基于新型混合模型的欠定盲分离方法

陈永强 [1, 2] 王宏霞 [1]

摘 要 针对欠定盲分离问题, 提出了一种新的源恢复方法. 在时频域局部区域采用复高斯分布对源信号进行建模, 将语音信号的稀疏性和局部平稳性结合在一起, 提出了一种新的混合模型来描述观测信号在局部区域的概率分布. 通过该模型, 将每个时频点的源信号状态的判断问题转换成模型的参数估计和后验概率的计算问题, 最后通过子混合矩阵的逆恢复出源信号. 实验结果表明, 该方法具有很快的收敛速度, 并且比已有方法具有更好的分离性能.

关键词 欠定盲分离, 混合模型, 稀疏性, 局部复高斯分布, 最大后验概率

引用格式 陈永强, 王宏霞. 基于新型混合模型的欠定盲分离方法. 自动化学报, 2014, **40**(7): 1412−1420

DOI 10.3724/SP.J.1004.2014.01412

A Method for Under-determined Blind Source Separation Based on New Mixture Model

CHEN Yong-Qiang[1, 2] WANG Hong-Xia[1]

Abstract To solve the problem of under-determined blind source separation, we propose a new source recovery method. By utilizing the complex valued Gaussian model to characterize the local distribution of source signals in each micro-region in the time-frequency domain and combining speech signals′ sparsity with their local stability, a new mixture model is derived to characterize the local distribution of observed signals. We convert the problem of judging the state of each source signal at each time-frequency point into a problem of model′s parameters estimation and posterior probability computation. Finally, the source signals are recovered by sub-mixing matrix′s inverse. Experiment results show that the proposed method converges very fast and has better separation performance compared with the existing methods.

Key words Under-determined blind source separation, mixture model, sparsity, local complex Gaussian distribution, maximum a posteriori

Citation Chen Yong-Qiang, Wang Hong-Xia. A method for under-determined blind source separation based on new mixture model. *Acta Automatica Sinica*, 2014, **40**(7): 1412−1420

盲源分离被广泛应用到语音识别、通信、生物医学信号处理等许多领域. 以往的研究多数针对适定或过定盲分离[1−2], 要求观测信号数目 M 不少于源信号数目 N, 从而使盲分离的应用受到了限制. 因此, 解决欠定混合 (即 M 小于 N) 下的盲分离问题就成为了盲分离领域的研究热点. 目前, 解决欠定盲分离的方法有很多, 大多数是利用信号的稀疏性. 文献 [3] 提出的退化分离估计技术 (Degenerate unmixing estimation technique, DUET) 假设源信号在时频域具有正交性, 以此来分离源信号. 文献 [4] 进一步将 DUET 算法拓展到多个传感器的情况. 以上文献对信号的稀疏性要求很高, 当源信号的稀疏度不够, 在某些时频点相互重叠时, 就会产生较大的音乐噪声.

文献 [5] 将信号的稀疏性与指数型分布联系起来, 在概率统计的框架下估计源信号. 为了获得更好的分离性能, 文献 [6−11] 相继从概率模型的角度, 提出了各种盲分离方法. 文献 [6] 提出拉普拉斯混合模型 (Laplacian mixture model, LMM) 来估计混合矩阵, 并通过 L_1-norm 最小化方法恢复源信号. 文献 [7−8] 分别采用柯西分布和方向拉普拉斯混合模型 (Mixtures of directional Laplacian distributions, MDLD) 获得二元时频掩蔽的最大后验估计. 文献 [9] 则采用高斯分布来描述源信号在局部区域的分布, 并同时估计混合矩阵和参数, 但收敛速度非常慢. 文献 [10−11] 分别采用

收稿日期 2012-12-31 录用日期 2013-08-01
Manuscript received December 31, 2012; accepted August 1, 2013
国家自然科学基金 (61170226), 中央高校基本科研业务费专项资金 (SWJTU11CX047, SWJTU12ZT02), 四川省青年科技创新研究团队项目 (2011JTD0007) 资助
Supported by National Natural Science Foundation of China (61170226), Fundamental Research Funds for the Central Universities (SWJTU11CX047, SWJTU12ZT02), and Young Innovative Research Team of Sichuan Province (2011JTD0007)
本文责任编委 张长水
Recommended by Associate Editor ZHANG Chang-Shui
1. 西南交通大学信息科学与技术学院 成都 610031 2. 成都信息工程学院电子实验中心 成都 610225
1. School of Information Science and Technology, Southwest Jiaotong University, Chengdu 610031 2. Electronic Experiment Center, Chengdu University of Information Technology, Chengdu 610225

高斯混合模型 (Gaussian mixture model, GMM) 和可伸缩高斯混合模型 (Gaussian scaled mixture model, GSMM) 描述源信号的分布, 其计算量会随源信号个数按指数关系急剧增加, 计算负担相当大. 文献 [12] 提出的基于广义高斯分布 (Generalized Gaussian distribution, GGD) 的盲分离算法, 每次迭代都需对函数进行优化, 同样存在计算复杂度大的问题.

相比基于概率模型的方法, 文献 [13−17] 的方法则具有较小的计算量, 并且允许源信号在时频域可以有一定的重叠, 从而减小音乐噪声. 文献 [13] 提出的子空间投影方法, 在每个时频点的主导源个数严格小于观测信号个数时, 可以判断该时频点有哪些源处于激活状态, 以此来恢复源信号. 文献 [14] 则提出了子空间投影的改进方法, 通过寻找使得观测矢量到混合矩阵列矢量张成的子空间距离为 0 (或接近 0) 时, 所需的最少列矢量数目来判断每个时频点处的实际主导源数. 文献 [15] 进一步将条件放宽, 允许每个时频点上的主导源个数最多可以达到观测信号的个数. 文献 [16] 则不限制主导源的个数, 只要源数目和观测信号数目满足 $N \leq 2M-1$ 条件, 就可以分离源信号. 但是, 文献 [15−16] 采用的是维格纳-威利分布方法, 是二次型时频表示, 因此含有不同源之间的交叉项, 当交叉项无法完全去除或者交叉项和自项在某些时频点重合时, 会对源信号的分离性能产生影响. 文献 [17] 则提出一种基于短时傅立叶变换 (Short time Fourier transform, STFT) 的观测信号协方差矩阵对角化方法, 来判断时频域上的局部区域有哪些源处于激活状态. 与文献 [15] 一样, 该方法同样允许每个时频点上主导源的个数最多可以为观测信号个数. 由于 STFT 是线性变换, 不存在交叉项的问题, 因此, 对语音信号有比较好的分离性能. 但该方法有一个假设前提, 即每个局部区域内所有时频点上起主导作用的源是完全一样的. 由于语音信号是非平稳的, 当局部区域内不同时频点上的主导源不一样时, 分离性能会下降. 针对以上文献在判断时频点主导源时所存在的问题, 本文提出了一种通过最大后验概率来判断各个时频点源信号状态的新方法. 本文对信号采用 STFT 变换, 将时频点上源信号的状态作为隐变量, 利用语音信号的弱稀疏性和局部平稳性, 构建了一个新的混合模型来描述局部区域内观测信号的分布; 然后, 通过对混合模型的参数估计, 用最大后验概率来判断每个时频点的源信号状态, 并由此恢复出源信号. 与文献 [17] 不同的是, 本文算法可以对每个时频点的源信号状态进行独立的判断, 并不要求局部区域内各时频点的主导源必须是完全一样的, 这与文献 [13−14] 类似. 但是, 本文算法允许主导源个数最多可以与观测信号数目相等, 因此优于文献 [13−14] 的算法. 虽然文献 [9] 也采用了局部高斯分布, 但是本文提出的是一种局部区域的混合模型. 本文算法并不将源信号本身作为隐变量, 而是将源信号的状态作为隐变量, 并利用了信号的稀疏性. 本文提出的混合模型可以有效克服 GMM、GSMM[10−11] 和 GGD 模型[12] 计算量过大的问题, 具有很快的收敛速度和更好的分离性能.

1 问题描述

本文考虑欠定盲分离, 即观测信号的个数 M 小于源信号的个数 N. 对观测信号采用短时傅里叶变换, 在每个时频点, 瞬时混合模型可写成

$$\boldsymbol{x}(t,f) = A\boldsymbol{s}(t,f) \tag{1}$$

其中, $\boldsymbol{x}(t,f) = [x_1(t,f), \cdots, x_M(t,f)]^{\mathrm{T}}$ 是 M 维观测数据向量; $\boldsymbol{s}(t,f) = [s_1(t,f), \cdots, s_N(t,f)]^{\mathrm{T}}$ 是 N 维源信号数据向量; A 是 $M \times N$ 维混合矩阵. 本文假设 A 行满秩, 且各源信号之间相互独立. 目前, 欠定盲分离混合矩阵的估计方法有很多[18−21], 为了减小计算量和估计精度, 很多文献在估计混合矩阵之前, 先检测那些只有一个主导源的时频点或时频区域 (即单源点或单源区)[12, 14, 18, 21], 然后通过各种聚类算法估计出混合矩阵的各个列矢量.

与适定 ($M = N$) 和超定 ($M > N$) 盲分离不同, 欠定盲分离即使估计出了混合矩阵, 也无法直接估计出源信号, 因为欠定盲分离的源估计其实是一个病态问题. 因此, 欠定盲分离包括混合矩阵估计和源恢复两个问题, 本文主要讨论源恢复问题, 下面介绍本文的源恢复方法.

2 观测信号的新型混合模型

2.1 混合模型的导出

先假设在任何时频点处的主导源个数等于观测信号个数, 根据这一假设, 本文给出一种新的混合模型来描述观测信号的分布. 为了引出这一混合模型, 先给出 4 个定义:

定义 1. M 个主导源对应 A 的 M 个列矢量, 这些列矢量按照其在 A 中的原排列次序构成的 $M \times M$ 维子矩阵称为主导源的顺序子矩阵, 记为 $B_q (q = 1, 2, \cdots, \mathrm{C}_N^M)$.

显然, 一共有 $Q = \mathrm{C}_N^M$ 个顺序子矩阵, B_q 表示第 q 个顺序子矩阵.

定义 2. 顺序子矩阵 B_q 的各个列矢量对应主导源信号, 将这些主导源信号的序号从小到大构成的集合称为 B_q 的顺序集合, 记为 $\mathbf{C}_q$.

例如, B_q 的各列分别对应源信号 s_1、s_3、s_4,

那么主导源信号的序号构成的顺序集合 $\mathbf{C}_q = \{1,3,4\}$.

定义 3. 如果顺序集合 $\mathbf{C}_q$ 包含序号 j, 那么将所有这样的 q 构成一个集合, 这个集合称为 j 的包含集合, 记为 $\mathbf{Z}_j$.

定义 4. 如果在时频点 (t,f), 主导源的顺序子矩阵为 B_q, 那么则称时频点 (t,f) 所处的状态 $\gamma(t,f)=q$.

从全局来看, 语音信号是弱稀疏信号, 其分布属于超高斯分布, 可以用指数型分布[5] 和柯西分布[7] 等来描述. 同时, 语音信号在局部又具有高斯分布的特征, 文献 [9] 在局部区域用高斯模型来描述其分布. 本文采用复高斯概率模型[22] 描述源信号在某个时频小区域内的分布, 即

$$P(s_j(t,f)) = \frac{1}{\pi\sigma_j^2}\exp\left(-\frac{|s_j(t,f)|^2}{\sigma_j^2}\right) \tag{2}$$

其中, σ_j^2 是该时频区域的方差, 代表 s_j 在该区域的能量强度. 对于不同时频区域, σ_j^2 是不同的, 这就是说语音信号在整个时频域上是非平稳的, 但是具有局部平稳性.

根据式 (1) 和式 (2) 以及各个源信号之间的相互独立性, 由概率论相关知识可知, 如果时频点所处的状态 $\gamma(t,f)=q$, 则观测信号在某个小区域内各个时频点的联合概率密度函数可表示为

$$P(\boldsymbol{x}(t,f)\,|\gamma(t,f)=q,\sigma_j^2\,,j\in\mathbf{C}_q) = [\![D_q]\!]\prod_{j\in\mathbf{C}_q}\frac{1}{\pi\sigma_j^2}\exp\left(-\frac{\left|\langle D_q\boldsymbol{x}(t,f)\rangle_{j_{\mathbf{C}_q}}\right|^2}{\sigma_j^2}\right) \tag{3}$$

其中, $[\![\cdot]\!]$ 表示行列式的绝对值, $D_q = B_q^{-1}$, $\langle\cdot\rangle_{j_{\mathbf{C}_q}}$ 表示向量的第 $j_{\mathbf{C}_q}$ 个元素, $j_{\mathbf{C}_q}$ 是 j 在集合 $\mathbf{C}_q$ 中的次序, 例如: $\mathbf{C}_3=\{1,3,4\}$, 4 在 $\mathbf{C}_3$ 中的次序是 3, 因此 $4_{\mathbf{C}_3}$=3.

由于时频点 (t,f) 可能处于 Q 种状态里的任何一种状态, 那么根据式 (3), 观测信号在某个小区上的分布可用下式来描述

$$P(\boldsymbol{x}(t,f)\,|\theta) = \sum_{q=1}^{Q} w_q\,[\![D_q]\!]\prod_{j\in\mathbf{C}_q}\frac{1}{\pi\sigma_j^2}\exp\left(-\frac{\left|\langle D_q\boldsymbol{x}(t,f)\rangle_{j_{\mathbf{C}_q}}\right|^2}{\sigma_j^2}\right) \tag{4}$$

其中, $\theta=\{w_q,\sigma_j^2\,|\,q=1,2,\cdots,Q,\ j=1,2,\cdots,N\}$; w_q 是时频点 (t,f) 的状态为 q 的概率, 且满足 $\sum_{q=1}^{Q} w_q = 1$, 本文将 w_q 称为权重因子. 很显然, 这是一个混合概率模型. 与 GMM 和 GSMM 混合模型[10−11] 不同的是, GMM 和 GSMM 描述源信号分布, 而本文的混合模型描述观测信号分布, 因此, 可以避免源信号个数较多所造成的计算瓶颈问题. LMM 和 MDLD 混合模型[6,8] 虽然描述观测信号, 但前者只能用于混合矩阵的估计, 后者则是一种时频掩蔽算法. 而本文的混合模型可以用来恢复源信号, 因为本文算法可以允许源信号之间有重叠, 所以具有比时频掩蔽算法更小的音乐噪声.

如果能够估计出式 (4) 中的 w_q 和 σ_j^2, 就可以知道每个时频点所处状态的后验概率, 即

$$P(\gamma(t,f)=l\,|\boldsymbol{x}(t,f),\theta) = \frac{P(\boldsymbol{x}(t,f)\,|\gamma(t,f)=l,\theta)P(\gamma(t,f)=l\,|\theta)P(\theta)}{\sum_{q=1}^{Q}P(\boldsymbol{x}(t,f)\,|\gamma(t,f)=q,\theta)P(\gamma(t,f)=q\,|\theta)P(\theta)} = \frac{P(\boldsymbol{x}(t,f)\,|\gamma(t,f)=l,\theta)w_l}{\sum_{q=1}^{Q}P(\boldsymbol{x}(t,f)\,|\gamma(t,f)=q,\theta)w_q} \tag{5}$$

其中, $P(\theta)$ 是参数的先验分布, 这里认为先验分布是均匀分布. 由式 (3), 上式进一步可以写成

$$P(\gamma(t,f)=l\,|\boldsymbol{x}(t,f),\theta) = \frac{[\![D_l]\!]\prod_{j\in\mathbf{C}_l}\frac{1}{\pi\sigma_j^2}\exp\left(-\frac{\left|\langle D_l\boldsymbol{x}(t,f)\rangle_{j_{\mathbf{C}_l}}\right|^2}{\sigma_j^2}\right)w_l}{\sum_{q=1}^{Q}[\![D_q]\!]\prod_{j\in\mathbf{C}_q}\frac{1}{\pi\sigma_j^2}\exp\left(-\frac{\left|\langle D_q\boldsymbol{x}(t,f)\rangle_{j_{\mathbf{C}_q}}\right|^2}{\sigma_j^2}\right)w_q} \tag{6}$$

其中, $l=1,2,\cdots,Q$. 当得到各个状态的后验概率分布, 就可以来判断该时频点所处的状态

$$\gamma(t,f) = \arg\max_{l} P(\gamma(t,f)=l\,|\boldsymbol{x}(t,f),\theta) \tag{7}$$

也就是说, 将后验概率最大的状态作为该时频点的状态. 为了求得每个时频点处的后验概率分布, 必须要估计出模型参数 θ, 这个问题将在本文第 3 节讨论.

利用式 (7) 判断出每个时频点所处的状态后, 就可以得到每个主导源信号在该时频点的系数, 即

$$\boldsymbol{s}_{\rm act}(t,f) = D_{\gamma(t,f)}\boldsymbol{x}(t,f) \tag{8}$$

并令非主导源信号在该点的系数为 0.

2.2 混合模型的进一步讨论

在导出式 (4) 所示的混合模型时, 假设主导源的个数等于观测信号的个数. 实际情况下, 由

于语音信号的稀疏性, 经常会出现时频点 (t,f) 处的主导源个数小于观测信号个数的情况. 这时, 时频点 (t,f) 所处的状态将不是唯一的. 例如, 当有 3 个观测信号和 4 个源信号时, 一共有 4 种状态. 这 4 种状态对应的顺序集合分别为 $\{1,2,3\}$、$\{1,2,4\}$、$\{1,3,4\}$、$\{2,3,4\}$. 如果在该时频点只有源信号 s_1 激活, 那么该时频点将有 3 种状态, 对应的顺序集合分别是 $\{1,2,3\}$、$\{1,2,4\}$、$\{1,3,4\}$. 只要后验概率最大的状态是这 3 种状态中的任意一个, 换句话说, 只要 $\{2,3,4\}$ 的后验概率不是最大的, 就可以保证能够正确的恢复出各个源信号, 这一点可以从式 (8) 看出.

3　混合模型参数估计和源信号的恢复

根据最大似然原理, 能够使 $P(\boldsymbol{x}\,|\theta)$ 最大的 θ 就是模型的参数. 但直接优化 $P(\boldsymbol{x}\,|\theta)$ 是非常困难的. 因此, 本文将时频点状态作为隐变量, 并通过期望最大化 (Expectation maxmization, EM) 算法[23] 来估计出每个局部区域的 θ. 为此, 引入辅助函数

$$\begin{aligned}H(\theta,\theta')=&\sum_{f=1}^{F}\sum_{t=1}^{T}\sum_{q=1}^{Q}\{P(\gamma(t,f)=q\,|\boldsymbol{x}(t,f),\theta')\times\\&\ln\left[P(\boldsymbol{x}(t,f)\,|\gamma(t,f)=q,\theta)P(\gamma(t,f)=q\,|\theta)\,\right]\}=\\&\sum_{f=1}^{F}\sum_{t=1}^{T}\sum_{q=1}^{Q}\{P(\gamma(t,f)=q\,|\boldsymbol{x}(t,f),\theta')\times\\&\ln\left[w_q\,[\![D_q]\!]\prod_{j\in\mathbf{C}_q}\frac{1}{\pi\sigma_j^2}\exp\left(-\frac{\left|\langle D_q\boldsymbol{x}(t,f)\rangle_{j_{\mathbf{C}_q}}\right|^2}{\sigma_j^2}\right)\right]\end{aligned}\tag{9}$$

式中, F、T 分别是小区域内的频率点和时间帧的个数; θ' 是上次迭代得到的参数估计, 最大化 $P(\boldsymbol{x}\,|\theta)$ 的问题就转化为最大化辅助函数的问题, 即找到一个新的 θ, 使得 $H(\theta,\theta')$ 最大化.

为了使上式最大化, 可以分别对 σ_j^2 和 w_q 求导, 并令导数为 0, 即可求得 σ_j^2 和 w_q. 只考虑式 (9) 中与 σ_j^2 有关的项 $H_{\sigma_j^2}(\theta,\theta')$, 并简化为

$$\begin{aligned}H_{\sigma_j^2}(\theta,\theta')=&\sum_{f=1}^{F}\sum_{t=1}^{T}\sum_{q=1}^{Q}\{P(\gamma(t,f)=q\,|\boldsymbol{x}(f,t),\theta')\times\\&\sum_{j\in\mathbf{C}_q}\left[-\frac{\left|\langle D_q\boldsymbol{x}(t,f)\rangle_{j_{\mathbf{C}_q}}\right|^2}{\sigma_j^2}-\ln(\sigma_j^2)\right]\end{aligned}\tag{10}$$

令 $\dfrac{\partial H(\theta,\theta')}{\partial\sigma_j^2}=\dfrac{\partial H_{\sigma_j^2}(\theta,\theta')}{\partial\sigma_j^2}=0$, 可以得到

$$\sigma_j^2=\frac{\sum\limits_{f,t,q}\left|\langle D_q\boldsymbol{x}(t,f)\rangle_{j_{\mathbf{C}_q}}\right|^2P(\gamma(t,f)=q\,|\boldsymbol{x}(t,f),\theta')}{\sum\limits_{f,t,q}P(\gamma(t,f)=q\,|\boldsymbol{x}(t,f),\theta')}\tag{11}$$

其中,

$$\sum_{f,t,q}=\sum_{f=1}^{F}\sum_{t=1}^{T}\sum_{q\in Z_j}\tag{12}$$

只考虑式 (9) 中与 $W=\{w_1,w_2,\cdots,w_Q\}$ 有关的项 $H_W(\theta,\theta')$, 可以简化为

$$H_W(\theta,\theta')=\sum_{f=1}^{F}\sum_{t=1}^{T}\sum_{q=1}^{Q}(P(\gamma(t,f)=q\,|\boldsymbol{x}(t,f),\theta')\ln w_q)\tag{13}$$

由于 Q 个权重因子要满足 $\sum_{q=1}^{Q}w_q=1$ 的约束条件, 因此, 可以用拉格朗日乘子法来求 w_q. 为此, 构造函数

$$\psi(w_1,w_2,\cdots,w_Q,\lambda)=H_W(\theta,\theta')+\lambda\left(\sum_{q=1}^{Q}w_q-1\right)\tag{14}$$

解方程组 $\begin{cases}\dfrac{\partial\psi}{\partial w_1}=0\\ \vdots\\ \dfrac{\partial\psi}{\partial w_Q}=0\\ \dfrac{\partial\psi}{\partial\lambda}=0\end{cases}$, 最后得到

$$\begin{aligned}w_l=&\frac{\sum\limits_{f=1}^{F}\sum\limits_{t=1}^{T}P(\gamma(t,f)=l\,|\boldsymbol{x}(t,f),\theta')}{\sum\limits_{f=1}^{F}\sum\limits_{t=1}^{T}\sum\limits_{q=1}^{Q}P(\gamma(t,f)=q\,|\boldsymbol{x}(t,f),\theta')}=\\&\frac{1}{FT}\sum_{f=1}^{F}\sum_{t=1}^{T}P(\gamma(t,f)=l\,|\boldsymbol{x}(t,f),\theta')\end{aligned}\tag{15}$$

本文提出的算法流程如下:

步骤 1. 首先估计出混合矩阵, 并将整个时频区划分为若干互不重叠的小区域;

步骤 2. 设置各小区的方差和权重因子初始值 $\theta^{(0)}=\left\{w_q^{(0)},\sigma_j^{(0)}\,|q=1,2,\cdots,Q,\ j=1,2,\cdots,N\right\}$, 并令迭代次数 $k=0$;

步骤 3. 通过式 (6) 计算小区中各个时频点的状态概率 $P(\gamma(t,f)=q\,|x(t,f),\theta^{(k)})$;

步骤 4. 通过式 (11) 和 (14) 分别计算小区的方差 $\sigma_j^{(k+1)}$ 和权重因子 $w_q^{(k+1)}$, 令 $k=k+1$, 并返回步骤 3, 一直到收敛;

步骤 5. 根据式 (7) 来决定每个时频点所处的状态, 然后根据式 (8) 分离出各个源信号的系数;

步骤 6. 当所有小区的源信号系数被估计出来后, 就可以通过傅里叶逆变换恢复出各个源信号的时域波形.

4 仿真实验

为了评价本文提出的算法性能, 将本文算法和基于协方差矩阵对角化 (Covariance matrix diagonalization, CMD)[17]、改进的子空间 (Improved subspace, IMSUB)[14]、GGD[12]、L_1-norm[6]、MDLD[8] 的源恢复算法进行比较. 在恢复源信号前, 先估计混合矩阵, 然后用估计出的混合矩阵来分离源信号. 除了 MDLD 外, 其他算法的混合矩阵都采用文献 [21] 提出的方法进行估计. 由于 MDLD 在估计混合矩阵列矢量的同时, 获得时频掩蔽的最大后验估计, 然后恢复源信号, 所以仿真时完全按文献 [8] 中的算法进行. 本文的源信号取自 http://www.speech.cs.cmu.edu/cmu_arctic/下载的 20 段语音信号 (男声 10 段, 女声 10 段, 采样率为 16 kHz), 这里将每一段截取为 2 s 作为源信号. MDLD 采用改进的离散余弦变换其他算法均采用短时傅立叶变换, 窗长都取为 1 024 点的汉宁窗, 重叠率为 50 %. 这里采用信干比 (Signal to interference ratio, SIR) 来衡量源信号的分离效果, 即

$$\mathrm{SIR\,(dB)}=\frac{1}{N}\sum_{i=1}^{N}10\lg\left\{\frac{\sum\limits_{t=1}^{P}s_i(t)^2}{\sum\limits_{t=1}^{P}\left[(s_i(t)-\hat{s}_i(t)\right]^2}\right\} \tag{16}$$

其中, $s_i(t)$ 和 $\hat{s}_i(t)$ 分别为时域中的源信号和估计信号, P 为采样个数. 实验中, 随机产生 20 次混合矩阵, 每次随机从 20 段语音中选取不同的源信号进行混合, 并保证男声和女声数量尽量相等; 然后, 计算 20 次 SIR 的平均值, 来比较各算法的性能. 以下仿真实验中的 SIR 都是指平均值.

4.1 局部区域的选择和算法性能分析

由于语音信号在整个时频域是非平稳的, 因此在时频域各处的方差是不一样的. 因此, 本文算法将时频平面划分为若干个小的局部区域, 这样方差就能够体现语音信号在局部区域的能量强度. CMD 算法的前提条件是各时频点的主导源完全一样, 因此, 也需要将局部区域取得较小.

为了有较小的计算开销, 同时也为了时频区的划分方便, 本文将小区域选择为互不重叠的矩形区域, 这和文献 [17] 是类似的. 以 2 个观测信号和 4 个源为例, 顺序集合有 6 个, 分别为 $\mathbf{C}_1=\{1,2\}$, $\mathbf{C}_2=\{1,3\}$, $\mathbf{C}_3=\{1,4\}$, $\mathbf{C}_4=\{2,3\}$, $\mathbf{C}_5=\{2,4\}$, $\mathbf{C}_6=\{3,4\}$, $\mathbf{C}_q$ 的下标 q 就是某时频点的状态, 即哪些源起主导作用. 在一个 1×4 (即包含 1 个频率点和 4 个时间帧) 的小区里, 各语音源信号在各个时频点的幅度 $|s_j(t,f)|$ 如图 1 (a) 所示. 可以看到, 在整个小区域内, 起主导作用的 2 个源始终是一样的, 也就意味着这 2 个源的方差较大. 通过实验, 我们发现存在着很多这样的小区域. 对于这样的小区域, 无论 CMD 算法还是本文算法, 对主导源的判断正确率都比较高. 但是, 还是有一些区域, 不满足这个条件, 如图 1 (b) 所示. 可以看到, 各时频点上的主导源是不完全一样的. 在第 1 个和第 3 个时频点, 4 个源的能量都比较小, 因此这 2 点的状态判断是无关紧要的; 在第 2 个时频点, 起主导作用的是 s_3, 其对应 3 种状态 "2"、"4"、"6", 也就是说, 将其判断为这 3 种状态中的任何一种都是对的; 而在第 4 个时频点, 起主导作用的则变成了 s_2 和 s_4, 其对应的状态为 "5". 因此, 从理论上讲, CMD 无法正确判断每个时频点的状态. 而本文算法与 CMD 算法不同的是, 本文算法并没有假设小区内各个时频点的状态必须是一样的, 本文算法可以根据观测信号对每个时频点的主导源进行 "动态" 判断, 而不是完全取决于这个小区的信号方差, 这一点可从式 (7) 看到. 如图 1 (c) 所示, 由于小区内各时频点的主导源不是完全一样的, CMD 算法出现了错误判断, 将整个小区各个时频点的状态都误判为 "1", 即 s_1 和 s_2 是主导源; 而本文算法在 2 个能量较大的时频点处, 都判断正确, 即第 2 个时频点的状态为 "2"; 第 4 个时频点的状态为 "5". 虽然对于图 1 (b) 这样的小区, 本文算法有时也会出现错误判断, 但仍可以在一定程度上提高主导源判断的准确率.

我们还可以做这样一个实验, 不用式 (7) 来判断各时频点的状态, 而仅用方差来判断, 即将本文算法收敛后的方差从大到小排序, 将前 M (M 为观测信号数目) 个源信号作为主导源, 然后通过式 (8) 得到源的估计. 通过 20 次实验, 可以发现 SIR 比原算法下降了 2.1 dB. 这也反映了本文算法对各个时频点的状态进行独立判断所具有的优点.

图 2 给出了不同尺寸的小区域对于本文算法和 CMD 算法的性能影响. 图中的小区尺寸用 $F\times T$ 表示, F 为频率点个数, T 为时间帧个数. 从图 2 可

以看出, 当局部区域较小时, 算法性能更好. 我们发现, $F < T$ 的小区要好于 $F > T$ 的小区, 同时, 选择 $1 \times T$ 这样的小区是比较理想的. 在所有小区中, 1×3 和 1×4 的小区对于 2 种算法来说, 都是最好的. 另外, 对于不同的小区, 本文算法都优于 CMD 算法. 以下仿真实验中, 小区都选为 1×4 的小区.

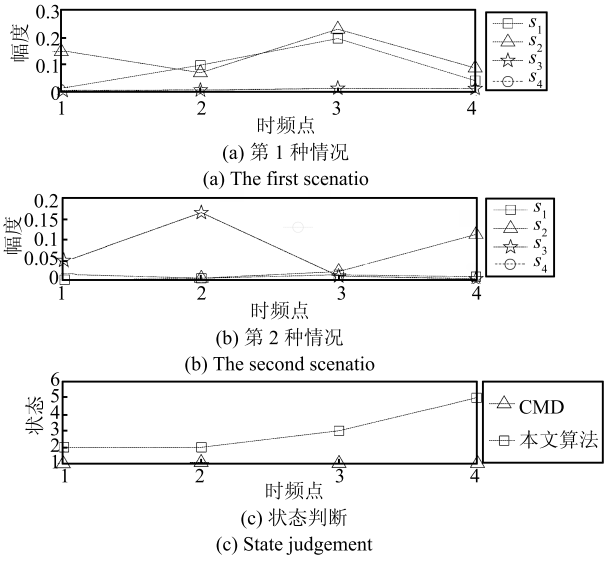


图 1　本文算法和 CMD 算法比较

Fig. 1　Comparison between the proposed algorithm and CMD

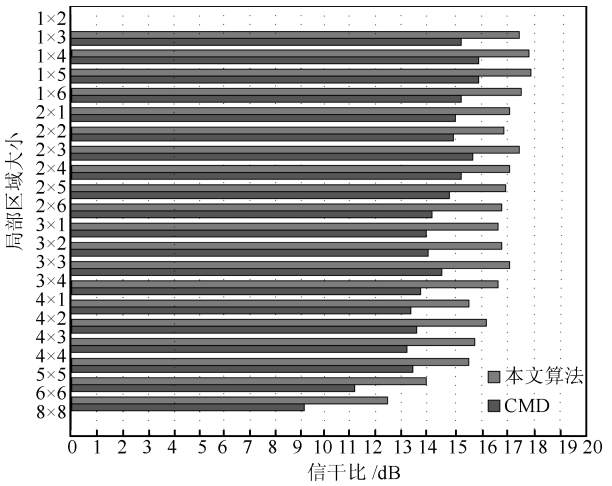


图 2　小区尺寸对算法性能的影响

Fig. 2　Size effect on the performance of the proposed algorithm and CMD

4.2　本文算法的收敛特性

图 3 是 3 个观测信号和 5 个源信号时, 本文算法模型参数的收敛曲线. 由于本文提出的算法在各个时频小区域中独立进行, 每个小区域里都要进行 EM 算法迭代. 为了看得清楚, 这里只画出了频率在 1.2 kHz～1.7 kHz 之间所有小区的方差 σ_1^2 和权重因子 w_1 与迭代次数的关系曲线, 其他参数收敛情况是类似的. 方差的初始值都取为 10^4, 权重因子的初始值都取为 $1/\mathrm{C}_N^M$. 仿真中发现, 方差初始值要取得很大, 这样可以非常有效地避免本文算法收敛到局部极值点. 从图 3 可以看出, 绝大多数区域的参数都在 10 步以内就收敛了, 表明本文算法具有很快的收敛速度. 有些小区的参数虽然在 10 步内还没有完全收敛, 但是对分离精度已经没有太大的影响了.

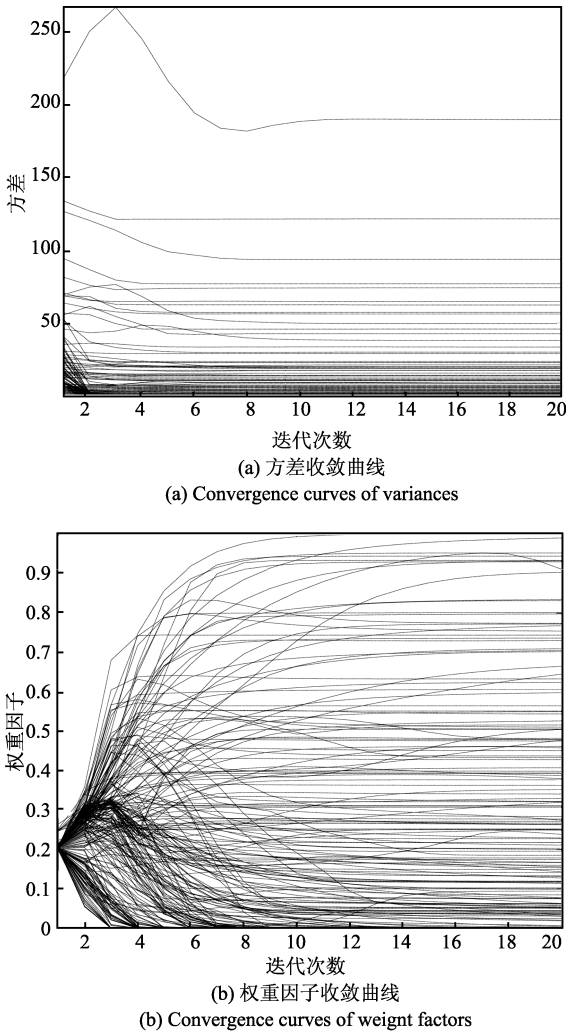


图 3　参数收敛曲线

Fig. 3　Convergence curves of parameters

图 4 是本文算法的迭代次数和 SIR 之间的关系曲线. 可以看出, 迭代次数比较少时, SIR 比较低, 这是因为算法尚未完全收敛. 不论是 4 个源还是 5 个源, 随着迭代过程的进行, 算法都迅速收敛, 到了第 7～8 步以后, 算法都趋于收敛.

4.3　各算法性能比较

图 5 是各算法的 SIR 和信噪比 (Signal to noise

ratio, SNR) 之间的关系曲线图. 这里考虑了噪声对各算法性能的影响. 本文算法的迭代次数在 4 个源时取为 8 步, 在 5 个源时取为 10 步. GGD 大多数情况下在 4 步内收敛, GGD 的仿真程序由 http://mmp.kaist.ac.kr/xe/?mid=software 提供, 网站上提供了 GGD 在 3 个观测信号和 4 个源信号时的源程序. 从图 5 可以看出, 随着信噪比的降低, 所有算法的分离性能都会变差. 这里有两个原因: 1) 信噪比降低, 混合矩阵的估计精度会下降; 2) 源恢复算法的性能也会受信噪比的影响. 当 SNR 较高时, 各算法中 MDLD 的 SIR 是最低的, 这是因为 MDLD 是一种时频掩蔽的方法, 只要在某个时频点的源信号相互重叠, MDLD 就无法分离源信号; IMSUB 比本文算法和 CMD 算法的分离效果差的原因, 是因为 IMSUB 要求主导源信号个数严格小于观测信号个数, 而本文算法和 CMD 算法可以在主导源信号个数等于观测信号个数的情况下分离源信号. 本文算法比 CMD 算法更具优势, 其原因在第 4.1 节已做了说明. 不过在各算法中, MDLD 的 SIR 随噪声增大而下降得最慢, 说明抗噪声能力是各算法中最好的, 这与 MDLD 算法是一种时频掩蔽方法有关.

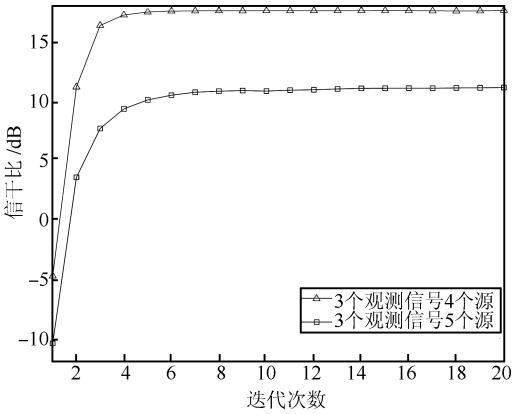


图 4 SIR 和迭代次数之间的关系曲线

Fig. 4 Relationship between the number of iterations and SIR

图 6 给出了 SNR = 35 dB, 观测信号数目为 2 时, 各算法的 SIR 与源数之间的关系曲线. 在这种情况下, 只要有 2 个及以上的源在某个时频点上重叠, IMSUB 就无法将其分离, 此时, 算法退化为二元时频掩蔽算法. 从图 6 可以看出, 当源数增加时, 各算法性能下降, 但本文算法仍然是优于其他算法的.

4.4 算法计算复杂度比较

本文算法每次迭代要对式 (6)、式 (11) 和式 (14) 进行计算, 当观测信号和源信号的个数分别为 M 和 N 时, 这 3 个式子的计算量都与 C_N^M 成正比. 因此, 本文的计算复杂度为 $O(C_N^M)$, 这显然比 GMM 和 GSMM 的计算复杂度小很多. 如果 GMM 和 GSMM 有 L 种状态, 其计算复杂度为 $O(L^N)$. 这显然是个指数关系. 而且 L 不能取的太小, 通常要在 8 以上. 即使只有 4 个源信号, 其计算量也是非常大的. 事实上, 本文算法比 GGD 算法也要简单得多. 虽然 GGD 迭代次数比本文算法迭代次数要少一些, 但是每一步都需要很大的计算量. 因为, GGD 每一次迭代都要进行一次随机抽样, 而且每一步要对函数进行优化.

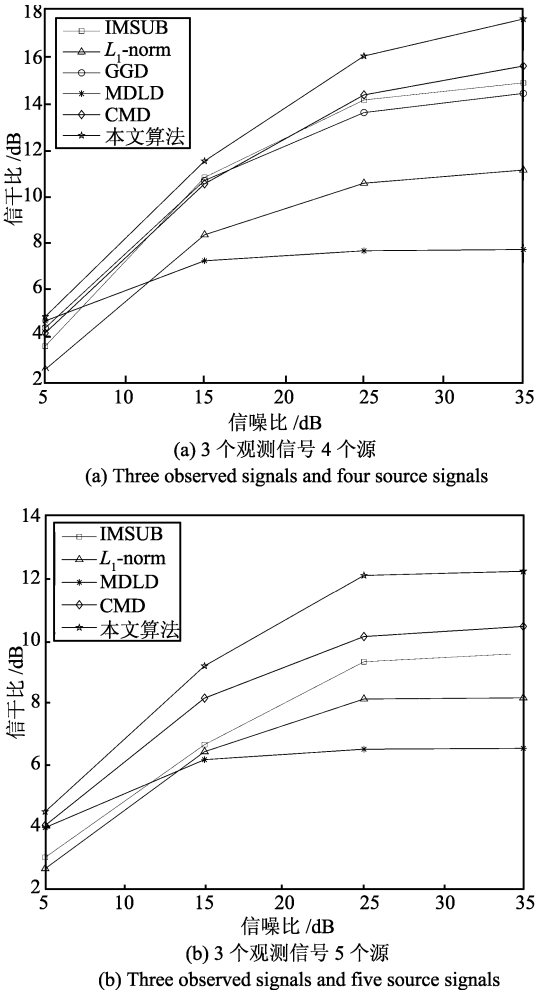


图 5 不同 SNR 时的各算法性能比较

Fig. 5 Comparison of performance according to SNR

当观测信号个数为 3, 源个数为 4 时, 对 2 s 语音信号 (32 000 个采样点) 进行处理时, 各算法运行一次所需的时间如表 1 所示. 所有算法都是在处理器为 Pentium (R) Dual-core T4500, 主频为 2.3 GHz, 内存为 2 G, 操作系统为 Windows XP 的平台下运行的, Matlab 的版本为 R2012a. 除 MDLD 外, 这里的运行时间不包括估计混合矩阵的时间. 本文算法和 GGD 算法的运行时间分别是迭代 8 步和 4 步

时所需的时间. 从表中可以看出, 本文算法在具有最好分离性能的同时, 仍然具有比较小的计算量.

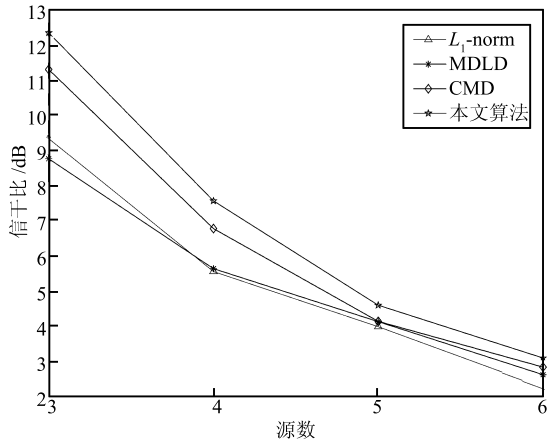


图 6　各算法的 SIR 与源数之间的关系曲线图

Fig. 6　Performance comparison according to the number of sources

表 1　各算法运行时间比较

Table 1　Running time comparison

算法	运行时间 (s)
IMSUB	7.3
CMD	2.36
GGD	343.24
MDLD	11.15
L_1-norm	42.77
本文算法	15.73

5　结论

对于语音等弱稀疏信号, 每个时频点主导源的个数一般不会超过观测信号的个数, 而且这类信号虽然是非平稳信号, 却具有局部平稳特性, 在局部区域服从复高斯分布. 本文将信号的稀疏性和源信号的局部复高斯分布结合起来, 建立了一种新的混合模型来描述观测信号在局部区域的分布, 提出了一种新的欠定盲分离方法. 通过 EM 算法估计出混合模型的参数, 并根据每个时频点所处状态的后验分布估计出每个时频点的状态, 并以此来获得源信号的估计. 本文方法并不假定局部区域内所有时频点所处状态是完全相同的, 与实际情况更加相符, 因此, 可以更好地恢复源信号.

References

1 Chien J T, Hsieh H L. Convex divergence ICA for blind source separation. *IEEE Transactions on Audio, Speech, and Language Processing*, 2012, **20**(1): 302−313

2 Kang Chun-Yu, Zhang Xin-Hua, Han Dong. DOA estimation and signal recovery combined blind source separation with high resolution. *Acta Automatica Sinica*, 2010, **36**(3): 442−445
(康春玉, 章新华, 韩东. 盲源分离与高分辨融合的 DOA 估计与信号恢复方法. 自动化学报, 2010, **36**(3): 442−445)

3 Yilmz O, Rickard S. Blind separation of speech mixtures via time-frequency masking. *IEEE Transactions on Signal Processing*, 2004, **52**(7): 1830−1847

4 Arakia S, Sawada H, Mukai R, Makino S. Underdetermined blind sparse source separation for arbitrarily arranged multiple sensors. *Signal Processing*, 2007, **87**(8): 1833−1847

5 Zibulevsky M, Pearlmutter B A. Blind source separation by sparse decomposition in a signal dictionary. *Neural Computation*, 2001, **13**(4): 863−882

6 O′Grady P D, Pearlmutter B A. The LOST algorithm: finding lines and separating speech mixtures. *EURASIP Journal on Advances in Signal Processing*, 2008, **784296**: 1−17

7 Cobos M, Lopez J J. Maximum a posteriori binary mask estimation for underdetermined source separation using smoothed posteriors. *IEEE Transactions on Audio, Speech, and Language Processing*, 2012, **20**(7): 2059−2064

8 Mitianoudis N. A generalized directional Laplacian distribution: estimation, mixture models and audio source separation. *IEEE Transactions on Audio, Speech, and Language Processing*, 2012, **20**(9): 2397−2408

9 Fevotte C, Cardoso J. Maximum likelihood approach for blind audio source separation using time-frequency Gaussian source models. In: Proceedings of IEEE Workshop on Applications of Signal Processing to Audio and Acoustics. New Paltz, New York, USA: IEEE, 2005. 78−81

10 Ozerov A, Philippe P, Bimbot F, Gribonval R. Adaptation of bayesian models for single-channel source separation and its application to voice/music separation in popular songs. *IEEE Transactions on Audio, Speech and Language Processing*, 2007, **15**(5): 1564−1578

11 Benaroya L, Bimbot F, Gribonval R. Audio source separation with a single sensor. *IEEE Transactions on Audio, Speech, and Language Processing*, 2006, **14**(1): 191−199

12 Kim S G, Yoo C D. Underdetermined blind source separation based on subspace representation. *IEEE Transactions on Signal Processing*, 2009, **57**(7): 2604−2614

13 Aissa-El-Bey A, Linh-Trung N, Abed-Meraimet K, Belouchrani A, Grenier Y. Underdetermined blind separation of nondisjoint sources in the time-frequency domain. *IEEE Transactions on Signal Processing*, 2007, **55**(3): 897−907

14 Lu Feng-Bo, Huang Zhi-Tao, Jiang Wen-Li. Underdetermined blind separation of time-delayed non-stationary signal based on single source region in the time-frequency domain. *Acta Electronica Sinica*, 2011, **39**(4): 854−858
(陆凤波, 黄知涛, 姜文利. 基于时频域单源区域的延迟欠定混合非平稳信号盲分离. 电子学报, 2011, **39**(4): 854−858)

15 Peng D Z, Xiang Y. Underdetermined blind source separation based on relaxed sparsity condition of sources. *IEEE Transactions on Signal Processing*, 2009, **57**(2): 809−814

16 Xie S, Yang L, Yang J M, Zhou G, Xiang Y. Time-frequency approach to underdetermined blind source separation. *IEEE Transactions on Neural Networks and Learning System*, 2012, **23**(2): 306−316

17 Lu F B, Huang Z T, Jiang W L. Underdetermined blind separation of non-disjoint signals in time-frequency domain based on matrix diagonalization. *Signal Processing*, 2011, **91**(7): 1568−1577

18 Reju V G, Koh S N, Soon I Y. An algorithm for mixing matrix estimation in instantaneous blind source separation. *Signal Processing*, 2009, **89**(9): 1762−1773

19 Zhou G X, Yang Z Y, Xie S L, Yang J M. Mixing matrix estimation from sparse mixtures with unknown number of sources. *IEEE Transactions on Neural Networks*, 2011, **22**(2): 211−221

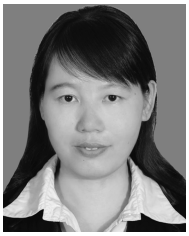
20 Xiao Ming, Xie Sheng-Li, Fu Yu-Li. Underdetermined blind source separation algorithm based on normal vector of hyperplane. *Acta Automatica Sinica*, 2008, **34**(2): 142−149
(肖明, 谢胜利, 傅予力. 基于超平面法矢量的欠定盲信号分离算法. 自动化学报, 2008, **34**(2): 142−149)

21 Chen Yong-Qiang, Wang Hong-Xia. A robust method for mixing matrix estimation in blind source separation. *Journal of Electronics & Information Technology*, 2012, **34**(9): 2039−2044
(陈永强, 王宏霞. 一种强鲁棒性的盲分离混合矩阵估计方法. 电子与信息学报, 2012, **34**(9): 2039−2044)

22 Van Den Bos A. The multivariate complex normal distribution-a generalization. *IEEE Transactions on Information Theory*, 1995, **41**(2): 537−539

23 McLachlan G J, Krishnan T. *The EM Algorithm and Extensions*. New York, USA: Wiley, 1997. 18−27

陈永强 西南交通大学信息科学与技术学院博士研究生. 成都信息工程学院电子实验中心副教授. 主要研究方向为盲源分离, 语音和音频信号处理. 本文通信作者. E-mail: chenyongq@cuit.edu.cn
(**CHEN Yong-Qiang** Ph. D. candidate at the School of Information Science and Technology, Southwest Jiaotong University, and also associate professor at the Electronic Experiment Center, Chengdu University of Information Technology. His research interest covers blind source separation, speech and audio processing. Corresponding author of this paper.)



王宏霞 西南交通大学信息科学与技术学院教授. 主要研究方向为多媒体信息安全, 数字水印与取证, 语音和音频信号处理. E-mail: hxwang@swjtu.edu.cn.
(**WANG Hong-Xia** Professor at the School of Information Science and Technology, Southwest Jiaotong University. Her research interest covers multimedia information security, digital watermarking and forensics, and speech and audio processing.)