文章编号:1004-9037(2012)02-0210-08

基于语音信号稀疏性的FDICA 初始化和后处理方法

马 峰 张 宁 戴礼荣

(中国科学技术大学电子工程与信息科学系,合肥,230027)

摘要:目前解决语音信号盲源分离(Blind source separation, BSS)的两大类方法分别为频域独立成分分析(Frequency domain independent component analysis, FDICA)和基于稀疏性的时频掩蔽(Time frequency masking, TF masking)。为此将两类方法优点相结合,利用TF masking 方法的结果,对FDICA 做初始化,在加快FDICA 收敛速度的同时也避免了次序不确定性问题。此外还提出了一种新的基于语音稀疏性FDICA 的BSS 后处理方法:基于局部最小比例控制(Local minimum ratio controlled, LMRC)谱减法,比常规的TF masking、维纳滤波等后处理方法,能够更有效地控制音乐噪声,提高分离性能。合成数据和实际采集数据的实验结果验证了所提方法的有效性。

关键词:盲源分离;独立成分分析;时频掩蔽;局部最小比例控制谱减法

中图分类号:TN912.35 **文献标识码**:A

FDICA Initialization and Post-Processing Method Based on Sparseness of Speech

Ma Feng, *Zhang Ning*, *Dai Lirong*

(IFlyTek Speech Lab, University of Science and Technology of China, Hefei, 230027, China)

Abstract: There are two approaches being widely studied and employed to solve the blind source separation (BSS) problem. One is based on independent component analysis (ICA) and the other relies on the sparseness of source signals time frequency masking (TF-masking). To speed up the convergence rate and to avoid permutation problems, a method combining the advantages of both methods is presented by using the results of TF masking to initialize the frequency domain ICA (FDICA). Moreover, a new post-processing method for FDICA is proposed, i.e. local minimum ratio control (LMRC) spectral subtraction. It is based on the sparse characteristics of speech. Compared with the conventional TF masking and Wiener filter post processing methods, the proposed method can control musical noise more effectively, and improve the separation performance. Experimental results with synthetic data and real data demonstrate the effectiveness of the proposed method.

Key words: blind source separation (BSS); independent component analysis (ICA); TF masking; LMRC spectral subtraction

引 言

盲源分离(Blind source separation, BSS)是指从若干观测的混合信号中提取出无法直接观测的源信号。这里的"盲"是指对源信号和传输信道没有先验知识。在移动通信、生物医学、语音识别中具有重要意义。从20 世纪80 年代Jutten C 和Herault J[1]提出BSS 问题以来变得到广泛的重视。目前主要的两类方法是独立成分分析(Independent component analysis, ICA)[1-3]和时频掩蔽(Time frequency masking, TF masking)方法[4-6]。根据传输信道复杂程度混合信号模型可以分为瞬时混合模型、无回声混合模型和卷积混合模型。目前对前面

收稿日期:2011-05-09;修订日期:2011-09-02

两个模型的BSS 问题已经很好解决，因此卷积混合模型是本文研究重点。

目前频域独立成分分析（Frequency domain ICA，FDICA）方法解决卷积混合型 BSS 的思路是将时域信号变换成频域信号，即把时域卷积近似为频域乘积，利用已瞬时混合模型BSS 的方法在每个频带上用ICA 做信号分离，然后解决幅度和次序不确定性问题，最后变换到时域信号。这种方法虽然简化了卷积模型 BSS，具有音乐噪声小，性能优于传统的波束形成方法等优点，但也引入了几个问题：(1)幅度不确定性问题；(2)次序不确定性问题；(3)由于语音信号是非平稳的，短时傅里叶变换（Short-time Fourier transform，STFT）必须注意窗长的选择，太长则帧内信号不满足平稳特性，太短则时域卷积近似为频域乘积误差增大[7]。同时 FDICA 方法还有收敛速度慢、非凸函数优化难以保证收敛到全局最优点等问题影响分离效果。

TF masking 方法则是在时频域内，基于语音信号的稀疏特性[4]选择合适的特征向量，通过聚类达到分离的目的。一个典型的方法是使用退化多传感器分离估计技术（Multiple sensor degenerate unmixing estimation technique，MEDUET）[6]，该方法对构造的特征向量进行 K-means 聚类，聚类中心对应于源的空间位置，根据聚类的结果对每个源构建一个TF Masking，然后用TF Masking 与时频域混合信号相乘得到每个源的时频域信号。最后经过短时傅里叶逆变换（Inverse short time Fourier transform，ISTFT）得到时域信号。虽然此种方法简单、快速，但 W-正交分离[5]假设过于理想。因为使用二元时频掩蔽方法采取一种“主导则全占”的策略，即某源在某个时频点占主导则认为这个时频点上信号全部是该源产生的，在不满足 W-正交分离条件的时频点上造成严重的信号失真，产生音乐噪声。针对此种情况，文献[8]中使用软时频掩蔽（Soft TF masking，STF marking）建立概率模型，用最大似然准则或者最大后验准则构建 STF masking，以减小信号失真和音乐噪声。

近几年，有学者将这两种方法结合到一起，用 TF masking 作为FDICA 的后处理步骤[8]。本文则从另一个角度出发，用 TF masking 的结果估计分离矩阵作为 FDICA 的初始化，试图减少 FDICA 迭代步数和避免次序不确定性问题。从文献[9,10]可以看出，一个较好的初始化可以使 FDICA 更快收敛，并且可以在一定程度上避免次序不确定性的问题发生。本文提出一种新的初始化方法可以提供一个更为准确的初始化点，从而使迭代步数减少。另外本文提出的基于局部最小比例控制（Local minimum ratio control，LMRC）谱减法的 BSS 后处理方法，能够有效地控制音乐噪声。

1　基于 FDICA 盲源分离基本算法

本文研究的对象是卷积混合模型BSS。假设有 N 个源信号为 $\boldsymbol{s}(t)=[s_1(t),s_2(t),\cdots,s_N(t)]^{\mathrm{T}}$，$M$ 个观测信号为 $\boldsymbol{x}(t)=[x_1(t),x_2(t),\cdots,x_M(t)]^{\mathrm{T}}$，本文中假设源 $M=N=2$。卷积混合模型数学表达式为

$$x_m(t)=\sum_{n=1}^{2}\sum_{p=0}^{\infty}A_{mn}(p)s_n(t-p)\quad m=1,2 \tag{1}$$

式中：A_{mn}表示第 n 个源到第 m 个麦克风的冲击响应，对式(1)的左右两端进行短时傅里叶变换，可以得到卷积混合模型的频域表示

$$\boldsymbol{x}(f,\tau)\approx[\boldsymbol{a}_1(f),\boldsymbol{a}_2(f)]\boldsymbol{s}(f,\tau)=\boldsymbol{A}(f)\boldsymbol{s}(f,\tau) \tag{2}$$

式中：f 为频率，τ 为帧的序号，$\boldsymbol{s}(f,\tau)=[s_1(f,\tau),s_2(f,\tau)]^{\mathrm{T}}$ 和 $\boldsymbol{x}(f,\tau)=[x_1(f,\tau),x_2(f,\tau)]^{\mathrm{T}}$ 分别为源向量和观察向量。$\boldsymbol{A}(f)=[\boldsymbol{a}_1(f),\boldsymbol{a}_2(f)]$为频率 f 处的混合矩阵，$\boldsymbol{a}_n(f)$为源 n 的混合向量。经过变换后，时域卷积近似变为频域乘积。本文采用信息最大化[11]结合自然梯度[12]的方法计算分离矩阵 $\boldsymbol{W}(f)$

$$\boldsymbol{y}(f,\tau)=\boldsymbol{W}(f)\boldsymbol{x}(f,\tau) \tag{3}$$

式中 $\boldsymbol{y}(f,\tau)=[y_1(f,\tau),y_2(f,\tau)]^{\mathrm{T}}$ 为在频率 f 上第 τ 帧分离出的源信号。分离矩阵迭代公式为

$$\begin{aligned}&\boldsymbol{W}_{i+1}(f)=\boldsymbol{W}_i(f)+\Delta\boldsymbol{W}\\&\Delta\boldsymbol{W}=\eta[\boldsymbol{I}-E_\tau(\varphi(\boldsymbol{y}(f,\tau)\boldsymbol{y}^{\mathrm{H}}(f,\tau)))]\end{aligned} \tag{4}$$

式中：$\boldsymbol{I}$ 为单位阵，$E_\tau(\cdot)$表示对帧求平均的算子，H 表示共轭转置，i 表示迭代过程的第 i 次，η 表示迭代步长，定义非线性向量函数$\phi(\cdot)$为

$$\begin{aligned}&\phi(\boldsymbol{y}(f,\tau))=[\phi(y_1(f,\tau)),\phi(y_2(f,\tau))]^{\mathrm{T}}\\&\phi(y_n(f,\tau))=\tanh(\mathrm{Re}(y_n(f,\tau)))+\\&\quad i\cdot\tanh(\mathrm{Im}(y_n(f,\tau)))\quad n=1,2\end{aligned} \tag{5}$$

式中Re(·)表示取实部运算，Im(·)表示取虚部运算。

由于FDICA 是在每个频带上独立地进行，所引起的输出幅度和次序的不确定问题必须得到解

决。本文采用反射投影(Projection back, PB)[12]方法解决幅度不确定性,次序不确定性问题则可由下面提到的初始化方法避免。

2 基于语音信号稀疏特性的分离矩阵初始化

2.1 基本原理

语音信号在时频域中具有很好的稀疏性[4-6]。在满足稀疏性要求的时频点则有以下关系式

$$\boldsymbol{x}(f,\tau) \approx \boldsymbol{a}_n(f)s_n(f,\tau), \exists\ n \in \{1,2\} \quad (6)$$

即在每个时频点中只有一个源占主导,其他源贡献很小或没有。因此可以根据此种特性,选取某种能区分不同源的特征向量将所有观测向量$\boldsymbol{x}(f,\tau)$进行聚类。根据聚类的结果$C_n(n=1,2)$,将每个时频点和源的对应关系找出,根据这种对应关系构建TF masking

$$M_n(f,\tau) = \begin{cases} 1 & x(f,\tau) \in C_n \\ 0 & x(f,\tau) \notin C_n \end{cases} \quad (7)$$

然后将$M_n(f,\tau)$应用于某个接收信号$x_m(f,\tau)$

$$\boldsymbol{y}_{mn}^{\text{mask}}(f,\tau) = M_n(f,\tau)x_m(f,\tau)$$

得到源n在第m个麦克风上的信号,以及源n在2个麦克上的信号矢量。

$$\boldsymbol{y}_n^{\text{mask}}(f,\tau) = [\boldsymbol{y}_{1n}^{\text{mask}}(f,\tau), \boldsymbol{y}_{2n}^{\text{mask}}(f,\tau)]^{\text{T}} \quad (8)$$

2.2 特征向量

本文采用文献[6]中的特征向量,实现观测向量$\boldsymbol{x}(f,\tau)$的聚类。由式(6)可以看出,可以根据$\boldsymbol{a}_n(f)$仅与源位置相关的特性对观测向量$\boldsymbol{x}(f,\tau)$进行分类。为消除源$s_n(f,\tau)$和频率的影响,首先对$\boldsymbol{x}(f,\tau)$做幅度和相位以及频率归一化

$$\bar{\boldsymbol{x}} \leftarrow \frac{x}{\|\boldsymbol{x}\|} \cdot \exp[-\mathrm{i}\arg(\boldsymbol{x}_J)] \quad (9)$$

$$\tilde{\bar{x}}_m(f,\tau) \leftarrow |\bar{x}_m(f,\tau)| \cdot \exp\left[\mathrm{i}\frac{\beta\arg[\bar{x}_m(f,\tau)]}{f}\right] \quad (10)$$

式(9)是幅度模值归一化,并取与参考麦克J上信号的相对相位(本文取$J=1$)。式(10)是对做频率归一化,消除相位差与频率的关系。β是一个常系数,调节相位信息和模值信息的相对重要程度。将所有归一化后的观测向量$\tilde{\bar{x}}$经行聚类,可以得到$N(=2)$个聚类中心

$$\bar{\boldsymbol{c}}_n = \begin{bmatrix} \bar{c}_{1n} \\ \bar{c}_{2n} \end{bmatrix} = \begin{bmatrix} \bar{\lambda}_{1n} \\ \bar{\lambda}_{2n} \cdot \exp(-\mathrm{i}2\pi\beta\,\bar{\tau}_{2n}) \end{bmatrix} \quad n = 1,2 \quad (11)$$

式中:$\bar{\lambda}_{mn}$表示第n个源到第m个麦克风的衰减,$\bar{\tau}_{2n}$表示第n个源在第2个麦克风与参考麦克(第1个麦克)的相对时延。从式(11)中可以算出相对延时

$$\tau_{mn} = -\frac{\arg(\bar{c}_{mn})}{2\pi\beta}, \quad \lambda_{mn} = |\bar{c}_{mn}|$$
$$m = 1,2; \quad n = 1,2 \quad (12)$$

计算出衰减和相对延时,则可以计算出每个频带各源的无回声混合模型的混合向量,作为卷积模型混合向量$\boldsymbol{a}_n(f)$的近似

$$\boldsymbol{c}_n(f) = \begin{bmatrix} \lambda_{1n} \\ \lambda_{2n} \cdot \exp(-\mathrm{i}2\pi f\tau_{2n}) \end{bmatrix} \quad n = 1,2 \quad (13)$$

然后求逆计算出无回声混合模型估计分离矩阵作为FDICA的初始化(文献[9]初始化方法)

$$\boldsymbol{W}_{\text{anechoic}}(f) = [\boldsymbol{c}_1(f), \boldsymbol{c}_2(f)]^{-1} \quad (14)$$

2.3 两步法FDICA初始化

文献[9]按式(13,14)进行FDICA的初始化,显然是将卷积混合模型近似为无回声混合模型。卷积模型比无回声模型复杂:无回声模型频率冲击响应的模值与频率无关,相对相位与频率呈线性关系,而卷积模型的频响幅度与频率相关,相对相位一般也不与频率呈线性关系,因此用无回声模型混合向量$\boldsymbol{c}_n(f)$去近似卷积模型混合向量$\boldsymbol{a}_n(f)$,会造成较大的误差。为实现更好的初始化性能,需要更合理地估计混合向量$\boldsymbol{a}_n(f)$。由于ICA存在幅度不确定性问题,因此对混合向量或混合矩阵$\boldsymbol{A}(f)$的估计等价于对$\boldsymbol{A}'(f)$估计

$$\boldsymbol{A}'(f) = \begin{bmatrix} 1 & 1 \\ \dfrac{a_{21}(f)}{a_{11}(f)} & \dfrac{a_{22}(f)}{a_{12}(f)} \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ r_1(f) & r_2(f) \end{bmatrix} \quad (15)$$

基于式(15),本节提出的两步法FDICA初始化方法如下:

(1)按2.1节和2.2节介绍的聚类方法,通过聚类构建TF masking $M_k(f,\tau)$,及通过式(8)估计第n源在2个麦克风上的信号矢量$\boldsymbol{y}_n^{\text{mask}}(f,\tau)$$(n=1,2)$。

(2)基于空间预测[13]的思想估计$\boldsymbol{A}'(f)$。由第一步估计的$\boldsymbol{y}_n^{\text{mask}}$可用卷积模型表示为

$$\boldsymbol{y}_n^{\text{mask}}(f,\tau) = \boldsymbol{a}_n(f)\boldsymbol{s}_n(f,\tau) + \boldsymbol{v}_n(f,\tau) \quad (16)$$

式中$\boldsymbol{v}_n(f,\tau)$为由于源信号不满足理想稀疏性(即不满足W-正交分离假设)所引起的估计误差。从式(16)可见,在时频点(f,τ)如果满足理想稀疏性假

设，则估计误差$\boldsymbol{v}_n(f,\tau)=0$，第$n$ 个信号在第2 个麦克风上的信号，可由该信号在第1 个麦克风上的信号预测，即有如下关系式

$$\boldsymbol{y}_{2n}^{\text{mask}}(f,\tau)=\frac{a_{2n}(f)}{a_{1n}(f)}\boldsymbol{y}_{1n}^{\text{mask}}(f,\tau) \tag{17}$$

因此可根据最小均方误差(Minimum mean square error，MMSE)准则估计式(15)中的r_n

$$r_n(f)=\underset{r_n}{\text{argmin}}E\{[\boldsymbol{y}_{2n}^{\text{mask}}-r_n\boldsymbol{y}_{1n}^{\text{mask}}]^2\} \tag{18}$$

$$r_n(f)=\frac{E\{\boldsymbol{y}_{2n}^{\text{mask}}(\boldsymbol{y}_{1n}^{\text{mask}})^{\text{H}}\}}{E\{\boldsymbol{y}_{1n}^{\text{mask}}(\boldsymbol{y}_{1n}^{\text{mask}})^{\text{H}}\}} \tag{19}$$

在时频点(f,τ)，如果不满足理想稀疏性假设，则估计误差$\boldsymbol{v}_{n2}\neq 0$，式(17)将不再成立。为了更准确的估计比值$r_n(f)$，需要对用于估计$r_n(f)$的样本点$\boldsymbol{y}_n^{\text{mask}}(f,\tau)$挑选，减少非稀疏性导致的$r_n(f)$估计误差。由式(16)可见，满足稀疏性时频点上的信号矢量$\boldsymbol{y}_n^{\text{mask}}$的归一化矢量$\tilde{\boldsymbol{y}}_n^{\text{mask}}$(按式(9)归一化)应该集中在真实混合向量附近，非稀疏点则随机散乱地分布在其他位置[6]。因此可以按照最小距离最小准则挑选稀疏点，具体算法步骤如下：

(1)对于每一频率点f($f=1,2,\cdots,\text{NFFT}/2$，NFFT 为STFT 的窗长)，采用TF masking 方法估计第n 个源$\boldsymbol{y}_n^{\text{mask}}(f,\tau)$($n=1,2;\tau=1,2,\cdots,T$；$T$ 为帧数)

(2) 如果$\boldsymbol{y}_n^{\text{mask}}(f,\tau)$中存在模值非零矢量，则去除$\boldsymbol{y}_n^{\text{mask}}(f,\tau)$中模值为零的矢量，并按式(9)做幅度和相位归一化得到B 个非零矢量$\tilde{\boldsymbol{y}}_n^{\text{mask}}(f,\tau)$；否则：由式(13)，$r_n(f)=\dfrac{c_{2n}(f)}{c_{1n}(f)}$，结束估计；

(3) 计算每一个$\tilde{\boldsymbol{y}}_n^{\text{mask}}(f,\tau)$到其他点的距离，取出$P$ 个最小距离，并求该P 个最小距离的平均值；一共得到B 个最小平均距离，对应给定频率点的B 个时频点；

(4) 根据B 个最小平均距离，保留Q 个最小的最小平均距离对应的时频点，利用这些时频点上的$\tilde{\boldsymbol{y}}_n^{\text{mask}}(f,\tau)$根据式(19)估计$r_n(f)$。

步骤(3)求出的最小P 个距离的均值是衡量一个点到其他点的离散程度，越小则是稀疏点的可能性越高，实验表明，P 值的选择对结果影响不大，本文取$P=3$；步骤(4)选出Q 个稀疏点可能性高的点。Q 如何选择将在实验部分给出。

最后得到的初始化分离矩阵为

$$\boldsymbol{W}_{\text{new}}(f)=\begin{bmatrix}1 & 1\\ r_1(f) & r_2(f)\end{bmatrix}^{-1} \tag{20}$$

由于用于估计r_n 的数据$\boldsymbol{y}_n^{\text{mask}}(f,\tau)$是已经解决次序不确定性的信号矢量，以式(20)作为FDICA 的初始化分离矩阵，保证了式(4)在迭代初始已不存在次序不确定性问题。如果式(20)作为FDICA 的初始化分离矩阵是一个很好的初始化，实验验证表明，式(4)迭代过程将保持初始化时分离信号的次序特性。

3　基于语音信号稀疏特性的FDICA 后处理方法

经过第1 节FDICA 和第2 节初始化，以及解决幅度不确定问题后，由于收敛问题和混响等原因，分离性能仍不够理想，因此需要对分离输出做后处理。本文提出的LMRC 谱减法是基于语音信号稀疏特性的自适应算法。该方法可以逐一对分离的各源信号进行增强处理，在以下讨论中，不妨设第1 个源为增强的目标信号，第2 个源为干扰信号。

3.1　混合-分离系统

将混合过程与分离过程看成一个综合系统：混合-分离系统

$$\boldsymbol{y}(f,\tau)=\boldsymbol{W}(f)\boldsymbol{A}(f)\boldsymbol{s}(f,\tau)=\boldsymbol{G}(f)\boldsymbol{s}(f,\tau) \tag{21}$$

$$\boldsymbol{G}(f)=\boldsymbol{W}(f)\boldsymbol{A}(f) \tag{22}$$

式中$\boldsymbol{y}(f,\tau)=[y_1(f,\tau),y_2(f,\tau)]^{\text{T}}$。在解决次序不确定性问题后，不妨设$y_1(f,\tau)$和$y_2(f,\tau)$的主要成分分别为是$s_1(f,\tau)$，$s_2(f,\tau)$，即分离混合矩阵$\boldsymbol{G}$ 的主对角元素模值远大于其他元素模值

$$|G_{11}(f)|\gg|G_{21}(f)|,|G_{22}(f)|\gg|G_{12}(f)| \tag{23}$$

将y_n($n=1$ 或2)分为目标信号和干扰信号

$$y_n(f,\tau)=y_n^t(f,\tau)+y_n^c(f,\tau)=G_{nn}(f)s_n(f,\tau)+\sum_{j\neq n}G_{nj}s_j(f,\tau) \tag{24}$$

式中：$y_n^t(f,\tau)$为目标信号，$y_n^c(f,\tau)$为干扰成分。如果混合-分离矩阵$\boldsymbol{G}$ 为对角阵则表明完全分离，如果存在非对角线上元素不为零，则表明没有完全分离。

3.2　局部最小比例控制谱减法

本文提出的后处理方法框图如图1 所示。基本原理是通过一个一阶自适应滤波器估计干扰成分，然后从分离后的信号中减去干扰成分达到增强的目的。为减小滤波器失调，需要检测干扰成分占主导的时间段，在这些时间段对滤波器更新。本文通过分析分离的两个信号的幅度谱比值的特性，提出

了基于LMRC的滤波器更新方法。由于本文后处理方法是在每个频带上进行的，为简便叙述，本节所有变量省略f。

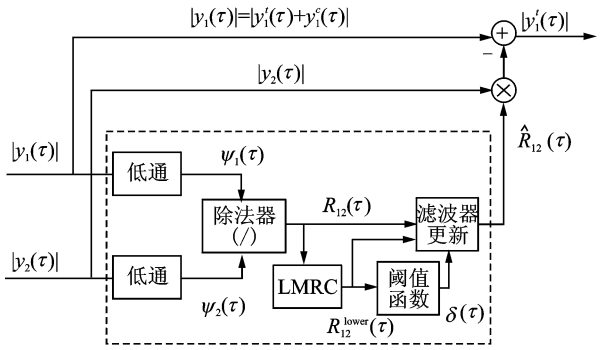


图1 对第1个分离信号做增强的后处理框图

由式(23,24)可得

$$R(\tau)=\frac{|y_1(\tau)|}{|y_2(\tau)|}=\frac{|G_{11}s_1(\tau)+G_{12}s_2(\tau)|}{|G_{21}s_1(\tau)+G_{22}s_2(\tau)|}\approx\frac{|G_{12}s_2(\tau)|}{|G_{22}s_2(\tau)|}\qquad s_2(\tau)\gg s_1(\tau)\tag{25}$$

式(25)表明当干扰源$s_2(\tau)$占主导时，分离信号的幅度比$R(\tau)\ll1$，随着$s_1(\tau)$幅度增大，$R(\tau)$有增大的趋势，当目标源$s_1(\tau)$占主导时，$R(\tau)\gg1$。因此根据$R(\tau)$这个特性可以看出，当$R(\tau)$小于某一个设定阈值$\delta(\tau)$时，可认为干扰源占主导。为减少由$R(\tau)$快速抖动造成的检测误差，本文采用分离信号平滑幅度谱比值$R_{12}(\tau)$代替$R(\tau)$

$$\begin{aligned}R_{12}(\tau)&=\frac{\psi_1(\tau)}{\psi_2(\tau)}\\\psi_1(\tau)&=\alpha\psi_1(\tau-1)+(1-\alpha)|y_1(\tau)|\\\psi_2(\tau)&=\alpha\psi_2(\tau-1)+(1-\alpha)|y_2(\tau)|\end{aligned}\tag{26}$$

式中$\psi_1(\tau)$，$\psi_2(\tau)$分别为$y_1(\tau)$，$y_2(\tau)$幅度经过平滑的信号。参数α控制平滑程度，越大平滑程度越高，反之则越低。选择合适的α，$R_{12}(\tau)$既能够保持$R(\tau)$的特性，又能够减小快速抖动。

进一步讨论如何由$R_{12}(\tau)$确定阈值$\delta(\tau)$。理论上，当计算出所有帧$\tau=1,2,\cdots,T$的$R_{12}(\tau)$时，可按式(27)确定阈值$\delta(\tau)$

$$\delta=\min\{R_{12}(\tau)|\tau=1,2,\cdots,T\}+\theta_T\tag{27}$$

式中：θ_T为正实数，T为帧数。考虑到在线后处理的需要，本文采用$R_{12}(\tau)$的下包络估计量$R_{12}^{\text{lower}}(\tau)$作为式(27)右边第1项求得的最小值的估计。显然，$R_{12}^{\text{lower}}(\tau)$是$R_{12}(\tau)$的局部最小值。估计$R_{12}^{\text{lower}}(\tau)$的具体递推算法步骤如下：

(1)定义一个临时变量$R_{12}^{\text{temp}}(\tau)$和一个局部最小值搜索范围$L$，并初始化$R_{12}^{\text{lower}}(1)=R_{12}(1)$，$R_{12}^{\text{temp}}(1)=R_{12}(1)$。

(2)如果τ不能整除L，更新规则如下：

$$\begin{aligned}R_{12}^{\text{lower}}(\tau)&=\min\{R_{12}^{\text{lower}}(\tau-1),R_{12}(\tau)\}\\R_{12}^{\text{temp}}(\tau)&=\min\{R_{12}^{\text{temp}}(\tau-1),R_{12}(\tau)\}\end{aligned}$$

如果τ能整除L，更新规则如下：

$$\begin{aligned}R_{12}^{\text{lower}}(\tau)&=\min\{R_{12}^{\text{temp}}(\tau-1),R_{12}(\tau)\}\\R_{12}^{\text{temp}}(\tau)&=R_{12}(\tau)\end{aligned}$$

临时变量R_{12}^{temp}用来搜索至多L帧范围内的最小值，$R_{12}^{\text{lower}}(\tau)$能够搜索到至少$L$至多$2L$帧内的最小值。选择合适的$L$，$R_{12}^{\text{lower}}(\tau)$既能够保持局部最小值又能够实时更新，因此可作为$R_{12}(\tau)$的下包络估计。估计出下包络后，则可设定阈值函数

$$\delta(\tau)=R_{12}^{\text{lower}}(\tau)+\theta_T\tag{28}$$

设定阈值$\delta(\tau)$后，可按式(29)更新滤波器的参数

$$\begin{cases}\hat{R}_{12}(\tau)=\gamma\hat{R}_{12}(\tau-1)+(1-\gamma)R_{12}(\tau) & R_{12}(\tau)<\delta(\tau)\\\hat{R}_{12}(\tau)=R_{12}^{\text{lower}}(\tau) & R_{12}(\tau)\geqslant\delta(\tau)\end{cases}\tag{29}$$

式中第1个式子是干扰源占主导时，用一阶递归方法更新滤波器。当目标源占主导时，信噪比高，为防止过减，滤波器更新采用式(29)第2个式子。最后，增强后的目标信号幅度谱为

$$|\bar{y}_1^t(\tau)|=\max\{|y_1(\tau)|-\hat{R}_{12}(\tau)|y_2(\tau)|,0\}\tag{30}$$

由式(30)结果再结合$y_1(\tau)$的相位再做ISTFT，得到最终的后处理结果。

由以上讨论可知，本文的后处理方法主要是通过平滑幅度谱比值$R_{12}(\tau)$的局部最小值$R_{12}^{\text{lower}}(\tau)$来设定滤波器参数更新的控制阈值$\delta(\tau)$，因此本文的后处理方法称作LMRC谱减法。

4 实验结果与分析

4.1 实验数据及性能指标

实验数据为3种环境下长度为10 s的语音信号，分别是BSS数据集SiSEC 2008[14]中T60为130，250 ms和实验室环境下(T60约为150 ms)实际录制的数据，实验数据采集配置见表1和图2。源到阵列中心距离为1 m，采用率为8 000 Hz。性能评测本文使用BSS评测工具包BSS EVAL Toolbox[15]中的信号失真比(Signal-to-distortion ratio，SDR)，信号干扰比(Signal-to-interference ratio，SIR)，信号人工误差比(Signal-to-artifact ratio，SAR)作

为性能指标。ISTFT 的窗长取64 ms(512 样本点)，帧移为16 ms(128 样本点)，ICA 的迭代步长为0.003，后处理中参数为：$\theta_T=1$，$\gamma=0.7$，$\alpha=0.3$，$L=75$。

表 1　实验配置参数

数据	T60/ms	麦克风间距/cm	θ/(°)
SiSEC	130，250	5	80，105
自录数据	150	4	45，135

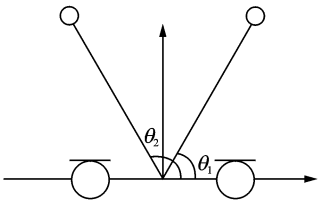
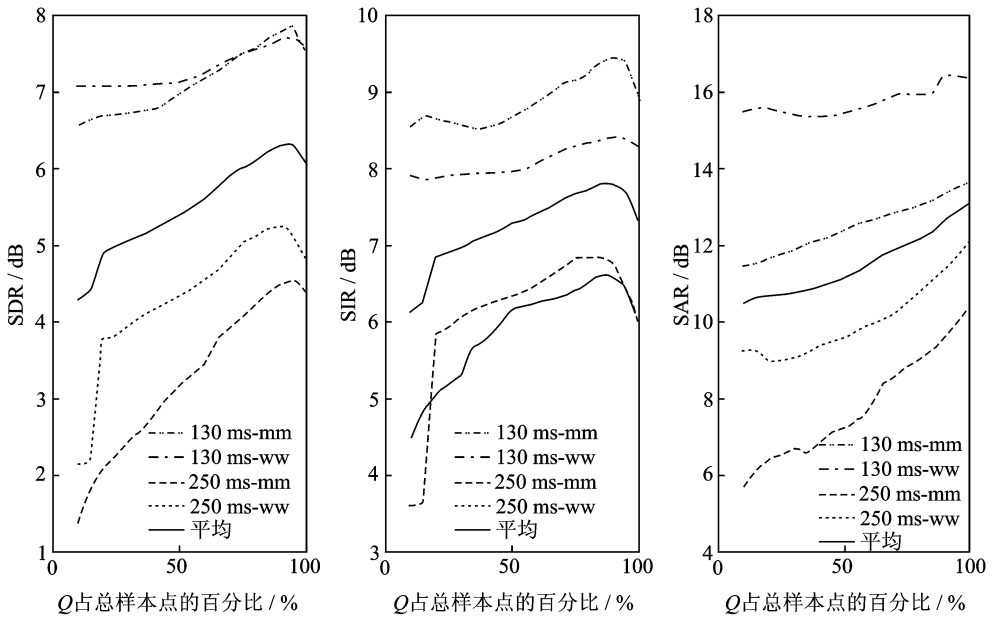


图 2　阵列结构

4.2　实验参数及结果

实验 1　对于初始化部分中 Q 如何选取。用SiSEC 数据做 4 组实验(不同T60，不同源组合)，Q 分别取非零点样本总数的10%～100%，将使用不同 Q 估计出的式(20)代入式(3)，不经过ICA 迭代得到的分离信号。实验结果3 个指标如图3 所示。从实验结果可以看出，4 组结果的SDR，SIR 对 Q 的变化趋势基本一致，在 Q 取非零样本点总数的90%附近取最大值，比不做挑选分别高 0.3 dB 和0.6 dB。SAR 随 Q 的增大而增大。综合 3 个指标 Q 选取非零样本点数的90%左右比较合适，太大不能够去除非稀疏点，太小则用于估计分离矩阵的数据减少都会造成初始化效果下降。因此后面的实验 Q 值都取非零样本点数的 90%。



mm：两个源都是男声；ww：两个源都是女声。

图 3　4 组实验分别是 T60 为 130，250 ms 环境下用不同性别的人的语音信号实验结果

实验 2　初始化方法对FDICA 收敛性能比较实验。FDICA 收敛分实验分别使用实验室环境录制和SiSEC 中的 3 组数据，分别用本文的初始化方法与文献[9]的方法对 FDICA 迭代 1 000 次，图 4 给出SDR 收敛情况(SIR，SAR 结果类似)。从迭代次数为 0 和不同迭代次数的SDR。实验结果表明，本文初始化方法在3 种环境下，SDR 均都高于无回声模型初始化方法，特别是代次数为0 时的SDR 远高于无回声模型初始化方法。该结果说明本文初始化方法减小了分离矩阵估计的误差，得到了更准确的初始点。同时，从图4 还可看出，本文初始化在达到相同的分离性能情况下，所需迭代次数比文献[9]中方法至少少 300 次，而本文的FDICA 初始化方法引入的额外的运算量(特别是稀疏点挑选需要计算距离矩阵)仅与FDICA 迭代100 次相当。结合实验 1 结果可以看出，即使不做非稀疏点挑选，本文的两步法FDICA 初始化0 次迭代结果的SDR 也远远高于无回声模型的初始化方法，由此可见本文初始化方法的改善主要是在于通过卷积模型的估计减小了混合矩阵估计误差。

实验 3　验证后处理方法的性能用了 3 种环境，每种环境各 3 组(男声-男声，女声-女声，男声-

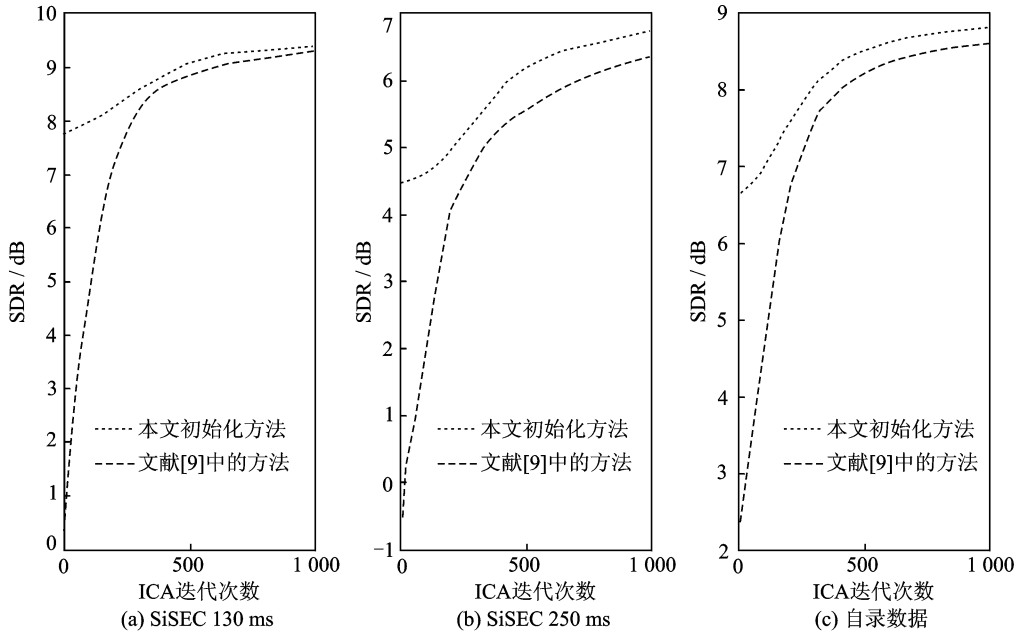


图4 3组环境下不同初始化ICA收敛图

女声)共9组数据,对经过FDICA迭代200次的输出分别用TF masking[8],和维纳滤波[16]和本文的LMRC谱减法作后处理。平均结果如图5所示。由于混合模型为时不变的线性系统,3种后处理方法都引入了时变或非线性处理,造成一定程度的信号失真,因此SAR指标都比FDICA输出结果低。本文的方法在SIR,SDR,SAR比TF masking方法均高2 dB左右,因为TF masking方法是对每个时频点分类,在T60较大或者源间距较小时,容易造成分类错误,对Masking等于1的时频点没有做噪声控制,并且没有考虑到每个时频点局部信息,容易造成谱的不连续,产生严重的音乐噪声(SAR最低)。维纳滤波虽然考虑了当前帧之前的局部信息,但是对噪声谱的估计不准确,因此在SIR上最低。本文后处理方法既考虑局部信息保证谱的连续性,又能对噪声幅度谱准确的估计,因此综合3个指标,本文的后处理方法要优于其他两种方法。

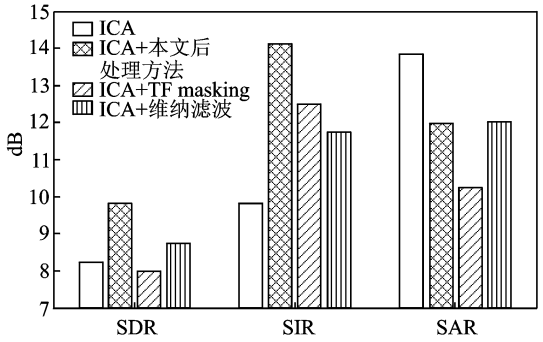


图5 不同后处理方法实验结果

5 结束语

本文基于语音信号稀疏特性,提出了一种两步法FDICA初始化方法以减少FDICA迭代次数,和一种后处理方法提高分离性能。所提出的初始化方法能在对源信号和阵列结构没有先验知识的情况下,利用从观察信号提取的特征向量的聚类结果做ICA初始化,既能提高最终分离性能又能减少迭代次数。本文提出的后处理方法能够在提高SIR的基础上有效控制音乐噪声。本文初始化方法计算量主要在稀疏点挑选步骤,而性能提升主要是基于卷积模型估计混合向量。今后的工作将研究如何快速有效地进行稀疏点挑选,既提高初始化性能,又不增加过多的计算量。

参考文献:

[1] Jutten C, Herault J. Blind separation of sources, part I: an adaptive algorithm based on neuron mimetic architecture [J]. Signal Processing, 1991, 24(1): 1-10.

[2] Cardoso J F, Comon P. Independent component analysis, a survey of some algebraic methods[C]// IEEE International Symposium Circuits and Systems. Atlana: IEEE, 1996: 93-96.

[3] Amari S, Cichoki A, Yang H H. A new learning algorithm for blind source separation [J]. Adv Neural Inf Process Syst, 1996, 8: 757-763.

[4] Yilmaz O, Rickard S. Blind separation of speech mixtures via time-frequency masking [J]. IEEE

Trans Signal Process, 2004,52(7):1830-1847.

[5] Araki S, Sawada H, Mukai R. A novel blind source separation method with observation vector clustering [C]// Proc 2005 Int Workshop Acoust Echo Noise Control (IWAENC 2005), 2005: 117-120.

[6] Araki S, Sawada H, Mukai R. Underdetermined blind sparse source separation for arbitrarily arranged multiple sensors [J]. Signal Processing, 2007, 87(8): 1833-1847.

[7] Duong N Q K, Vincent E, Gribonval R. Under-determined reverberant audio source separation using a full-rank spatial covariance model [J]. IEEE Trans on Audio, Speech, and Language Processing, 2010, 18 (7): 1830-1840.

[8] Sawada H, Araki S, Mukai R, et al. Blind extraction of dominant target sources using ICA and time-frequency masking [J]. IEEE Trans Audio, Speech, Lang Process, 2006,14(6): 2165-2173.

[9] 吴奇昌. 基于频域的卷积语音信号盲分离研究[D]. 合肥:中国科学技术大学电子工程与信息科学系, 2011.
Wu Qichang. Research on blind source separation in frequency domain for convolutive speech signals [D]. Hefei: University of Science and Technology of China, 2011.

[10] Sawada H, Araki S, Makino S. A two-stage frequency-domain blind source separation method for underdetermined convolutive mixture [C] // Proc IEEE Workshop Applicat Signal Process Audio Acoust. [S.l.]: IEEE, 2007:139-142.

[11] Bell A, Sejnowski T. An information-maximizati-on approach to blind separation and blind deconvo-lution [J]. Neural Computation, 1995, 7(6): 1129-1159.

[12] Kondo K, Yamada M, Kenmochi H. A semi-blind source separation method with a less amount of computation suitable for tiny DSP modules [C]// Proc Interspeech 2009. Brighton, UK: [s. n.], 2009: 1339-1342.

[13] Chen Jingdong, Benesty J, Huang Yiteng. A minimum distortion noise reduction algorithm with multiple microphones [J]. Audio, Speech, and Language Processing, IEEE Transactions, 2008,16(3): 481-493.

[14] Vincent E, Araki S, Bofill P. The 2008 signal separation evaluation campaign: a community-based approach to large-scale evaluation [EB/OL]. (2009-05-09)[2010-12-11]. http://sisec2008.wiki.irisa.fr/tiki-index.php.

[15] Gribonval R, Fevotte C, Vincent E. BSS EVAL toolbox user guide revision 2.0 [R]. IRISA Technical Report 1706, 2005.

[16] Park K S, Park J S, Son K S, et al. Post processing with Wiener filtering technique for reducing residual crosstalk in blind source separation [J]. IEEE Signal Processing Letters, 2006, 13(12): 749-751.

作者简介:马峰(1986-),男,硕士研究生,研究方向:语音信号处理,E-mail:mafeng@mail.ustc.edu.cn;张宁(1988-),男,硕士研究生,研究方向:语音信号处理;戴礼荣(1962-),男,博士,教授,研究方向:数字信号处理和模式识别。