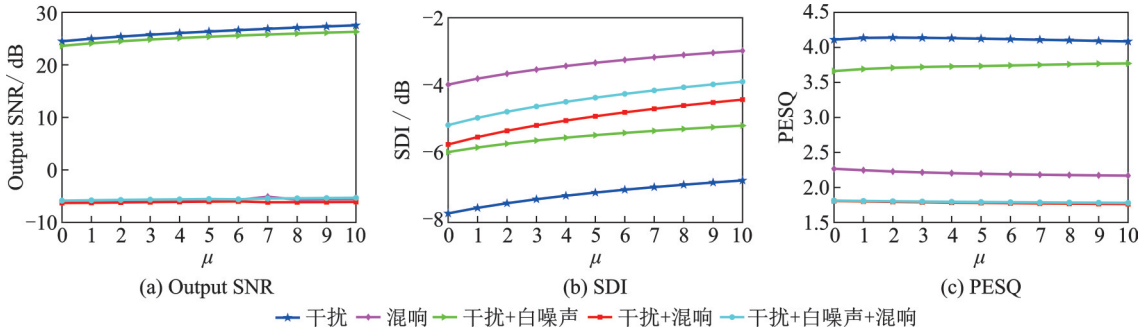

图1 语音评价指标与语音失真权重因子μ的关系

Fig.1 Relationship between speech evaluation index and speech distortion weight factor μ

4.2 不同混响时间下算法性能对比

本次仿真在纯混响环境下进行,即只有混响无干扰和白噪声。观察混响时间从100 ms到500 ms变化过程中语音评价指标的变化,从而探究混响时间变化对各波束形成方法滤波性能的影响,结果如图2所示。图2显示了4种波束形成方法在混响时间从100 ms到500 ms变化时,输出信噪比、语音失真指数和PESQ的变化情况。由图2可知,输出信噪比和PESQ随混响时间的增大而减小,降噪效果逐渐变差;语音失真指数随混响时间的增大而增大,语音失真程度逐渐变大。GEV-SDW-MMSE具有最大的输出信噪比和最大的PESQ值,GEV-MMSE具有最小的语音失真指数,这充分说明了引入语音失真权重因子后能够增强降噪效果,但也会导致语音失真程度加重,但从PESQ看整体语音质量有所提升。与文献[10]提出的GEV-BAN和文献[13]提出的GEV-PAN相比,本文提出的GEV-MMSE和GEV-SDW-MMSE输出信噪比和PESQ值明显较大,语音失真指数值较小,较前文有所提升,证明了在混响情况下本文做出的工作。

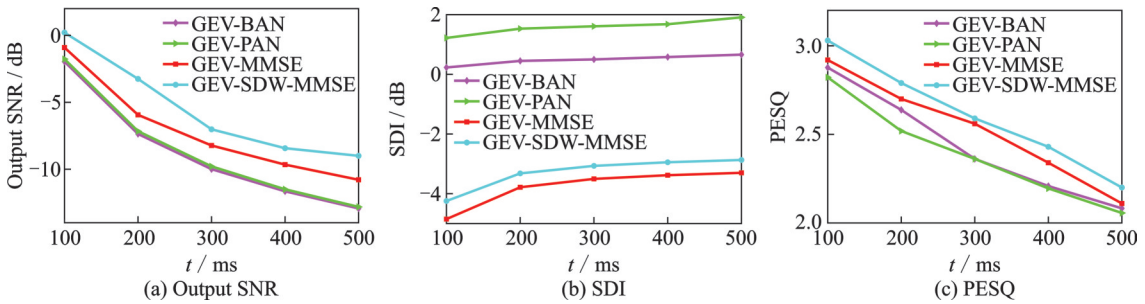

图2 语音评价指标与混响时间的关系

Fig.2 Relationship between speech evaluation index and reverberation time

4.3 不同噪声类型和不同输入信噪比下算法性能对比

本次仿真实验在不同噪声类型和不同信噪比下进行,通过更改噪声类型和输入信噪比观察输出信噪比、语音失真指数和PESQ的变化,分析算法的不同性能表现,仿真结果如表1所示,其中噪声都随机选取自NOISEX-92噪声数据库。babble噪声指嘈杂餐厅内讲话声,factory1噪声指工程中切割和电焊设备附近处噪音,m109噪声指移动坦克噪声,machinegun噪声指枪开火时噪声,volvo噪声指雨天行驶中的车内噪声。输入信噪比从−5 dB到15 dB依次增大。

表1 语音评价指标与噪声类型和输入信噪比的关系

Table 1 Relationship of speech evaluation index with noise type and input SNR

噪声类型	输入SNR/dB	GEV-BAN			GEV-PAN			GEV-MMSE			GEV-SDW-MMSE		
		SNR/dB	SDI/dB	PESQ	SNR/dB	SDI/dB	PESQ	SNR/dB	SDI/dB	PESQ	SNR/dB	SDI/dB	PESQ
babble	−5	0.09	6.05	1.31	0.48	3.40	1.27	20.06	**−4.83**	3.38	**21.42**	−4.50	**3.49**
	0	5.40	5.64	1.67	5.84	3.09	1.61	24.34	**−5.75**	3.72	**25.46**	−5.40	**3.75**
	5	11.18	5.30	2.09	11.62	2.74	1.99	28.81	**−6.71**	4.00	**29.75**	−6.35	**4.02**
	10	17.46	5.00	2.49	17.95	2.19	2.33	33.29	**−7.72**	4.20	**34.16**	−7.34	**4.20**
	15	24.51	4.69	2.85	25.06	1.56	2.67	37.69	**−9.02**	4.33	**38.52**	−8.54	**4.32**
factory1	−5	−1.21	6.79	1.38	−1.04	4.05	1.40	13.16	**−4.87**	3.39	**15.31**	−4.52	**3.49**
	0	3.65	6.31	1.72	3.93	3.64	1.72	17.94	**−5.86**	3.72	**19.88**	−5.48	**3.78**
	5	8.94	5.86	2.10	9.29	3.19	2.08	22.74	**−6.88**	4.01	**24.51**	−6.49	**4.04**
	10	14.61	5.50	2.51	15.00	2.75	2.40	27.54	**−7.93**	4.21	**29.18**	−7.51	**4.22**
	15	20.86	5.20	2.86	21.30	2.17	2.69	32.44	**−9.21**	4.34	**33.98**	−8.71	**4.34**
m109	−5	−1.76	4.88	1.72	−1.40	2.56	1.64	15.56	**−5.14**	3.36	**17.78**	−4.77	**3.47**
	0	3.27	4.56	2.12	3.71	2.20	2.01	20.24	**−6.12**	3.71	**22.28**	−5.73	**3.75**
	5	8.55	4.29	2.52	9.02	1.70	2.35	24.90	**−7.14**	3.96	**26.81**	−6.73	**4.00**
	10	14.08	4.09	2.89	14.55	1.04	2.69	29.45	**−8.20**	4.18	**31.25**	−7.76	**4.20**
	15	20.00	3.91	3.20	20.47	0.29	2.99	33.94	**−9.51**	4.33	**35.64**	−8.98	**4.33**
machinegun	−5	12.56	3.83	1.37	15.53	1.17	1.30	21.52	**−5.26**	3.32	**22.49**	−4.92	**3.40**
	0	18.35	3.84	1.83	21.08	0.66	1.73	25.82	**−6.15**	3.62	**26.60**	−5.79	**3.69**
	5	24.25	3.77	2.27	26.59	0.12	2.10	30.19	**−7.11**	3.91	**30.85**	−6.73	**3.97**
	10	30.18	3.67	2.66	31.97	−0.48	2.46	34.61	**−8.15**	4.16	**35.19**	−7.74	**4.19**
	15	36.06	3.55	2.98	37.35	−1.12	2.84	39.14	**−9.41**	4.32	**39.66**	−8.93	**4.32**
volvo	−5	−4.25	3.11	2.37	−3.90	0.77	2.29	5.91	**−5.65**	3.33	**8.03**	−5.26	**3.40**
	0	0.79	3.13	2.74	1.21	0.10	2.62	10.87	**−6.60**	3.64	**12.89**	−6.18	**3.69**
	5	5.89	3.15	3.07	6.33	−0.62	2.96	15.72	**−7.60**	3.92	**17.68**	−7.16	**3.98**
	10	11.11	3.16	3.44	11.57	−1.36	3.27	20.71	**−8.69**	4.16	**22.71**	−8.22	**4.19**
	15	16.75	3.16	3.70	17.25	−1.95	3.53	26.32	**−10.03**	4.31	**28.09**	−9.47	**4.33**

从表1可以看出，在不同噪声类型下，各种波束形成方法在machinegun噪声下滤波性能最好，在volvo噪声下滤波性能最差。各种算法的输出信噪比和PESQ都随着输入信噪比的增大而增大，语音失真指数随着输入信噪比的增大而减小，这说明在噪声更小的情况下各波束形成方法性能越好，同时也说明不同波束形成方法在不同噪声类型和不同输入信噪比下，滤波性能有所不同。但无论哪种情况，本文提出的GEV-MMSE和GEV-SDW-MMSE在各项性能指标上都优于文献[10]提出的GEV-BAN和文献[13]提出的GEV-PAN，这充分说明本文算法的有效性和优越性。输出信噪比和PESQ都是GEV-SDW-MMSE最大，其中输出信噪比明显提高，而语音失真指数总是GEV-MMSE最小，这也说明了语音失真权重因子在提高降噪效果的同时也加重了语音失真，但两者相权衡下，PESQ的值变高，说明语音质量得到了改善。

4.4 波形图和语谱图分析

波形图与语谱图能直观反映出噪声的去除情况，现仿真分析babble噪声加高斯白噪声的高噪声环

境下各算法的滤波性能,其中输入信噪比为−5 dB。图3展示了4种波束形成方法在输入信噪比为−5 dB滤波后的波形图和语谱图。从图3(c)和图3(d)中的语谱图可以看出,GEV-BAN和GEV-PAN能消除部分高频段噪声,但在低频段和存在语音时间段噪声残留明显,同时波形图中存在明显的毛刺和严重的语音失真现象。从图3(e)可以看出,GEV-MMSE不仅能消除高频段噪声,还能在一定程度上消除存在语时间段的噪声,但在低频段滤波效果较差。从图3(f)可以看出,GEV-SDW-MMSE能消除大部分的噪声,并且在波形图中不存在明显的毛刺现象,语音失真程度也较小。而低频段仍然残留有一些噪声,这说明该算法在低频段性能有限。综上所述,在不同频段,本文提出的GEV-MMSE和GEV-SDW-MMSE在抑制噪声和降低语音失真程度上相较于文献[10]提出的GEV-BAN和文献[13]提出的GEV-PAN都有所改善。

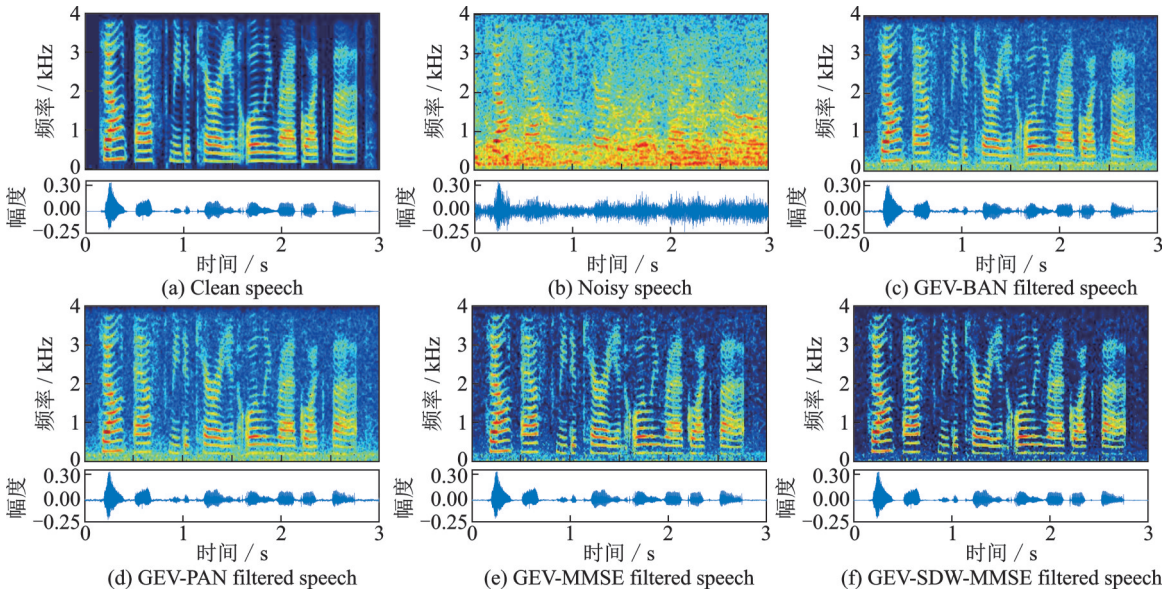

图3 实际情况下不同波束形成方法的语谱图和波形图

Fig.3 Spectrograms and waveforms of different beamforming methods under actual situation

5 结束语

本文提出了一种基于MMSE的广义特征值波束形成方法,该方法通过最小化波束形成滤波后输出语音信号和目标语音信号的均方误差,在广义特征向量已知的情况下求得复系数。同时,将语音失真权重因子引入到复系数的最小均方误差估计中,以调节降噪效果和语音失真之间的权重关系,并提出了基于SDW-MMSE的广义特征值波束形成方法。在多噪声和混响等复杂环境下,仿真实验表明,相较于前人提出的方法,本文提出的GEV-MMSE和GEV-SDW-MMSE输出信噪比和PESQ值较高,语音失真指数值较低,能够有效地消除部分混响、相干噪声和非相干噪声,提高输出信噪比,减少语音失真,从而提高语音质量,达到更好的语音增强效果。

参考文献:

[1] PALIWAL K, WÓJCICKI K, SCHWERIN B. Single-channel speech enhancement using spectral subtraction in the short-time modulation domain[J]. Speech Communication, 2010, 52(5): 450-475.

[2] CHEN J, BENESTY J, HUANG Y, et al. New insights into the noise reduction Wiener filter[J]. IEEE Transactions on Audio, Speech, and Language Processing, 2006, 14(4): 1218-1234.

[3] BENESTY J, CHEN J, HUANG Y. Microphone array signal processing[M]. Berlin, German: Springer, 2008.

[4] 李康宁, 郭永刚, 王肃静, 等. 一种并行主偏度分析算法及其在盲源分离上的应用[J]. 数据采集与处理, 2020, 35(5): 910-919.
LI Kangning, GUO Yonggang, WANG Sujing, et al. A parallel principal skewness analysis algorithm and its application in blind source separation[J]. Journal of Data Acquisition and Processing, 2020, 35(5): 910-919.
[5] 叶中付, 朱星宇. 基于协方差矩阵重构的稳健自适应波束形成算法综述[J]. 数据采集与处理, 2019, 34(6): 962-973.
YE Zhongfu, ZHU Xingyu. Review of robust adaptive beamforming algorithms based on covariance matrix reconstruction[J]. Journal of Data Acquisition and Processing, 2019, 34(6): 962-973.
[6] SOUDEN M, BENESTY J, AFFES S. A study of the LCMV and MVDR noise reduction filters[J]. IEEE Transactions on Signal Processing: A publication of the IEEE Signal Processing Society, 2010, 58(9): 4925-4935.
[7] ALI R, BERNARDI G, WATERSCHOOT T V, et al. Methods of extending a generalized sidelobe canceller with external microphones[J]. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2019, 27(9): 1349-1364.
[8] PARK J, HONG J, CHOI J, et al. Determinant-based generalized sidelobe canceller for dual-sensor noise reduction[J]. IEEE Sensors Journal, 2022, 22(9): 8858-8868.
[9] ASANO F, ASOH H. Blind speech enhancement using generalized eigenvalue decomposition[C]//Proceedings of 2002 11th European Signal Processing Conference. Piscataway, NJ: IEEE, 2002: 1-4.
[10] WARSITZ E, HAEB-UMBACH R. Blind acoustic beamforming based on generalized eigenvalue decomposition[J]. IEEE Transactions on Audio, Speech, and Language Processing, 2007, 15(5): 1529-1539.
[11] KRUEGER A, WARSITZ E, HAEB-UMBACH R. Speech enhancement with a GSC-like structure employing eigenvector-based transfer function ratios estimation[J]. IEEE Transactions on Audio, Speech, and Language Processing, 2011, 19(1): 206-219.
[12] HUANG G, BENESTY J, LONG T, et al. A family of maximum SNR filters for noise reduction[J]. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2014, 22(12): 2034-2047.
[13] PFEIFENBERGER L, ZÖHRER M, PERNKOPF F. Eigenvector-based speech mask estimation for multi-channel speech enhancement[J]. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2019, 27(12): 2162-2172.
[14] ANDERSEN K T, MOONEN M. Robust speech-distortion weighted interframe wiener filters for single-channel noise reduction[J]. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2018, 26(1): 97-107.
[15] 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.
[16] COHEN I. Multichannel post-filtering in nonstationary noise environments[J]. IEEE Transactions on Signal Processing, 2004, 52(5): 1149-1160.
[17] GANNOT S, COHEN I. Speech enhancement based on the general transfer function GSC and postfiltering[J]. IEEE Transactions on Speech and Audio Processing, 2004, 12(6): 561-571.
[18] CAMPBELL D R, PALOMAEKI K J, BROWN G J. A Matlab simulation of 'Shoebox' room acoustics for use in research and teaching[J]. Computing and Information Systems, 2005, 9(3): 59-62.
[19] BENESTY J, COHEN I, CHEN J. Fundamentals of signal enhancement and array signal processing[M]. Newark: Wiley, 2017.
[20] THUNE P, ENZNER G. Maximum-likelihood approach with Bayesian refinement for multichannel-wiener postfiltering[J]. IEEE Transactions on Signal Processing, 2017, 65(13): 3399-3413.

作者简介:

李海龙(1999-),男,硕士研究生,研究方向:麦克风阵列声源定位与语音增强,E-mail:lhlong9123@whu.edu.cn。

杨飞(1981-),**通信作者**,男,博士,副教授,研究方向:机器视觉、智能传感器与检测技术和信号处理,E-mail:f.yang@whu.edu.cn。

杨诗童(1998-),男,硕士研究生,研究方向:阵列信号处理和语音增强算法,E-mail: 2016301470005@whu.edu.cn。

路晓庆(1983-),女,博士,教授,研究方向:智能电网控制、AI赋能信息技术等,E-mail:luxiaoqing2012@hotmail.com。

(编辑:刘彦东)