部分W-分离正交语音信号的盲分离方法

刘郁林，王国鹏，罗颖光

(重庆通信学院DSP研究室　重庆 沙坪坝区　400035)

【摘要】在W-分离正交性假设的语音盲分离方法中，由于没有考虑多个源信号同时存在的情况，导致分离信号中不可避免地存在音乐噪声。针对这种部分W-分离正交情况，提出了基于信道估计的语音盲分离方法。该方法先检测只有一个源信号存在的时频点并进行归一化处理，使得处理后的结果与频率无关，克服了W-分离正交性假设的不足以及频率置换问题，通过K-means聚类估计出信道，再结合信号子空间方法重构源信号。仿真结果表明，提出的方法可以有效减少分离语音中的音乐噪声，与典型的时频二元掩蔽方法相比，其平均信号失真比提高3.02 dB，同时平均信干比提高4.61 dB。

关　键　词　信道估计;　部分W-分离正交;　主分量分析;　语音分离

中图分类号　TN911　　**文献标识码**　A　　doi:10.3969/j.issn.1001-0548.2010.02.007

Blind Separation for Partial W-Disjoint Orthogonal Speech Signals

LIU Yu-lin, WANG Guo-peng, and LUO Ying-guang

(DSP Lab, Chongqing Communication College　Shapingba Chongqing　400035)

Abstract　In blind speech separation methods based on the assumption of W-disjoint orthogonality (W-DO), musical noise is inevitable in separated signals because the assumption does not include the case of existing multiple source signals in the time-frequency domain. A blind speech separation method based on channel estimation is proposed for partial approximate W-disjoint orthogonality. The time-frequency cells with only one source are detected and normalized to be independent of frequency, which overcomes not only the shortcoming of W-DO property but also the frequency permutation problem, and then the channel estimation is obtained by K-means clustering. Finally, signal subspace method is exploited to reconstruct sources. Simulation results demonstrate that the novel method can effectively reduce the musical noise in the separated speech signals, and it outperforms the typical time frequency binary masking method, the averaged signal to distortion ratio (SDR) is improved by 3.02 dB and the averaged signal to interference ratio (SIR) is improved by 4.61 dB.

Key words　channel estimation;　partial W-disjoint orthogonality;　principal component analysis;　speech separation

盲源分离是在源信号和混合过程未知的情况下，仅根据观测信号恢复源信号的一种方法。盲源分离的应用十分广泛，可应用于语音信号的分离和识别、生物医学信号处理、数字通信以及地震探测等诸多领域。近年来，盲源分离[1-2](blind source separation，BSS)已成为信号处理领域的一个研究热点。

对于语音盲分离问题[3]，当前研究的难点是欠定卷积混合模型，该模型假设源信号的数目大于传感器的数目，并且考虑了语音信号在传播过程中的反射及时延。近几年，基于稀疏性的语音盲分离方法越来越受到研究人员的关注[4-5]，该类方法利用语音信号的W-分离正交性假设，可以容易地解决欠定卷积混合问题，其中常用的一类方法是时频二元掩蔽，而比较有代表性的是MENUET(multiple sENsor degenerate unmixing estimation technique)方法[5]。MENUET方法的优点在于不用估计信道(混合矩阵)，从而避免了频率置换问题[6]。但是W-分离正交性的假设过于理想，使得掩蔽后的信号存在音乐噪声问题[7]，影响分离语音的质量。由于音乐噪声是时频二元掩蔽的必然产物，对这一缺点MENUET方法本身无法克服，但通过估计信道可以有效克服[8]。但欠定情况下估计信道的方法鲜有报道，常用的一个方法是对观测向量进行聚类，将聚类中心作为信道的估计[8-9]，但是该类方法主要利用语音信号的W-分离正交性，其估计精度往往不高。同时，估计出信道以后，必须解决频率置换问题。

本文针对语音信号的部分W-分离正交性，提出

收稿日期：2008 – 08 – 23；修回日期：2009 – 04 – 02
基金项目：国家自然科学基金(60672157, 60672158)
作者简介：刘郁林(1971 –)，男，博士，教授，主要从事盲信号处理、超宽带通信、认知无线电、无线传感器网络及DSP技术等方面的研究.

了一种基于PCA的时频点检测方法，用于检测只有一个源信号存在的时频点(其他源信号在该时频点的能量为零)，也即满足W-分离正交性的时频点。根据近声场模型，选定一个参考传感器并对检测出的时频点进行归一化处理，处理后的结果与频率无关，仅与源信号相对于传感器的位置以及传感器之间的距离有关，实际是相应信道进行归一化处理的结果，因此同一个源信号在不同频率点处的信道经过归一化处理后是相等的。对处理结果进行K-means聚类，将聚类中心作为信道的估计，经归一化逆处理，估计出各个频率点处的信道，此时不存在频率置换问题。最后结合信号子空间法重构源信号。仿真结果表明，与MENUET方法相比，本文方法的平均信干比增加了4.61 dB，而平均信号失真比增加了3.02 dB，因此具有更优的分离性能，并且减少了音乐噪声，改善了分离语音的质量。

1　问题描述

假设 M 个传感器和 N 个语音源信号 $s_1(t),s_2(t),\cdots,s_N(t)$ 进行卷积混合，传感器的观测信号为：

$$x_i(t)=\sum_{k=1}^{N}\sum_{l}h_{ik}(l)s_k(t-l)\qquad i=1,2,\cdots,M \tag{1}$$

式中　$x_i(t)$ 表示第 i 个传感器的观测信号；h_{ik} 表示从第 k 个源信号到第 i 个传感器的信道冲激响应。本文的目标是在源信号 $s_k(t)$ 为稀疏信号的假设下，经过信道估计，从 M 个观测信号中得到其估计 $\hat{s}_k(t)$。

首先通过短时傅里叶变换(STFT)将观测信号转换到时频域：

$$x_i(f,\tau)=\sum_{r=-L/2}^{L/2-1}x_i(\tau+r)\,\mathrm{win}(r)\mathrm{e}^{-\mathrm{j}2\pi f\tau} \tag{2}$$

式中　$f\in\left\{0,\frac{1}{L}f_s,\cdots,\frac{L-1}{L}f_s\right\}$，$f_s$ 为抽样频率；$\mathrm{win}(r)$ 为两端平滑过渡到零的窗函数；L 为窗长；τ 为时间。将信号转换到时频域有两个优点：(1) 时域上的卷积混合转化为频域内每个频率点上的瞬时混合，使得计算复杂度降低；(2) 较好呈现了语音信号的稀疏性，有利于算法处理。

将式(2)写成向量形式：

$$\boldsymbol{X}(f,\tau)=\sum_{k=1}^{N}\boldsymbol{h}_k(f)s_k(f,\tau)=\boldsymbol{H}(f)\boldsymbol{S}(f,\tau) \tag{3}$$

式中　$\boldsymbol{X}(f,\tau)=[x_1(f,\tau),x_2(f,\tau),\cdots,x_M(f,\tau)]^{\mathrm{T}}$ 和 $\boldsymbol{S}(f,\tau)=[s_1(f,\tau),s_2(f,\tau),\cdots,s_N(f,\tau)]^{\mathrm{T}}$ 分别为观测信号和源信号的时频域；$\boldsymbol{h}_k(f)=[h_{1k}(f),h_{2k}(f),\cdots,h_{Mk}(f)]^{\mathrm{T}}$ 表示第 k 个源信号到所有传感器的信道冲激响应(简称信道)，也称之为混合向量；$\boldsymbol{H}(f)=[\boldsymbol{h}_1(f),\boldsymbol{h}_2(f),\cdots,\boldsymbol{h}_N(f)]$ 表示信道矩阵，也称为混合矩阵。

2　单源时频点的检测

2.1　语音信号的W-分离正交性

设有 N 个语音源信号 $s_1(t),s_2(t),\cdots,s_N(t)$，其时频域为 $s_i(f,\tau)$ ($i\in I$，$I=\{1,2,\cdots,N\}$ 是源信号指标集合)。在时频域 (f,τ) 中，如果任意两个语音信号之间满足：

$$s_i(f,\tau)s_j(f,\tau)=0\qquad i,j\in I\text{ 且 }i\neq j \tag{4}$$

则称 $s_i(t)$ 和 $s_j(t)$ 满足严格W-分离正交性[4]。

在实际情况中，由于语音信号的傅里叶变换通常不为零，如果某些信号的时频变换系数很小，式(4)可以近似表示为：

$$s_i(f,\tau)s_j(f,\tau)\approx 0\qquad i,j\in I\text{ 且 }i\neq j \tag{5}$$

则称 $s_i(t)$ 和 $s_j(t)$ 满足近似W-分离正交性[4]。

上述两种情况隐含的假设是，在任意时频点至多存在(或近似存在)一个源信号。

如果部分时频点上同时存在多个源信号，即：

$$\begin{cases}|s_i(f,\tau)|\approx 0 & i\in I_1,I_1\subset I\\ |s_j(f,\tau)|\neq 0 & j\in I_2,I_2\subset I\\ I_1\cap I_2=\varnothing,I_1\cup I_2=I & I_2\text{中至少包含两个元素}\end{cases} \tag{6}$$

则称 N 个语音源信号 $s_1(t),s_2(t),\cdots,s_N(t)$ 满足部分W-分离正交性。

在基于W-分离正交性假设的语音盲分离方法中[5,8]，通常假设所有的时频点均满足严格或近似W-分离正交性条件(即假设满足部分W-分离正交性的时频点也至多存在一个源信号)，因此导致分离信号中存在音乐噪声的问题。本文提出一种基于PCA的时频点检测方法，用于检测出时频域中只有一个源信号存在的时频点，称该类时频点为单源时频点。检测出的单源时频点满足近似W-分离正交性，可用于降低分离信号中的音乐噪声。

2.2　基于PCA的单源时频点检测

假设 $\boldsymbol{X}(f,\tau)$ 的均值为零，对其协方差矩阵进行特征值分解：

$$\begin{aligned}\boldsymbol{R}_{\boldsymbol{X}}(f,\tau)&=E[\boldsymbol{X}(f,\tau)\boldsymbol{X}(f,\tau)^{\mathrm{H}}]=\\&\boldsymbol{H}(f)\boldsymbol{R}_{\boldsymbol{S}}(f,\tau)\boldsymbol{H}(f)^{\mathrm{H}}=\boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^{\mathrm{H}}\end{aligned} \tag{7}$$

式中　$\boldsymbol{U}=[\boldsymbol{u}_1(f),\boldsymbol{u}_2(f),\cdots,\boldsymbol{u}_k(f),\cdots,\boldsymbol{u}_M(f)]$ 为特征向量矩阵；$(\cdot)^{\mathrm{H}}$ 表示共轭转置；$\boldsymbol{\Lambda}=\mathrm{diag}[\sigma_1^2(f,\tau),$

$\sigma_2^2(f,\tau),\cdots,\sigma_k^2(f,\tau),\cdots,\sigma_M^2(f,\tau)]$为一个对角矩阵，$\sigma_i^2(f,\tau)$ $(i=1,2,\cdots,M)$ 为特征值，表示第i个源信号的方差(信号功率)，$\boldsymbol{u}_i(f)$ 为其对应的特征向量。假设语音信号满足W-分离正交性，如果在该时频点处只有第 k 个源信号存在，则其他信号的功率均为零，即：

$$\sigma_k^2(f,\tau)\neq 0\ ,\sigma_i^2(f,\tau)=0 \qquad i=1,2,\cdots,M,i\neq k \tag{8}$$

考虑噪声的影响，$\sigma_i^2(f,\tau)$ 的值可能很小但不为零，即：

$$\sigma_k^2(f,\tau)>>\sigma_i^2(f,\tau) \qquad i=1,2,\cdots,M,i\neq k \tag{9}$$

对所有的特征值进行排序，使得：

$$\sigma_1^2\geqslant\sigma_2^2\geqslant\cdots\geqslant\sigma_M^2 \tag{10}$$

引入一个阈值：

$$\varepsilon=1-\frac{\sqrt{\sigma_2^2}}{\sqrt{\sigma_1^2}} \tag{11}$$

根据式(9)可知，$0\leqslant\varepsilon<1$，当 ε 取值近似为1时，说明此时只有一个源信号的能量非常大，表示只有一个源信号存在。因此，利用上述方法可以检测出单源时频点。

3 信道估计

3.1 归一化处理

对于检测出的单源信号时频点，假设第 k 个源信号存在，式(3)近似为：

$$\boldsymbol{X}(f,\tau)\approx\boldsymbol{h}_k(f)s_k(f,\tau) \tag{12}$$

如图1所示，选定一个参考传感器 J，根据近声场模型可以将多路径混合模型近似为直达路径模型，第 k 个源信号到第 i 个传感器的信道为[5]：

$$h_{ik}(f)\approx\frac{q(f)}{d_{ik}}\exp[\mathrm{j}2\pi fc^{-1}(d_{ik}-d_{Jk})] \tag{13}$$

式中 $d_{ik}>0$ 为源信号 k 与传感器 i 之间的距离；c 为声音传播的速度；$q(f)$ 为与频率有关的常数。第 k 个源信号到达第 i 个传感器相对于到达参考传感器的相位差为 $2\pi fc^{-1}(d_{ik}-d_{Jk})$。

对式(12)进行相位归一化处理：

$$\bar{x}_i(f,\tau)=|x_i(f,\tau)|\exp\left(\mathrm{j}\frac{\arg[x_i(f,\tau)/x_J(f,\tau)]}{4fc^{-1}d_{\max}}\right)=$$

$$|h_{ik}(f)s_k(f,\tau)|\exp\left(\mathrm{j}\frac{\pi(d_{ik}-d_{Jk})}{2d_{\max}}\right) \tag{14}$$

式中 $d_{\max}$ 为参考传感器与其他传感器之间的最大距离。将式(14)写成向量形式：

$$\bar{\boldsymbol{X}}(f,\tau)=|s_k(f,\tau)|\bar{\boldsymbol{h}}_k(f) \tag{15}$$

其中，

$$\bar{\boldsymbol{h}}_k(f)=\left[|h_{1k}(f)|\exp\left(\mathrm{j}\frac{\pi(d_{1k}-d_{Jk})}{2d_{\max}}\right),\right.$$

$$|h_{2k}(f)|\exp\left(\mathrm{j}\frac{\pi(d_{2k}-d_{Jk})}{2d_{\max}}\right),\cdots,$$

$$\left.|h_{Mk}(f)|\exp\left(\mathrm{j}\frac{\pi(d_{Mk}-d_{Jk})}{2d_{\max}}\right)\right]$$

$$\bar{\boldsymbol{X}}(f,\tau)=\left[\bar{x}_1(f,\tau),\bar{x}_2(f,\tau),\cdots,\bar{x}_M(f,\tau)\right]^{\mathrm{T}}$$

对式(12)进行幅度归一化处理：

$$\bar{\bar{\boldsymbol{X}}}(f,\tau)=\bar{\boldsymbol{X}}(f,\tau)/\left\|\bar{\boldsymbol{X}}(f,\tau)\right\| \tag{16}$$

此时有：$\bar{\bar{x}}_i(f,\tau)=\dfrac{1}{d_{ik}D}\exp\left(\mathrm{j}\dfrac{\pi(d_{ik}-d_{Jk})}{2d_{\max}}\right)$，

$D=\sqrt{\sum_{i=1}^{M}\dfrac{1}{d_{ik}^2}}$，且：

$$\bar{\bar{\boldsymbol{X}}}(f,\tau)=\left[\bar{\bar{x}}_1(f,\tau),\bar{\bar{x}}_2(f,\tau),\cdots,\bar{\bar{x}}_M(f,\tau)\right]^{\mathrm{T}} \tag{17}$$

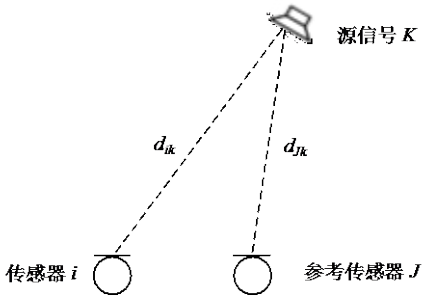

图1 近声场模型

3.2 K-means聚类

从式(17)可以看出，经过归一化处理以后，单源时频点的幅度和相位均与频率无关，与源信号的幅度和相位也无关，只与源信号相对于传感器的位置以及传感器之间的距离有关，实际上是对相应信道处理的结果。因此，不同频率点处同一个源信号的信道，在归一化处理以后与频率无关且相等，所有单源时频点归一化处理的结果形成 N 个信道的集合。对归一化处理的结果进行K-means聚类，每个聚类中心即为其中一个信道的估计。

将归一化处理的结果聚成 $C_1,C_2,\cdots,C_N$ 共 N 类，对应的聚类中心分别为 $\boldsymbol{c}_1,\boldsymbol{c}_2,\cdots,\boldsymbol{c}_N$。聚类中心由下式计算：

$$\boldsymbol{c}_k=\frac{1}{|C_k|}\sum_{\bar{\bar{\boldsymbol{X}}}(f,\tau)\in C_k}\bar{\bar{\boldsymbol{X}}}(f,\tau),\quad \boldsymbol{c}_k=\frac{\boldsymbol{c}_k}{\|\boldsymbol{c}_k\|} \tag{18}$$

K-means聚类要满足如下代价函数：

$$J=\min\ \sum_{k=1}^{N}J_k\ ,\quad J_k=\sum_{\bar{\bar{\boldsymbol{X}}}(f,\tau)\in C_k}\left\|\bar{\bar{\boldsymbol{X}}}(f,\tau)-\boldsymbol{c}_k\right\| \tag{19}$$

3.3　估计各频率点处的信道

聚类结束后，对聚类中心进行归一化逆处理，从而估计出各个频率点处的信道：

$$\hat{\boldsymbol{h}}_k(f)=|\boldsymbol{c}_k|\exp[-\mathrm{j}4\pi f d_{\max}c^{-1}\arg(\boldsymbol{c}_k)] \tag{20}$$

在式(23)的估计方法中，各频率点处信道的排序是固定的，因此不存在频率置换问题。由于 $q(f)$ 未知，所以上述方法存在比例模糊问题，但从实际分离效果来看，这种影响并不大，也即是说不同频率处的 $q(f)$ 近似相等，因此可将 $q(f)$ 近似为1。

4　基于信号子空间的源信号重构

在信道已知的情况下，本文采用信号子空间方法重构源信号。时频二元掩蔽方法在重构源信号时，假设每个时频点至多存在一个源信号，因此导致了音乐噪声问题。而信号子空间法允许每个时频点存在多个源信号，但存在源信号的数目要小于传感器的数目，这在一定程度上减少了音乐噪声。

在同一频率点处，由 K 个信道向量构成的信号子空间为：

$$\boldsymbol{W}(f)=[\hat{\boldsymbol{h}}_{\alpha_1}(f),\hat{\boldsymbol{h}}_{\alpha_2}(f),\cdots,\hat{\boldsymbol{h}}_{\alpha_K}(f)]\quad K\leqslant M-1 \tag{21}$$

式中　$[\hat{\boldsymbol{h}}_{\alpha_1}(f),\hat{\boldsymbol{h}}_{\alpha 2}(f),\cdots,\hat{\boldsymbol{h}}_{\alpha_K}(f)]$ 是 $[\hat{\boldsymbol{h}}_1(f),\hat{\boldsymbol{h}}_2(f),\cdots,\hat{\boldsymbol{h}}_N(f)]$ 的一个子集。为表达简洁，省略频率指示 f 。设 $\boldsymbol{Q}_W$ 为 $\boldsymbol{W}$ 对应的信号子空间的正交投影矩阵，则：

$$\boldsymbol{Q}_W=\boldsymbol{I}-\boldsymbol{W}(\boldsymbol{W}^{\mathrm{H}}\boldsymbol{W})^{-1}\boldsymbol{W}^{\mathrm{H}} \tag{22}$$

式中　$\boldsymbol{I}$ 为单位矩阵。根据 $\boldsymbol{Q}_W$ 的性质有：

$$\begin{cases}\boldsymbol{Q}_W\hat{\boldsymbol{h}}_i=0 & i\in\{\alpha_1,\alpha_2,\cdots,\alpha_K\}\\ \boldsymbol{Q}_W\hat{\boldsymbol{h}}_i\neq 0 & \text{其他}\end{cases} \tag{23}$$

考虑到噪声的影响，可采取如下方法确定组成信号子空间的信道向量，即确定一个时频点处存在的源信号：

$$\{\alpha_1,\alpha_1,\cdots,\alpha_K\}=\min_{\beta_1,\beta_2,\cdots,\beta_K}\left\{\left\|\boldsymbol{Q}_{W_\beta}\boldsymbol{X}(f,\tau)\right\|\ \middle|\ \boldsymbol{W}_\beta\right\} \tag{24}$$

$\{\alpha_1,\alpha_2,\cdots,\alpha_K\}$ 确定以后，就可以确定 $\boldsymbol{W}(f)$ 并用于分离信号：

$$\tilde{\boldsymbol{S}}(f,\tau)=\boldsymbol{W}^{\#}(f)\boldsymbol{X}(f,\tau) \tag{25}$$

式中　$\tilde{\boldsymbol{S}}(f,\tau)=[y_{\alpha_1}(f,\tau),y_{\alpha_2}(f,\tau),\cdots,y_{\alpha_K}(f,\tau)]^{\mathrm{T}}$，其他源信号在该时频点处的估计值均为零，$(\cdot)^{\#}$ 表示伪逆。对所有时频点进行类似处理后，得到重构的源信号 $\hat{\boldsymbol{s}}_k(f,\tau)$，$k=1,2,\cdots,N$，最后进行短时傅里叶逆变换(ISTFT)得到时域信号：

$$\hat{\boldsymbol{s}}_k(\tau+r)=\frac{1}{L\cdot\mathrm{win}(r)}\sum_f\hat{\boldsymbol{s}}_k(f,\tau)\mathrm{e}^{\mathrm{j}2\pi f r} \tag{26}$$

式中　$f\in\left\{0,\frac{1}{L}f_s,\cdots,\frac{L-1}{L}f_s\right\}$。

5　仿真实验

本文对3个传感器和4个语音源信号的卷积混合情况进行了仿真，仿真环境如图2所示。采用线性传感器阵列，传感器间距均为4 cm，源信号距离阵列中心均为1.1 m，语音长度为5 s，采样频率 $f_s=10$ kHz，进行短时傅里叶变换时采用Hamming窗，窗长(帧长)512点，帧移256点，房间的回响时间 $\mathrm{RT}_{60}\approx 100\,\mathrm{ms}$，单源时频点检测阈值 $\varepsilon=0.99$。

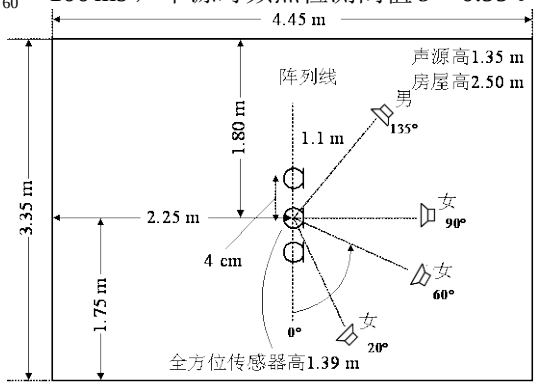

图2　仿真环境的设置

定义两个性能评价指标[5]：

(1) 分离信号的信干比(signal to interference rRatio，SIR)：

$$\mathrm{SIR}_i=10\lg\frac{\sum_t|y_{ii}(t)|^2}{\sum_t\left|\sum_{k\neq i}y_{ik}(t)\right|^2} \tag{27}$$

式中　$y_{ik}(t)$ 为分离信号，$y_i(t)$ 中包含的源信号 $s_k(t)$ 的成分；SIR用于评价算法的分离性能。

(2)分离信号的信号失真比(signal to distortion ratio，SDR)：

$$\mathrm{SDR}_i=10\lg\frac{\sum_t|x_{Ji}(t)|^2}{\sum_t|x_{Ji}(t)-\beta y_{ii}(t-D)|^2} \tag{28}$$

式中　$x_{Ji}(t)$ 为第 J 个传感器接收到的第 i 个源信号 $s_i(t)$；β 和 D 分别用于补偿 $x_{Ji}(t)$ 与 $y_{ii}(t)$ 之间幅度和相位的差异；SDR用于评价分离语音的质量。

本文采用BSS_EVAL Toolbox[10]计算上述两种性能评价指标。分别对MENUET方法和本文方法进行了仿真，同时利用提出的信道估计方法，对不采用信号子空间的源信号重构方法进行了仿真(即式(21)中令 $K=1$，简记为非子空间方法)，3种方法的性能如图3和图4所示。

从图3和图4中可以看出，非子空间方法与MENUET方法的性能接近，即在语音信号满足严格W-分离正交性的假设下，基于信道估计的分离方法能够达到MENUET方法的分离性能。由于MENUET方法假设每个时频点至多存在一个源信号，如果时频点存在多个源信号而仍然将其提取出来，则会增加分离信号中的干扰成分，从而降低信干比(SIR)。本文的方法是针对部分W-分离正交性提出的，减小了对语音稀疏性的依赖，假设每个时频点至多可以存在 $M-1$ 个源信号，能够在一定程度上减少分离信号中的干扰成分，从而提高分离性能。与MENUET方法相比，本文方法的平均SIR增加了4.61dB，同时平均SDR增加了3.02 dB，并且从实际试听方面来讲，分离效果也更好。

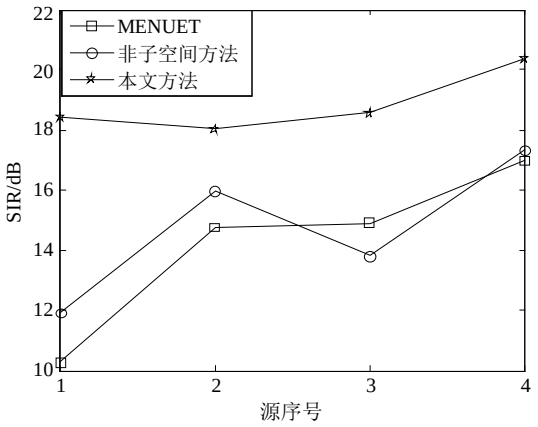

图3 SIR性能比较

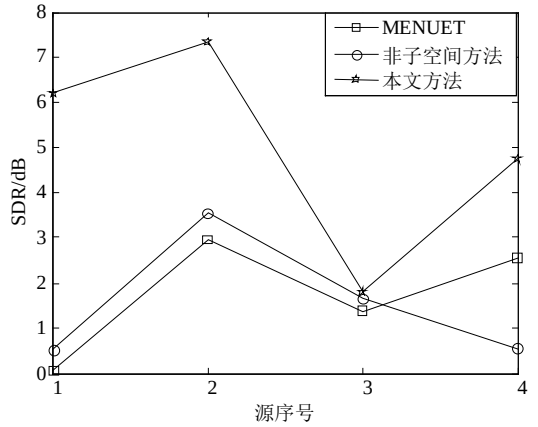

图4 SDR性能比较

由于本文的方法需要通过检测单源时频点估计信道，所以具有一定的局限性。因为随着回响情况的加剧，各源信号时频点的重叠越来越多，而单源时频点越来越少，对检测造成了一定的困难，所以本文方法适用于回响不严重的情况。从另一个方面讲，如果回响非常严重，使得不存在单源时频点，则基于稀疏性方法的前提也就不成立，因此需要寻找新的方法和理论解决严重回响的问题。

6 结 论

在语音盲分离中，W-分离正交性的假设可以使问题简化，但也由此导致了音乐噪声问题。本文针对语音信号的部分W-分离正交性，通过检测单源时频点来估计信道，估计的信道不存在频率置换问题，并且由于减少了对稀疏性的依赖，估计精度得到了提高。在结合信号子空间方法重构源信号时，可以有效减少音乐噪声，并且提高分离性能。

参 考 文 献

[1] HAYKIN S. Unsupervised adaptive filtering, volume 1: blind source separation[M]. Canada: John Wiley & Sons, 2000.

[2] 傅 彦, 周俊临. 基于无监督学习的盲信号源分离技术研究[J]. 电子科技大学学报, 2004, 33(1): 63-66.
FU Yan, ZHOU Jun-lin. Research of blind source Technology which based on unsupervised learning[J]. Journal of University of Electronic Science and Technology of China, 2004, 33(1): 63-66.

[3] 徐 舜, 刘郁林, 朱行涛. 采用二阶特征窗的语音盲分离方法[J]. 电子科技大学学报, 2008, 37(3): 374-377.
XU Shun, LIU Yu-lin, ZHU Xing-tao. Second order characteristic window approach to blind speech signal separation[J]. Journal of University of Electronic Science and Technology of China, 2008, 37(3): 374-377.

[4] RICKARD S, YILMAZ Z. On the approximate W-disjoint orthogonality of speech[C]//Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing. Orlando: IEEE Computer Society, 2002: 529-532.

[5] ARAKI S, SAWADA H, MUKAI R, et al. Underdetermined blind sparse source separation for arbitrarily arranged multiple sensors[J]. Signal Processing, 2007, 87(8): 1833-1847.

[6] KIM M, CHOI S. ICA-based clustering for resolving permutation ambiguity in frequency domain convolutive source separation[C]//Proceedings of the 18th International Conference on Pattern Recognition. Washington: IEEE Computer Society, 2006: 950-954.

[7] RAKI S A, MAKINO S, SAWADA H, et al. Reducing musical noise by a fine-shift overlap-add method applied to source separation using a time-frequency mask[C]// Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing. Philadelphia: IEEE Computer Society, 2005, 3: 81-84.

[8] AÏSSA-EL-BEY A, ABED-MERAIM K, GRENIER Y. Underdetermined blind source separation of audio sources in time-frequency domain[C]//Proceedings of the Signal Processing with Adaptive Sparse Structured Representation. Rennes: ACTES, 2005: 100-113.

[9] WINTER S, KELLERMANN W, SAWADA H, et al. MAP-based underdetermined blind source separation of convolutive mixtures by hierarchical clustering and l1-norm minimization[J]. EURASIP Journal on Applied Signal Processing, 2007, 2007(1): 81-81.

[10] FEVOTTE C, GRIBONVAL R, VINCENT E. BSS EVAL toolbox user guide[DB/OL]. [2008-06-21]. http://www.irisa.fr/metiss/bss eval/ .

编 辑 张 俊