双微阵列语音增强算法在说话人识别中的应用

毛 维，曾庆宁，龙 超
(桂林电子科技大学信息与通信学院，广西桂林 541004)

摘要：针对复杂噪声环境下识别性能显著降低的问题，提出一种用于说话人识别系统前端的双微阵列语音增强算法。该算法采用的是相干滤波和频域宽带最小方差无畸变响应波束形成器后置结合改进的维纳滤波器。其基本原理是首先求出双微麦克风阵列信号中两个相邻通道间的相干函数，再利用通道间信号的相干性来进行初始噪声抑制。其次，通过一个频域宽带最小方差无畸变响应(Minimum Variance Distortionless Response, MVDR)波束形成器保留目标声源方向的信号并抑制其他方向的信号干扰，再通过改进的维纳滤波器去除噪声残留提升语音质量。最后，使用梅尔频率倒谱系数(Mel Frequency Cepstral Coefficients, MFCC)和伽马通滤波器组频率倒谱系数(Gammatone Filter-bank Frequency Cepstral Coefficients, GFCC)对增强后的语音信号做特征参数提取并进行说话人识别。仿真过程采用声学人工头模拟双耳采集数据，实验结果表明，该语音增强算法在复杂噪声环境下能够获得较好的增强效果，能有效提升说话人识别系统的识别率。

关键词：双微阵列；语音增强；相干滤波；最小方差无畸变响应；改进维纳滤波；说话人识别
中图分类号：TN912.34 **文献标识码：**A **文章编号：**1000-3630(2018)-03-0253-08
DOI 编码：10.16300/j.cnki.1000-3630.2018.03.011

Application of dual-mini microphone array speech enhancement algorithm in speaker recognition

MAO Wei, ZENG Qing-ning, LONG Chao
(*School of Information and Communication*, *Guilin University of Electronic Technology*, *Guilin* 541004, *Guangxi*, *China*)

Abstract: Aiming at the problem of lowering recognition performance in noisy speech environment, a dual-mini microphone array speech enhancement algorithm is proposed for the front-end processing of recognition system. The speech enhancement algorithm based on Coherent Filter and MVDR-wiener is presented. First, the dual-mini microphone array signals are collected to derive the coherence function between adjacent channels and to carry out the initial noise suppression by using the coherence between elements. Then, the information of target speech is processed by the broad-band MVDR algorithm to keep the signal in the desired sound source direction and suppress the interference signals in other directions. The improved Wiener filter which can get better voice quality by removing residual noise is utilized to process the enhanced signal. Finally, a speaker recognition system using Mel frequency cepstral coefficients (MFCC) and GFCC for feature extraction is used to recognize the enhanced speech. Binaural data are acquired with acoustic artificial head in simulations, the experimental results show that the speech enhancement algorithm can obtain better enhanced effect in noisy environment and effectively improve the recognition rate.

Key words: dual-mini array; speech enhancement; coherence filtering; minimum variance distortionless response; modified Wiener filter; speaker recognition

0 引 言

说话人识别是利用说话人的语音特征来确定或鉴定说话人身份的技术。实际的说话人识别系统受到周围环境噪声的影响，系统识别率严重下降，语音增强[1]是解决这一问题的一种方式，其主要目的是从带噪语音中尽可能地恢复出原有的纯净语音。语音增强算法种类繁多，其中大多数算法是使用各种噪声消除方法结合语音信号的特征来进行研究，如基于短时谱估计的单通道语音增强算法有：谱减法[2]，最小均方误差方法[3-4]，维纳滤波方法[5]等。常见的多通道语音增强算法[6]有固定波束形成法、广义旁瓣抵消器等。在噪声抑制方面，固定波束形成法有着明显的缺陷，且处理后的语音信号中仍存在较多的噪声残留；而广义旁瓣抵消器则是

收稿日期：2017-06-21；**修回日期：**2017-08-18
基金项目：国家自然科学基金项目(61461011)、教育部重点实验室 2016 年主任基金项目资助(CRKL160107)、桂林电子科技大学研究生科研创新项目(2017YJCX16、2017YJCX20)
作者简介：毛维(1992－)，男，湖南岳阳人，硕士研究生，研究方向为语音增强、语音识别等。
通讯作者：龙超，E-mail: bishe006@163.com

在抑制混响和非相干噪声方面有一定的限制。近年来，说话人识别技术得到进一步的发展。在无噪声干扰的情况下，当前说话人识别系统可以获得较高的识别率。但是在电噪声、房间混响等复杂噪声情况下，说话人识别系统的识别效果较差，而且存在适用性差以及实用性不足等诸多缺点。而采用麦克风阵列，与单个阵元相比，可以利用麦克风阵列信号的时域、频域以及空域信息更好地处理说话人的语音并提升语音质量。本文针对此问题，利用声学人工头模拟头部双耳距离，采用一种双微麦克风阵列来获取说话人的语音，用相干滤波先将各通道间的含噪语音进行初步的信噪比提升，再通过最小方差无畸变响应(Minimum Variance Distortionless Response, MVDR)波束形成器对非目标声源方向的信号进行噪声抑制，最后通过一个改进的维纳后置滤波器去除残留噪声并对语音做增益补偿提升可懂度，以此来提升说话人识别系统的识别率。

1 语音增强

1.1 双微阵列

双微阵列是由2个子阵元个数为M的麦克风阵列构成，其中单个子阵列阵元间距大约为1～2 cm，而两个子阵列之间的间距一般设置为15～18 cm，本文采用的子阵列之间间距为16 cm，其中双微阵列中单个子阵列中的阵元个数都为2，组成一个2×2的4元阵，双微阵列的简易结构如图1所示。

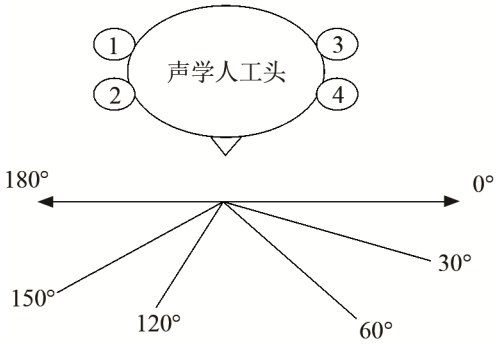


图1 双微阵列的简易结构

Fig.1 Simple structure of dual-mini microphone array

1.2 相干滤波器

Allen 等[6]在 1977 年首先提出将相干滤波器引入到麦克风阵列语音增强算法中，而该相干滤波器主要是以相邻通道间的含噪语音信号的相干性为前提，对非相干性噪声有一定的抑制能力。Yousefian 等[7]提出了一种改进相干函数的增强算法，该算法对相干噪声也有一定的抑制效果。相干滤波器的基本处理流程：首先，初始化一个阈值，然后求出通道间相干滤波增益函数的值，如果大于某个特定的值λ_2时，那么经过该相干滤波器的含噪语音信号可以无失真地通过，而当通道间相干滤波增益函数的值小于某个特定的值λ_1时，对含噪语音信号进行阻塞，通过该滤波器的信号可以滤除非相干噪声。本文针对由两个微型阵列组成的双微阵列，采用双微阵列中子阵列相邻较近的两个子阵元进行相干滤波，且只在噪声段进行滤波处理，同时结合了语音活动检测[8]的方法，相干滤波后可以获得四路已抑制非相干噪声的信号，具体过程如图2所示，其中d为麦克风1和2以及麦克风3和4之间的阵元间距，$8d$为由麦克风1、2组成的子阵列和麦克风3、4组成的子阵列之间的阵列间距。

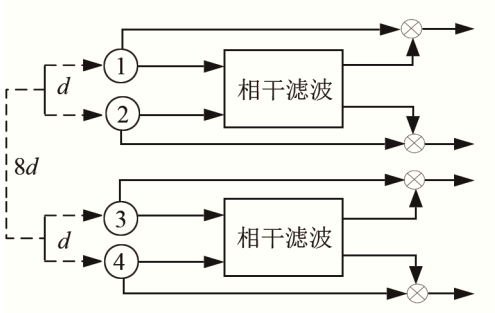


图2 相干滤波原理图

Fig.2 Schematic diagram of coherent filtering

首先，对各个麦克风接收到的语音信号进行加窗分帧，接着对每个麦克风信号中的每帧信号进行傅里叶变换，然后求出求出各个麦克风中输入信号的互功率谱密度(Cross-power Spectral Density, CSD)[9]为

$$\Phi_{y_iy_j}(\omega,k)=\Phi_{x_ix_j}(\omega,k)+\Phi_{n_in_j}(\omega,k) \tag{1}$$

其中：i，j分别为1、2和3、4，表示为通道索引位置，$\Phi_{n_in_j}(\omega,k)$表示噪声信号的功率谱密度，$\Phi_{x_ix_j}(\omega,k)$表示含噪语音信号的功率谱密度[10]。

两个相邻麦克风接收到的信号的相干函数可用式(2)定义：

$$\Gamma_{y_iy_j}(\omega,k)=\frac{\Phi_{y_iy_j}(\omega,k)}{\sqrt{\Phi_{y_iy_i}(\omega,k)\Phi_{y_jy_j}(\omega,k)}} \tag{2}$$

通过式(3)判决式来设置相干滤波阈值，得到相干滤波的增益函数：

$$\tilde{\Gamma}_{y_iy_j}(\omega,k)=\begin{cases}0.001, & \Gamma_{y_iy_j}<\lambda_1\\ 0.99, & \Gamma_{y_iy_j}>\lambda_2\\ \Gamma_{y_iy_j}, & \text{其他}\end{cases} \tag{3}$$

其中，λ_1设定为0.35，λ_2设定为0.8。最后将相邻麦克风之间的信号通过相干滤波器，该滤波器可以

抑制弱相关噪声，得到初步增强的语音信号。

$$\tilde{Y}_i(\omega,k)=Y_i(\omega,k)\times\widetilde{\Gamma}_{y_iy_j}(\omega,k) \tag{4}$$

式(4)中的 $Y_i(\omega,k)$ 为第 i 个麦克风接收的信号经过傅里叶变换之后的频域表达式，$\tilde{Y}_i(\omega,k)$ 为第 i 个麦克风通过相干滤波器后的频域表达式。

1.3　广义旁瓣抵消器

Griffiths 等[11]在 1982 年首次提出了广义旁瓣抵消器(Generalized Sidelobe Canceller, GSC)的概念，基于线性约束最小方差条件下，首次引入了阻塞矩阵这个概念。广义旁瓣抵消器将采集到的信号分为两个通道来处理，即上通道固定波束形成部分和下通道阻塞矩阵以及自适应噪声抵消部分。上通道部分，对 n 路采集到的信号起初步增强的作用，得到增强后的语音信号 $y'(i)$ 。下通道部分，采用构建的阻塞矩阵 $\boldsymbol{B}(i)$ 来滤除 n 路信号中的语音部分，经过阻塞矩阵后得到的信号 $N(i)$ 通过一个自适应噪声抵消器可以得到噪声估计值 $N'(i)$ 。本文采用的自适应噪声抵消器是基于归一化最小均方误差算法。最后在上通道固定波束形成部分得到的信号 $y'(i)$ 与下通道自适应噪声抵消器部分得到的噪声估计信号 $N'(i)$ 相减后可以得到期望的语音信号 $y''(i)$；GSC 基本原理如图 3 所示。

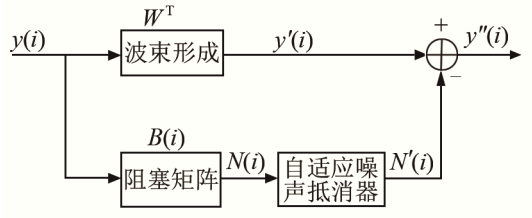


图 3　广义旁瓣抵消器原理图

Fig.3　Schematic diagram of Generalized Sidelobe Canceller

图 3 中 $\boldsymbol{y}'(i)$、$\boldsymbol{y}(i)$、$\boldsymbol{W}^{\mathrm{T}}$、$\boldsymbol{N}(i)$ 分别为

$$\boldsymbol{y}'(i)=\boldsymbol{y}(i)\times\boldsymbol{W}^{\mathrm{T}} \tag{5}$$

$$\boldsymbol{y}(i)=[y_1(i),y_2(i),\cdots,y_n(i)] \tag{6}$$

$$\boldsymbol{W}^{\mathrm{T}}=[\omega_1,\omega_2,\cdots,\omega_n] \tag{7}$$

$$\boldsymbol{N}(i)=\boldsymbol{y}(i)\times\boldsymbol{B}(i) \tag{8}$$

1.4　频域宽带最小方差无畸变响应

MVDR 是 Capon 于 1967 年提出的一种自适应的空间波数谱估计算法[12]，通过 MVDR 算法得到的权系数可以使在期望方向上的阵列输出功率最小，同时信噪比最大，与固定波束形成算法相比，MVDR 算法[13]在很大程度上提高了波束谱估计的分辨率，有效地抑制了干扰和噪声。本文采集到的双微阵列中各通道的含噪语音信号为 $y_m(i)$，$m=1,2,3,4$, 首先对每个通道的语音信号加窗分帧，然后对每一帧信号进行短时傅里叶变换，那么第 m 个麦克风采集到的含噪语音信号中第 k 个频率点第 λ 帧的离散傅里叶变换(Discrete Fourier Transform, DFT)系数可表示为

$$Y_m(k,\lambda)=\sum_{i=0}^{L-1}y_m(i)\mathrm{e}^{-\mathrm{j}2\pi ki/L} \tag{9}$$

式(9)获得每一帧的 DFT 系数 $Y_m(k)$ 与复权因子 $\boldsymbol{W}_m^*$ 相乘即可实现 MVDR 波束形成算法[9,14]，如式(10)所示。为了使得输出信号的信噪比最大，同时使得非目标方向信号输出能量最小，需要满足以下约束条件：

$$\boldsymbol{X}(k)=\sum_m\boldsymbol{W}_m^*\boldsymbol{Y}_m(k)=\boldsymbol{W}^{\mathrm{H}}\boldsymbol{Y} \tag{10}$$

$$\min_w\{\boldsymbol{W}^{\mathrm{H}}(\theta_s,k)\boldsymbol{\Phi}_{MM}\boldsymbol{W}(\theta_s,k),\boldsymbol{W}^{\mathrm{H}}(\theta_s,k)D(\theta_s,k)\}=1 \tag{11}$$

式(10)中，$\boldsymbol{W}$ 为最优权矢量，式(11)中的 θ_s 为麦克风声源角度，$\boldsymbol{\Phi}_{MM}$ 表示阵列输入的协方差矩阵，输入的协方差矩阵可以由一段自带数据快拍来估计完成，通常情况下，该估计的数据越长，则精确度越高。根据拉格朗日乘子法，可以计算出 MVDR 波束形成的系数矢量为[13]

$$\boldsymbol{W}(\theta_S,k)=\frac{\boldsymbol{\Phi}_{MM}^{-1}(k)\boldsymbol{D}(\theta_S,k)}{\boldsymbol{D}^H(\theta_S,k)\boldsymbol{\Phi}_{MM}^{-1}(k)\boldsymbol{D}(\theta_S,k)} \tag{12}$$

式(12)中，$\boldsymbol{D}(\theta_S,k)$ 为导向矢量。

1.5　改进的维纳滤波

维纳滤波法是通过计算最小均方值获得最优估计结果的算法，其中 Scalart[15]提出的基于先验信噪比的维纳滤波法在较小计算量下实现了实时处理。但是这种方法需要估计先验信噪比，因此无法在试验中直接得到纯净语音的功率。若将 $\xi(k,\lambda)$ 定义为先验信噪比，将 $\gamma(k,\lambda)$ 定义为后验信噪比，则根据“直接判决”思想，计算过程为

$$\xi(k,\lambda)=a\cdot G_\omega^2(k,\lambda-1)\cdot\gamma(k,\lambda-1)+(1-a)\cdot\max[\gamma(k,\lambda)-1,0] \tag{13}$$

其中，$G_\omega(k,\lambda-1)$ 表示前一帧的频谱增益函数：

$$G_\omega(k,\lambda)=\frac{\xi(k,\lambda)}{1+\xi(k,\lambda)} \tag{14}$$

研究表明[16]，对先验信噪比和后验信噪比的高估和低估会对语音的可懂度有一定的影响，在低于 −10 dB 区域，估计的先验信噪比往往会高于理想情况，此区域所产生的误差将会明显降低增强语音的可懂度；在大于 6.02 dB 区域，则存在低估的情况，因此，进一步的处理方法如下：

$$\hat{\xi}(k,\lambda)=\frac{\xi(k,\lambda)}{1+\xi(k,\lambda)}\cdot(1+\frac{\xi(k,\lambda)}{1+\xi(k,\lambda)}\cdot\gamma(k,\lambda)) \tag{15}$$

式(15)中，$\hat{\xi}(k,\lambda)$ 为修正的信噪比估计。

在小于−10 dB 以下区域，引入偏差值修正频谱增益函数：

$$\hat{G}(k,\lambda)=(G_{\omega}(k,\lambda)+c)\cdot\frac{1}{1+c} \tag{16}$$

式(16)中，$c=b/(1-b)$，b 为先验信噪比的修正系数，$0<b<1$，当 $b=0.02$ 时，噪声抑制效果最佳。

在大于 6.02 dB 区域，有：

$$\mathrm{E}\left\{\left|\hat{S}(k,\lambda)\right|\right\}>4\cdot\mathrm{E}\left\{\left|S(k,\lambda)\right|^2\right\} \tag{17}$$

式(17)中，$S(k,\lambda)$ 为原始的语音信号频谱，$\hat{S}(k,\lambda)$ 为修正的语音信号频谱。

最终可推得：

$$\hat{G}(k,\lambda)^2\cdot(1+\frac{1}{\hat{\xi}(k,\lambda)})>4 \tag{18}$$

通过式(17)～(18)中先验信噪比和增益函数，可以判定幅度谱大于 6.02 dB 的放大畸变区域，对此区域的幅度谱进行一定的约束为

$$\hat{S}_1(k,\lambda)=d\cdot\hat{S}(k,\lambda) \tag{19}$$

其中，d 为衰减系数，$d=0.08$ 且 $0<d<1$。

本文采用的是双微麦克风阵列系统，模拟头部双耳距离，其中通道 1、2 和通道 3、4 距离为 d，d 为 2 cm，而由通道 1、2 构成的子阵列和通道 3、4 构成的子阵列之间的距离为 $8d$；因通道 1、2 和通道 3、4 相距较远，麦克风子阵列之间的时空域信息可以获得更多的有效信息，相干滤波后将输出信号通过 MVDR 波束形成器，用于抑制非目标声源方向上的干扰，并将输出信号通过改进可懂度的维纳滤波器，去除噪声残留提高语音质量。本文算法结构如图 4 所示。

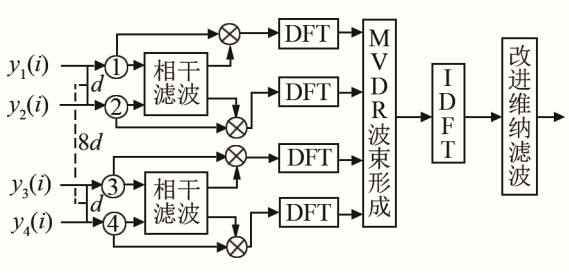


图 4 本文算法结构图

Fig.4 The structure diagram of the proposed algorithm in the paper

为了验证该双微阵列对于说话人识别系统的可行性，给出了单个麦克风阵列、双麦克风阵列以及双微阵列在 f16 噪声和 m109 噪声情况下的主观语音质量评估(Perceptual Evaluation of Speech Quality, PESQ)得分，如表 1 和表 2 所示。

根据表 1 和表 2 的 PESQ 得分可知，双微阵列情况下的去噪效果是较好的，因此，将该双微阵列应用在说话人识别系统前端是可行的。

表 1 f16 噪声情况下，不同信噪比下的 PESQ 得分

Table 1 PESQ scores under the condition of different signal to noise ratio with f16 noise

信噪比/dB	单麦阵列	双麦阵列	双微阵列
−5	0.57	1.03	1.47
0	0.74	1.32	1.70
5	0.97	1.67	2.02
10	1.23	1.99	2.30

表 2 m109 噪声情况下，不同信噪比下的 PESQ 得分

Table 2 PESQ scores under the condition of different signal to noise ratio with m109 noise

信噪比/dB	单麦阵列	双麦阵列	双微阵列
−5	0.67	1.13	1.40
0	0.86	1.48	1.75
5	1.10	1.78	2.15
10	1.36	2.02	2.45

2 说话人识别

本文使用的说话人识别系统是基于高斯混合模型[17-18](Gaussian mixture model, GMM)的文本无关的说话人辨认系统，系统分为两个阶段，一是训练阶段，二是测试阶段。在判定阶段，对测试语音的特征与训练模型中的各个特征进行对比，并计算对数似然度，得分最高的说话人模型为识别结果，简易流程图如图 5 所示。

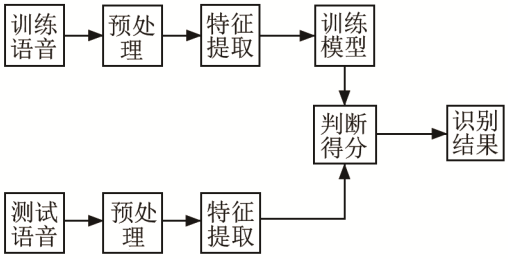


图 5 说话人识别系统简易流程图

Fig.5 The flow chart of the speaker recognition system

2.1 梅尔频率倒谱系数(Mel Frequency Cepstral Coefficients, MFCC)

线性预测倒谱系数(Linear Prediction Cepstrum Coefficient, LPCC)是一种谱估计方法，语音信号在获取线性预测倒谱系数后在频域上是线性的，但这与人耳的实际听觉属性有所不符合，并且 LPCC 中的大量的噪声细节都出现在语音信号的高频区域，这些都会影响系统的性能。基于 MFCC 系数对于人耳更为敏感这一听觉属性，在说话人识别中得到了广泛的应用。MFCC 系数的基本原理是将经过采样后的信号进行预加重处理，提升语音信号中的高频分量，然后对预加重处理后的语音信号进行加窗分帧，求每一帧信号的快速傅里叶变换，得到信号的

能量，接着通过一个梅尔滤波器组并计算每个滤波器组输出信号的对数能量，最后进行离散余弦变换(Discrete Cosine Transform, DCT)，最终得到 MFCC 系数[19]。

2.2　伽玛通滤波器组倒谱系数(Gammatone Filter-bank Frequency Cepstral Coefficients, GFCC)

伽马通(Gammatone)滤波器组[20]最初用在模拟听觉神经响应，是一种常见的耳蜗听觉滤波器[21]。本文采用的滤波器组的公式如下：

$$g_i(t)=A\cdot t^{a-1}e^{-2\pi BW(f_i)t}\cos(2\pi f_i t+\varphi_i)U(t),\quad t\geq 0,\ 1\leq i\leq a \tag{20}$$

其中，A 和 a 分别为滤波器增益和阶数；$U(t)$ 为阶跃信号，f_i 和 φ_i 分别为中心频率和相位；B_w 是等效矩形带宽，B_w 与中心频率的关系式为

$$B_w(f_i)=24.7+0.108f_i \tag{21}$$

GFCC 系数提取过程如图 6 所示，首先对语音信号进行预加重处理提升高频分量，然后对其进行加窗分帧，对每帧语音信号进行快速傅里叶变换，得到的幅度谱通过一个 Gammatone 滤波器组；最后对其进行离散余弦变换(Discrete Cosine Transform, DCT)。

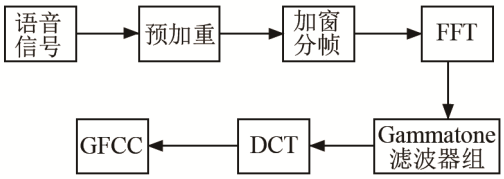
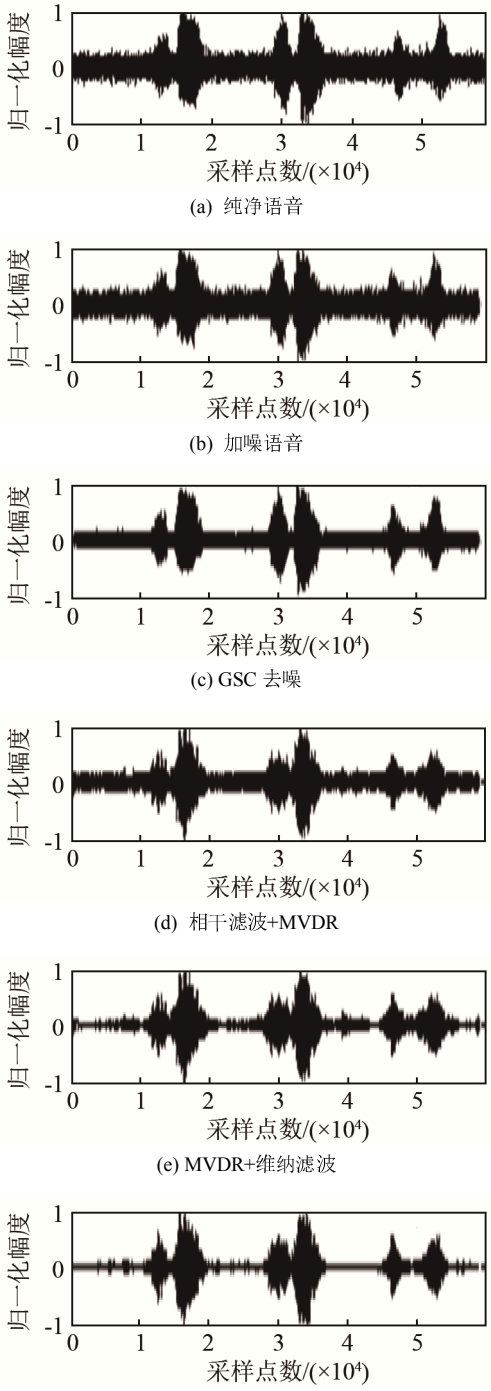


图 6　GFCC 提取过程
Fig.6　GFCC extraction process

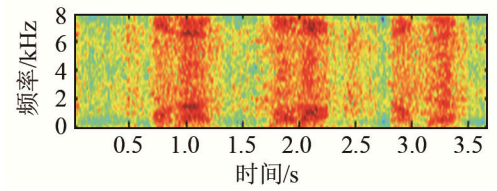
3　仿真实验

本次实验数据录制环境为学校教学楼楼顶空旷的天台，采集设备为 M-AUFIO 多路音频采集器，双耳采用双微阵列来模拟，且单个微型阵列中的阵元间距为 2 cm，两个微型阵列之间的距离为 16 cm，说话人距离阵列的距离为 1～1.5 m。语音和噪声采集都是在同样的录制环境下，除了说话人位置，在麦克风正前方的任意角度设置了若干噪声源。本次实验采集了 20 位同学的声音数据，其中男生 15 人，女生 5 人，每个说话人有 10 句语音，每段语音时间约 3 s，从每个说话人语音中选出 5 句语音作为训练模型，同时对每个说话人使用 5 句语音来测试，总共测试 100 句，语种为汉语普通话。采集语音和噪声数据时均采用 16 kHz 的采样率，采样精度为 16 bit，帧长为 32 ms，帧移为 16 ms，窗函数为 Hamming 窗。为了验证本文算法在说话人识别系统前端的可行性，对比了以下四种方法：广义旁瓣抵消(GSC)、相干滤波结合 MVDR、MVDR 结合改进维纳滤波以及本文使用的相干滤波结合 MVDR 波束形成后置改进维纳滤波。图 7 为 0 dB 白噪声情况下各算法增强后的时域仿真图，图 8 为 0 dB 白噪声情况下各算法增强后的时频图。

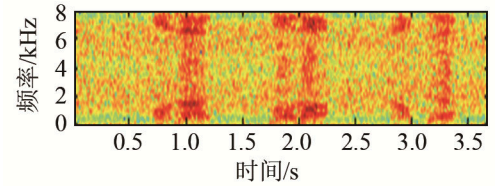


(a) 纯净语音
(b) 加噪语音
(c) GSC 去噪
(d) 相干滤波+MVDR
(e) MVDR+维纳滤波
(f) 本文算法

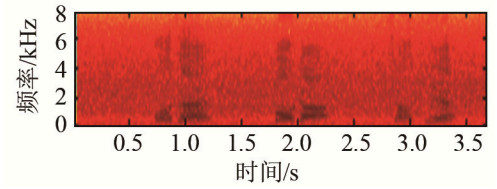
图 7　白噪声环境下各算法增强后时域仿真图
Fig.7　Time domain simulation chart enhanced by each algorithm in white noise environment



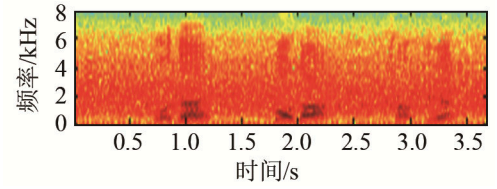
(a) 纯净语音



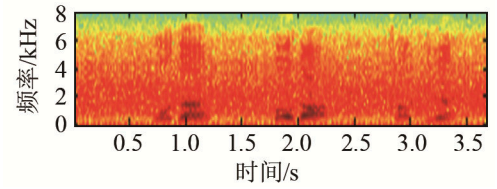
(b) 加噪语音



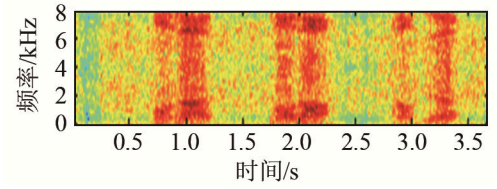
(c) GSC 去噪



(d) 相干滤波+MVDR



(e) MVDR+维纳滤波



(e) 本文算法

图 8 白噪声情况下各算法增强后的时频图

Fig.8 Spectrum map enhanced by each algorithm in white noise environment

从图 7 的时域仿真图中可见，广义旁瓣抵消器的去噪效果较差。相干滤波结合 MVDR 虽有一定的效果，但是噪声残留较大；而 MVDR 结合改进的维纳滤波增强后信噪比得到一定的提升，但仍然存在一定的毛刺和部分噪声残留。本文提出的相干滤波结合 MVDR 波束形成后置改进维纳滤波方法不仅在去除噪声残留方面得到了提升，同时还通过改进维纳滤波器降低了语音的失真度，改善了语音质量，在低信噪比环境下，这些影响是可以忽略的。因此本文提出的双微阵列语音增强算法适合用于说话人识别系统的前端处理。

3.1 实验 1

在说话人识别系统中选取了 LPC、MFCC、GFCC 这三种特征参数对白噪声进行测试，仿真结果如表 3 所示。

表 3 三种特征参数在白噪声环境下识别率(%)

Table 3 Recognition rate of three characteristic parameters under white noise environment(%)

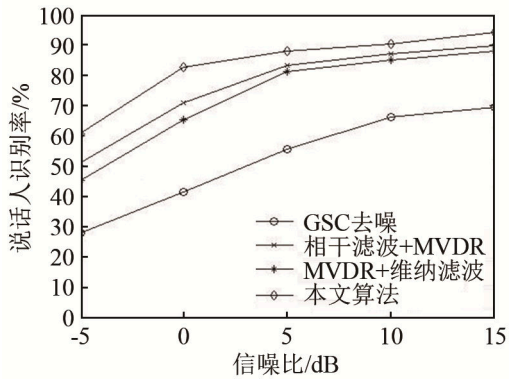
信噪比/dB	LPC	MFCC	GFCC
纯净语音	95.0	97.0	98.5
15	27.5	48.5	75.0
6	11.0	26.5	45
3	8.0	15.5	25.5
0	3.5	13.0	14.5
-3	3.5	5.5	5.5

如表 3 所示，在无白噪声干扰的环境下，三种特征参数的识别率都达到了较高的水平，但总体上，GFCC 的识别率明显高于其他两种特征参数提取的识别率。

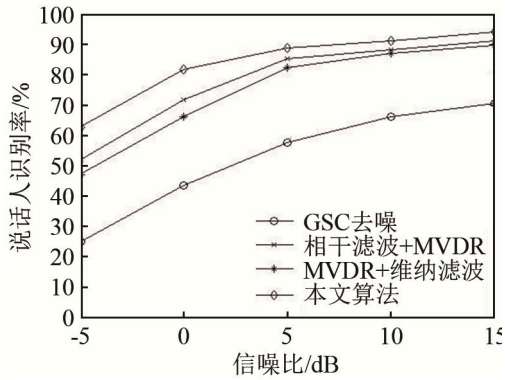
3.2 实验 2

为了验证本文算法在不同噪声环境下的可行性，分别在 babble 噪声、volvo 噪声、白噪声、f16 噪声的环境下进行测试，计算四种增强算法在说话人识别系统中的识别率。结合实验 1 的仿真结果，选取 26 维 GFCC 作为说话人识别特征，四种噪声的说话人识别率如图 9 所示。

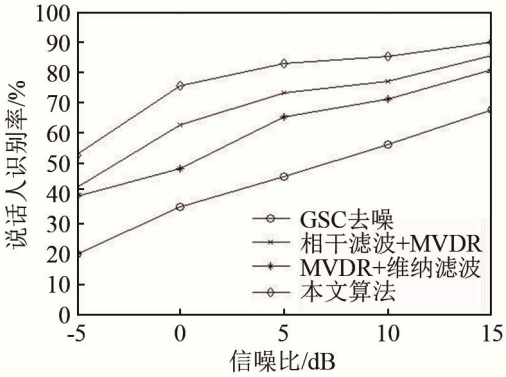
从图 9 中可以看到，广义旁瓣抵消器由于去噪效果的限制，表现出较低的识别率；而 MVDR 结合改进维纳滤波算法由于信噪比的提升比 GSC 去噪算法的识别率更高，但是其去噪效果仍然存在残留噪声和部分毛刺。本文提出的双微阵列语音增强算法通过对相邻通道间的信号进行初步的相干滤波增强，而后通过一个 MVDR 波束形成后置改进



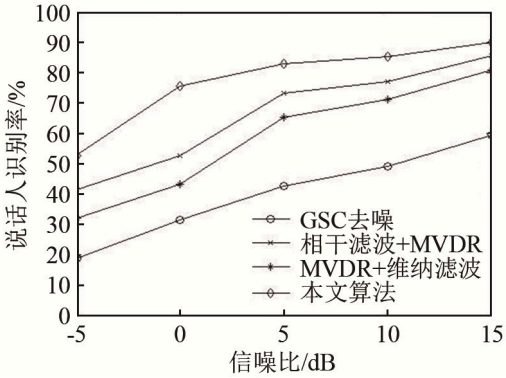
(a) babble 噪声



(b) volvo 噪声



(c) 白噪声



(d) f16 噪声

图 9　不同噪声环境下的说话人识别率

Fig.9　Recognition rates in different noise environments

维纳滤波器，使其增强效果更明显，同时减少残留噪声。在后置改进维纳滤波部分，对有产生失真的语音段进行了增益补偿，进一步提高了说话人识别系统的识别率。因此，证明了本文算法的可行性。

4　结束语

针对复杂噪声环境识别性能显著下降的问题，提出双微阵列语音增强算法，该算法先在基于两个相邻麦克风通道基础上，进行相干滤波增强，再通过一个 MVDR 结合改进的维纳滤波器，抑制非目标声源方向的干扰信号，同时去除噪声残留，最后得到增强后的信号。采用声学人工头模拟头部双耳距离采集数据，实验结果表明通过该算法对说话人识别系统前端进行初步增强后的说话人识别率提升明显，可应用到人工智能机器人的信号采集和说话人识别身份确认系统中。

参　考　文　献

[1] LOIZOU P C. Speech enhancement: Theory and Practice[M]. America: The Chemical Rubber Company Press, 2013: 75-109.

[2] STEVEN F B. A Spectral Substraction Algorithm for Suppression of Acoustic Noise in Speech[J]. IEEE International Conference on Acoustics Speech & Signal Processing, 1979, **27**(2): 200-203.

[3] EPHRAIM Y, MALAH D. Speech enhancement using a minimum-mean square error short-time spectral amplitude estimator [J]. IEEE Transactions on Acoustics Speech and Signal Processing, 1984, **32**(6): 1109-1121

[4] 张鑫琪，冯海泓，徐海东. 改进的最小均方误差语音增强算法的研究[J]. 声学技术, 2008, **27**(2): 230-234.
ZHANG Xinqi, FENG Haihong, XU Haidong. A study of an improved minimum mean-square error speech enhancement algorithm[J]. Technical Acoustics, 2008, **27**(2):230-234.

[5] 李宁，蒋建中，郭军利. 一种听觉掩蔽效应和维纳滤波的语音增强算法[J]. 计算机工程与应用, 2011, **47**(29): 161-163.
LI Ning, JIANG jianzhong, GUO Junli. Speech enhancement algorithm based on auditory masking effect and Wiener filter[J]. Computer Engineering and Applications, 2011, **47**(29):161-163.

[6] ALLEN J B, BERKLEY D A, BLAUERT J. Multimicrophone Signal-Processing technique to remove room reverberation from speech signals[J]. J. Acoust. Soc. Am., 1977, **62**(4): 912-915.

[7] YOUSEFIAN N, LOIZOU P C. A dual-microphone speech enhancement algorithm based on the coherence function[J]. IEEE Transactions on Audio Speech & Language Processing, 2011, **20**(2): 599-609.

[8] GHOSH P K, TSIARTAS A, NARAYANAN S. Robust voice activity detection using long-term signal variability[J]. IEEE Transactions on Audio Speech & Language Processing, 2011, **19**(3): 600-613.

[9] 马金龙，曾庆宁，胡丹，等. 基于麦克风小阵的多噪声环境语音增强算法[J]. 计算机应用, 2015, **35**(8): 2341-2344.
MA Jinlong, ZENG Qingning, HU Dan, et al. Speech enhancement algorithm based on microphone array under multiple noise envrionments[J]. Journal of Computer Applications, 2015, **35**(8): 2341-2344.

[10] 王群，曾庆宁，郑展恒. 低信噪比环境下的麦克风阵列语音识别算法研究[J]. 科学技术与工程, 2017, **17**(31): 101-107.
WANG Qun, ZENG Qingning, ZHENG Zhanheng. Speech recognition based on microphone array in low SNR[J]. Science Technolpgy and Engineering, 2017, **17**(31): 101-107.

[11] GRIFFIITHS L J, JIM C W. An alternative approach to linearly constrained adaptive beamforming[J]. IEEE Transactions on Antennas & Propagation, 1982, **30**(1): 27-34.

[12] CAPON J, GREEENFIELD R J, KOLKER R J. Multidimensional maximum-likelihood processing of a large aperture seismic array[J]. Proceedings of the IEEE, 1967, **55**(2): 192-211.

[13] 郑恩明，黎远松，陈新华，等. 改进的最小方差无畸变响应波束形成方法[J]. 上海交通大学学报, 2016, **50**(2): 188-193.
ZHENG Enming, LI Yuansong, CHEN Xinhua, et al. Improved bearing resolution approach for MVDR beam-forming[J]. Journal of Shanghai Jiaotong University, 2016, **50**(2): 188-193.

[14] 马金龙, 曾庆宁, 龙超, 等. 多噪声环境下可懂度提升的助听器语音增强[J]. 计算机工程与设计, 2016, **37**(8): 2160-2164.
MA Jinlong, ZENG Qingning, LONG Chao, et al. Intelligibility improved speech enhancement for hearing aids in complex noise envrionment[J]. Computer Engineering and Design, 2016, **37**(8): 2160-2164.
[15] SCALART P, FILHO J V. Speech enhancement based on a prior signal to noise estimation[C]//IEEE International Conference on Acoustics, 1996, 629-632.
[16] 郭利华, 马建芬. 具有高可懂度的改进的维纳滤波的语音增强算法[J]. 计算机应用与软件, 2014, **31**(11): 155-157.
GUO Lihua, MA Jianfen. An improved wiener filtering speech enhancement algorithm with high intelligibility[J]. Computer Applications and Software, 2014, **31**(11): 155-157.
[17] 蒋晔, 唐振民. GMM 文本无关的说话人识别系统研究[J]. 计算机工程与应用, 2010, **46**(11): 179-182.
JIANG Ye, TANG Zhenmin. Research on GMM text-independent speaker recognition[J]. Computer Engineering and Applications, 2010, **46**(11): 179-182.
[18] 程小伟, 王健, 曾庆宁, 等. 基于调制域谱减法的鲁棒性说话人识别[J]. 科学技术与工程, 2017, **17**(3): 252-257.
CHENG Xiaowei, WANG Jian, ZENG Qingning, et al. Robust speaker recognition based on modulation domain spectral subtraction[J]. Science Technology and Engineering, 2017, **17**(3): 252-257.
[19] 余建潮, 张瑞林. 基于 MFCC 和 LPCC 的说话人识别[J]. 计算机工程与设计, 2009, **30**(5): 1189-1191.
YU Jianchao, ZHANG Ruilin. Speaker recognition method using MFCC and LPCC features[J]. Computer Engineering and Design, 2009, **30**(5): 1189-1191.
[20] 王玥, 钱志鸿, 王雪, 等.基于伽马通滤波器组的听觉特征提取算法研究[J].电子学报, 2010, **38**(3): 525-528
WANG Yue, QIAN Zhihong, WANG Xue, et al. An auditory feature extraction algorithm based on gammatone filter-banks[J]. Acta Electronica Sinica, 2010, **38**(3): 525-528.
[21] 林琳, 陈虹, 陈建. 基于鲁棒听觉特征的说话人识别[J]. 电子学报, 2013, **41**(3): 619-624.
LIN Lin, CHEN Hong, CHEN Jian. Speaker recognition based on robust auditory feature[J]. Acta Electronica Sinica, 2013, **41**(3): 619-624.

“中国超声诊断 60 周年学术大会暨周永昌超声医学讲坛”即将召开

超声技术在我国已较早地应用于医学领域，如理疗等。1959 年 9 月，上海市第六人民医院率先将军工声呐技术应用于人体检测，揭开了中国超声诊断的历史序幕。经全国几代超声人的不懈努力，已建成颇具中国特色的三级学科，并且也是医学影像学科中执业注册医师最多的学科。目前我国是超声诊断装机量最大、受益人群最多的国家。为传承我国诊断超声创业、协同攻关精神，展示各个时期的丰硕成果，不忘初心，面向未来，推进行业的进一步发展，特举办“中国超声诊断 60 周年学术大会暨周永昌超声医学讲坛”，并围绕热点、难点等设立分论坛。会议由 11 个全国性超声专业学会、协会和上海 4 个专业学会及科技部数字诊疗重点专项“超声成像综合评价及培训”项目办公室联合发起举办，上海交通大学附属第六人民医院、上海超声医学研究所主办。热诚欢迎国内外同道及厂家协同参会。

会议宗旨： 创新引领、传承融合

会议时间： 2018 年 9 月 27～30 日

会议注册费： 1200 元/人(含资料费、餐费，不含住宿费，参会者可获得国家级继续教育 I 类学分 10 分)
前 100 名报名者优惠价：1000 元/人

报名方式： 本次会议只接受网上报名
网址：上海超声医学研究所专用网站 http://www.sium.org.cn

联系人： 姜立新 18930173496
陈 磊 18930172789
诸晨琳 13918453953

《声学技术》编辑部