

## 基于小波变换和独立分量分析的含噪混叠语音盲分离

赵彩华 刘 璐 孙建德 闫 华  
(山东大学信息科学与工程学院 济南 250100)

**摘 要** 含噪混叠语音的分离是语音信号处理中的重要研究问题。该文针对语音信号的非平稳特性与不同语音源之间的相互独立性,提出用小波变换与独立分量分析相结合的方法来进行分离。首先利用小波变换分别对各含噪混叠语音进行消噪,然后用独立分量分析的方法对消噪后的混叠信号进行分离,最后进一步对分离信号作矢量归一和再消噪处理,得到各个语音源信号的最终估计。仿真结果表明这种方法取得了很好的分离效果。

**关键词** 语音分离, 小波变换, 独立分量分析, 噪声消除

中图分类号: TN912.3

文献标识码: A

文章编号: 1009-5896(2006)09-1565-04

## Blind Separation of Noisy Speech Mixtures Based on Wavelet Transform and Independent Component Analysis

Zhao Cai-hua Liu Ju Sun Jian-de Yan Hua  
(School of Information Science and Engineering, Shandong University, Jinan 250100, China)

**Abstract** A vital issue in speech processing is to extract source speeches from noisy mixtures. A method is presented based on wavelet transform and independent component analysis in this paper. Firstly, de-noise the noisy mixtures with discrete wavelet transform. Secondly, get them separated by independent component analysis. Finally, do the post-processing to the separated signals, then the estimated source speeches are got. Simulation results exhibit a high level of separating performance.

**Key words** Speech separation, Wavelet transform, Independent Component Analysis(ICA), De-noising

### 1 引言

多个说话者的混叠语音分离是语音处理领域中的重要研究方向。目前的混叠语音分离主要以盲源分离和计算声场景分析两类方法为主<sup>[1]</sup>,前者是根据信号的统计特性从几个观测信号中恢复出未知的独立源成分<sup>[2]</sup>;后者则是利用人耳的听觉感知要素从混叠语音中分离出感兴趣的目标语音<sup>[3]</sup>。通常为了研究方便,这些方法大多都不考虑环境噪声。然而,在实际语音通信中不可避免会受到周围噪声的影响,因此寻找有效的含噪混叠语音分离方法具有非常重要的理论价值和实际意义。

含噪混叠语音包含了多个说话者和环境噪声,其分离较为困难。目前,一些学者正致力于含噪盲源分离算法的研究,但总体研究成果不多<sup>[4,5]</sup>;而噪声环境中的感知要素检测困难使计算声场景分析方法受到一定的局限性<sup>[6,7]</sup>。国内外现有的关于含噪混叠语音分离的研究成果非常少。

独立分量分析(Independent Component Analysis, ICA)是一种新兴的盲源分离技术,近年来已成为信号处理领域的热点研究问题。ICA有效的分离作用使其在语音处理、生物医学信号处理、图像处理、电子通讯等方面都有着非常重要的应用,但目前的ICA理论算法大多忽略噪声的影响。本文根

据语音信号的非平稳特性和不同语音源之间的相互独立性,提出了基于小波变换和独立分量分析的含噪混叠语音分离方法。文中首先采用小波变换对各个含噪混叠语音进行消噪处理,然后用独立分量分析的方法对消噪后的混叠语音进行分离,最后根据分离信号的特点进一步提出对其进行矢量归一和再消噪处理,最终得到各个语音源信号的估计。计算机仿真结果表明本文所提出的方法取得了很好的含噪混叠语音分离效果。

### 2 问题描述

独立分量分析(ICA),是指在输入源信号和传输信道参数未知时,根据输入源信号的统计特性,仅由观测信号恢复出源信号的各个独立成分的过程<sup>[5]</sup>。ICA的应用要求:(1)观测信号的数目大于等于源信号的数目;(2)各个源信号之间相互统计独立;(3)至多有一个源信号是高斯的。由于不同的语音源之间具有相互独立性,并且语音信号具有超高斯分布的特点,因此可用ICA方法进行有效分离。

#### 2.1 含噪混叠语音信号模型

噪声环境下传感器接收的线性瞬时混叠语音系统模型为

$$\mathbf{x} = \mathbf{A}\mathbf{s} + \mathbf{N} \quad (1)$$

式中,  $\mathbf{x} = [x_1(t), x_2(t), \dots, x_m(t)]^T$  为  $m$  维观测矢量,  $\mathbf{s} = [s_1(t), s_2(t), \dots, s_n(t)]^T$  为  $n$  维未知语音源信号,  $\mathbf{A}$  为  $m \times n$  维未知混叠矩阵,  $\mathbf{N} = [n_1(t), n_2(t), \dots, n_m(t)]^T$  为  $m$  维

2004-12-13 收到, 2005-05-19 改回  
国家自然科学基金(30000041)和教育部留学回国人员科研启动基金  
(2005)55)资助课题

加性高斯白噪声,一般设  $m=n$ 。每个观测信号  $x_i(t)$  的语音成分都是由  $n$  个未知源信号  $s_i(t)$  的瞬时混叠组成。我们的目的就是要从  $m$  个观测信号  $x_i(t)$  中提取出各个未知源  $s_i(t)$  来。

## 2.2 独立分量分析理论算法

近年来,独立分量分析理论算法取得了许多重要的发展:Cardoso提出了基于高阶统计的联合对角化方法;Sejnowski和Bell基于信息论,通过输出输入之间熵最大得到了Infomax算法;Amari和Cichocki基于信息理论中的概率密度的Cram-Charlier展开,提出了最小互信息算法;Hyvarinen和Oja基于近似负熵原理给出了一种快速定点ICA(Fast ICA)算法;Lee提出了扩展Infomax算法。目前的这些算法大多不考虑环境噪声的影响,即式(1)中不含噪声  $N$ ,此时设观测信号为  $\mathbf{y}$ ,则

$$\mathbf{y} = \mathbf{A}\mathbf{s} \quad (2)$$

ICA的目的就是要找到一个分离矩阵  $\mathbf{W}$ ,使  $\mathbf{W} = \mathbf{A}^{-1}$ ,然后由  $\mathbf{s} = \mathbf{W}\mathbf{y}$  得到各个独立源信号。

ICA分离之前需要对  $\mathbf{y}$  进行预处理:去均值和预白化。去均值使  $\mathbf{y}$  成为零均值矢量,预白化则用以去除观测信号之间的相关性,得到具有单位方差的信号。白化过程采用对协方差矩阵进行特征值分解的方法来进行<sup>[4]</sup>:

$$E\left\{(\mathbf{y} - E\{\mathbf{y}\})(\mathbf{y} - E\{\mathbf{y}\})^T\right\} = \mathbf{D}\mathbf{A}\mathbf{D}^T \quad (3)$$

$\mathbf{A}$  为特征值对角阵,  $\mathbf{D}$  为相应的特征向量矩阵,则白化向量为

$$\mathbf{z} = \mathbf{A}^{-\frac{1}{2}}\mathbf{D}^T(\mathbf{y} - E\{\mathbf{y}\}) \quad (4)$$

Fast ICA 算法<sup>[8]</sup>具有计算简单、收敛速度快等突出优点,本文重点介绍 FastICA 算法。定义  $g(\cdot) = G'(\cdot)$ ,  $G$  为任意非二次函数,  $\mathbf{w}$  为行矢量,则基本的 FastICA 算法步骤如下:

- (1) 随机选择初始权矢量  $\mathbf{w}$ ;
- (2) 令  $\mathbf{w}^+ = E\{y g(\mathbf{w}\mathbf{y})\} - E\{g'(\mathbf{w}\mathbf{y})\}\mathbf{w}$ ;
- (3) 令  $\mathbf{w} = \mathbf{w}^+ / \|\mathbf{w}^+\|$ ;
- (4) 若不收敛,返回(2)。

若要估计  $n$  个独立分量,则需要运行  $n$  次上述基本Fast ICA算法,过程中每分离出一个独立分量,就要从观测信号中减去这一分量,为了防止收敛相同,在每次迭代后都要对输出  $\mathbf{w}_1\mathbf{y}, \dots, \mathbf{w}_n\mathbf{y}$  进行去相关,如此反复,直到分离出所有的独立分量。

## 3 含噪混叠语音盲分离

以上独立分量分析的方法多用于解决无噪环境或噪声能量很低时的混叠问题,一般能够较好地分离,但是当噪声影响较大时,分离效果则不好,因此我们提出了基于小波变换和独立分量分析的含噪混叠语音盲分离方案。

### 3.1 含噪混叠语音盲分离方案

本文实验中对两个观测信号进行分离,即

$$\begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} = \mathbf{A} \begin{bmatrix} s_1(t) \\ s_2(t) \end{bmatrix} + \begin{bmatrix} n_1(t) \\ n_2(t) \end{bmatrix} \quad (5)$$

首先利用小波变换对含噪混叠语音信号  $x_1, x_2$  进行消噪,得到尽可能“纯净”的混叠语音  $y_1, y_2$ ,所谓“纯净”是指没有噪声干扰的完全语音信号。然后对  $y_1, y_2$  进行去值和预白化,经 FastICA 分离后,得到独立语音分量  $ic_1, ic_2$ 。由于 ICA 问题中存在着幅值和顺序不确定性<sup>[4]</sup>,其中顺序的不确定性对分离结果没有影响,可以不予考虑,而幅值的不确定性使信号的能量变化较大,因此需要对幅值进行调整,本文采用矢量归一的方法进行幅值调整,即令 FastICA 分离信号:

$$ic_i = \frac{ic_i}{\|ic_i\|}, \quad i=1,2 \quad (6)$$

在以上小波消噪和 FastICA 分离过程中必须要注意:为了从  $y_1, y_2$  中分离出较好的语音源信号,在对  $x_1, x_2$  消噪时不能太彻底,以尽可能地不损失观测信号中的有用语音成分,否则 ICA 分离出来的语音信号就会严重失真,从而语音质量下降,语音识别率降低。然而这样带来的另一个问题就是:由于观测信号中的噪声消除不彻底,在 ICA 分离信号中会存在一些明显的残留噪声。针对这一情况,本文进一步提出了对 ICA 分离信号作再消噪处理,即再次利用小波变换的方法对矢量归一后的分离语音进行消噪,最终得到更加清晰的估计信号  $\hat{s}_1, \hat{s}_2$ 。具体分离框图如图 1 所示:

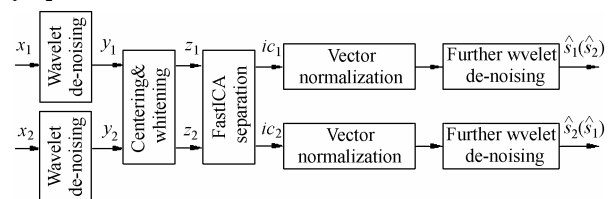


图 1 基于小波变换和独立分量分析的含噪混叠语音盲分离框图

Fig.1 Flow chart of blind speech separation based on wavelet transform and ICA

### 3.2 小波消噪处理

在我们提出的含噪混叠语音盲分离方案中,有先后两次利用小波变换方法进行噪声消除:第 1 次是为了从含噪语音中得到尽可能“纯净”的混叠语音;第 2 次则是为了去除分离信号中的残留噪声。含噪信号经多次小波分解以后,噪声和一些高频细节分布在小波系数上,对每层小波细节都进行阈值处理,可以较大程度地抑制噪声<sup>[9]</sup>。然后将分解信号经小波逆变换进行重构,便可得到消噪后的信号。

在具体应用时,我们选择常用的软阈值处理方法对观测信号进行去噪,阈值规则采用 SURE 无偏估计法,取正交小波“sym8”作为实验中的最优小波基。为得到更合适的阈值大小,在每个分解层都重新估计噪声,以调整阈值。另外,为了保留混叠信号中的有用语音成分,设对观测信号进行  $L_1$  次小波分解;为了去除 ICA 分离信号中的残留噪声,设在再消噪处理时进行  $L_2$  次分解。 $L_1$  和  $L_2$  的取值对分离结果的影

响非常大, 尤其是  $L_1$ , 它直接影响着分离语音信号的质量。针对这一情况, 在后面的实验仿真中, 我们对不同的  $L_1$  和  $L_2$  下的分离结果进行了具体的分析比较, 以获取合适的前后两次小波分解次数。

### 3.3 FastICA 分离

我们采用FastICA算法对去噪后经预处理的混叠语音进行分离。FastICA分离中, 合适的函数  $G$  可以使估计结果更好, 由于语音信号具有超高斯分布的特点, 且鲁棒性要求较高, 根据文献[8]中对  $G$  的分析, 我们选择:

$$G(u) = -\frac{1}{a} \exp(-au^2/2) \quad (7)$$

相应地

$$g(u) = u \exp(-au^2/2) \quad (8)$$

式中取  $a=1$ 。迭代过程采用 Gram-Schmidt 标准正交化方法进行去相关<sup>[8]</sup>, 逐个求出  $w_1, \dots, w_n$ , 从而得到各个独立语音源信号。

## 4 实验仿真及结果分析

实验对人工混叠的观测信号进行分离。原始语音  $s_1, s_2$  如图 2(a)所示(来自 TIMIT 语音数据库), 采样频率为 8kHz, 字长为 16bit。随机产生  $1 \times 2$  的混叠矩阵对  $s_1, s_2$  进行混叠, 并加入一定幅度的高斯白噪声, 得到一般环境中的模拟观测信号  $x_1, x_2$  如图 2(b)所示。考虑在同一背景下噪声分布基本相同, 实验中设所加噪声  $n_1 = n_2$ 。按照图 1 所示的流程, 对  $x_1, x_2$  逐步进行处理, 最终得到各个语音源信号的估计  $\hat{s}_1, \hat{s}_2$ 。

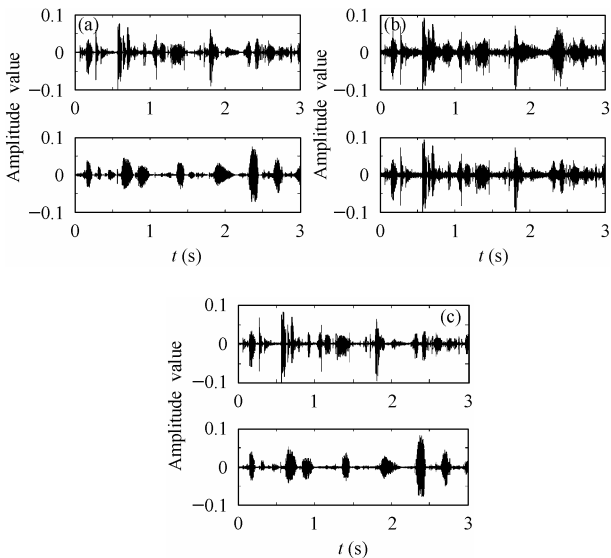


图2 一般噪声环境下的语音分离结果

(a)源语音信号  $s_1, s_2$  (b)混叠加噪信号  $x_1, x_2$  (c)分离信号  $\hat{s}_1, \hat{s}_2$

Fig.2 Speech separation result in general noisy environment

(a) Source speeches  $s_1, s_2$  (b) Noisy speech mixtures  $x_1, x_2$

(c) Separated speeches  $\hat{s}_1, \hat{s}_2$

我们用分离信号的信噪比 SNR 和原始信号与估计信号之间的均方根误差 RMSE 来评价分离结果的性能。当得到的

分离信号的信噪比越高和均方根误差越小时, 分离效果越好。

用下式表示分离信号的信噪比:

$$SNR_i = 10 \lg \frac{\sum_t s_i^2(t)}{\sum_t (s_i(t) - \hat{s}_i(t))^2} (\text{dB}), \quad i=1,2 \quad (9)$$

原始信号与估计信号之间的均方根误差为

$$RMSE_i = \sqrt{\frac{1}{n} \sum_{t=1}^n [s_i(t) - \hat{s}_i(t)]^2}, \quad i=1,2 \quad (10)$$

首先来看一下  $SNR_i (i=1,2)$  随前后两次小波分解次数  $L_1$  和  $L_2$  的变化关系。 $L_1$  的取值对分离结果有着至关重要的影响,  $L_1$  太小则影响ICA分离, 太大则容易语音失真; 而  $L_2$  的变化对分离结果的影响基本成正比, 但  $L_2$  过大会使信号严重失真。图3(a)所示为同一  $L_2$  下  $SNR_i$  随  $L_1$  的变化曲线, 由图可见, 当  $L_1=2$  时, 两分离信号的信噪比最大。 $L_1=2$  时  $SNR_i$  随  $L_2$  的变化如图3(b), 可见再消噪时5次小波分解后信噪比的提高已不明显。因此我们取  $L_1=2, L_2=5$  分别进行前后两次小波消噪, 此时得到的估计信号  $\hat{s}_1, \hat{s}_2$  如图2(c)所示。

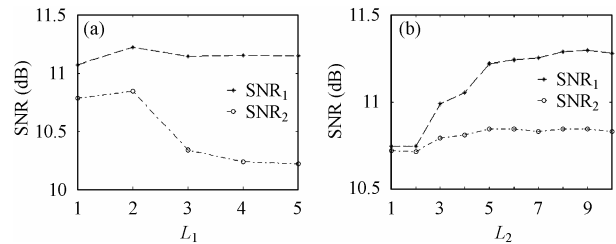


图3  $SNR_i$  随前后两次小波分解次数  $L_1$  和  $L_2$  的变化曲线

(a)  $SNR_i$  随  $L_1$  的变化曲线 (b)  $SNR_i$  随  $L_2$  的变化曲线

Fig.3 Varying curve of  $SNR_i$  respectively with  $L_1$  and  $L_2$

(a) Varying curve of  $SNR_i$  with  $L_1$  (b) Varying curve of  $SNR_i$  with  $L_2$

针对各种不同的情况, 我们进行了以下分析比较:

(1) 再消噪前后分离信号的性能改善 再消噪处理使分离语音信号的质量得到了进一步的改善。在图2实验中的一般噪声环境下, 再消噪前后分离信号的信噪比和均方根误差比较如表1所示, 听音结果也显示再消噪处理后, 分离信号中的“嘶嘶”声得到了很大程度地抑制。

表1 图2实验中再消噪前后分离信号的SNR 和 RMSE 比较

评价指标	再消噪前	再消噪后	增加
$SNR_1$ (dB)	10.7468	11.2633	0.5165
$SNR_2$ (dB)	10.5230	10.8307	0.3077
$RMSE_1$	0.0019	0.0018	-0.0001
$RMSE_2$	0.0020	0.0019	-0.0001

(2) 强噪环境下的语音分离结果 在图2实验相同的混叠情况下, 将噪声幅度提升为其2.5倍, 得到强噪环境下的观测信号如图4(a)所示, 最终分离信号如图4(b)所示, 其信噪比分别为6.1209dB和6.4191dB, 均方根误差分别为0.0033和0.0032, 人耳能较好地听辨出各个语音源信号, 可见本文方

法对强噪环境下的混叠也可实现有效地分离。

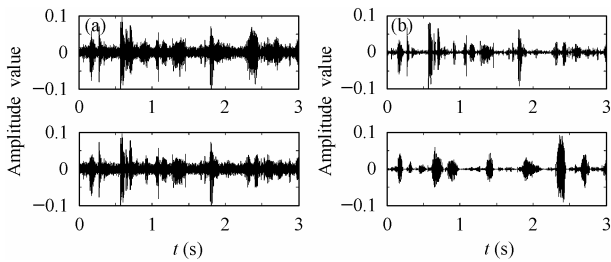


图4 强噪环境下的语音分离结果

(a) 混叠加信号  $x_1, x_2$  (b) 分离信号  $\hat{s}_1, \hat{s}_2$

Fig.4 Speech separation result in strong noisy environment

(a) Noisy speech mixtures  $x_1, x_2$  (b) Separated speeches  $\hat{s}_1, \hat{s}_2$

(3) 相同噪声强度、不同混叠信道参数下的分离效果统计 对5000组相同噪声强度、不同随机混叠矩阵下的含噪混叠语音进行分离, 统计结果表明: 约85%能得到有效地分离; 约15%由于混叠中一方语音太低而被噪声严重淹没, 只能分离出另一语音信号; 无一例完全分离失败情况。

另外, 将本文方法应用于TIMIT数据库中不同的语音样本均表现出了良好的分离性能。

由以上图表及统计结果可见, 本文所提出的小波变换和独立分量分析相结合的方法取得了令人满意的含噪混叠语音分离效果。

## 5 结束语

含噪混叠语音信号的分离在耳科医学、军事等语音通信领域中具有非常重要的实用价值。本文根据语音信号的特点, 提出了基于小波变换和独立分量分析的含噪混叠语音盲分离方法。文中首先介绍了独立分量分析算法思想, 然后详细阐述了具体分离方案, 通过模拟实验和结果分析说明了本文方法的有效性, 听音结果也表明分离效果令人满意。实验中我们进行的是两个说话者的含噪混叠语音分离, 在保证观测信号的数目不少于源信号数目的情况下, 可以将其推广到3个及3个以上说话者的含噪混叠语音分离。然而, 由于真实

环境中的多说话者情况较为复杂, 例如信道噪声、环境卷积声等的存在, 本文方法对实测信号的分离效果并不理想, 下一步我们将尝试对其进行完善和补偿以期能实现实际环境中的混叠语音分离。

## 参考文献

- [1] 赵鹤鸣, 朱祺, 陈雪勤等. 临界频带子波变换用于混叠语音分离的研究. 声学学报, 2004, 29(2): 177-181.
- [2] Cardoso J F. Blind signal separation: statistical principles. *Proc. IEEE*, 1998, 86(10): 2009-2025.
- [3] Brown G J, Cooke M. Computational auditory scene analysis. *Computer Speech and Language*, 1994, 8(4): 297-336.
- [4] Hyvarinen A, Karhunen J, Oja E. Independent Component Analysis. New York: John Wiley & Sons, Inc, 2001: 147-161, 293-304.
- [5] 刘璐, 何振亚. 盲源分离和盲反卷积. 电子学报, 2002, 30(4): 570-576.
- [6] 陈雪勤, 赵鹤鸣, 陈小平. 基于计算听觉场景分析的强噪声背景下基音检测方法. 电路与系统学报, 2003, 8(3): 128-131.
- [7] van der Kouwe A J W, Wang D L, Brown G J. A comparison of auditory and blind separation techniques for speech segregation. *IEEE Trans. on Speech and Audio Proceeding*, 2001, 9(3): 189-195.
- [8] Hyvarinen A. Fast and robust fixed-point algorithms for independent component analysis. *IEEE Trans. on Neural Networks*, 1999, 10(3): 626-634.
- [9] 彭玉华. 小波变换与工程应用. 北京: 科学出版社, 1999: 59-62.

赵彩华: 女, 1981年生, 硕士生, 研究方向为语音信号处理、盲信号处理。

刘璐: 男, 1965年生, 教授, 博士生导师, 从事盲信号处理、通信信号处理和多媒体通信等研究。

孙建德: 男, 1978年生, 博士生, 研究方向为多媒体通信。

闫华: 女, 1973年生, 博士生, 研究方向为多媒体图像处理。