一种新的基于稀疏分解的单通道混合语音分离方法

郭海燕[1,2],杨 震[2],朱卫平[2]
(1.南京农业大学工学院,江苏南京 210031; 2.南京邮电大学"宽带无线通信与传感网技术"教育部重点实验室,江苏南京 210003)

摘 要: 论文以新的语音信号稀疏基—准KLT基的构造为基础,提出了一种新的基于稀疏分解的单通道混合语音分离方法.论文首先以理想准KLT基的构造为基础,从理论上提出并证明了基于各源语音信号的理想准KLT基,利用 l exp(0)-范数优化算法,可实现单通道混合语音的完美分离.鉴于单通道混合语音分离时,无法精确求取各源语音信号的理想准KLT基,论文提出先基于正交匹配追踪算法,以混合语音信号为已知条件,构造各源语音信号的正交匹配追踪模板匹配准KLT基,再由 l exp(0)-范数优化算法来分离单通道混合语音.仿真实验表明论文所提理论的正确性,和基于正交匹配追踪模板匹配准KLT基来分离单通道混合语音信号的有效性.

关键词: 语音分离; 稀疏分解; l exp (0) -范数优化; 正交匹配追踪; Karhunen-Loeve 变换
中图分类号: TN912 **文献标识码:** A **文章编号:** 0372-2112 (2012)04-0762-07
电子学报 URL: http://www.ejournal.org.cn **DOI:** 10.3969/j.issn.0372-2112.2012.04.023

A New Single-Channel Speech Separation Method Based on Sparse Decomposition

GUO Hai-yan[1,2], YANG Zhen[2], ZHU Wei-ping[2]
(1. *College of Engineering, Nanjing Agricultural University, Nanjing, Jiangsu* 210031, *China*;
2. *Key Laboratory of Broadband Wireless Communication and Sensor Network Technology (Ministry of Education), Nanjing University of Posts & Telecommunications, Nanjing, Jiangsu* 210003, *China*)

Abstract: This paper proposes a new single-channel speech separation method based on sparse decomposition. The new method is based on the construction algorithms of quasi-KLT bases proposed in this paper. First, it is proved that based on the ideal quasi-KLT bases generated from the sources, both sources can be perfectly separated from a single mixture by l exp(0) optimization. Then, considering the fact that the ideal quasi-KLT bases cannot be obtained in practice, orthogonal matching pursuit (OMP) template-matching quasi-KLT bases generated from the single mixture are first proposed, and then perform single-channel speech separation by l exp(0) optimization using the OMP template-matching quasi-KLT bases. Finally, simulation results demonstrating the perfect separation of both sources from a single-channel speech mixture by l exp(0) optimization using ideal quasi-KLT bases, as well as the effective performance of single-channel separation of two speech sources using the OMP template-matching quasi-KLT bases are presented.

Key words: speech separation; sparse decomposition; l exp(0) optimization; orthogonal matching pursuit (OMP); Karhunen-Loeve transform (KLT)

1 引言

在混合语音的分离研究中,当已知混合信号数目小于未知源信号数目情况时,即欠定条件下时,分离难度较大.这是因为与过(适)定情况相比,欠定情况下已知条件最少,我们无法精确求解出混合矩阵,进而来重构各源语音信号.然而,这种欠定情况在现实中经常出现,例如在现实环境中,语音信号往往通过一个麦克风采集得到,以此种方式采集得到的语音,在很多情况下都包含有多个源语音信号.稀疏分解理论的出现和发展,使得解决欠定语音分离问题有了新的思路,在一定程度下降低了欠定语音分离的难度.这是因为:一方面,语音信号在稀疏域下投影向量的非零元素个数,远小于时域语音信号的非零元素个数,这就意味着在稀疏域中,不

收稿日期:2011-03-22;修回日期:2011-06-13
基金项目:国家自然科学基金(No.60971129);中国博士后基金(No.20100481167);南京农业大学工学院引进人才科研启动基金(No.rcqd11-02)

同源语音信号重叠的位置更少,这无疑降低了分离的难度与计算量;另一方面,从数学的角度看,可以将欠定混合语音的分离问题看作一组欠定方程组的求解问题,已知量为各混合语音信号,未知量为各源语音信号,在稀疏域求解此类问题,相当于在原有已知条件的基础上又加入了稀疏性这一限制,缩小了欠定问题解的范围,有利于问题的求解.

本论文研究的对象为一种特殊的欠定混合语音分离类型:单通道混合语音分离,即混合语音信号个数为1,其难度由于已知条件更少而进一步加大.文献[1~3]都是基于稀疏分解理论来分离单通道混合语音信号的,具体先训练构造与说话人相关的过完备字典来组成联合字典,再在此基础上,采用优化算法重构分离语音信号.文献[1]基于稀疏非负矩阵分解(sparse non-negative matrix factorization,SNMF)算法训练构造字典,文献[2]采用 l^0-范数优化算法,在 mel 谱域训练构造字典,文献[3]选择具有代表性的各说话人语音归一化幅度谱作为各说话人的码本.虽然上述论文基于稀疏分解提出了有一定效果的分离方法,但都未从理论层面对什么情况下能达到完美分离进行分析,缺乏完整理论依据.文献[4]证明,当字典为若干个基的联合,且字典满足一定的条件时,通过 l^0-范数优化算法,可以唯一地求取信号在字典下的稀疏投影向量.试想如果能对各源语音信号构造满足条件的基函数,将各基函数联合起来组成联合字典,则可以唯一地求取混合语音信号在此联合字典下的投影向量,以此投影向量和各基函数为条件,有望获得分离质量很好的各源语音信号.但问题是满足此种条件的基函数的联合不易求取,且到目前为止,也很少有论文涉及此问题.从理论上探讨如何完美分离单通道混合语音信号,是本文的一个重要工作.

本文针对源语音信号数目为2的情况,从探讨如何完美分离单通道混合语音入手,首先在理论上提出并证明了,基于作者以前提出的理想准 KLT 基,采用 l^0-范数优化算法,可实现单通道混合语音的完美分离.并以此为理论依据,提出实用的基于正交匹配(Orthogonal Matching Pursuit,OMP)模板匹配准 KLT 基的单通道混合语音的分离算法,以克服在实际分离单通道混合语音时,由于各源语音信号未知而无法精确求取各源语音信号理想准 KLT 基的问题.仿真实验证明这种算法的分离性能优于已有的两种经典算法.

2　基于理想准 KLT 基的单通道混合语音分离

本小节主要考虑从理论上提出并证明,基于理想准 KLT 基,采用 l^0-范数优化算法,可实现单通道混合语音的完美分离.首先我们考虑一般信号的线性混合模型,并不局限于语音信号.已知两源信号 $\boldsymbol{x}_i=[x_i(1),x_i(2),\cdots,x_i(\mathrm{N})]^{\mathrm{T}}(i=1,2)$

$$\boldsymbol{x}=\sum_{i=1}^{2}\boldsymbol{x}_i \tag{1}$$

假设源信号 $\boldsymbol{x}_i$ 在基 $\boldsymbol{\Psi}_i$ 下具有稀疏性,即

$$\left\|\boldsymbol{\Theta}_{\boldsymbol{x}_i}\right\|_0=\left\|\boldsymbol{\Psi}_i\boldsymbol{x}_i\right\|_0=K_i\ll\mathrm{N},\ i=1,2 \tag{2}$$

其中 $\boldsymbol{\Theta}_{x_i}$ 为 $\boldsymbol{x}_i$ 在基 $\boldsymbol{\Psi}_i$ 下的投影向量.我们可证明当 $\boldsymbol{\Psi}_i$ 满足一定条件时,通过解下列 l^0-范数优化问题可以实现单通道混合信号 x 的完美分离,即

$$\begin{aligned}&\min f(\boldsymbol{\Theta}_1,\boldsymbol{\Theta}_2)=\min\sum_{i=1}^{2}\|\boldsymbol{\Theta}_i\|_0\\&\text{s.t.}\quad\boldsymbol{x}=\sum_{i=1}^{2}\boldsymbol{\Psi}_i^{-1}\boldsymbol{\Theta}_i\end{aligned}, \tag{3}$$

我们推导得出如下定理:

定理 1　已知源信号 $\boldsymbol{x}_1$ 和 $\boldsymbol{x}_2$ 的混合信号 $\boldsymbol{x}$,如果 $\boldsymbol{\Psi}_1$ 和 $\boldsymbol{\Psi}_2$ 均可逆,且满足式(2)及下式

$$\|\boldsymbol{\Psi}_1\boldsymbol{z}\|_0+\|\boldsymbol{\Psi}_2(\boldsymbol{x}-\boldsymbol{z})\|_0>\sum_{i=1}^{2}K_i,\quad\forall\,\boldsymbol{z}\neq\boldsymbol{x}_1 \tag{4}$$

则问题(3)有唯一解 $\boldsymbol{\Theta}^{\mathrm{opt}}=[\boldsymbol{\Theta}_{x_1}\quad\boldsymbol{\Theta}_{x_2}]$,其中 $\boldsymbol{\Theta}_{x_i}$ 与式(2)中相同;由 $\boldsymbol{x}_i=\boldsymbol{\Psi}_i^{-1}\boldsymbol{\Theta}_{x_i}(i=1,2)$ 可从混合信号 $\boldsymbol{x}$ 中完美分离出源信号 $\boldsymbol{x}_1,\boldsymbol{x}_2$.

证明:

由于 $\boldsymbol{\Psi}_i$ 可逆,根据式(1)和(2)可得,

$$\boldsymbol{x}=\sum_{i=1}^{2}\boldsymbol{\Psi}_i^{-1}\boldsymbol{\Theta}_{x_i} \tag{5}$$

假设存在 $\boldsymbol{\Theta}'=[\boldsymbol{\Theta}_1'\quad\boldsymbol{\Theta}_2']\neq[\boldsymbol{\Theta}_{x_1}\quad\boldsymbol{\Theta}_{x_2}]$ 为问题(3)的最优解,则 $[\boldsymbol{\Theta}_1'\quad\boldsymbol{\Theta}_2']$ 必满足

$$\boldsymbol{x}=\sum_{i=1}^{2}\boldsymbol{\Psi}_i^{-1}\boldsymbol{\Theta}_i' \tag{6}$$

且

$$\sum_{i=1}^{2}\|\boldsymbol{\Theta}_i'\|_0\leqslant\sum_{i=1}^{2}\|\boldsymbol{\Theta}_{x_i}\|_0=\sum_{i=1}^{2}K_i \tag{7}$$

其中 $\boldsymbol{\Theta}_i'$ 为 $\mathrm{N}\times1$ 维的列向量.定义 $\boldsymbol{z}=\boldsymbol{\Psi}_1^{-1}\boldsymbol{\Theta}_1'$,我们有

$$\boldsymbol{\Theta}_1'=\boldsymbol{\Psi}_1\boldsymbol{z} \tag{8}$$

同时,由式(6)可得

$$\boldsymbol{\Theta}_2'=\boldsymbol{\Psi}_2(\boldsymbol{x}-\boldsymbol{z}) \tag{9}$$

结合式(8)和(9),可得

$$\sum_{i=1}^{2}\|\boldsymbol{\Theta}_i'\|_0=\|\boldsymbol{\Psi}_1\boldsymbol{z}\|_1+\|\boldsymbol{\Psi}_2(\boldsymbol{x}-\boldsymbol{z})\|_0 \tag{10}$$

由于 $\boldsymbol{\Theta}'=[\boldsymbol{\Theta}_1'\quad\boldsymbol{\Theta}_2']\neq[\boldsymbol{\Theta}_{x_1}\quad\boldsymbol{\Theta}_{x_2}]$,故 $\boldsymbol{z}\neq\boldsymbol{x}_1$,根据式(4)有

$$\sum_{i=1}^{2}\|\boldsymbol{\Theta}_i'\|_0>\sum_{i=1}^{2}K_i=\sum_{i=1}^{P}\|\boldsymbol{\Theta}_{x_i}\|_0 \tag{11}$$

这与式(7)相矛盾,故假设不成立,即不存在 $\boldsymbol{\Theta}'=[\boldsymbol{\Theta}_1'\quad\boldsymbol{\Theta}_2']\neq[\boldsymbol{\Theta}_{x_1}\quad\boldsymbol{\Theta}_{x_2}]$ 为问题(3)的最优解.由式(5)可知,$[\boldsymbol{\Theta}_{x_1}\quad\boldsymbol{\Theta}_{x_2}]$ 满足问题(3)的约束条件,故问

题(3)有最优解 $\boldsymbol{\Theta}^{\text{opt}}=[\boldsymbol{\Theta}_{x_1}\quad\boldsymbol{\Theta}_{x_2}]$.

根据上述唯一最优解 $\boldsymbol{\Theta}^{\text{opt}}=[\boldsymbol{\Theta}_{x_1}\quad\boldsymbol{\Theta}_{x_2}]$,由

$$\boldsymbol{x}_i=\boldsymbol{\Psi}_i^{-1}\boldsymbol{\Theta}_{x_i},\quad i=1,2 \tag{12}$$

可以获得完美分离的源信号 $\boldsymbol{x}_i$.

在作者以前的研究工作中,本文作者提出了准 KLT 基的构造[5],下面证明理想情况下(源语音信号已知时)构造的准 KLT 基满足定理 1 中的完美分离条件.

已知源语音信号 $\boldsymbol{s}_i=[s_i(1),s_i(2),\cdots,s_i(\mathrm{N})]^{\mathrm{T}}(i=1,2)$,通过对 $\boldsymbol{s}_i$ 的自相关矩阵 $\boldsymbol{R}_{s_i}$ 进行特征值分解,我们可构造 $\boldsymbol{s}_i$ 的理想准 KLT 基 $\boldsymbol{U}_{s_i}^{\mathrm{T}}$ 和理想准 IKLT 基 $\boldsymbol{U}_{s_i}$,

$$\boldsymbol{R}_{s_i}=\boldsymbol{s}_i\boldsymbol{s}_i^{\mathrm{T}},i=1,2 \tag{13}$$

$$\boldsymbol{R}_{s_i}=\boldsymbol{U}_{s_i}{}^{\mathrm{T}}\boldsymbol{\Lambda}_{s_i}\boldsymbol{U}_{s_i},i=1,2 \tag{14}$$

其中 $\boldsymbol{\Lambda}_{s_i}$ 为对角阵,$\boldsymbol{U}_{s_i}$ 为正交矩阵,$\boldsymbol{U}_{s_i}^{-1}$ 表示 $\boldsymbol{U}_{s_i}$ 的逆.由于 $\boldsymbol{U}_{s_i}$ 为正交矩阵,我们有

$$\boldsymbol{U}_{s_i}^{\mathrm{T}}=\boldsymbol{U}_{s_i}^{-1} \tag{15}$$

其中 $\boldsymbol{U}_{s_i}^{\mathrm{T}}$ 表示 $\boldsymbol{U}_{s_i}$ 的转置.我们将 $\boldsymbol{U}_{s_i}^{\mathrm{T}}$ 称为 $\boldsymbol{s}_i$ 的理想准 KLT 基,将 $\boldsymbol{U}_{s_i}$ 称为 $\boldsymbol{s}_i$ 的理想准 IKLT(inverse KLT)基.显然,$\boldsymbol{U}_{s_i}$ 可逆.信号 $\boldsymbol{s}_i$ 在其理想准 KLT 基 $\boldsymbol{U}_{s_i}^{\mathrm{T}}$ 下的投影向量 $\boldsymbol{\Theta}_{s_i}$ 为,

$$\boldsymbol{\Theta}_{s_i}=\boldsymbol{U}_{s_i}^{\mathrm{T}}\boldsymbol{s}_i \tag{16}$$

下面来证明信号 $\boldsymbol{s}_i$ 在其理想准 KLT 基 $\boldsymbol{U}_{s_i}^{\mathrm{T}}$ 下的投影向量 $\boldsymbol{\Theta}_{s_i}$ 是绝对稀疏的.

定理 2 信号 $\boldsymbol{s}_i$ 在其理想准 KLT 基 $\boldsymbol{U}_{s_i}^{\mathrm{T}}$ 下的投影向量 $\boldsymbol{\Theta}_{s_i}$ 满足

$$\boldsymbol{\Theta}_{s_i}(j)=\begin{cases}\neq 0, & j=j_\Delta\\ =0, & j\neq j_\Delta,1\leqslant j\leqslant N\end{cases}$$

其中 $\boldsymbol{\Theta}_{s_i}(j)$ 为 $\boldsymbol{\Theta}_{s_i}$ 的第 j 个元素.

证明 由式(13),(14),(15),(16)得

$$\begin{aligned}\boldsymbol{\Theta}_{s_i}\boldsymbol{\Theta}_{s_i}^{\mathrm{T}}&=(\boldsymbol{U}_{s_i}^{\mathrm{T}}\boldsymbol{s}_i)(\boldsymbol{U}_{s_i}^{\mathrm{T}}\boldsymbol{s}_i)^{\mathrm{T}}=\boldsymbol{U}_{s_i}^{\mathrm{T}}\boldsymbol{s}_i\boldsymbol{s}_i^{\mathrm{T}}\boldsymbol{U}_{s_i}\\&=\boldsymbol{U}_{s_i}^{\mathrm{T}}\boldsymbol{R}_{s_i}\boldsymbol{U}_{s_i}=\boldsymbol{\Lambda}_{s_i}\end{aligned} \tag{17}$$

由于 $\boldsymbol{\Lambda}_{s_i}$ 为元素非全零的对角阵,故 $\exists j_\Delta$ 使

$$\boldsymbol{\Lambda}_{s_i}(j_\Delta,j_\Delta)=\boldsymbol{\Theta}_{s_i}(j_\Delta)\boldsymbol{\Theta}_{s_i}^{\mathrm{T}}(j_\Delta)\neq 0 \tag{18}$$

$$\boldsymbol{\Lambda}_{s_i}(j,j_\Delta)=\boldsymbol{\Theta}_{s_i}(j)\boldsymbol{\Theta}_{s_i}^{\mathrm{T}}(j_\Delta)=0,\quad \text{其中 } j\neq j_\Delta,1\leqslant j\leqslant \mathrm{N} \tag{19}$$

其中 $\boldsymbol{\Lambda}_{s_i}(j,j_\Delta)$ 为 $\boldsymbol{\Lambda}_{s_i}$ 的第 j 行第 j_Δ 列元素,$\boldsymbol{\Lambda}_{s_i}(j_\Delta,j_\Delta)$ 为 $\boldsymbol{\Lambda}_{s_i}$ 的第 j_Δ 行第 j_Δ 列元素.由式(18),(19)得

$$\boldsymbol{\Theta}_{s_i}(j)=\begin{cases}\neq 0, & j=j_\Delta\\ =0, & j\neq j_\Delta,1\leqslant j\leqslant \mathrm{N}\end{cases} \tag{20}$$

综上所述,定理 1 得证.

由定理 1 得,$\|\boldsymbol{\Theta}_{s_i}\|_0=1$,即信号 $\boldsymbol{s}_i$ 在 $\boldsymbol{U}_{s_i}^{\mathrm{T}}$ 下的投影向量 $\boldsymbol{\Theta}_{s_i}$ 中只有一个非零元素,故 $\boldsymbol{\Theta}_{s_i}$ 是绝对稀疏的.

最后,我们来证明 $\boldsymbol{U}_{s_i}^{\mathrm{T}}$ 满足定理 1 中的不等式条件.

由于 $\boldsymbol{s}_1$ 和 $\boldsymbol{s}_2$ 表示有确切物理含义的语音信号,故一般情况下,$\boldsymbol{s}_1$ 不可能由 $\boldsymbol{U}_{s_2}$ 的某单独一行表示,同样 $\boldsymbol{s}_2$ 不可能由 $\boldsymbol{U}_{s_1}$ 的某单独一行表示.所以,我们可做下面的合理假设.

假设 源语音信号 $\boldsymbol{s}_i(i=1,2)$ 在另一源语音信号 $\mathrm{s}_j(j=1,2$ 且 $j\neq i)$ 的理想准 KLT 基 $\boldsymbol{U}_{s_j}^{\mathrm{T}}$ 下的投影向量中的非零元素个数大于 1,即

$$\|\boldsymbol{U}_{s_i}^{\mathrm{T}}\boldsymbol{s}_j\|_0>1,i\neq j \tag{21}$$

在上述假设下,我们可推导得到以下另外一个定理.

定理 3 $\boldsymbol{U}_{s_1}^{\mathrm{T}}$ 和 $\boldsymbol{U}_{s_2}^{\mathrm{T}}$ 满足

$$\|\boldsymbol{U}_{s_1}^{\mathrm{T}}z\|_0+\|\boldsymbol{U}_{s_2}^{\mathrm{T}}(\boldsymbol{x}-z)\|_0>2,\forall\, z\neq\boldsymbol{s}_1 \tag{22}$$

其中 $\boldsymbol{x}$ 为 $\boldsymbol{s}_1$ 和 $\boldsymbol{s}_2$ 的单通道混合语音信号.

证明 采用反证法证明.假设存在 $z\neq\boldsymbol{s}_1$ 满足 $\|\boldsymbol{U}_{s_1}^{\mathrm{T}}z\|_1+\|\boldsymbol{U}_{s_2}^{\mathrm{T}}(\boldsymbol{x}-z)\|_0\leqslant 2$,那么只可能为下面三种情况之一:

$$\|\boldsymbol{U}_{s_1}^{\mathrm{T}}z\|_0=0,\|\boldsymbol{U}_{s_2}^{\mathrm{T}}(\boldsymbol{x}-z)\|_0\leqslant 2 \tag{23}$$

或
$$\|\boldsymbol{U}_{s_1}^{\mathrm{T}}z\|_0=1,\|\boldsymbol{U}_{s_2}^{\mathrm{T}}(\boldsymbol{x}-z)\|_0\leqslant 1 \tag{24}$$

或
$$\|\boldsymbol{U}_{s_1}^{\mathrm{T}}z\|_0=2,\|\boldsymbol{U}_{s_2}^{\mathrm{T}}(\boldsymbol{x}-z)\|_0=0 \tag{25}$$

首先证明情况(23)不成立.

假设情况(23)成立,则必有

$$z=0 \tag{26}$$

$$\boldsymbol{x}=\boldsymbol{s}_1+\boldsymbol{s}_2=d_1\boldsymbol{U}U_{s_2}(j_1)+d_2\boldsymbol{U}_{s_2}(j_2) \tag{27}$$

其中 $\boldsymbol{U}_{s_2}(j_1)$ 表示 $\boldsymbol{U}_{s_2}$ 的第 j_1 列,$\boldsymbol{U}_{s_2}(j_2)$ 表示 $\boldsymbol{U}_{s_2}$ 的第 j_2 列,$\boldsymbol{d}_1$ 和 $\boldsymbol{d}_2$ 为向量 $\boldsymbol{U}_{s_2}^{\mathrm{T}}x$ 中的非零元素值.

由于 $\boldsymbol{s}_1$ 和 $\boldsymbol{s}_2$ 相互独立,根据式(27)可得

$$\boldsymbol{s}_1=d_1\boldsymbol{U}_{s_2}(j_1)$$

$$\text{或 } \boldsymbol{s}_1=d_2\boldsymbol{U}_{s_2}(j_2) \tag{28}$$

否则,$\boldsymbol{s}_2$ 可由 $\boldsymbol{s}_1$ 线性表示.由式(28)可得

$$\|\boldsymbol{U}_{s_2}^{\mathrm{T}}\boldsymbol{s}_1\|_0=1 \tag{29}$$

这与式(21)相矛盾,故情况(23)不成立.

同理,可证明情况(24)和情况(25)也不成立.因此,不存在 $z\neq\boldsymbol{s}_1$ 满足 $\|\boldsymbol{U}_{s_1}^{\mathrm{T}}z\|_1+\|\boldsymbol{U}_{s_2}^{\mathrm{T}}(\boldsymbol{x}-z)\|_0\leqslant 2$,定理得证.

综合定理 1、定理 2 和定理 3,可得出下述定理.

定理 4 已知单通道混合语音信号 $\boldsymbol{x}=\sum_{i=1}^{2}\boldsymbol{s}_i$ 和源语音信号 $\boldsymbol{s}_1,\boldsymbol{s}_2$ 的理想准 KLT 基 $\boldsymbol{U}_{s_1}^{\mathrm{T}},\boldsymbol{U}_{s_2}^{\mathrm{T}}$,且源语音信号满足假设(21),则下述 l^0 – 范数优化问题

$$\begin{aligned}&\min f(\boldsymbol{\beta}_1,\boldsymbol{\beta}_2)=\min\sum_{i=1}^{2}\|\boldsymbol{\beta}_i\|_0,\\&\text{s.t.}\quad \boldsymbol{x}=\sum_{i=1}^{2}\boldsymbol{U}_i\boldsymbol{\beta}_i\end{aligned} \tag{30}$$

有唯一最优解 $\boldsymbol{\beta}^{\text{opt}}=[\boldsymbol{\Theta}_{s_1}\quad\boldsymbol{\Theta}_{s_2}]$,其中 $\boldsymbol{\Theta}_{s_i}$ 与式(16)中相同;基于问题(30)的最优解,由 $\boldsymbol{s}_i=\boldsymbol{U}_{s_i}\boldsymbol{\Theta}_{s_i}(i=1,2)$ 可从混合信号 $\boldsymbol{x}$ 中完美分离出源信号 $\boldsymbol{s}_1,\boldsymbol{s}_2$.

文献[6]证明,在一定条件下,类似式(30)的 l^0-范数问题可转化为 l^1-范数优化问题解决,

$$\begin{aligned}&\min f'(\boldsymbol{\beta}_1,\boldsymbol{\beta}_2)=\min\sum_{i=1}^{2}\|\boldsymbol{\beta}_i\|_1,\\&\text{s.t.}\quad\boldsymbol{x}=\sum_{i=1}^{2}\boldsymbol{U}_s\boldsymbol{\beta}_i\end{aligned}\tag{31}$$

由于式(30)为 NP-hard 问题,难以求解,所以在实际中我们通过求解式(31)获得 $\boldsymbol{\beta}^{\text{opt}}=[\boldsymbol{\Theta}_{s_1},\boldsymbol{\Theta}_{s_2}]$,进而重构源语音信号 $\boldsymbol{s}_1,\boldsymbol{s}_2$.式(31)可通过线性规划(linear programming,LP)方法方便求解.

3　基于 OMP 模板匹配准 KLT 基的单通道混合语音分离

虽然基于理想准 KLT 基可以实现单通道混合语音的完美分离,但是由于在实际进行单通道混合语音分离前,源语音信号未知,因此实际中难以精确获取源语音信号的理想准 KLT 基.为解决此问题,本节提出基于理想准 KLT 基的单通道混合语音分离理论思路,先以混合信号为已知条件,基于 OMP 算法,构造各源语音信号的 OMP 模板匹配准 KLT 基,再利用 l^0-范数优化算法来设计实用化的单通道混合语音分离算法.

由于理想准 KLT 基是通过对源语音信号的自相关矩阵进行特征值分解得到,我们考虑先对各源语音信号的自相关矩阵近似,再对近似的自相关矩阵进行特征值分解,来构造各源语音信号的 OMP 模板匹配准 KLT 基.在构造近似自相关矩阵时,我们计算若干训练语音信号的自相关矩阵作为模板,用这些模板的线性组合来构造近似自相关矩阵,其中模板的匹配系数及模板对应原子的选取,我们都基于 OMP 算法来求解.

为产生模板,我们预先对每个源说话人训练,各自构造两个字典.以源说话人 1 为例:记 $\boldsymbol{d}_{s_1}^k$ 为源说话人 1 的第 k 帧训练语音,计算 $\boldsymbol{d}_{s_1}^k$ 的自相关矩阵对角向量,记为 $\boldsymbol{g}_{s_1}^k$,即

$$\boldsymbol{g}_{s_1}^k=\text{diag}(\boldsymbol{d}_{s_1}^k(\boldsymbol{d}_{s_1}^k)^{\text{T}}),\quad k=1,2,\cdots,\text{N}_{1,\text{train}}\tag{32}$$

其中 $\text{N}_{1,\text{train}}$ 为源说话人 1 的训练语音帧数.将 $\boldsymbol{g}_{s_1}^k(k=1,2,\cdots,\text{N}_{1,\text{train}})$ 归一化,使其 l^2-范数均为 1,记为 $\tilde{\boldsymbol{g}}_{s_1}^k(k=1,2,\cdots,\text{N}_{1,\text{train}})$.将 $\tilde{\boldsymbol{g}}_{s_1}^k$ 作为原子构造字典 $\tilde{\boldsymbol{G}}_{s_1}=\{\tilde{\boldsymbol{g}}_{s_1}^k\}_{k=1}^{\text{N}_{1,\text{train}}}$.然后,将各训练语音帧 $\boldsymbol{d}_{s_1}^k$ 的各样点值进行修正,使得修正后的各训练语音帧 $\tilde{\boldsymbol{d}}_{s_1}^k$ 满足

$$\tilde{\boldsymbol{g}}_{s_1}^k=\text{diag}(\tilde{\boldsymbol{d}}_{s_1}^k(\tilde{\boldsymbol{d}}_{s_1}^k)^{\text{T}}),\quad k=1,2,\cdots,\text{N}_{1,\text{train}}\tag{33}$$

将 $\tilde{\boldsymbol{d}}_{s_1}^k$ 作为原子,构造另一个字典 $\tilde{\boldsymbol{D}}_{s_1}=\{\tilde{\boldsymbol{d}}_{s_1}^k\}_{k=1}^{\text{N}_{1,\text{train}}}$.采用类似的方法,对源说话人 2 训练构造字典 $\tilde{\boldsymbol{G}}_{s_2}=\{\tilde{\boldsymbol{g}}_{s_2}^k\}_{k=1}^{\text{N}_{2,\text{train}}}$ 和 $\tilde{\boldsymbol{D}}_{s_2}=\{\tilde{\boldsymbol{d}}_{s_2}^k\}_{k=1}^{\text{N}_{2,\text{train}}}$.

基于相同矩阵对角向量相同的事实,我们考虑用两自相关矩阵对角向量的匹配度来衡量这两个自相关矩阵的匹配度.由于在进行单通道混合语音分离时,各源语音信号未知,我们计算混合语音信号 x 的自相关矩阵对角向量 $\boldsymbol{d}_{R_x}$

$$\boldsymbol{d}_{R_x}=\text{diag}(\boldsymbol{R}_x)=\text{diag}(\boldsymbol{x}\boldsymbol{x}^{\text{T}})\tag{34}$$

作为参照,从 $\tilde{\boldsymbol{G}}_{s_1}=\{\tilde{\boldsymbol{g}}_{s_1}^k\}_{k=1}^{\text{N}_{1,\text{train}}}$ 和 $\tilde{\boldsymbol{G}}_{s_2}=\{\tilde{\boldsymbol{g}}_{s_2}^k\}_{k=1}^{\text{N}_{2,\text{train}}}$ 中选取合适的原子,以尽量近似各源语音信号自相关矩阵的对角向量.具体考虑如何从联合字典 $\tilde{\boldsymbol{G}}_{s_1\cup s_2}=[\tilde{\boldsymbol{G}}_{s_1}\quad\tilde{\boldsymbol{G}}_{s_2}]$ 中选取尽量少的原子,来尽可能地无损表示 $\boldsymbol{d}_{R_x}$.我们采用 OMP 算法[7~9]来迭代选取原子和计算原子匹配系数,下面详述具体步骤:

①输入:联合字典 $\tilde{\boldsymbol{G}}_{s_1\cup s_2}=[\tilde{\boldsymbol{G}}_{s_2}\quad\tilde{\boldsymbol{G}}_{s_2}]$,初始信号 $\boldsymbol{d}_{R_x}$

②输出:字典 $\tilde{\boldsymbol{G}}_{s_1\cup s_2}$ 的子集 $\tilde{\boldsymbol{G}}_{I_{s_1\cup s_2}}$(以选中的 $\tilde{\boldsymbol{G}}_{s_1\cup s_2}$ 中的原子为各列构成的矩阵),信号 d_{R_x} 在矩阵 $\tilde{\boldsymbol{G}}_{I_{s_1\cup s_2}}$ 下的稀疏表示 $\gamma_{I_{s_1\cup s_2}}$

③初始化:残差信号 $\boldsymbol{r}=\boldsymbol{d}_{R_x}$,$\boldsymbol{\gamma}_{I_{s_1\cup s_2}}=0$,$\boldsymbol{I}_{s_1\cup s_2}=[\quad]$,$\tilde{\boldsymbol{G}}_{I_{s_1\cup s_2}}=[\quad]$

④迭代:当残差信号 $\boldsymbol{r}$ 与 $\tilde{\boldsymbol{G}}_{s_1\cup s_2}$ 中余下的原子(即 $\tilde{\boldsymbol{G}}_{s_1\cup s_2}$ 中除去 $\tilde{\boldsymbol{G}}_{I_{s_1\cup s_2}}$ 中的原子)的最大相关系数不小于某阈值 ε 时,进行下列操作(⑤—⑧).

⑤将 $\tilde{\boldsymbol{G}}_{s_1\cup s_2}$ 记为 $\tilde{\boldsymbol{G}}_{s_1\cup s_2}=\{\boldsymbol{g}^k\}_{k=1}^{\text{N}_{1,\text{train}}+\text{N}_{2,\text{train}}}$

其中 $\boldsymbol{g}^k=\begin{cases}\boldsymbol{g}_{s_1}^k & k\leqslant\text{N}_{1,\text{train}}\\ \boldsymbol{g}_{s_2}^{k-\text{N}_{1,\text{train}}} & \text{N}_{1,\text{train}}<k\leqslant\text{N}_{1,\text{train}}+\text{N}_{2,\text{train}}\end{cases}$,

计算字典 $\tilde{\boldsymbol{G}}_{s_1\cup s_2}$ 中各原子与残差信号 $\boldsymbol{\beta}$ 的最大相关系数,得到最大相关系数对应的原子序号 $\hat{k}=\underset{k}{\text{argmax}}\,|(\boldsymbol{g}^k)^{\text{T}}\boldsymbol{r}|$.

⑥更新 $\tilde{\boldsymbol{G}}_{I_{s_1\cup s_2}}$:$\tilde{\boldsymbol{G}}_{I_{s_1\cup s_2}}=[\tilde{\boldsymbol{G}}_{I_{s_1\cup s_2}},\boldsymbol{g}^{\hat{k}}]$

⑦计算信号 $\boldsymbol{d}_{R_x}$ 在字典 $\tilde{\boldsymbol{G}}_{I_{s_1\cup s_2}}$ 下的稀疏表示 $\boldsymbol{\gamma}_{I_{s_1\cup s_2}}=(\tilde{\boldsymbol{G}}_{I_{s_1\cup s_2}}^{\text{T}}\tilde{\boldsymbol{G}}_{I_{s_1\cup s_2}})^{-1}\tilde{\boldsymbol{G}}_{I_{s_1\cup s_2}}^{\text{T}}\boldsymbol{d}_{R_x}$

⑧更新 $\boldsymbol{r}$:$\boldsymbol{r}=\boldsymbol{r}-\tilde{\boldsymbol{G}}_{I_{s_1\cup s_2}}\boldsymbol{\gamma}_{I_{s_1\cup s_2}}$

⑨迭代终止

从上述迭代过程可看出,$\tilde{\boldsymbol{G}}_{I_{s_1\cup s_2}}$ 中的原子按来源可分为两组,一组来自 $\tilde{\boldsymbol{G}}_{s_1}$,记为 $\tilde{\boldsymbol{g}}_{s_1}^{j_1},\tilde{\boldsymbol{g}}_{s_1}^{j_2},\cdots,\tilde{\boldsymbol{g}}_{s_1}^{j_{s_1'}}$;一组来自 $\tilde{\boldsymbol{G}}_{s_2}$,记为 $\tilde{\boldsymbol{g}}_{s_2}^{k_1},\tilde{\boldsymbol{g}}_{s_2}^{k_2},\cdots,\tilde{\boldsymbol{g}}_{s_2}^{k_{s_2'}}$.根据选取的这些原子,结合字典 $\tilde{\boldsymbol{D}}_{s_1}$ 和字典 $\tilde{\boldsymbol{D}}_{s_2}$,我们可构造用以近似各源语音信号自相关矩阵的模板.同时,我们将上述迭代过程中计算

得到的 $\boldsymbol{\gamma}_{I_{s_1}\cup s_2}$ 中各元素作为相应的模板匹配系数.基于所构造的模板和计算的模板匹配系数,我们可以构造各源语音信号的 OMP 模板匹配准 KLT 基,具体步骤如下:

① 从字典 $\widetilde{\boldsymbol{D}}_{s_1}=\{\widetilde{\boldsymbol{d}}_{s_1}^k\}_{k=1}^{\mathrm{N}_{1,\mathrm{train}}}$ 中选取与 $\widetilde{\boldsymbol{g}}_{s_1}^{j_1},\widetilde{\boldsymbol{g}}_{s_1}^{j_2},\cdots,\widetilde{\boldsymbol{g}}_{s_1}^{j_{n_1'}}$ 对应的原子,计算其自相关矩阵,作为模板 $\widetilde{\boldsymbol{R}}_{s_1}^{j_1},\widetilde{\boldsymbol{R}}_{s_1}^{j_2},\cdots,\widetilde{\boldsymbol{R}}_{s_1}^{j_{n_1'}}$;

$$\widetilde{\boldsymbol{R}}_{s_1}^{j_n}=\boldsymbol{d}_{s_1}^{k_n}(\boldsymbol{d}_{s_1}^{k_n})^{\mathrm{T}},\quad n=1,2,\cdots,n_1' \tag{35}$$

② 从字典 $\widetilde{\boldsymbol{D}}_{s_2}=\{\widetilde{\boldsymbol{d}}_{s_2}^k\}_{k=1}^{\mathrm{N}_{2,\mathrm{train}}}$ 中选取与 $\widetilde{\boldsymbol{g}}_{s_2}^{k_1},\widetilde{\boldsymbol{g}}_{s_2}^{k_2},\cdots,\widetilde{\boldsymbol{g}}_{s_2}^{k_{n_2'}}$ 对应的原子,计算其自相关矩阵,作为模板 $\widetilde{\boldsymbol{R}}_{s_2}^{k_1},\widetilde{\boldsymbol{R}}_{s_2}^{k_2},\cdots,\widetilde{\boldsymbol{R}}_{s_2}^{k_{n_2'}}$,

$$\widetilde{\boldsymbol{R}}_{s_2}^{j_n}=\boldsymbol{d}_{s_2}^{k_n}(\boldsymbol{d}_{s_2}^{k_n})^{\mathrm{T}},\quad n=1,2,\cdots,n_2' \tag{36}$$

③根据生成的模板及匹配系数,计算各源语音信号的近似自相关矩阵 $\hat{\boldsymbol{R}}_{s_i}$.为方便表述,将 $\boldsymbol{\gamma}_{I_{s_1}\cup s_2}$ 中与 $\widetilde{\boldsymbol{g}}_{s_1}^{j_1},\widetilde{\boldsymbol{g}}_{s_1}^{j_2},\cdots,\widetilde{\boldsymbol{g}}_{s_1}^{j_{n_1'}}$ 相对应的系数记为 $\gamma_{j_1},\gamma_{j_2},\cdots,\gamma_{j_{n_1'}}$,与 $\widetilde{\boldsymbol{g}}_{s_2}^{k_1},\widetilde{\boldsymbol{g}}_{s_2}^{k_2},\cdots,\widetilde{\boldsymbol{g}}_{s_2}^{k_{n_2'}}$ 相对应的系数记为 $\gamma_{j_{n_1'+1}},\gamma_{j_{n_1'+2}},\cdots,\gamma_{j_{n_1'+n_2'}}$.

$$\hat{\boldsymbol{R}}_{s_1}=\sum_{i=1}^{n_1'}\gamma_{j_i}\widetilde{\boldsymbol{R}}_{s_1}^{j_i} \tag{37}$$

$$\hat{\boldsymbol{R}}_{s_2}=\sum_{i=1}^{n_2'}\gamma_{j_{n_1'+i}}\widetilde{\boldsymbol{R}}_{s_2}^{k_i} \tag{38}$$

④对 $\hat{\boldsymbol{R}}_{s_i}$ 进行特征值分解,求取 $\boldsymbol{s}_i$ 的 OMP 模板匹配准 KLT 基 $\widetilde{\boldsymbol{U}}_{s_i}^{\mathrm{T}}$ 和 OMP 模板匹配准 IKLT 基 $\widetilde{\boldsymbol{U}}_{s_i}$,

$$\hat{\boldsymbol{R}}_{s_i}=\widetilde{\boldsymbol{U}}_{s_i}^{\mathrm{T}}\widetilde{\boldsymbol{\Lambda}}_{s_i}\widetilde{\boldsymbol{U}}_{s_i} \tag{39}$$

其中,$\widetilde{\boldsymbol{U}}_{s_i}^{\mathrm{T}}$ 和 $\widetilde{\boldsymbol{U}}_{s_i}$ 均为正交阵,$\widetilde{\boldsymbol{\Lambda}}_{s_i}$ 为对角阵.

基于所构造的 OMP 模板匹配准 KLT 基,采用 l^0-范数优化算法,进行单通道混合语音分离.具体通过解 l^0-范数优化问题

$$\begin{aligned}&\min f_1(\widetilde{\boldsymbol{\chi}}_1,\widetilde{\boldsymbol{\chi}}_2)=\min\sum_{i=1}^{2}\|\widetilde{\boldsymbol{\chi}}_i\|_0,\\&\text{s.t.}\quad \boldsymbol{x}=\sum_{i=1}^{2}\widetilde{\boldsymbol{U}}_{s_i}\widetilde{\boldsymbol{\chi}}_i\end{aligned} \tag{40}$$

得到 $\boldsymbol{s}_i$ 在 $\widetilde{\boldsymbol{U}}_{s_i}$ 下稀疏表示的估计 $\hat{\boldsymbol{\Theta}}_{s_i}$,其中 $\boldsymbol{\chi}^{\mathrm{opt}}=[\hat{\boldsymbol{\Theta}}_{s_1}\quad\hat{\boldsymbol{\Theta}}_{s_2}]$ 为问题(40)的最优解.同样,将问题(40)转化为对应的 l^1-范数优化问题来求解.最后通过下式重构得到分离语音信号 $\hat{\boldsymbol{s}}_i$,

$$\hat{\boldsymbol{s}}_i=\widetilde{\boldsymbol{U}}_{s_i}\hat{\boldsymbol{\Theta}}_{s_i},\quad i=1,2 \tag{41}$$

4 实验结果及分析

本论文采用的数据库为对称的语音库,一共包含四位说话人(两男两女)的语音,每位说话人的发音内容是一样的,且朗读方式也尽量一样.语句主要为短语结构的汉语,覆盖了大部分汉语音节.所有语音在信噪比不低于 30dB 的实验环境下录音,采样率为 16kHz.

首先我们通过实验来验证定理 4 的正确性.每位说话人随机选取 5 句汉语短语结构的语音作为测试数据,单通道混合语音由两个源测试语音叠加获取(如式(2)所示).我们对不同输入信噪比条件下的混合语音进行实验:调节各说话人测试语音语句的能量,将不同说话人的各测试语音语句两两混合,得到输入信噪比为 -10dB,-5dB,0dB,5dB,10dB 的男男混合语音各 25 句,女女混合语音各 25 句,男女混合语音各 100 句.(将一说话人的语音作为信号,另一说话人的语音作为噪声).语音信号采用矩形窗分帧,每帧 160 样点(10ms),帧间无重叠.对混合语音分帧,逐帧进行分离,再将各分离语音帧顺序连接,得到分离后的语音信号.我们用两说话人的平均语句改进信噪比(improved signal to noise ratio,ISNR)$\mathrm{ISNR}_{\mathrm{av}}$,来衡量不同语句输入信噪比下的单通道混合语音分离性能,结果如表 1 所示.$\mathrm{ISNR}_{\mathrm{av}}$ 定义为

$$\mathrm{ISNR}_{\mathrm{av}}=\frac{1}{2}\sum_{i=1}^{2}\mathrm{ISNR}_i \tag{42}$$

$$\mathrm{ISNR}_i=\frac{1}{K}\sum_{k=1}^{K}\left\{\begin{array}{l}10\lg\left[\dfrac{(\boldsymbol{r}_i^k)^{\mathrm{T}}\boldsymbol{r}_i^k}{(\boldsymbol{r}_i^k-\hat{\boldsymbol{r}}_i^k)^{\mathrm{T}}(\boldsymbol{r}_i^k-\hat{\boldsymbol{r}}_i^k)}\right]\\-10\lg\left[\dfrac{(\boldsymbol{r}_i^k)^{\mathrm{T}}\boldsymbol{r}_i^k}{(\boldsymbol{r}_{\mathrm{mix}}^k-\hat{\boldsymbol{r}}_i^k)^{\mathrm{T}}(\boldsymbol{r}_{\mathrm{mix}}^k-\hat{\boldsymbol{r}}_i^k)}\right]\end{array}\right\} \tag{43}$$

其中 ISNR_i 为第 i 个说话人的 ISNR,$\boldsymbol{r}_{\mathrm{mix}}^k$ 为由 $\boldsymbol{r}_1^k$ 和 $\boldsymbol{r}_2^k$ 叠加得到的混合语音信号.

表 1 基于理想准 KLT 基的分离性能—$\mathrm{ISNR}_{\mathrm{av}}$

语句输入	两说话人的平均语句改进信噪比 $\mathrm{ISNR}_{\mathrm{av}}$(dB)		
SNR(dB)	男男	男女	女女
-10	467.2	473.2	472.7
-5	467.8	477.3	474.0
0	471.3	472.8	473.9
5	468.7	473.2	473.1
10	468.0	472.8	471.3

从表 1 可以看出,无论源语音信号以何种信噪比进行混合,基于理想准 KLT 基的分离算法,得到的两说话人的平均语句 $\mathrm{ISNR}_{\mathrm{av}}$ 均远远大于 50dB,可认为源语音信号被完美重构.因此,我们可以得出结论:基于源语音信号的理想准 KLT 基,采用 l^0 - 范数优化算法可以实现单通道混合语音的完美分离,即定理 4 成立.

下面通过仿真实验来分析基于 OMP 模板匹配准 KLT 基的单通道混合语音分离性能.所用语音库同上.对四位说话人分别进行训练,每位男性说话人各取 250 句汉语短语结构的语音作为训练数据,每位女性说话人各取 300 句汉语短语结构的语音作为训练数据.每位说话人随机选取 5 句与训练语音不同的汉语短语结构的语音作为测试数据.训练语音和测试语音均采用矩

形窗分帧,帧长为 160 样点(10ms),训练语音帧间重叠 120 样点,测试语音帧间无重叠.我们用同样的方法得到输入信噪比为 -10dB, -5dB,0dB,5dB,10dB 的男男混合语音各 25 句,女女混合语音各 25 句,男女混合语音各 100 句,对其进行实验.从构造 OMP 模板匹配准 KLT 基的步骤中,我们可以看出,构造的 OMP 模板匹配准 KLT 基与设置的最大相关系数阈值密切相关.故我们首先衡量不同最大相关系数阈值下的单通道混合语音分离性能,最大相关系数阈值 ε 分别设为 10^{-1},5×10^{-2},10^{-2},10^{-3}.以 $ISNR_{av}$为指标衡量算法的分离性能,结果如表 2 示.

表 2　基于 OMP 模板匹配准 KLT 基的分离性能—$ISNR_{av}$

语句输入 SNR(dB)	两说话人的平均语句改进信噪比 $ISNR_{av}$(dB)(男男/男女/女女)			
	最大相关系数阈值为 10^{-1}	最大相关系数阈值为 5×10^{-2}	最大相关系数阈值为 10^{-2}	最大相关系数阈值为 10^{-3}
-10	1.7/4.4/2.2	1.7/4.5/2.3	1.8/4.6/2.4	1.8/4.7/2.4
-5	1.6/4.5/2.4	1.6/4.6/2.5	1.7/4.7/2.6	1.7/4.8/2.7
0	1.3/4.5/2.6	1.4/4.6/2.7	1.4/4.8/2.9	1.4/4.7/3.0
5	0.9/4.5/2.7	1.0/4.6/2.9	1.0/4.7/2.9	1.0/4.8/3.1
10	0.3/4.4/2.8	0.3/4.5/2.9	0.4/4.6/3.0	0.4/4.6/3.0

从表 2 可看出,无论输入信噪比如何,无论混合语音类型如何,采用基于 OMP 模板匹配准 KLT 基的分离算法都能获得一定的分离性能,同样地,当混合语音为男女混合语音时,平均语句改进信噪比最大,这是由于男女语音差别最大造成的.从表 2 还可看出,随着最大相关系数阈值的减小,平均语句改进信噪比随之增大,但当最大相关系数阈值减小到一定程度后,平均语句改进信噪比的变化不再明显.由于最大相关系数阈值越小,构造源语音信号的 OMP 模板匹配准 KLT 基所需的迭代次数越多,运行时间越长,故综合考虑分离性能和算法的运行时间,选取最大相关系数阈值 $\varepsilon=10^{-2}$.

为了说明基于 OMP 模板匹配准 KLT 基的分离算法的有效性,我们将选取最大相关系数阈值 $\varepsilon=10^{-2}$时的男/女混合语音分离效果,与文献[10]中基于独立成分分析(independent component analysis,ICA)基的单通道男女混合语音分离效果,文献[11]中的采用形分析技术改进后的计算机场景分析(computational auditory scene analysis,CASA)方法得到的单通道男女混合语音分离效果进行比较,如表 3 所示,比较指标为平均语句改进信噪比 $ISNR_{av}$.

表 3　基于 OMP 模板匹配准 KLT 基的分离性能和其它几种分离算法的性能比较

男女说话人的平均语句改进信噪比 $ISNR_{av}$(dB)		
基于 OMP 模板匹配准 KLT 基的分离算法	基于 ICA 基的分离算法	采用形分析技术改进后的 CASA 算法
4.7	2.6	4.2

从表 3 可以看出,三种算法相比较,论文提出的基于 OMP 模板匹配准 KLT 基的分离算法的分离效果最佳.同样,为了衡量分离语音的主观听觉质量,我们用 P.862 标准分别对上述输入信噪比为 0dB,最大相关系数阈值 $\varepsilon=10^{-2}$时的分离语音质量进行评估,见表 4.表中数值为各说话人的平均语句的平均意见评分(mean opinion score,MOS).值得注意的是,P.862 标准测试所得的 MOS 分范围为 -0.5~4.5.

表 4　基于 OMP 模板匹配准 KLT 基的分离性能—平均语句 MOS 分

混合语音类型	平均语句 MOS 分	
	信号 1	信号 2
男男	1.4	1.4
男女	1.5	1.4
女女	1.0	0.9

最后,我们给出输入信噪比为 0dB,最大相关系数阈值 $\varepsilon=10^{-2}$时,采用基于 OMP 模板匹配准 KLT 基的分离算法获得的分离语音波形,如图 1 所示.图 1 中的混合语音信号由男声语音信号和女声语音信号线性叠加得到.

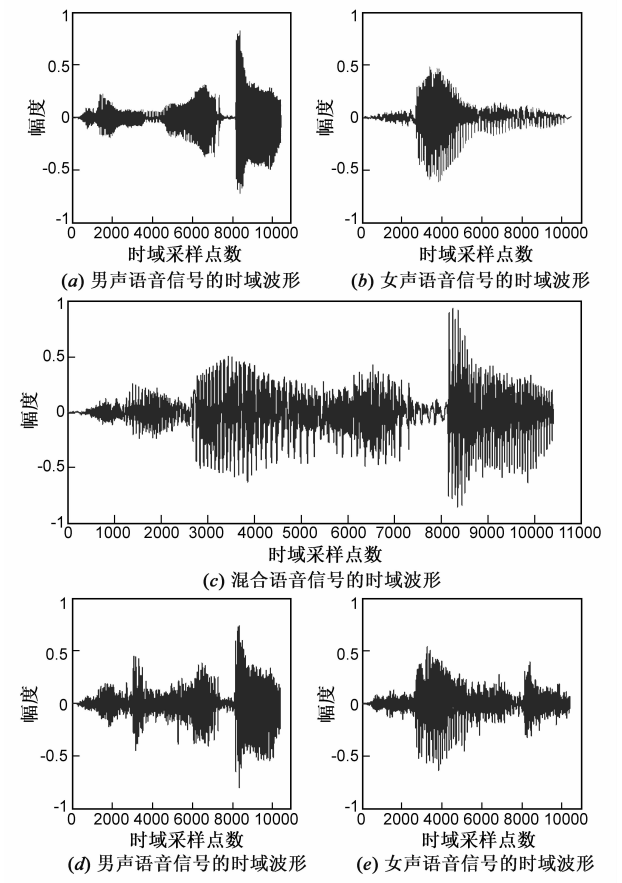

图1　基于OMP模板匹配准KLT基的分离语音波形

5　结论

本论文在构造新的语音信号稀疏基—理想准 KLT 基和 OMP 模板匹配准 KLT 基的基础上,提出了一种新

的基于稀疏分解的单通道混合语音分离方法.论文首先定义和构造了理想准 KLT 基,并在此基础上,从理论上提出且证明了基于理想准 KLT 基和 l_0-范数优化算法,可实现单通道混合语音的完美分离.此理论证明分两步进行:首先,证明了当基函数满足一定条件时,基于 l_0-范数优化算法可实现单通道混合语音的完美分离;然后,证明了理想准 KLT 基满足上述完美分离单通道混合语音的条件.鉴于在实际语音分离时,无法获得确切的理想准 KLT 基,论文提出了采用模板匹配的思路,基于 OMP 算法同步选取原子和计算原子的匹配系数,来构造 OMP 模板匹配准 KLT 基.在此基础上,提出了基于 OMP 模板匹配准 KLT 基的单通道混合语音分离算法.仿真实验表明,此算法的分离性能优于已有的两种经典算法性能.

参考文献

[1] Schmidt M N, Olsson R K. Linear regression on sparse features for single-channel speech separation[A]. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics[C]. NY, USA, 2007. 26 – 29.

[2] Pearlmutter B, Olsson R. Linear program differentiation for single – channel speech separation[A]. 16th IEEE Signal Processing Society Workshop on Machine Learning for Signal Processing[C]. Maynooth, Ireland, 2006. 421 – 426.

[3] Nakashizuka N, Okumura H, Iiguni Y. Single-channel speech separation by using a sparse decomposition with periodic structure[A]. 2008 International Symposium on Intelligent Signal Processing and Communications Systems[C]. Bangkok, Thailand, 2008. 1 – 4.

[4] Elad M, Bruckstein A. A generalized uncertainty principle and sparse representation in pairs of bases[J]. IEEE Transactions on Information Theory, 2002, 48: 2558 – 2567.

[5] 郭海燕,杨震.基于近似 KLT 的语音信号压缩感知[J].电子与信息学报,2009,31(12):2948 – 2952.
Guo Hai-yan, Yang Zhen. Compressed speech signal sensing based on approximate KLT[J]. Journal of Electronics & Information Technology. 2009, 31(12): 2948 – 2952. (in Chinese)

[6] Donoho D L. For Most Large Underdetermined Systems of Equations, the Minimal l_1 – norm Near – Solution Approximates the Sparsest Near-Solution[R]. http://www-stat.stanford.edu/~donoho/Reports/2004.

[7] Y Pati, R Rezaiifar, and P Krishnaprasad. Orthogonal matching pursuit: recursive function approximation with applications to wavelet decomposition[A]. Proceedings of 27th Annual Asilomar Conference on Signals, Systems and Computers[C]. CA, USA, 1993. 40 – 44.

[8] 石光明,刘丹华,高大化,刘哲,林杰,王良君.压缩感知理论及其研究进展[J].电子学报,2010,37(5): 1070 – 1081.
Shi Guang-ming, Liu Dan-hua, Gao Da-hua, Liu Zhe, Lin Jie, Wang Liang-jun. Advances in theory and application of compressed sensing[J]. Acta Electronica Sinica, 2010, 37(5): 1070 – 1081. (in Chinese)

[9] 孙玉宝,吴敏,韦志辉,肖亮,冯灿.基于稀疏表示的脑电棘波检测算法研究[J].电子学报,2009,37(9): 1971 – 1976.
Sun Yu-bao, Wu Min, Wei Zhi-hui, Xiao Liang, Feng Can. EEG spike detection using sparse representation[J]. Acta Electronica Sinica, 2010, 37(9): 1971 – 1976. (in Chinese)

[10] Gil-Jin Jang, Te-Won Lee, Yung-Hwan Oh. Single-channel signal separation using time-domain basis functions[J]. IEEE Signal Processing Letters, 2003, 10(6): 168 – 171.

[11] Yun-Kyung Lee, Oh-Wook Kwon. Application of shape analysis techniques for improved CASA-based speech separation [J]. IEEE Transactions on Consumer Electronics, 2009, 55 (1): 146 – 149.

作者简介

郭海燕 女,1983 年 10 月出生于湖北钟祥.2011 年于南京邮电大学获得博士学位.现为南京农业大学讲师.研究方向为语音盲源分离和压缩感知等.
E-mail: guohaiyan198311@163.com

杨 震 男,1961 年 11 月出生于江苏苏州.现为南京邮电大学教授、博士生导师.研究方向为无线通信与网络信号处理、语音处理与现代语音通信和信息安全技术.