一种新的基于稀疏表示的单通道盲源分离算法

田元荣* 王 星 周一鹏
(空军工程大学航空航天工程学院 西安 710038)

摘 要：该文针对稀疏表示应用于单通道盲源分离中存在字典间互干扰的问题，通过在常规联合字典中引入一个新的子字典——“共同子字典”，提出一种新的基于稀疏表示的单通道盲源分离算法。新的字典学习目标函数中单个源的保真度由对应子字典和共同子字典构成，共同子字典的存在可以有效避免某一源信号在其他子字典上寻求成份而带来的互干扰问题。目标函数的求解通过交替执行稀疏表示、字典更新和比例系数优化 3 个步骤来实现。在测试阶段，通过收集单个源所对应子字典和共同子字典上的分量可以估计出混合信号中的单个源信号，从而达到盲源分离的目的。在语音数据库上进行的对比实验发现，所提算法较传统算法和前沿算法在两个通用评价指标上最高有近 1 dB 的提高。

关键词：稀疏表示；单通道盲源分离；字典学习；鉴别力；保真度
中图分类号： TP391 **文献标识码**： A **文章编号**：1009-5896(2017)06-1371-08
DOI: 10.11999/JEIT160888

Novel Single Channel Blind Source Separation Algorithm Based on Sparse Representation

TIAN Yuanrong WANG Xing ZHOU Yipeng
(*Institute of Aeronautics and Astronautics Engineering, Air Force Engineering University, Xi'an* 710038, *China*)

Abstract: The main drawback of sparse representation based Single Channel Blind Source Separation (SCBSS) is the interference between sub-dictionaries. To alleviate this drawback, an extra sub-dictionary, named common sub-dictionary, is proposed to add into traditional union dictionary. The single source is reconstructed by linear combining sparsely activity atoms of its corresponding sub-dictionary and common sub-dictionary. The common sub-dictionary can pure discriminative information in each source's specified sub-dictionary since the common information different sources shared together is gathered in common sub-dictionary. The optimization of objective function involves three steps: sparse representation, dictionary updating and weight coefficients optimization, the three steps are iteratively performed for a specified number of times or until convergence. In test stage, single source separation is achieved by combining atoms in source corresponding sub-dictionary and common sub-dictionary with the sparse coefficients of single mixed signal over union dictionary. Experimental results on speech dataset show that, when compared with traditional and state of art algorithms, the proposed algorithm can improve the performance 1 dB at most.

Key words: Sparse representation; Single channel blind source separation; Dictionary learning; Discrimination; Fidelity

1 引言

单通道盲源分离(Single Channel Blind Source Separation, SCBSS)是指仅通过一个观测到的混合信号恢复出各个源信号的过程[1]。由于问题存在的普遍性和挑战性，单通道盲源分离成为理论界和工程应用界的研究热点，很多极富创新的成果不断涌现[2–7]，但寻求一种高精度、高鲁棒、高效率的方法依然是解决单通道源分离问题的迫切需求。稀疏表示应用于单通道盲源分离的关键在于联合字典的构建，即混合信号表达的大空间由多个子空间构成，每个子空间能够最大可能地表达某一个源的信号，那么通过稀疏系数和所对应子空间的基就可以重构出单个源的信号，实现信号分离[8–10]。通常而言联合字典都为冗余字典，即使施加稀疏性约束，也难以保证某个源的信号在其他源对应的子字典上不产

收稿日期：2016-09-02；改回日期：2017-01-22；网络出版：2017-03-21
*通信作者：田元荣 yrtian_mail@126.com
基金项目：国家自然科学基金(61372167)，航空科学基金(20152096019)
Foundation Items: The National Natural Science Foundation of China (61372167), The Aviation Science Foundation of China (20152096019)

生响应，因此近年来基于稀疏表示的单通道盲源分离集中于学习一个优良的字典来避免上述情况的发生。Grais 等人[11]通过正交化手段惩罚各子字典间的重叠部分实现联合字典具有鉴别力的目的，但其训练和测试的过程是分离的，所以其算法效果更多地依赖于各个源数据本身的非相关性；Weninger 等人[12]提出了将训练和测试统一在一个框架下构建目标函数，但是其结构复杂；Bao 等人[13]将稀疏表示联合字典学习策略用于单通道盲源分离，但 Bao 等人的工作仍然没有彻底摆脱人脸识别框架下的字典学习，训练的时候混合数据的各分量独立表达，然后将表达结果统一结合进一个优化函数中进行字典更新，从机理上并不是信号分离任务驱动的字典学习；Wang 等人[14]通过构建众多的信号分离任务进行联合字典训练，属于典型的基于信号分离任务驱动的字典学习算法，其结果比经典的稀疏非负矩阵分解方法有一定的提高。

事实上，任何两个源的信号总会含有各自独特的成份，同时也会含有共同的成份。如果只强调鉴别力，即期望只用区别于其他源的成份来近似重构源信号，保真度就会比较差；相反，在一定的保真度下，如果某个源的信号在自身子字典上无法满足保真度，就难免会在其他源对应的子字典上产生响应，从而引起干扰。为了平衡这一矛盾，本文在各个源子字典构成的联合字典中加入一个公共子字典，用来表达各个源之间的共同成份，在训练的过程中，除了更新字典原子外，增加学习每个源信号的共同成份所占的比例，这样将混合信号在训练好的联合字典上稀疏表示后，就可以用子字典上的响应加共同子字典上一定比例的响应来恢复各个源信号。

2 稀疏表示模型应用于单通道盲源分离问题的描述

设 N 个信号源对应的子字典为 $\boldsymbol{D}_1,\boldsymbol{D}_2,\cdots,\boldsymbol{D}_i,\cdots,\boldsymbol{D}_N$，其中 $\boldsymbol{D}_i\in \mathrm{R}^{d\times n_i}$，$N$ 个子字典次序排列起来形成联合字典 $\boldsymbol{D}=[\boldsymbol{D}_1,\boldsymbol{D}_2,\cdots,\boldsymbol{D}_i,\cdots,\boldsymbol{D}_N]$。记混合信号 $\boldsymbol{y}=\boldsymbol{x}^{(1)}+\cdots+\boldsymbol{x}^{(i)}+\cdots+\boldsymbol{x}^{(N)}$，那么 $\boldsymbol{y}$ 在联合字典 $\boldsymbol{D}$ 上的表示系数就可以写为 $\boldsymbol{\alpha}=\left[\boldsymbol{\alpha}^{(1)},\boldsymbol{\alpha}^{(2)},\cdots,\boldsymbol{\alpha}^{(i)},\cdots,\boldsymbol{\alpha}^{(N)}\right]^{\mathrm{T}}$，其中 $\boldsymbol{\alpha}^{(i)}=\left[\alpha_1^{(i)},\alpha_2^{(i)},\cdots,\alpha_{n_i}^{(i)}\right]$ 表示混合信号在子字典 $\boldsymbol{D}_i$ 上的系数。在稀疏约束下，$\boldsymbol{y}$ 在联合字典 $\boldsymbol{D}$ 上总是寻求最有效的表达，因此 $\hat{\boldsymbol{x}}_i=\boldsymbol{D}_i{\boldsymbol{\alpha}^{(i)}}^{\mathrm{T}}$ 以极大的概率是 $\boldsymbol{x}^{(i)}$ 的估计，从而实现混合信号的分离，文献[8–10]中的结果印证了该原理的正确性与优势。

为了论述的方便且不失一般性，本文只针对两个源 $(N=2)$ 混合的情况进行讨论。

3 具有鉴别力和保真度的字典学习

3.1 目标函数的构建

记 $\boldsymbol{D}=[\boldsymbol{D}_1,\boldsymbol{D}_2,\boldsymbol{D}_c]$ 为联合字典，$\boldsymbol{D}_s(s=1,2)$ 为第 s 个源对应的子字典，$\boldsymbol{D}_c$ 为共同子字典，为了简化叙述，假设子字典 $\boldsymbol{D}_s$ 和 $\boldsymbol{D}_c$ 含有相同的原子个数 l；$\aleph=\{\boldsymbol{X}_1,\boldsymbol{X}_2\}$ 为两个源训练数据的集合，其中 $\boldsymbol{X}_s=\left[\boldsymbol{x}_s^{(1)},\boldsymbol{x}_s^{(2)},\cdots,\boldsymbol{x}_s^{(n_s)}\right]$；记 $\boldsymbol{z}_i=\boldsymbol{x}_1^{(i_1)}+\boldsymbol{x}_2^{(i_2)}$ 为某个混合信号，则总共可以构建 $m=n_1\cdot n_2$ 个混信号进行字典学习。字典学习的目标函数如式(1)所示。

$$J(\boldsymbol{D},\boldsymbol{\alpha})=\sum_{i=1}^{m}g_i(\boldsymbol{D},\boldsymbol{\alpha}) \tag{1}$$

其中，

$$\begin{aligned}g_i(\boldsymbol{D},\boldsymbol{\alpha})=\min_{\boldsymbol{D},\boldsymbol{\alpha}}\Bigg(&\left\|\boldsymbol{z}_i-\boldsymbol{D}\boldsymbol{C}_i^{\mathrm{T}}\right\|_F^2\\&+\frac{\gamma}{2}\Bigg(\left\|\boldsymbol{x}_1^{i_1}-\boldsymbol{D}_1\left(\boldsymbol{C}_i^{(1)}\right)^{\mathrm{T}}-\alpha_1\boldsymbol{D}_c\left(\boldsymbol{C}_i^{(c)}\right)^{\mathrm{T}}\right\|_F^2\\&+\left\|\boldsymbol{x}_2^{i_2}-\boldsymbol{D}_2\left(\boldsymbol{C}_i^{(2)}\right)^{\mathrm{T}}-\alpha_2\boldsymbol{D}_c\left(\boldsymbol{C}_i^{(c)}\right)^{\mathrm{T}}\right\|_F^2\Bigg)\Bigg),\\&\text{s.t. }\left\|\boldsymbol{\varphi}_j\right\|_2=1,\quad \alpha_1+\alpha_2=1,\quad \alpha_1>0,\quad \alpha_2>0\end{aligned} \tag{2}$$

式中，$\|\bullet\|_F^2$ 代表矩阵的 Frobenius 范数，$\boldsymbol{\varphi}_j$ 是 $\boldsymbol{D}$ 的任意一列，$\boldsymbol{C}_i=\left[\boldsymbol{C}_i^{(1)},\boldsymbol{C}_i^{(2)},\boldsymbol{C}_i^{(c)}\right]$ 是混合信号在字典 $\boldsymbol{D}$ 上的稀疏表示系数，α_s 是权重系数，$\boldsymbol{\alpha}=[\alpha_1,\alpha_2]^{\mathrm{T}}$。式(2)通过图 1 进行解释，为了便于在平面上显示图 1 中假定数据的维度 $d=2$。

式(2)等号右边的第 1 项为混合信号的总体保真度误差项，对应图 1 中的 $\boldsymbol{z}$，$\hat{\boldsymbol{x}}_1$，$\hat{\boldsymbol{x}}_2$ 分别是 $\boldsymbol{x}_1$，$\boldsymbol{x}_2$ 的估计用虚线表示。从图 1 可以看出，虽然图 1(a)，图 1(b)和图 1(e)都可以作为式(2)的解，但是最优解当属图 1(e)，因为在图 1(e)中，不仅总体保真度误差很小，而且各个源信号与其估计之间的误差也很小。为了使式(2)总能取得如图 1(e)所示的解，需要约束各个源信号和其估计之间的误差，如式(2)右边第 2 项所示，称之为鉴别力误差项。另外值得注意的是，式(2)与文献[13,14]等人不同的是，式(2)中的联合字典包含 $\boldsymbol{D}_c$，其优势可以通过图 1(c)和图 1(d)反映出来。图 1(c)中我们假定 $\boldsymbol{x}_1$ 和 $\boldsymbol{x}_2$ 之间不同的成份是互相垂直的 $\boldsymbol{x}_1^{(d)}$ 和 $\boldsymbol{x}_2^{(d)}$，相同的成份是同方向的 $\boldsymbol{x}_1^{(c)}$ 和 $\boldsymbol{x}_2^{(c)}$，很显然 $\boldsymbol{x}_1^{(d)}$ 和 $\boldsymbol{x}_1^{(c)}$ 共同才能完整地构成 $\boldsymbol{x}_1$。若只用各个源之间不同的成份 $\boldsymbol{x}_1^{(d)}$ 和 $\boldsymbol{x}_2^{(d)}$ 去恢复各个源信号如图 1(d)所示，误差将会非常大。

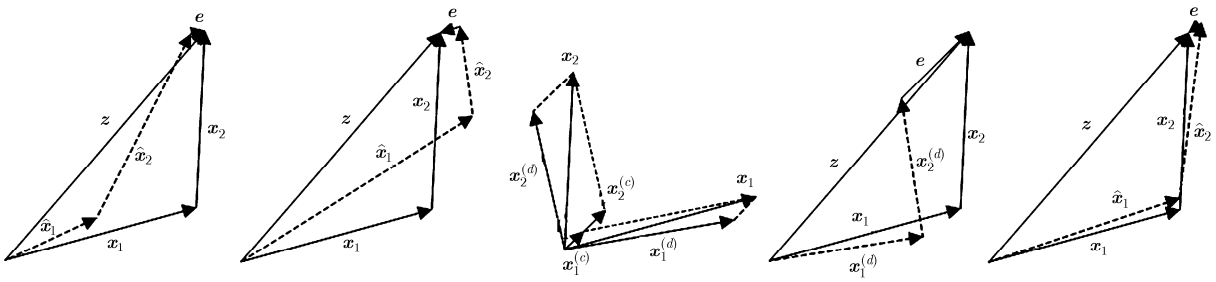


图 1 两个 2 维源信号混合及分解示意图

3.2 目标函数的优化

优化方程式(2)可以划分为两个子问题：固定 $\boldsymbol{\alpha}$ 更新 $\boldsymbol{D}$ 、固定 $\boldsymbol{D}$ 求 $\boldsymbol{\alpha}$ 。在优化 $\boldsymbol{D}$ 和 $\boldsymbol{\alpha}$ 之前，需要先求得稀疏系数 $\boldsymbol{C}$ 。

(1)固定 $\boldsymbol{D}$ 求 $\boldsymbol{C}$ ：　之所以在求 $\boldsymbol{C}$ 时没有固定 $\boldsymbol{\alpha}$ ，是因为虽然式(2)右边两项都含有 $\boldsymbol{C}$ ，但是第 2 项中 $\boldsymbol{C}_i^{(1)}$ ， $\boldsymbol{C}_i^{(2)}$ 和 $\boldsymbol{C}_i^{(c)}$ 是第 1 项求出的 $\boldsymbol{C}_i$ 的部分。所以 $\boldsymbol{C}$ 可以通过求解式(3)来实现。

$$\min_{\boldsymbol{C}_i}\left\|\boldsymbol{z}_i-\boldsymbol{D}\boldsymbol{C}_i^{\mathrm{T}}\right\|_F^2,\quad \text{s.t.}\quad \left\|\boldsymbol{C}_i\right\|_0\le K \tag{3}$$

K 是 $\boldsymbol{z}_i$ 在 $\boldsymbol{D}$ 上表达后稀疏系数的个数。式(3)是一个标准的稀疏表达问题，可以通过 MP, OMP 和 BP 算法来实现，本文选用 OMP[15]来分解 $\boldsymbol{z}_i$ 。

(2)固定 $\boldsymbol{\alpha}$ 更新 $\boldsymbol{D}$ ：　在求得 $\boldsymbol{C}$ 并固定 $\boldsymbol{\alpha}$ 的情况下，更新 $\boldsymbol{D}$ 时式(2)便退化为

$$\begin{aligned}&\min_{\boldsymbol{D}}\left\|\boldsymbol{z}_i-\boldsymbol{D}\boldsymbol{C}_i^{\mathrm{T}}\right\|_F^2\\&\quad+\frac{\gamma}{2}\left(\left\|\boldsymbol{x}_1^{i_1}-\boldsymbol{D}_1\left(\boldsymbol{C}_i^{(1)}\right)^{\mathrm{T}}-\alpha_1\boldsymbol{D}_c\left(\boldsymbol{C}_i^{(c)}\right)^{\mathrm{T}}\right\|_F^2\right.\\&\quad\left.+\left\|\boldsymbol{x}_2^{i_2}-\boldsymbol{D}_2\left(\boldsymbol{C}_i^{(2)}\right)^{\mathrm{T}}-\alpha_2\boldsymbol{D}_c\left(\boldsymbol{C}_i^{(c)}\right)^{\mathrm{T}}\right\|_F^2\right),\\&\text{s.t.}\quad\left\|\boldsymbol{\varphi}_j\right\|_2=1\end{aligned} \tag{4}$$

引入指示矩阵 $\boldsymbol{P}_1$, $\boldsymbol{P}_2$ 和 $\boldsymbol{P}_c$ 如式(5)所示。

$$\boldsymbol{P}_1=\begin{bmatrix}\mathbf{1}&\mathbf{0}&\mathbf{0}\\\mathbf{0}&\mathbf{0}&\mathbf{0}\\\mathbf{0}&\mathbf{0}&\mathbf{0}\end{bmatrix},\ \boldsymbol{P}_2=\begin{bmatrix}\mathbf{0}&\mathbf{0}&\mathbf{0}\\\mathbf{0}&\mathbf{1}&\mathbf{0}\\\mathbf{0}&\mathbf{0}&\mathbf{0}\end{bmatrix},\ \boldsymbol{P}_c=\begin{bmatrix}\mathbf{0}&\mathbf{0}&\mathbf{0}\\\mathbf{0}&\mathbf{0}&\mathbf{0}\\\mathbf{0}&\mathbf{0}&\mathbf{1}\end{bmatrix} \tag{5}$$

式中，$\mathbf{1}$ 代表 $l\times l$ 的单位矩阵，$\mathbf{0}$ 代表 $l\times l$ 的零矩阵。因此，式(4)可以进一步改写为

$$\begin{aligned}&\min_{\boldsymbol{D}}\left\|\left[\boldsymbol{z}_i,\frac{\gamma^2}{4}\boldsymbol{x}_1^{i_1},\frac{\gamma^2}{4}\boldsymbol{x}_2^{i_2}\right]\right.\\&\quad-\boldsymbol{D}\left[\boldsymbol{C}_i^{\mathrm{T}},\frac{\gamma^2}{4}\left(\boldsymbol{P}_1\boldsymbol{C}_i^{\mathrm{T}}+\alpha_1\boldsymbol{P}_c\boldsymbol{C}_i^{\mathrm{T}}\right),\right.\\&\quad\left.\left.\frac{\gamma^2}{4}\left(\boldsymbol{P}_2\boldsymbol{C}_i^{\mathrm{T}}+\alpha_2\boldsymbol{P}_c\boldsymbol{C}_i^{\mathrm{T}}\right)\right]\right\|_F^2,\ \text{s.t.}\ \left\|\boldsymbol{\varphi}_j\right\|_2=1\end{aligned} \tag{6}$$

式中，除了 $\boldsymbol{D}$ 未知外，其他的参数均已知，是 K-SVD[16]解决的标准问题，可以采用 K-SVD 来更新字典的原子。

(3)固定 $\boldsymbol{D}$ 求 $\boldsymbol{\alpha}$ ：　已知 $\boldsymbol{C}$ 并固定 $\boldsymbol{D}$ 的情况下，借助指示矩阵式(5)，优化 $\boldsymbol{\alpha}$ 是一个典型的二次规划问题，如式(7)所示。

$$\left.\begin{aligned}&\min_{\boldsymbol{\alpha}}\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{H}\boldsymbol{\alpha}+\boldsymbol{c}^{\mathrm{T}}\boldsymbol{\alpha},\quad\text{s.t.}\ \alpha_1+\alpha_2=1,\\&\qquad\alpha_1>0,\ \ \alpha_2>0\\&\boldsymbol{H}=\begin{bmatrix}\left\|\boldsymbol{D}_c\left(\boldsymbol{C}_i^{(c)}\right)^{\mathrm{T}}\right\|_F^2&0\\0&\left\|\boldsymbol{D}_c\left(\boldsymbol{C}_i^{(c)}\right)^{\mathrm{T}}\right\|_F^2\end{bmatrix}\\&\boldsymbol{c}=\begin{bmatrix}\sum\sum\left(\boldsymbol{x}_1^{i_1}-\boldsymbol{D}_1\left(\boldsymbol{C}_i^{(1)}\right)^{\mathrm{T}}\right)\odot\left(\boldsymbol{D}_c\left(\boldsymbol{C}_i^{(c)}\right)^{\mathrm{T}}\right)\\\sum\sum\left(\boldsymbol{x}_2^{i_2}-\boldsymbol{D}_1\left(\boldsymbol{C}_i^{(1)}\right)^{\mathrm{T}}\right)\odot\left(\boldsymbol{D}_c\left(\boldsymbol{C}_i^{(c)}\right)^{\mathrm{T}}\right)\end{bmatrix}\end{aligned}\right\} \tag{7}$$

式中，$\odot$ 代表哈达马乘积，本文采用内点法来求解式(7)。

式(2)的求解就是通过交替执行以上 3 个步骤来达到最优的 $\boldsymbol{D}$ 和 $\boldsymbol{\alpha}$ 。

3.3 源信号的估计

设某个待分离的信号为 $\boldsymbol{z}=\sum_{i\in\Gamma}\boldsymbol{x}_i$, Γ 为输入源序号集合 Π 的某个非空子集，可知 Π 共有 $\zeta=\sum_{c=1}^{N}\boldsymbol{C}_N^c$ 个非空子集(记为 $\{\Lambda_1,\Lambda_2,\cdots,\Lambda_\zeta\}$)。在本文考虑的 $N=2$ 的情况中， Λ_1={1} , Λ_2={2} , Λ_3 = {1, 2}，对应的混合信号分别为 $\boldsymbol{z}=\boldsymbol{x}_1$ ， $\boldsymbol{z}=\boldsymbol{x}_2$ 和 $\boldsymbol{z}=\boldsymbol{x}_1+\boldsymbol{x}_2$ 。测试阶段的盲源分离算法步骤为

(1)稀疏表示混合信号 $\boldsymbol{z}$ ，如式(8)所示。

$$\min_{\boldsymbol{C}}\left\|\boldsymbol{z}-\boldsymbol{D}\boldsymbol{C}\right\|_2^2,\quad\text{s.t.}\ \left\|\boldsymbol{C}\right\|_0\le K \tag{8}$$

与 3.1 节的叙述一致， $\boldsymbol{C}=\left[\boldsymbol{C}_1,\boldsymbol{C}_2,\boldsymbol{C}_c\right]$ 。

(2)根据稀疏系数和对应的原子估计各个源的信号为

$$\left.\begin{aligned}\hat{\boldsymbol{x}}_1&=\boldsymbol{D}_1\boldsymbol{C}_1^{\mathrm{T}}+\alpha_1\boldsymbol{D}_c\boldsymbol{C}_c^{\mathrm{T}}\\\hat{\boldsymbol{x}}_2&=\boldsymbol{D}_2\boldsymbol{C}_2^{\mathrm{T}}+\alpha_2\boldsymbol{D}_c\boldsymbol{C}_c^{\mathrm{T}}\end{aligned}\right\} \tag{9}$$

(3)根据式(10)确定输入信号的源的编号集合。

$$\Gamma^* = \widehat{\Lambda}_1 \bigcap \widehat{\Lambda}_2 \cdots \bigcap \widehat{\Lambda}_\hbar \tag{10}$$

文献[15]证明，输入信号经过 T 次分解后误差小于 $\left[(1-1/K)(1+\mu)^{\mathrm{T}}\right]\|\boldsymbol{z}\|_2^2$，其中 $\mu = \sup_{i,j,\ i\neq j}\left|\left\langle \boldsymbol{\varphi}_i, \boldsymbol{\varphi}_j \right\rangle\right|$ 为字典 $\boldsymbol{D}$ 的互相关系数[15]。所以经过式(8)分解后，式(10)中

$$\left\{\widehat{\Lambda}_1, \widehat{\Lambda}_2, \cdots, \widehat{\Lambda}_\hbar\right\} = \arg\min_{\Lambda_j \in \Lambda}\left(\left\|\sum_{i\in\Lambda_j} \boldsymbol{x}_i - \boldsymbol{z}\right\| \le \left[(1-1/K)(1+\mu)^K\right]\|\boldsymbol{z}\|_2^2\right)$$

最后输出式(9)中估计出的各输入源的信号 $\hat{\boldsymbol{x}}_i$，其中 $i \in \Gamma^*$。

4 实验分析

为了验证本文算法的有效性和先进性，在 Grid corpus 语音数据库[17]上进行了一系列实验，Grid corpus 语音数据库共收集了 34 名发音者每人所说出的 1000 个短句。本文的实验中，随机选取了 6 名发音者(3 女、3 男)每人 500 个短句进行实验，500 个短句中 350 句用于训练字典 $\boldsymbol{D}$ 和系数 $\boldsymbol{\alpha}$，150 句用于测试算法的性能。本文选用梅尔谱来表达语音信号，根据梅尔谱的提取算法，每个短句平均可以提取 100 个 80 维的梅尔谱向量，训练样本需要不同源的数据两两混合，所以每个源 350 个短句足以保证训练数据的充分性，同时 150 个短句也可以充分满足测试的多样性。

信号分离优劣评价选用文献[18]中的 SIR，以及文献[14]中的 SNR 指标实验分析中的数据指标均为多次实验的平均结果。值得说明的是，本文选择在梅尔谱域评价算法主要是因为梅尔谱是一种强有力的语音信号特征，语音理解和语音识别等问题都可以很好地通过梅尔谱来实现[19,20]。对比算法选用文献[8]中提出的稀疏非负矩阵分解算法(SNMF)，和文献[14]中提出的鉴别力字典学习算法(DDL)。

4.1 有效性验证

为了验证本文提出的字典学习算法的有效性，本小节给出一个基于本文算法的单通道盲源分离的例子。选用的测试短句为‘id2’发音者的‘sgai8a.wav’和‘id31’的‘bgid7s.wav’，值得说明的是，以上这组数据无论是发音者还是具体的短句完全是随机选出的。字典训练中选用的参数为 $l=100$，$K=30$ 和 $\gamma=0.85$。由于从单一观测到的混合信号中，各个源信号的相位信息是未知的，我们采用随机的相位信息来将梅尔谱域的信号返回到时间域。结果如图 2 所示。

从图 2(a)和图 2(d)以及图 2(b)和图 2(e)对比可以看出，本文提出的字典学习算法分离的信号与源信号从总体轮廓来看几乎一致，说明本文算法可以有效地实现单通道盲源分离，但也存在着一定的误差，误差主要体现在信号的细节部分主要原因有以下 3 个：(1)稀疏表示原理是提取输入信号的总体轮廓信息，所以细节信息有所忽略；(2)将能量谱映射到梅尔谱过程中，低频部分滤波器宽度要大于高频部分的滤波器宽度，即梅尔谱更关注于高频部分；(3)将能量谱信号返回到频谱域过程中采用的随机相位也是不可忽略的影响因素。

4.2 字典原子个数、稀疏度影响分析

稀疏表示中字典的原子个数决定着字典的规模，字典规模越大所包含源信号的成份就越多，在同样多的稀疏系数下重构信号就越精确，但字典规

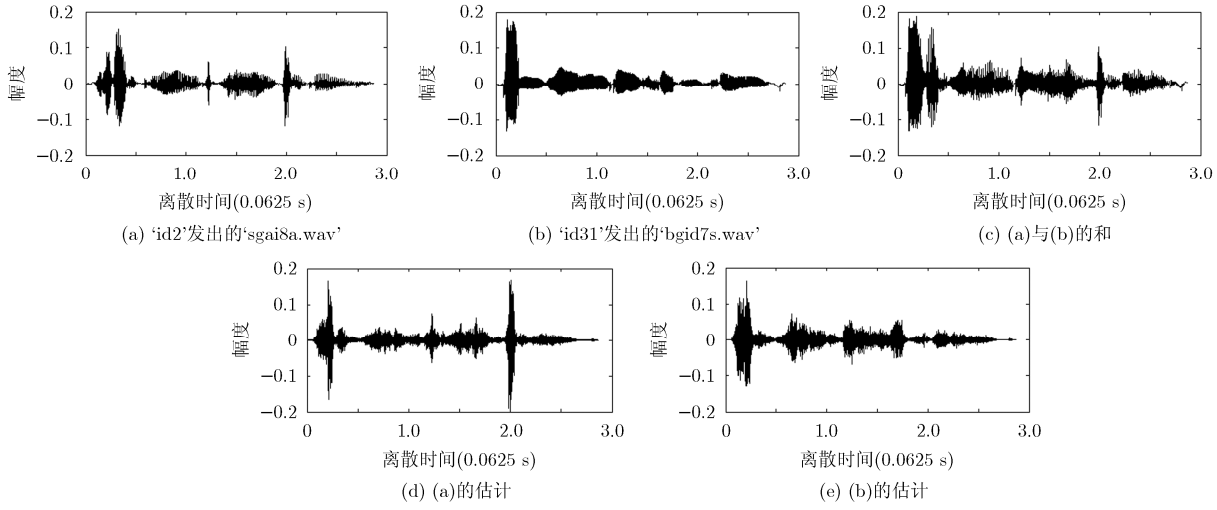


图 2 混合信号分解过程图

模越大各个子字典所包含成份相似的概率增大，子字典间互干扰问题也就越严重，表 1 给出了当 $\lambda = 0.85$ 时不同算法在不同 l 和 K 取值时的 SIR 和 SNR。表 1 中使用稀疏度 η 代表 K 来统计实验结果，其中 $K = \lfloor \eta L \rfloor$。

从表 1 每一行(横向)比较来看，SNMF 的 SNR 和 SIR 随着 η 的增加都在不断地减小，这主要是由于 SNMF 的各个子字典是独立训练的，没有考虑其他子字典的影响，所以随着稀疏系数的增多干扰会有所加大，导致 SNR 和 SIR 下降。考虑了字典间的互干扰影响后，DDL 和本文算法在 SNR 和 SIR 上下降速度明显减缓，在一些参数上甚至出现了上升趋势。例如本文算法的 SIR 指标除了在 $l = 100, 120$ 和 $\eta = 0.10, 0.15$ 时出现下跌，在其余参数上均是上升趋势，上升的原因是，在 l 比较小的时候增大 η 有助于稀疏表示选择更多的信息来拟合输入信号，但当 l 和 η 都较大时，字典本身是冗余的，过多的稀疏系数会增大子字典间的干扰概率，因此会出现 SIR 下降的现象。在 $l = 120$ 和 $\eta = 0.15$ 时，本文算法的 SIR 甚至降到了 7.2 以下。DDL 的 SIR 指标和本文算法基本相同，但是在一些参数上有波动，主要是由于本文算法联合字典中含有共同子字典。在 SNR 指标上，本文算法除了在 $l = 60$ 时呈现的基本是上升趋势，其余的均是下降趋势。而 DDL 的 SNR 在 $l = 60, 70$ 时呈现是上升趋势，其余基本都是下降趋势。对于 SNR 和 SIR 呈现出的不同趋势，主要是由于 SIR 反映的是其他源对当前源恢复信号干扰，而 SNR 反映的是真实信号与总干扰之比，这些干扰可能还包括噪声干扰，人为干扰等等。

从表 1 每一列(纵向)比较来看，在 $\eta = 0.06, 0.07$ 时，本文算法和 SNMF 的 SNR 指标均呈现先上升后下降趋势，但是 DDL 却呈现出了单调上升趋势。这主要是由于 DDL 字典的每个原子包含的不仅仅是区别于其他源的信息，还包含有和其他源相同的信息，所以 SNR 会增大。对本文算法在 $\eta = 0.06, 0.07$ 时下降趋势的一种可能解释是，过大的 l 会导致共同字典对各个源恢复信号的干扰加大，从而降低 SNR，在 $\eta > 0.7$ 时这种情况更是如此。对于 SNMF 的上升趋势，主要是由于在 η 较小时适当地增加 l 有助于丰富字典的信息，但是当 η 较大时，这种优势被字典过大干扰会越大的劣势取代，所以呈现了先增大再减小的趋势。当在 SIR 指标上对比时，本文算法的优势相对于 SNMF 和 DDL 体现的比较明显，主要是由于本文提出的字典包含有共同字典。

为了公平的对比，选择了每一种算法表现均优良的参数 $l = 80,\ \eta = 0.06$ 进行后续实验。最后值得指出的是 SIR 需要计算每一个估计信号在源信号构成空间上的投影，所以 SIR 是一帧一帧计算得到的，而 SNR 是计算所有帧构成的矩阵。所以表 1 中 SNR 总是小于 SIR，但是这并不影响对比分析。

4.3 目标函数权重系数 γ 的分析

目标函数中权衡总体保真度和单独源重构误差的系数 γ 是一个非常重要的参数，它表明了字典的性能更偏重哪一方面。经典的 SNMF 算法是单独训

表 1　不同算法在子字典原子个数 l 和稀疏度 η 不同时的 SIR 和 SNR 值(dB)

		SNR					SIR				
		η					η				
	l	0.06	0.07	0.08	0.10	0.15	0.06	0.07	0.08	0.10	0.15
SNMF[8]	60	1.9812	1.8207	1.8091	1.3943	0.7510	**9.3506**	8.9260	8.6734	7.8278	6.5565
	70	2.0151	1.9118	1.7216	1.2457	0.4764	9.0587	8.7883	8.3195	6.9817	5.7745
	80	**2.0658**	1.9074	1.6596	1.3146	0.9233	9.0209	8.3449	7.6121	7.0103	5.3982
	100	1.8429	1.6673	1.4684	1.3312	1.0254	7.8888	7.4993	7.2101	6.4852	5.5509
	120	1.7859	1.6673	1.5309	1.3263	1.0867	7.5746	7.1593	6.9140	6.5061	5.4975
DDL[14]	60	1.9474	2.0006	2.0253	2.0896	2.1463	9.3448	9.2493	9.0734	8.9464	9.2146
	70	2.0692	2.0928	2.1306	2.1748	2.1162	9.5017	9.3448	9.1807	9.3194	9.3973
	80	2.1183	2.1053	2.1748	2.1779	2.1241	9.3786	9.0659	9.3029	9.2307	9.6849
	100	2.1648	2.1588	2.1143	2.1220	2.0796	9.2474	9.2908	9.2334	9.4190	**9.9624**
	120	**2.2376**	2.2160	2.1716	2.1081	1.9998	9.6857	9.7015	9.7238	9.7069	9.8718
本文算法	60	2.8301	2.8287	2.8739	2.9003	2.5625	8.7897	8.7504	8.9841	9.2578	9.4516
	70	**2.9073**	2.8991	2.8041	2.7953	2.6355	9.1568	9.2931	8.9902	9.9279	10.478
	80	2.8951	2.8426	2.7729	2.6808	2.4604	9.3647	9.3421	9.7199	10.0238	**10.494**
	100	2.8052	2.7405	2.6297	2.5817	2.1597	9.9327	9.7930	9.9514	10.3215	7.7783
	120	2.7229	2.6357	2.5879	2.3857	2.1339	10.2116	10.3843	10.2168	9.6330	7.1948

练各个子字典后再排列在一起，并不涉及这个参数，因此，图 3 只对比了本文算法和 DDL 算法在 γ 取不同值时的 SIR 和 SNR。

通过图 3 可以发现，随着 γ 的增大，本文算法和对比算法在两个指标上的变化均呈现了先迅速增大再逐渐减小的趋势，本文算法在 SNR 上减小速度先是很剧烈然后趋于缓和。这一趋势表明，在 $\gamma=0.5$ 的基础上适当的增加 γ 值有助于提高本文算法和对比算法的性能。但是当 γ 值过大的时候将会放大字典间互干扰的影响，尤其是 γ 过大时会迅速地降低本文算法在 SNR 上的表现，主要是 γ 放大了共同字典对算法的影响。图 3 还表明，在 $\gamma<2$ 时本文算法要明显优于对比算法，尤其是在 $\gamma=0.75$ 时，本文算的 SNR 接近于 3，而 DDL 的 SNR 却低于 2。总体而言，当 $\gamma=0.85$ 时，本文算法和对比算法都可以达到最优。

4.4 不同性别发音者对算法的影响

一种更全面地分析本文算法性能的方法是，测试本文方法分离男发音者与男发音者的混合信号(Male and Male, M+M)、女发音者与男发音者的混合信号(Female and Male, F+M)、女发音者与女发音者的混合信号(Female and Female, F+F)时的性能，根据 4.2 节和 4.3 节的实验结果，取 $l=80$，$\eta=0.06$，$\gamma=0.85$，图 4 给出了不同算法在分离 3 种混合信号时的 SIR 和 SNR。

通过图 4 可以看出，在每一种混合情况下，本文算法在两个指标上的性能均优于两种对比算法，尤其是在性别相同的情况下，本文算法的优势更为明显，如在 M+M 情况下，本文算法在 SNR 和 SIR 上比 DDL 分别高出了 0.9 dB 和 1.8 dB。这主要是因为本文算法的字典中的共同子字典可以很好地吸收输入源信号的共同成份，但是对比算法的这部分共同成份只能在其他子字典上寻求表达从而会引起不小的干扰。图 4(a)和图 4(b)表明，最好的 SNR 和 SIR 在 F+M 的情况下取得，且与对比算法之间的差别不大。主要是因为不同的源之间本身差别很大，即构成信号的主要成份是不同源信号所特有的成份，所以本文提出的共同字典的优势体现不明显。

4.5 算法复杂度分析

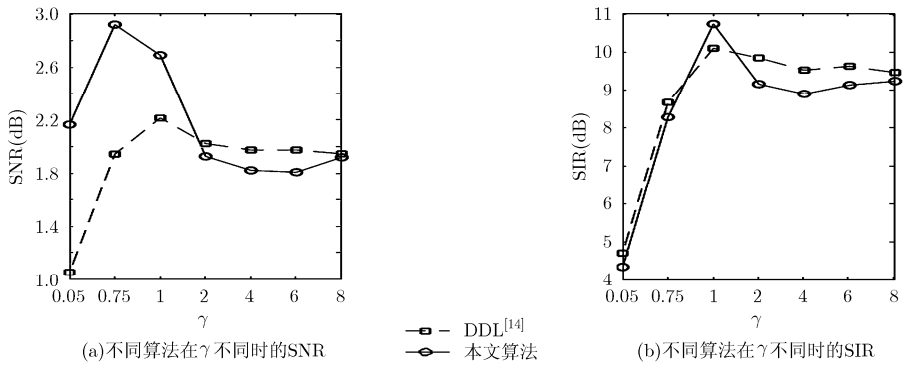
根据 3.2 节描述的优化过程可知，本文算法和对比算法都涉及到两个的步骤，利用 OMP 求解稀疏表示系数和用 K-SVD 更新每个原子。OMP 和 K-SVD 的复杂度主要在于矩阵求伪逆和 SVD 分解。对于一个 m 行 n(假定 $n>m$)列的矩阵，基于 QR 的 SVD 分解算法复杂度约为 $O(d^3)+2mn^2+3n^2+2mn-m-n$。基于 SVD 的矩阵伪逆的复杂度只比 SVD 多 $O(n^2)$，所以本文讨论中矩阵的伪逆和 SVD 的复杂度都取 $O(n^3+2mn^2)$。

基于 SVD 和矩阵求逆的复杂度，可得 OMP 得复杂度为 $Q_{\text{OMP}}=O(\eta l\cdot d^3)$，K-SVD 的复杂度为 $Q_{K\text{-SVD}}=t(\eta lO(d^3)+lO(d^3))$，其中 t 为字典更新的次数，可以发现 $Q_{K\text{-SVD}}$ 中有一项等于 Q_{OMP}，这是由于在每次迭代开始前需要利用新更新的字典重新表示输入信号。内点法求解二次规划问题的非常成熟，其复杂度远小于 $O(d^3)$，可以忽略。

所以在本文设定源个数为 2 的情况下，本文算法，SNMF[8]和 DDL[14]的复杂度分别为 $3tl(\eta+1)\cdot O(d^3)$，$tl(\eta+1)O(d^3)$ 和 $2tl(\eta+1)O(d^3)$。可以看出，本文算法由于增加了一个子字典比其他对比算法的复杂度要高。但是值得指出的是，以上分析只是针对训练算法而言的，对于实际的应用情况，训练基本在线下完成，对时效性要求并不高，要求比较高的是数据连续输入的在线测试。对于测试过程，因为只涉及 OMP，没有 K-SVD 的迭代次数 t，本文算法只比对比算法慢一点，是可以接受的。

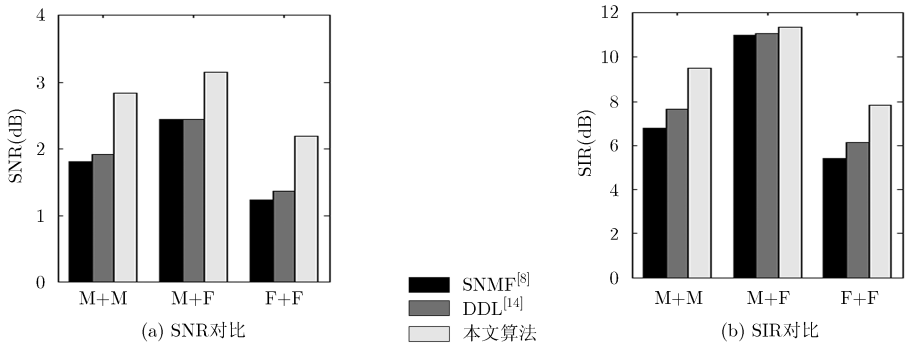
5 结束语

针对联合字典在稀疏表达时存在的子字典互干



(a)不同算法在γ不同时的SNR　(b)不同算法在γ不同时的SIR

图 3 不同算法在 γ 不同时的 SIR 和 SNR



(a) SNR对比　　(b) SIR对比

图 4 不同算法在不同性别发音者源信号混合时的 SNR 和 SIR 对比

扰的问题，通过引入共同子字典，本文提出了一新的字典构建方法，即在原来各个源对应子字典构成联合字典的基础上，引入了共同子字典。并通过 3 个步骤顺序执行优化了目标函数，实现了基于信号分离任务的字典学习。通过在语音数据库上的实验表明，引入共同子字典后训练得到的联合字典在鉴别力和保真度两方面均有提高，改善了基于联合字典稀疏表示的单通道盲源分离性能。但是训练过程中字典的原子个数，稀疏表达的稀疏系数以及权衡总体保真度误差和鉴别力误差的权重，对算法的性能均有不同程度的影响。虽然本文只实验和分析了两个源信号混合时的盲源分离性能，但是从文中字典学习原理可以发现，本文所提方法不难推广到多个源混合时的情况，后续可以做进一步的讨论和实验。另外本文为了论述的方便只讨论了各个子字典原子个数相同的情况，因此子字典大小不同的情况也是以后的一个研究方向。

参 考 文 献

[1] VANEPH A, MCNEIL E, RIGAUD F, *et al*. An automated source separation technology and its practical applications[C]. Audio Engineering Society Convention 140. Audio Engineering Society, Paris, France, 2016: 181–182.

[2] 杜健, 巩克现, 葛临东. 基于单路定时准确的低复杂度成对载波复用多址信号盲分离算法[J]. 电子与信息学报, 2014, 36(8): 1872–1877. doi: 10.3724/SP.J.1146.2013.01459.
DU Jian, GONG Kexian, and GE Lindong. Low complexity algorithm on blind separation of paired carrier multiple access signals based on single way timing accuracy[J]. *Journal of Electronics & Information Technology*, 2014, 36(8): 1872–1877. doi: 10.3724/SP.J.1146.2013.01459.

[3] LOPEZ A R, ONO N, REMES U, *et al*. Designing multichannel source separation based on single-channel source separation[C]. 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, Brisbane, Australia, 2015: 469–473. doi: 10.1109/ICASSP.2015.7178013.

[4] 吴迪, 陶智, 张晓俊, 等. 感知听觉场景分析的说话人识别[J]. 声学学报, 2016, 41(2): 260–272. doi: 10.15949/j.cnki.0371-0025.2016.02.015.
WU Di, TAO Rui, and ZHANG Xiaojun, *et al*. Perception auditory scene analysis for speaker recognition[J]. *Acta Acustica*, 2016, 41(2): 260–272. doi: 10.15949/j.cnki.0371-0025.2016.02.015.

[5] 杨立东, 王晶, 谢湘, 等. 基于低秩张量补全的多声道音频信号恢复方法[J]. 电子与信息学报, 2016, 38(2): 394–399. doi: 10.11999/JEIT150589.
YANG Lidong, WANG Jing, and XIE Xiang, *et al*. Low rank tensor completion for recovering missing data in multi-channel audio signal[J]. *Journal of Electronics & Information Technology*, 2016, 38(2): 394–399. doi: 10.11999/JEIT150589.

[6] JANG G J, LEE T W, and OH Y H. Single-channel signal separation using time-domain basis functions[J]. *IEEE Signal Processing Letters*, 2003, 10(6): 168–171. doi: 10.1109/LSP.2003.811630.

[7] 王钢, 孙斌. 盲信号分离技术及算法研究[J]. 航天电子对抗, 2015, 31(4): 53–56. doi: 10.16328/j.htdz8511.2015.04.015.
WANG Gang and SUN Bin. Research on blind signal separation technology and algorithm[J]. *Aerospace Electronic Warfare*, 2015, 31(4): 53–56. doi: 10.16328/j.htdz8511.2015.04.015.

[8] SCHMIDT M N and OLSSON R K. Single-channel speech separation using sparse non-negative matrix factorization[C]. ISCA International Conference on Spoken Language Proceesing, (INTERSPEECH), Pittsburgh, Pennsylvania, 2006: 2614–2617.

[9] KING B J and ATLAS L. Single-channel source separation using complex matrix factorization[J]. *IEEE Transactions on Audio, Speech, and Language Processing*, 2011, 19(8): 2591–2597. doi: 10.1109/TASL.2011.2156786.

[10] GRAIS E M and ERDOGAN H. Single channel speech music separation using nonnegative matrix factorization with

sliding window and spectral masks[C]. Annual Conference of the International Speech Communication Association (INTERSPEECH), Florence, Italy, 2011: 1773–1776.

[11] GRAIS E M and ERDOGAN H. Discriminative nonnegative dictionary learning using cross-coherence penalties for single channel source separation[C]. INTERSPEECH, Lyon, France, 2013: 808–812.

[12] WENINGER F, LE Roux J, HERSHEY J R, *et al*. Discriminative NMF and its application to single-channel source separation[C]. Annual Conference of the International Speech Communication Association (INTERSPEECH), Singapore, 2014: 865–869.

[13] BAO G, XU Y, and YE Z. Learning a discriminative dictionary for single-channel speech separation[J]. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 2014, 22(7): 1130–1138. doi: 10.1109/TASLP.2014.2320575.

[14] WANG Z and SHA F. Discriminative non-negative matrix factorization for single-channel speech separation[C]. 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Florence, Italy, 2014: 3749–3753. doi: 10.1109/ICASSP.2014.6854302.

[15] 张春梅, 尹忠科, 肖明霞. 基于冗余字典的信号超完备表示与稀疏分解[J]. 科学通报, 2006, 51(6): 628–633.
ZHANG Chunmei, YIN Zhongke, and XIAO Mingxia. Signal over-complete representation and sparse decomposition based on redundant dictionary[J]. *Chinese Science Bulletin*, 2006, 51(6): 628–633.

[16] AHARON M, ELAD M, and BRUCKSTEIN A. K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation[J]. *IEEE Transactions on Signal Processing*, 2006, 54(11): 4311–4322. doi: 10.1109/TSP.2006.881199.

[17] COOKE M, BARKER J, CUNNINGHAM S, *et al*. An audio-visual corpus for speech perception and automatic speech recognition[J]. *The Journal of the Acoustical Society of America*, 2006, 120(5): 2421–2424. doi: 10.1121/1.2229005.

[18] VINCENT E, GRIBONVAL R, and FEVOTTE C. Performance measurement in blind audio source separation[J]. *IEEE Transactions on Audio, Speech, and Language Processing*, 2006, 14(4): 1462–1469. doi: 10.1109/TSA.2005.858005.

[19] THOMAS S, SAON G, KUO H, *et al*. The IBM BOLT speech transcription system[C]. Sixteenth Annual Conference of the International Speech Communication Association, Dresden, Germany, 2015: 3150–3153.

[20] NORRIS D, MCQUEEN J M, and CUTLER A. Prediction, Bayesian inference and feedback in speech recognition[J]. *Language, Cognition and Neuroscience*, 2016, 31(1): 4–18. doi: 10.1080/23273798.2015.1081703.

田元荣： 男，1989 年生，博士生，研究方向为稀疏表示理论与信号分离.

王 星： 男，1965 年生，博士，教授，研究方向为电子对抗理论与技术.

周一鹏： 男，1992 年生，硕博连读生，研究方向为电子侦察与信号处理技术.