

基于两步单源点筛选的改进退化解混和估计算法

吴礼福^{1,2}, 马思佳¹, 孙康¹

(1. 南京信息工程大学电子与信息工程学院, 南京 210044; 2. 江苏省大气环境与装备技术协同创新中心, 南京 210044)

摘要: 退化解混和估计 (Degenerate unmixing estimation technique, DUET) 算法是一种典型的欠定盲源分离算法, 其采用的二进制时频掩蔽会保留部分干扰信号。提出了基于两步单源点筛选的改进 DUET 算法, 首先使用余弦角算法进行单源点筛选, 再采用计算相似度的方法进行第二步单源点筛选。通过两步单源点筛选获得更精确的目标信号和干扰信号后, 设计用于抵消干扰信号的滤波器取代 DUET 中的二进制时频掩蔽, 达到抑制干扰信号和提取目标信号的目的。仿真实验结果表明, 该方法在正定盲源分离和欠定盲源分离两种情况下都有较优的盲源分离性能。

关键词: 盲源分离; 退化解混和估计算法; 单源点筛选; 抵消核; 语音信号

中图分类号: TN912.3 **文献标志码:** A

Improved Degenerate Unmixing Estimation Technique Algorithm Based on Two-Step Single-Source Point Screening

WU Lifu^{1,2}, MA Sijia¹, SUN Kang¹

(1. School of Electronic & Information Engineering, Nanjing University of Information Science & Technology, Nanjing 210044, China; 2. Jiangsu Collaborative Innovation Center of Atmospheric Environment and Equipment Technology, Nanjing 210044, China)

Abstract: The degenerate unmixing estimation technique (DUET) algorithm is a typical underdetermined blind source separation algorithm. However, as a binary time-frequency mask-based method, DUET erroneously results in some interference signals retention. This paper proposes an improved DUET algorithm based on two-step single-source point screening. The cosine angle algorithm is used for the first-step single-source point screening, and then a similarity calculation method is employed for the second-step single-source point screening. After obtaining more accurate target and interference signals through two-step single-source point screening, the filter designed to cancel the interference signals replaces the binary time-frequency mask of DUET, achieving interference signal suppression and target signal extraction. Simulation results show that the proposed method has good performance in both positive definite and underdetermined blind source separation.

Key words: blind source separation; degenerate unmixing estimation technique (DUET) algorithm; single-source point screening; cancellation kernel; speech signal

引言

盲源分离(Blind source separation, BSS)是一种无需先验信息,能够通过观测得到的混合信号进行识别、分析并分离源信号的热门方法。近年来,在语音信号处理领域,BSS方法得到广泛研究应用^[1],但主要针对解决超定(观测得到的信号个数大于源信号个数)和正定(观测得到的信号个数等于源信号个数)问题^[2-4],而在现实场景中需要考虑多方面因素,欠定盲源分离(Underdetermined blind source separation, UBSS),即观测得到信号个数小于源信号个数的情况也是常见的,具有更大的技术难度。

在欠定盲源分离中,从语音信号在变换域的稀疏特性入手实现盲源分离是近年来的主要研究方向^[5]。例如,Hassan等^[6]提出一种自适应的方法,引入自适应时频阈值检测识别单源点(Single-source points, SSPs),也就是在某个时频点只有一个源信号占主导地位,而其余源信号占比很小以至于可忽略的点,增强了时频域的稀疏性,并使用最小二乘法(Least squares methods, LSM)恢复源信号,缩短了分离源的时间。Zhu等^[7]就欠定盲源分离方法中的混合矩阵估计问题展开研究,提出了一种结合余弦角算法和 L_1 范数优化算法的两阶段单源点筛选方法。第一阶段通过余弦角算法提取主要单源点,第二阶段从原始混合信号中通过 L_1 范数优化算法提取少量SSPs作为参考单源点,对其进行聚类并获取聚类中心作为参考中心,随后通过和参考中心的相似度对经过第一阶段筛选的单源点再次筛选,最后通过对经过两阶段筛选的SSPs进行聚类实现混合矩阵的估计,在保证准确混合矩阵估计的同时,也增强了混合矩阵估计的鲁棒性。

退化解混和估计(Degenerate unmixing estimation technique, DUET)算法^[8]作为一种高效的BSS方法,使用短时傅里叶变换(Short-time Fourier transform, STFT)将语音信号从时域变换到时频域,并根据语音信号在时频域的稀疏性,认为语音信号在时频域近似符合窗口互斥正交(W-disjoint orthogonality, W-DO)性^[9],随后通过二进制时频掩蔽实现了欠定情况下的盲源分离。由于DUET算法计算复杂度低,更适用于实际情况,引起了许多学者的关注。考虑到W-DO的假设与时频域的稀疏度有关,陈斌杰^[10]提出多分辨率共同命运变换(Multi-resolution common fate transform, MCFT)代替STFT,提高时频域稀疏度,并且结合Gammatone滤波器组将DUET算法中原始的二进制时频掩蔽进行平滑处理,从而更好地分离出源信号。Abdulla等^[11]使用双树复小波变换(Dual-tree complex wavelet transform, DTCWT)作为预处理器,同步提取变换(Synchro-extracting transform, SET)作为后处理器,结合DUET算法,提高了UBSS的抗噪性能。另外,DUET算法的混合参数对估计不准确以及二进制时频掩蔽自身的局限性,都会造成分离性能的降低。Lian等^[12]针对在混合语音信号中出现瞬时信号或混合语音信号信噪比低两种情况下,传统DUET算法通过二维直方图峰值估计混合参数对不准确这一问题,提出在建立子直方图的基础上,分别使用最大分布和方差分布两种方法,更精确地估计了混合参数对,提高了分离的准确性。孙庆强^[13]提出采用分水岭算法并结合一种自适应阈值策略进行混合参数对的估计,提高了混合参数对估计的精度,同时在二进制时频掩蔽的基础上引入按比例“软分配”策略,实现了分离性能的提高。He等^[14]提出LSFs-DUET算法,在DUET算法的基础上构造多个线性空间滤波器(Linear spatial filters, LSFs)代替二进制时频掩蔽,提高了UBSS的干扰抑制比。

近年来,基于深度学习的UBSS方法也取得了很大的进展,Luo等^[15]提出了通过同步变换-平均-拼接(Transform-average-concatenate, TAC)模块结合滤波求和网络(Filter-and-sum network, FasNet),提高了在未知麦克风数量和位置情况下的分离性能,使得分离系统应用更加广泛。Gruttadauria等^[16]提出了一种新的语音分离引导日志(Speech separation-guided diarization, SSGD),通过全卷积时域语音分离网络(Convolutional time-domain audio separation network, ConVer TasNet)和基于时域的双路径循环神经网络(Dual-path recurrent neural network, DPRNN)分别进行语音分离,然后对分离出的语音信号进行语音活动检测(Voice activity detection, VAD),最后通过嵌入和聚类的方法对语音信号进行标记、分

配和拼接,有效地解决了盲源分离中语音重叠的问题。

本文研究了一种对单源点进行两步筛选的改进DUET算法。首先,考虑到在传统DUET算法中二进制时频掩蔽过于绝对,故采用抵消的思想以去除干扰信号,改进目标信号的提取质量;其次,通过在时频域进行单源点的两步筛选来获取干扰信号和目标信号的信息:第一步使用余弦角算法进行单源点筛选,随后为消除第一步单源点筛选后仍保留的干扰点和偏差点,采用计算相似度的方法进行第二步单源点筛选。

1 算法描述

1.1 DUET算法

DUET算法解决2个麦克风收集到 N 个源信号的盲源分离问题,其基本模型如图1所示,其中 $x_1(t)$ 表示麦克风1收集到的混合语音信号, $x_2(t)$ 表示麦克风2收集到的混合语音信号, $s_j(t)$, $j \in 1, 2, \dots, N$ 表示第 j 个源发出的信号。

为了降低计算复杂度,采用线性无回声混合模型^[13],忽略了回声效应和非线性因素等复杂因素,但考虑到源信号到两个麦克风的距离不同,引入了时延,表达式为

$$\begin{cases} x_1(t) = \sum_{j=1}^N s_j(t) \\ x_2(t) = \sum_{j=1}^N \alpha_j s_j(t - \delta_j) \end{cases} \quad (1)$$

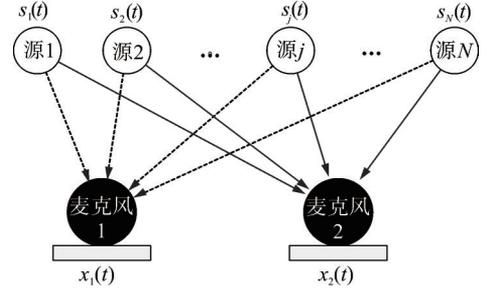


图1 DUET模型

Fig.1 DUET model

式中: $t=1, 2, \dots, T$ 表示离散时刻;以麦克风1作为参考信

号, α_j 和 δ_j 分别表示针对第 j 个源信号,麦克风2采集到的语音信号相对于麦克风1采集到的语音信号的相对衰减和相对时延。随后通过STFT将信号从时域变换到时频域, $s_j(t)$ 的STFT变换表示为

$$S_j(\tau, \omega) = \mathcal{F}^W[s_j(t)](\tau, \omega) \quad (2)$$

式中: $S_j(\tau, \omega)$, $j \in 1, 2, \dots, N$ 表示第 j 个源信号经过STFT后的时频域表达, $\mathcal{F}^W[\cdot]$ 为STFT变换, τ 表示时间帧索引, ω 表示离散频率点索引。

根据语音信号具有局部平稳性,选取合适的STFT窗长度后,式(1)中的 $\alpha_j s_j(t - \delta_j)$ 可以近似表示为

$$\mathcal{F}^W[\alpha_j s_j(t - \delta_j)](\tau, \omega) \approx \alpha_j e^{-i\omega\delta_j} \cdot \mathcal{F}^W[s_j(t)](\tau, \omega) \quad j \in 1, 2, \dots, N \quad (3)$$

结合式(2,3)可以得到时频域中的无回声混合数学表达式为

$$\begin{bmatrix} X_1(\tau, \omega) \\ X_2(\tau, \omega) \end{bmatrix} = \begin{bmatrix} 1 & \dots & 1 & \dots & 1 \\ \alpha_1 e^{-i\omega\delta_1} & \dots & \alpha_j e^{-i\omega\delta_j} & \dots & \alpha_N e^{-i\omega\delta_N} \end{bmatrix} \begin{bmatrix} S_1(\tau, \omega) \\ \vdots \\ S_j(\tau, \omega) \\ \vdots \\ S_N(\tau, \omega) \end{bmatrix} \quad j \in 1, 2, \dots, N \quad (4)$$

式中: $X_1(\tau, \omega)$ 、 $X_2(\tau, \omega)$ 分别对应两个麦克风收集到的混合语音信号 $x_1(t)$ 、 $x_2(t)$ 经过STFT后的时频表示。

由于语音信号在时频域中的能量分布具有稀疏性,通常情况下,多个源信号不会在同一个时频点具有较大能量。换句话说,针对大多数时频点,存在一个源信号的系数远大于其他源信号,即

$$\Omega_j = \{(\tau, \omega) \mid |S_j(\tau, \omega)| \gg |S_k(\tau, \omega)|, \forall k \neq j\} \quad j \in 1, 2, \dots, N \quad (5)$$

式中: Ω_j 为第 j 个源信号占主导地位的组合。DUET算法假设语音信号近似符合W-DO特性,即对任意属于 Ω_j 的时频点,只有一个源信号占主导地位,多个源信号的作用域互不重叠^[17],因此式(4)可简化为

$$\begin{bmatrix} X_1(\tau, \omega) \\ X_2(\tau, \omega) \end{bmatrix} = \begin{bmatrix} 1 \\ \alpha_j e^{-i\omega\delta_j} \end{bmatrix} S_j(\tau, \omega) \quad j \in 1, 2, \dots, N \quad (6)$$

式中: j 表示在该时频点占主导地位所对应源序号。观察式(6)可知,相对衰减 α_j 和相对时延 δ_j 可以通过两个混合语音信号的比间接获得,因此根据式(7)计算每个时频点的相对衰减-相对时延混合参数对 $(\tilde{\alpha}(\tau, \omega), \tilde{\delta}(\tau, \omega))$ 为

$$(\tilde{\alpha}(\tau, \omega), \tilde{\delta}(\tau, \omega)) = \left(\left| \frac{X_2(\tau, \omega)}{X_1(\tau, \omega)} \right|, \angle \left(\frac{X_2(\tau, \omega)}{X_1(\tau, \omega)} \right) \right) \quad (7)$$

式中: $|\cdot|$ 为取复数的模值, $\angle(\cdot)$ 为取复数的角。

将所有时频点的相对衰减-相对时延混合参数对结合最大似然(Maximum likelihood, ML)法构建二维直方图^[18],估计源信号混合参数对,假定有 N 个峰值,通过 N 个峰值估计源信号的混合参数对,即 $(\tilde{\alpha}_j(\tau, \omega), \tilde{\delta}_j(\tau, \omega)), j \in 1, 2, \dots, N$,然后通过式(8)求得每个时频点到峰值的接近度,并寻找最小值。

$$\Lambda_j = \left\{ (\tau, \omega) \mid j = \arg \min_j \frac{|\tilde{\alpha}_j e^{-i\omega\tilde{\delta}_j} X_1(\tau, \omega) - X_2(\tau, \omega)|^2}{1 + \tilde{\alpha}_j^2} \right\} \quad (8)$$

根据式(8)创建二进制时频掩蔽 $\tilde{M}_j(\tau, \omega), j \in 1, 2, \dots, N$,表达式为

$$\tilde{M}_j(\tau, \omega) = \begin{cases} 1 & (\tau, \omega) \in \Lambda_j \\ 0 & \text{其他} \end{cases} \quad (9)$$

最后,将麦克风1收集到的混合信号 $X_1(\tau, \omega)$ 分离,表达式为

$$\tilde{S}_j(\tau, \omega) = \tilde{M}_j(\tau, \omega) X_1(\tau, \omega) \quad (10)$$

1.2 DUET-ISR

在DUET算法中,W-DO的假设对于分离源太过绝对,导致采用式(9)的二进制时频掩蔽时会保留部分干扰信号,造成分离性能下降,本文通过抵消核思想^[19]在每个时频点获取干扰信号的信息来设计滤波器,并去除干扰信号,通过求解式(11)的优化问题得到干扰抑制响应(Interference suppression response, ISR)滤波器 \mathbf{w} 为

$$\begin{cases} \min_{\mathbf{w}} \mathbf{w}^H \sum_j \mathbf{w} \\ \text{s.t. } \omega_1 = 1 \end{cases} \quad (11)$$

式中 $(\cdot)^H$ 表示为矩阵的共轭转置运算, \sum_j 表示为

$$\sum_j = \frac{1}{\Gamma} \sum_{\tau=1}^{\Gamma} \mathbf{u}_j(\tau, \omega) \mathbf{u}_j^H(\tau, \omega) \quad \tau \in \{1, 2, \dots, \Gamma\} \quad (12)$$

式中: $\mathbf{u}_j = \sum_{i=1, i \neq j}^N \alpha_i S_i(t - \delta_i)$ 为第 j 个源信号对应的干扰信号样本协方差矩阵, \mathbf{w} 中的 $\omega_1 = 1$ 即保留第1个麦克风收集到的信号作为参考信号,以此来求解每个源信号对应的干扰抑制响应滤波器权值。

由于在盲信号分离中,干扰信号是未知的,不能直接求得 \mathbf{u}_j ,因此可以通过筛选单源点间接求得,如式(13)所示,采用DUET算法中求得的接近度直接筛选SSPs的方法

$$\Lambda_j = \left\{ (\tau, \omega) \mid j = \arg \min_j \frac{|\tilde{\alpha}_j e^{-i\omega\tilde{\delta}_j} X_1(\tau, \omega) - X_2(\tau, \omega)|^2}{1 + \tilde{\alpha}_j^2} \right\} < \mu \quad j \in 1, 2, \dots, N \quad (13)$$

式中： μ 为给定接近度误差。由于这种方法并未直接进行单源点的筛选计算，而是根据接近度判断，故将这种方法称为DUET-ISR。

通过DUET-ISR筛选的单源点会保留部分接近度相近的干扰信号，故提出采用两步单源点筛选方法，与上述方法作区分，将此方法称为DUET-SSPs。如果某个时频点 (τ, ω) 是单源点，则观测信号 $X(\tau, \omega)$ 在时频域实部和虚部方向是一致的^[20]，故第一步单源点筛选通过余弦角算法实现，如式(14)所示。

$$\rho(\tau, \omega) = \left| \frac{R\{X(\tau, \omega)\}^T I\{X(\tau, \omega)\}}{\|R\{X(\tau, \omega)\}\| \|I\{X(\tau, \omega)\}\|} \right| > \cos(\Delta\theta) \quad (14)$$

式中： $R\{\cdot\}$ 表示实部， $I\{\cdot\}$ 表示虚部， $|\cdot|$ 表示绝对值， $\|\cdot\|$ 表示模值， $\cos(\Delta\theta)$ 为设定的单源点筛选阈值，当给定时频点的 $\rho(\tau, \omega)$ 大于筛选阈值，将该点视为单源点。

文献[7]指出余弦角算法存在局限性，运用到单源点筛选后会存在一定的偏差点和干扰点，故提出第二步单源点筛选，如式(15)所示，比较经过第一步单源点筛选后的单源点所对应的观测信号和各峰值中心点所对应的观测信号的相似度。

$$D_j = 1 - \frac{X_i(\tau, \omega) X_j(\tau, \omega)}{\|X_i(\tau, \omega)\| \|X_j(\tau, \omega)\|} < \sigma \quad (15)$$

式中： $X_j(\tau, \omega)$ ， $j=1, 2, \dots, N$ 为DUET算法中通过二维直方图选取的第 j 个峰值中心点对应的观测信号， $X_i(\tau, \omega)$ 为除 $X_j(\tau, \omega)$ 外，经过第一步单源点筛选确定的单源点观测信号，即符合式(15)的单源点， σ 为设定的筛选阈值。

当时频点同时满足式(14,15)时，筛选为单源点，从而间接确定干扰信号，最后由ISR滤波输出第 j 个源信号，如式(16)所示。

$$\tilde{S}_j(\tau, \omega) = \mathbf{w}_j^H(\tau, \omega) X(\tau, \omega) \quad (16)$$

式中： $(\cdot)^H$ 表示矩阵的共轭转置运算， $\mathbf{w}_j(\tau, \omega)$ 为第 j 个源对应的滤波器权值， $X(\tau, \omega)$ 为两个麦克风采集的混合信号时频域表示。DUET-SSPs的算法流程如表1所示。

表1 DUET-SSPs算法
Table 1 DUET-SSPs algorithm

输入： $x_1(t), x_2(t)$ ，源个数 N ， $\cos(\Delta\theta)$ ， σ ，输出： N 个分离信号
(1) 对输入信号进行STFT；
(2) 在式(7)的基础上结合ML法建立二维直方图；
(3) 在直方图上找到 N 个峰值，并通过式(8)计算接近度，同时得到峰值中心点对应的 $X_j(\tau, \omega)$ ；
(4) 通过式(14, 15)进行两步单源点筛选；
(5) 筛选出属于第 j 个源的单源点，间接获得干扰信号并代入式(11, 12)，从而估计第 j 个源对应的ISR滤波器权值；
(6) 通过式(16)从混合信号中分离出第 j 个源信号，最后进行逆短时傅里叶变换，输出时域分离源信号。

2 仿真实验

2.1 实验设置

实验中使用的语音源信号来自于SiSEC2011数据集^[21]中的dev1开发数据库 (https://www.irisa.fr/metiss/SiSEC11/underdetermined/underdetermined_dev1_all.html)，选取female4文件和male4文件，

混合语音信号参考文献[14]提供的混合语音信号(<https://ydcnanhe.github.io/demo-icassp2022/>),包括语音源信号个数 $N=2,3,4$,在混响时间 130 ms 和 250 ms 的会议室现场录制获得的混合语音信号,混合语音信号采样率为 16 kHz,会议室的尺寸为 4.45 m×3.55 m×2.5 m,其中 4 个扬声器和两个麦克风中心相距 1 m,两个麦克风的间距为 5 cm,扬声器的方位角分别为 $-50^\circ, -10^\circ, 15^\circ$ 和 45° 。

实验选取 $N=2,3,4$ 三种情况,当 $N=2$,即正定盲源分离,比较了 4 种算法:DUET、独立低秩矩阵分析(Independent low-rank matrix analysis, ILRMA)^[22]、DUET-ISR 和 DUET-SSPs;当 $N=3,4$,即欠定盲源分离,比较了 4 种算法:DUET、DUET-ISR、DUET-SSPs 和多通道非负矩阵分解(Multichannel non-negative matrix factorization, MULTINMF)^[23],已有算法的参数根据原始文献设置,分别如表 2、3 所示,其中 STFT 变换均采用 Hann 窗。

表 2 $N=2$ 时实验参数设置

Table 2 Parameters setting with simulation environment of $N=2$

算法	窗长	帧移	基的数量	迭代次数	μ	$\cos(\Delta\theta)$	σ
DUET	1 024	256					
DUET-ISR	1 024	256			0.05		
DUET-SSPs	1 024	256				0.6	0.001
ILRMA	4 096	2 048	2	200			

表 3 $N=3,4$ 时实验参数设置

Table 3 Parameters setting with simulation environment of $N=3,4$

算法	窗长	帧移	源的组件个数	μ	$\cos(\Delta\theta)$	σ
DUET	1 024	256				
DUET-ISR	1 024	256		0.05		
DUET-SSPs	1 024	256			0.7,0.9	0.001
MULTINMF	2 048	512	10			

采用盲源分离评估工具包(Blind source separation evaluation toolbox, BSS_EVAL toolbox)^[24]评估分离后的语音源信号,选取源失真比(Source-to-distortion ratio, SDR)、源干扰比(Source-to-interference ratio, SIR)、源伪影比(Source-to-artifacts ratio, SAR)作为分离的语音信号评估指标,具体计算公式分别如下

$$\tilde{s} = s_{\text{target}}(t) + e_{\text{interf}}(t) + e_{\text{noise}}(t) + e_{\text{target}}(t) \quad (17)$$

$$\text{SDR} = 10 \lg \frac{\|s_{\text{target}}(t)\|^2}{\|e_{\text{interf}}(t) + e_{\text{noise}}(t) + e_{\text{target}}(t)\|^2} \quad (18)$$

$$\text{SIR} = 10 \lg \frac{\|s_{\text{target}}(t)\|^2}{\|e_{\text{interf}}(t)\|^2} \quad (19)$$

$$\text{SAR} = 10 \lg \frac{\|s_{\text{target}}(t) + e_{\text{interf}}(t) + e_{\text{noise}}(t)\|^2}{\|e_{\text{target}}(t)\|^2} \quad (20)$$

式(17)表示将分离出的待评估语音信号 \tilde{s} 分为 4 部分,其中: s_{target} 为目标信号,是估计信号中与参考信号匹配的部分, e_{interf} 为干扰分量,是来自其他源信号的干扰信号, e_{noise} 为噪声分量, e_{target} 为伪影分量,是在信号分离过程中引入的伪影。式(18)为 SDR 的计算公式,反映了总体的分离效果。式(19)为 SIR 的计算

公式,反映了分离算法对干扰信号的抑制性能。式(20)为SAR的计算公式,反映了算法在分离过程中引入的伪影量。SDR、SIR、SAR的值越高,表明算法分离语音信号的能力越强,同时使用语音感知质量评估(Perceptual evaluation of speech quality, PESQ)^[25]针对分离后语音信号进行质量评测,PESQ得分越高,表明分离出的语音信号质量越好。

2.2 结果和讨论

图2给出了 $N=2$ 的盲源分离性能对比,即源信号个数为2的盲源分离性能对比,比较了DUET、ILRMA、DUET-ISR和DUET-SSPs四种算法的性能。其中图2(a)为混响时间130 ms和混响250 ms情况下的SDR性能对比,可以看出在正定盲源分离的情况下,引入ISR滤波器后,DUET-ISR对比DUET、ILRMA,SDR的数值有明显上升,约2.95 dB。DUET-SSPs在增加单源点筛选过程后分离效果也有明显提高,在DUET-ISR基础上增加约1.15 dB;在混响时间为250 ms条件下,DUET-SSPs相比DUET有3.75 dB的提高,相比ILRMA提高了5.46 dB,与DUET-ISR比较SDR增加0.43 dB。图2(b)为混响时间130 ms和混响时间250 ms情况下的SIR对比,与DUET、ILRMA、DUET-ISR相比,DUET-SSPs在两种混响时间下抑制干扰信号的能力都更优秀。图2(c)为SAR数值对比,同样可以看出在两种混响情况下DUET-SSPs相较于其他方法,源伪影比有一定提升。

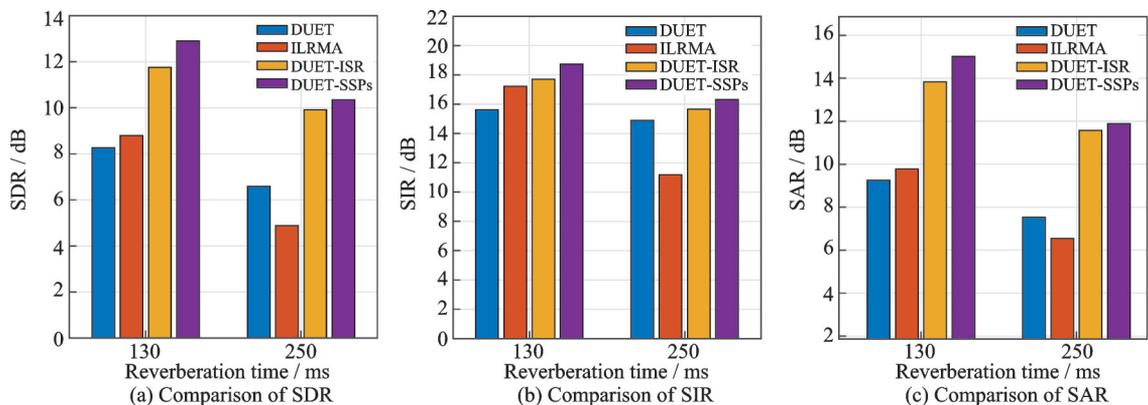


图2 130 ms和250 ms混响时间下两个声源混合的性能对比

Fig.2 Performance comparison of two-source mixture in the presence of 130 ms and 250 ms reverberation time

图3为源信号个数 $N=3$ 的欠定盲源分离性能对比,比较了DUET、MULTINMF、DUET-ISR和DUET-SSPs四种算法的性能。图3(a)、图3(b)和图3(c)分别给出了两种混响情况下的SDR、SIR和SAR对比,可以看出DUET-SSPs相比其他方法,在3个评比指标中都有优势。在250 ms混响环境下,DUET-SSPs相较于DUET、MULTINMF、DUET-ISR,分离效果更显著。同时通过比较SIR数值可以看出,相较于DUET-ISR,在混响时间为130 ms时,DUET-SSPs提高了约2.12 dB,在混响时间为250 ms时提高了约1.88 dB,即所提出的DUET-SSPs对干扰信号的抑制能力较强,这与进行两步单源点筛选的目的相一致:通过筛选出单源点,确定干扰信号并抵消。

图4为源信号个数 $N=4$ 的欠定盲源分离性能对比。图4(a)、图4(b)和图4(c)分别给出了两种混响情况下的4个声源混合的SDR、SIR和SAR数值对比,同样比较了DUET、MULTINMF、DUET-ISR和DUET-SSPs四种算法的分离混合语音信号的性能。从图中可以看出,相比DUET的二进制时频掩蔽,通过ISR进行干扰信号抑制后,盲源分离效果在各指标上有显著提高。

同时对比图3(b)和图4(b)可以看出,在不同混响时间下源个数的增加也对干扰抑制比有一定影响,通过接近度进行单源点筛选的DUET-ISR,对干扰信号的抑制能力明显下降,而所提出的

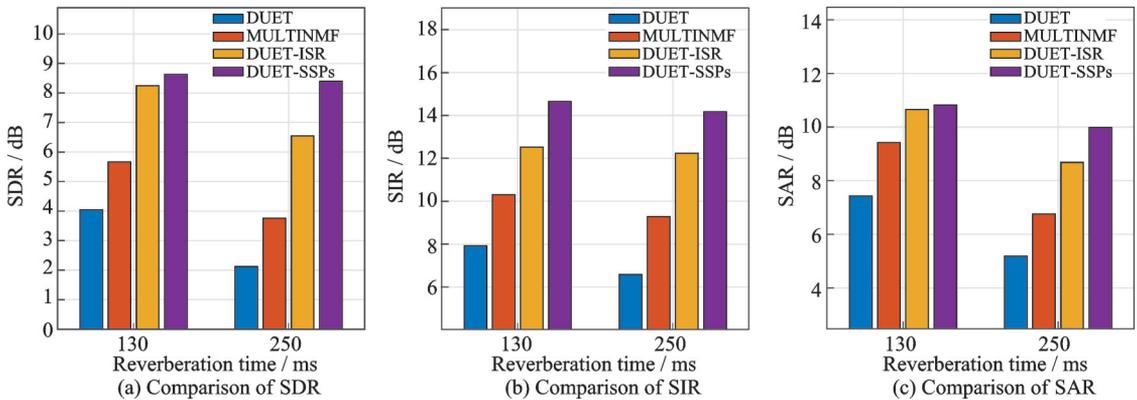


图3 130 ms和250 ms混响时间下3个声源混合的性能对比

Fig.3 Performance comparison of three-source mixture in the presence of 130 ms and 250 ms reverberation time

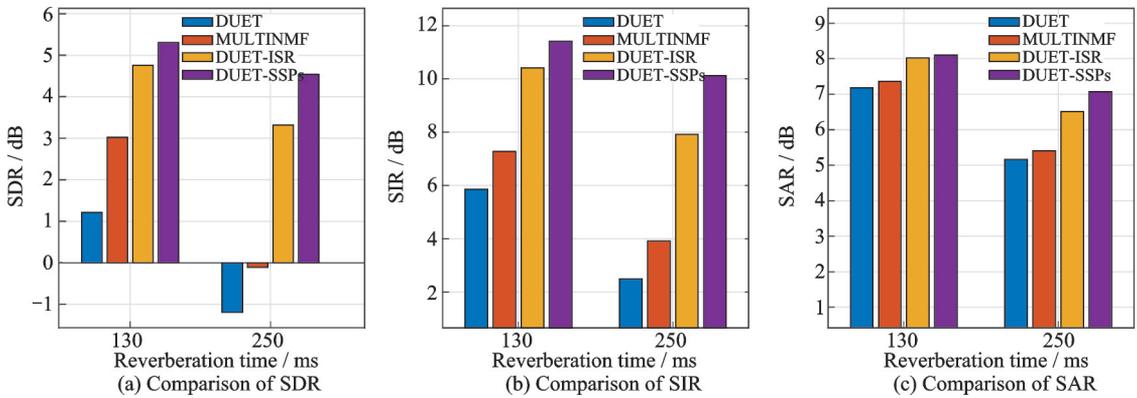


图4 130 ms和250 ms混响时间下4个声源混合的性能对比

Fig.4 Performance comparison of four-source mixture in the presence of 130 ms and 250 ms reverberation time

DUET-SSPs通过两步法进行单源点筛选后,受到的影响相对较小,能较精准地筛选单源点,从而抑制干扰信号。

图5展示了源信号个数 $N=4$ 、混响时间250 ms的混合信号经过DUET、MULTINMF、DUET-ISR和DUET-SSPs四种算法分离出的同一个源信号的语谱图对比。图5(a)为混合信号语谱图,图5(b)为纯净的语音信号语谱图,图5(c)为采用DUET分离的语音信号语谱图,图5(d)为采用MULTINMF分离的语音信号语谱图,图5(e)为采用DUET-ISR分离的语音信号语谱图,图5(f)为采用DUET-SSPs分离的语音信号语谱图。通过对比可以看出,DUET-SSPs分离出的语音信号语谱图更接近纯净语音信号语谱图,说明本文提出的改进算法在UBSS方面能够较好地实现源分离任务。同时比较图5(e)和图5(f)的高频段可以看出,在抑制干扰信号方面,DUET-SSPs相比DUET-ISR,通过两步对单源点进行筛选后,能够对干扰信号起到更好的抑制作用。

图6展示了两种混响情况下,不同源个数的混合信号经过DUET、ILRMA、MULTINMF、DUET-ISR、DUET-SSPs算法分离后源信号的平均PESQ得分,即对分离后的每个语音信号都进行PESQ后所计算得出的平均值。在声源个数 $N=2$ 时,比较了DUET、ILRMA、DUET-ISR和DUET-SSPs;在声源个数 $N=3,4$ 时,比较了DUET、MULTINMF、DUET-ISR和DUET-SSPs四种算

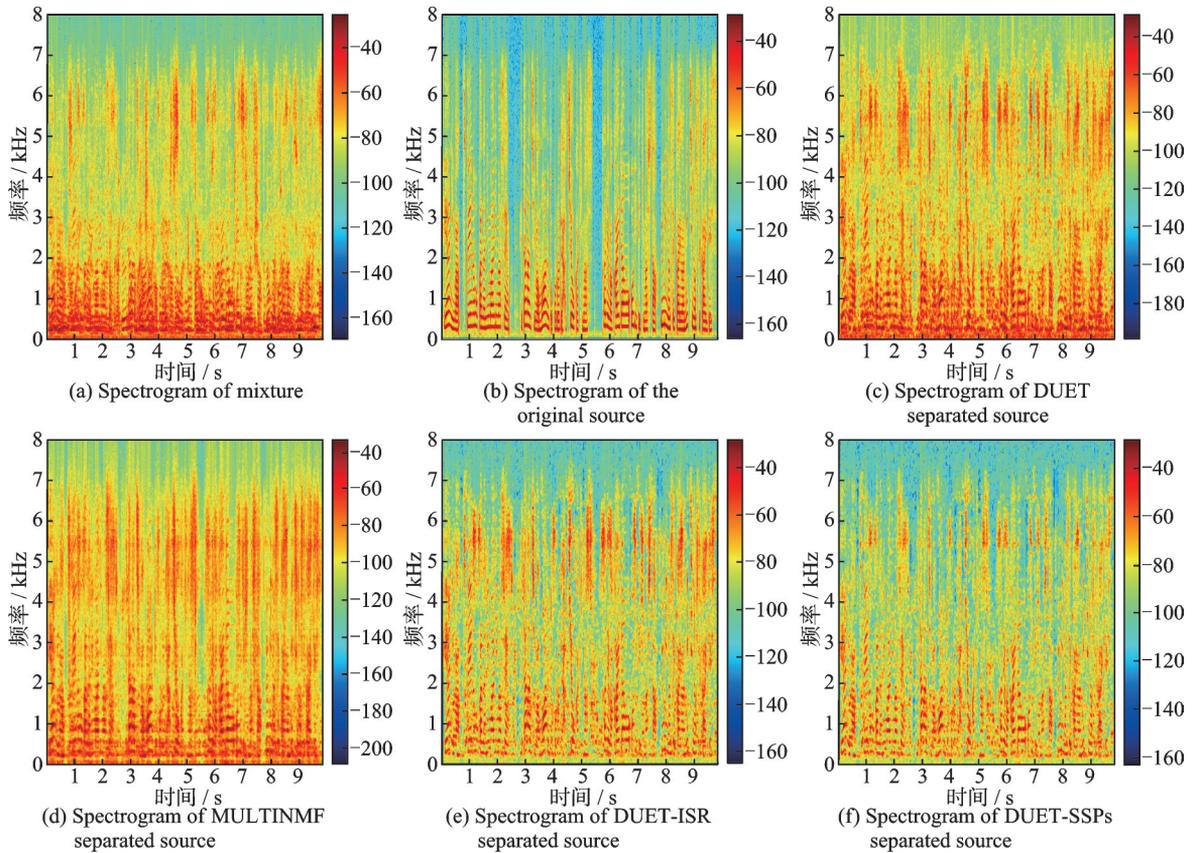


图5 250 ms混响时间下4个语音信号混合分离出的语音信号语谱图对比

Fig.5 Comparison of a separated source spectrograms of four-source mixture in the presence of 250 ms reverberation time

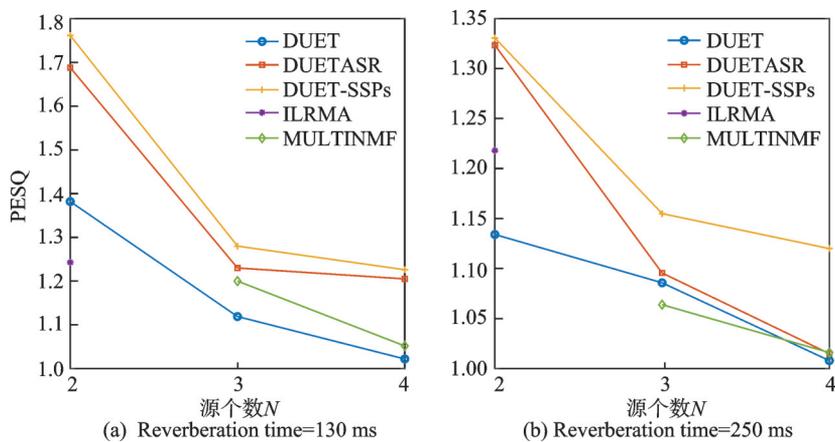


图6 130 ms和250 ms混响时间下不同声源个数的PESQ得分

Fig.6 PESQ scores for different numbers of source mixture in the presence of 130 ms and 250 ms reverberation time

法。图6(a)为混响时间130 ms,图6(b)为混响时间250 ms。可以看出,基于两步单源点筛选的DUET-SSPs在两种混响时间下,语音质量相较于其他方法都更好。

图7是对单源点分别进行一步筛选和两步筛选后,进行盲源分离的性能比较,图7(a)为源信号个数为4的分离指标对比图,图7(b)为源信号个数为3的分离指标对比图,图7(c)为源信号个数为2的分离指标对比图。通过对比各数据,可以看出在不同源个数或者在不同混响时间的情况下,经过两步单源点筛选后的盲源分离效果都较好。同时通过对比可以看出SIR数值有一定的提高,即干扰抑制比提高,这也说明通过第二步单源点筛选筛选掉了一部分第一步单源点筛选后仍存在的偏差点和干扰点,即对单源点的筛选更准确,从而能够更好地实现对干扰信号的抑制。

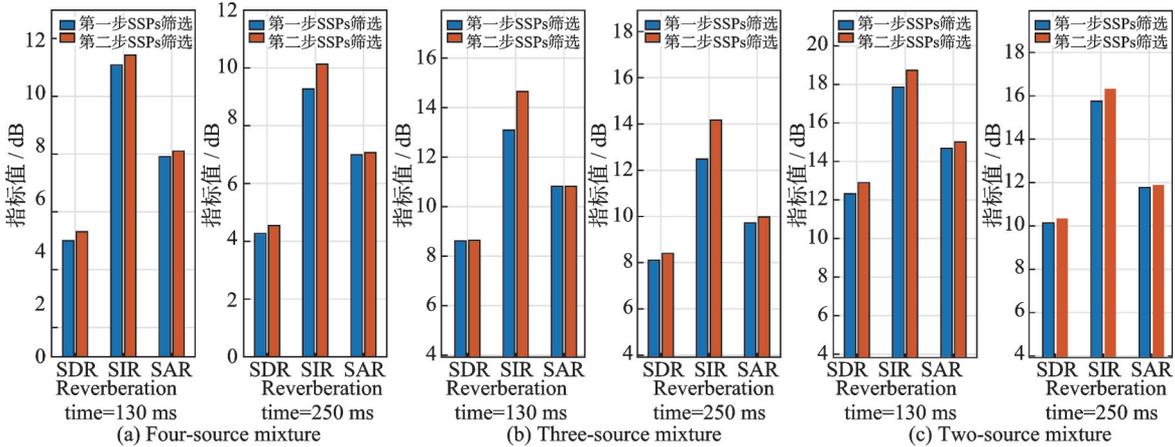


图7 130 ms 和 250 ms 混响时间下不同混合源个数的两次单源点筛选的性能对比

Fig.7 Average BSS performance comparison of different numbers of source mixture in the presence of 130 ms and 250 ms reverberation time

文献[20]提出对于包含单源点的混合信号,在时频域单源点会出现线性聚类,为了更清晰地对比两步单源点的筛选效果,图8展示了混合信号进行两步单源点筛选时的时频域散点图,实验选用的源个数 $N=3$,图8(a)为原始混合信号的散点图,图8(b)为进行第一步单源点筛选后的散点图,图8(c)为进行第二步单源点筛选之后的散点图。可以看出经过第一步单源点筛选后,有一部分点可以很好地在特定方向聚集,但仍然有一些偏差点和干扰点,而进行第二步单源点筛选后,进一步减少了一部分干扰点和偏差点的存在,混合信号的散点图更加清晰紧凑。

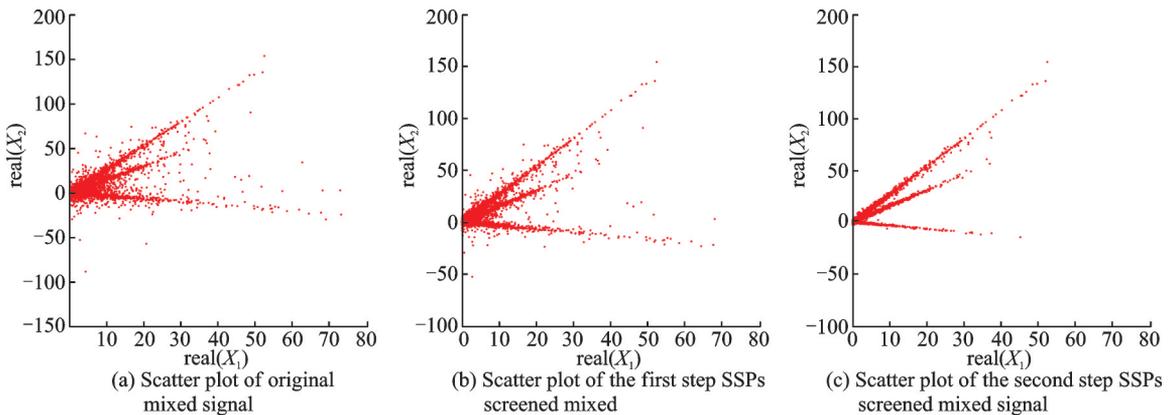


图8 混合信号散点图

Fig.8 Scatter plot of mixed signal

3 结束语

本文研究了一种基于两步单源点筛选的DUET改进算法。不同于DUET算法中的二进制时频掩蔽,本文利用抵消思想,根据干扰信号设计滤波器,进行干扰信号的抑制。为获得干扰信号的准确信息,通过两步法进行单源点的筛选,第一步采用余弦角算法筛选单源点,第二步将筛选出来的单源点与DUET算法中确定的峰值中心进行相似度比较,去除第一步遗留下的偏差点和干扰点,随后通过筛选单源点间接获得的干扰信号信息,计算滤波器的权值,实现源信号的分离。仿真实验结果表明,面对源个数不同的混合语音信号,相较已有的BSS算法,所提出基于两步单源点筛选的改进DUET方法能更有效地解决欠定盲源分离问题,且分离的语音质量也相对较高,同时在正定盲源分离情况下也有一定的优势。

参考文献:

- [1] GANNOT S, VINCENT E, MARKOVICH-GOLAN S, 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.
- [2] WANG L, CAVALLARO A. A blind source separation framework for ego-noise reduction on multi-rotor drones[J]. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 2020, 28: 2523-2537.
- [3] ALBATAINEH Z, SALEM F M. A robust ICA-based algorithmic system for blind separation of convolutive mixtures[J]. *International Journal of Speech Technology*, 2021, 24: 701-713.
- [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 to blind source separation[J]. *Journal of Data Acquisition and Processing*, 2020,35(5): 910-919.
- [5] 李雅凤.基于时频稀疏性的欠定盲源语音信号分离技术研究[D].成都:电子科技大学,2016.
LI Yafeng. Research on underdetermined blind speech separation based on sparsity in time-frequency domain[D]. Chengdu: School of Electronic Engineering, 2016.
- [6] HASSAN N, RAMLI D A. Sparse component analysis (SCA) based on adaptive time-frequency thresholding for underdetermined blind source separation (UBSS)[J]. *Sensors*, 2023, 23(4): 2060.
- [7] ZHU Zhanyu, CHEN Xingjie, LV Zhaomin. Underdetermined blind source separation method based on a two-stage single-source point screening[J]. *Electronics*, 2023, 12(10): 2185.
- [8] JOURJINE A, RICKARD S, YILMAZ O. Blind separation of disjoint orthogonal signals: Demixing N sources from 2 mixtures[C]//*Proceedings of 2000 IEEE International Conference on Acoustics, Speech, and Signal Processing*. [S. l.]: IEEE, 2000: 2985-2988.
- [9] RICKARD S. The DUET blind source separation algorithm[M]. Dordrecht: [s. n.], 2007: 217-241.
- [10] 陈斌杰.基于双麦克风的室内语音分离与声源定位研究[D].宁波:宁波大学,2019.
CHEN Binjie. Indoor speech separation and sound source localization based on dual microphone[D].Ningbo: Ningbo University, 2019.
- [11] ABDULLA S M, JAYAKUMARI J. Improving time-frequency sparsity for enhanced audio source separation in degenerate unmixing estimation technique algorithm[J]. *Journal of Control and Decision*, 2022, 9(4): 502-515.
- [12] LIAN K Y, LIN J H. Unique methods for determining the attenuation and delay in blind source separation based on the degenerate unmixing estimation technique[J]. *IEEE Access*, 2021, 9: 129460-129470.
- [13] 孙庆强.基于稀疏性的欠定语音盲分离方法研究[D].南京:南京大学,2018.
SUN Qingqiang. Research on underdetermined speech blind separation based on sparsity[D]. Nanjing: Nanjing University, 2018.
- [14] HE Yudong, HE Wang, CHEN Qifeng, et al. Harvesting partially-disjoint time-frequency information for improving degenerate unmixing estimation technique[C]//*Proceedings of 2022 IEEE International Conference on Acoustics, Speech and*

- Signal Processing (ICASSP). [S. l.]: IEEE, 2022: 506-510.
- [15] LUO Y, HAN C, MESGARANI N, et al. FaSNet: Low-latency adaptive beamforming for multi-microphone audio processing [C]//Proceedings of 2019 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU). [S.l.]:IEEE, 2019: 260-267.
- [16] GRUTTADAURIA E, FONTAINE M, ESSID S. Online speaker diarization of meetings guided by speech separation[C]// Proceedings of 2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). [S. l.] IEEE, 2024: 11356-11360.
- [17] 陈纯增. 基于宽间距麦克风阵列的多声源测向方法研究[D]. 南京: 南京理工大学, 2017.
CHEN Chunzeng. Research on the direction-of-arrival estimation method for multiple sound sources based on wide-spaced microphone arrays [D]. Nanjing: Nanjing University of Science and Technology, 2017.
- [18] YILMAZ O, RICKARD S. Blind separation of speech mixtures via time-frequency masking[J]. IEEE Transactions on Signal Processing, 2004, 52(7): 1830-1847.
- [19] WANG Yang, ZHOU Zhengfang. Source extraction in audio via background learning[J]. Inverse Problems & Imaging, 2013, 7 (1): 283.
- [20] REJU V G, KOH S N, SOON Y. An algorithm for mixing matrix estimation in instantaneous blind source separation[J]. Signal Processing, 2009, 89(9): 1762-1773.
- [21] ARAKI S, NESTA F, VINCENT E, et al. The 2011 signal separation evaluation campaign (SiSEC2011): -audio source separation [C]//Latent Variable Analysis and Signal Separation. [S. l.]: Springer Berlin Heidelberg, 2012: 414-422.
- [22] KITAMURA D, ONO N, SAWADA H, et al. Determined blind source separation unifying independent vector analysis and nonnegative matrix factorization[J]. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2016, 24(9): 1626-1641.
- [23] OZEROV A, FÉVOTTE C. Multichannel nonnegative matrix factorization in convolutive mixtures for audio source separation [J]. IEEE Transactions on Audio, Speech, and Language Processing, 2009, 18(3): 550-563.
- [24] FÉVOTTE C, GRIBONVAL R, VINCENT E. BSS_EVAL toolbox user guide—Revision 2.0[Z]. 2005.
- [25] RIX A W, BEERENDS J G, HOLLIER M P, et al. Perceptual evaluation of speech quality (PESQ)—A new method for speech quality assessment of telephone networks and codecs[C]//Proceedings of 2001 IEEE International Conference on Acoustics, Speech, and Signal Processing. [S. l.]: IEEE, 2001: 749-752.

作者简介:



吴礼福(1981-), 通信作者, 男, 工学博士, 教授, 博士生导师, 研究方向: 噪声控制, 语音去混响、识别、增强, 传声器阵列及声学检测等, E-mail: wulifu@nuist.edu.cn。



马思佳(2000-), 女, 硕士研究生, 研究方向: 语音信号盲源分离, E-mail: 202212490117@nuist.edu.cn。



孙康(1998-), 男, 硕士研究生, 研究方向: 声场控制, E-mail: 202212490151@nuist.edu.cn。

(编辑: 夏道家)