

大连理工大学

硕士学位论文

多语音分离语音增强方法的研究

姓名：杨捷

申请学位级别：硕士

专业：信号与信息处理

指导教师：马晓红

20091216

摘 要

语音是人类进行信息传递和交流的重要载体,语音的质量能否得到保证不仅会影响到人耳的听觉效果,还会影响到语音处理系统的各个环节。在实际环境中,由于各种类型干扰的存在,语音信号往往会受到各种各样的污染,造成语音质量的明显下降。语音增强则是以去除各类干扰为出发点,尽可能地恢复出原始的纯净语音信号。针对不同类型的干扰,会有不同的语音增强方法。用于去除语音干扰的语音分离技术是当前语音增强研究领域的热点之一

本文主要研究多语音分离语音增强问题,具体包括以下三方面内容:

(1) 基于独立分量分析的语音信号盲源分离方法。在观测信号个数不少于源信号个数的情况下,采用独立分量分析技术可以很好地解决盲源分离问题,有效去除语音干扰。独立分量分析技术的核心在于对解混矩阵(混合矩阵的逆矩阵)的求解,源信号则是通过解混矩阵与观测信号向量直接相乘求得。本文在独立分量分析快速算法的基础上,研究了更为有效的改进型快速算法和结合语音信号短时平稳性的改进算法,以提高解混矩阵的精度及源信号的恢复质量。

(2) 基于聚类的欠定盲辨识方法。对于观测信号个数少于源信号个数的欠定盲源分离问题,独立分量分析方法不再适用,此时估计系统混合矩阵需要利用源信号的稀疏性。本文研究了基于聚类的欠定盲辨识算法,给出一种采用迭代自组织数据分析技术来估计混合矩阵的方法,并采用去除野点的预处理步骤及逐步去除类中最大偏差样本点的后置处理步骤,进一步提高了算法的稳定性和混合矩阵的估计精度。

(3) 欠定语音信号分离的逐层分离方法。在欠定盲源分离问题中,通常利用源信号的稀疏性,采用统计学的方法来分离源信号。由于语音信号本身的稀疏性还不够理想,并不严格满足正交不重叠条件,因此分离出的源信号之间存在相互干扰和音乐噪声比较明显。本文通过对混合矩阵进行逐次变换,采用从混合信号中逐步消去各个源信号,并由各混合信号中逐次产生的零值点来构造多层二值掩蔽模板的方法,将源信号进行逐层分离,在一定程度上抑制了分离出的源信号之间的相互干扰和音乐噪声,提高了源信号的分离质量。

计算机仿真实验结果表明了上述方法的有效性。

关键词: 语音增强; 独立分量分析; 欠定盲源分离; 聚类; 逐层分离

Study on the Speech Enhancement Method of the Multiple Speech Signals Separation

Abstract

Speech is very important for information transmission and communication of human. The quality of speech is not only influence the human hearing but also the other steps in speech processing system. For the interference pollution in real world, speech quality always become bad. Speech enhancement is based on removing the interference in speech to recover the original pure speech. Different method could be designed to eliminate different kinds of interference. The multiple speech separation technology to remove the speech interference is one of the focus in the field of speech enhancement.

This paper study on the separation method of multiple speech signals and mainly include three aspects as follows:

(1) The blind source separation method of speech signal based on independent component analysis. When the number of observed signals which is linear mixed by sources is not less than the number of sources, the method of independent component analysis could resolve the blind source separation problem effectively and eliminate the noise of speech interference well. The key of independent component analysis method is to calculate the demixing matrix. Then sources could be recovered from the observed signals by multiplying the demixing matrix with the observed signal vector. This paper focus on the fast independent component analysis algorithm, and studied two more efficient version of the algorithm to improve the accuracy of the demixing matrix and the quality of recovered sources.

(2) The underdetermined blind identification method based on clustering. When the number of observed signals is less than the number of sources, called as the underdetermined condition, the independent component analysis method is not suitable for the separation any more. It is necessary to utilize the sparse of signals to evaluate the mixing matrix for further separation. This paper studied the underdetermined blind identification algorithm based on clustering, and give a method by using the Iterative self organizing data analysis techniques algorithm to obtain the mixing matrix. Then, a pre-proceeding step and a post-proceeding step are proposed to improve the robust of the algorithm and the accuracy of the evaluated mixing matrix.

(3) The underdetermined speech separation by a step-wise method. In the underdetermined condition, methods based on statistics are always adopted for the separation by using the sparse of signals. However, there are still some mutual interference and music

noise in the separated sources, because speech sources do not satisfy the W-disjoint orthogonality condition strictly in time frequency domain which led to the overlapping in some extent between sources. this paper utilize the mixing matrix that get from some clustering method, and cancel the sources one by one from each mixed signal to produce corresponding zero value points in the mixed signals. Then, construct multiple binary mask from zero points to extract the disjoint or overlapped sources from the mixture and separate them step by step. The interference and music noise of the separated signals is depressed in some extent by this method, and the quality of the separation become better.

Experimental result reveal the efficiency of the methods above in this paper.

Key Words: Speech Enhancement; Independent Component Analysis; Underdetermined Blind Source Separation; Clustering; Step-wise Separation

大连理工大学学位论文版权使用授权书

本人完全了解学校有关学位论文知识产权的规定，在校攻读学位期间论文工作的知识产权属于大连理工大学，允许论文被查阅和借阅。学校有权保留论文并向国家有关部门或机构送交论文的复印件和电子版，可以将本学位论文的全部或部分内容编入有关数据库进行检索，可以采用影印、缩印、或扫描等复制手段保存和汇编本学位论文。

学位论文题目： 多语音分离语音增强方法的研究

作者签名： 杨捷 日期： 2009 年 12 月 30 日

导师签名： 马晓红 日期： 2009 年 12 月 30 日

大连理工大学学位论文独创性声明

作者郑重声明：所呈交的学位论文，是本人在导师的指导下进行研究工作所取得的成果。尽我所知，除文中已经注明引用内容和致谢的地方外，本论文不包含其他个人或集体已经发表的研究成果，也不包含其他已申请学位或其他用途使用过的成果。与我一同工作的同志对本研究所做的贡献均已在论文中做了明确的说明并表示了谢意。

若有不实之处，本人愿意承担相关法律责任。

学位论文题目： 多语音分离语音增强方法的研究

作者签名： 杨捷 日期： 2009 年 12 月 30 日

1 绪论

1.1 语音增强应用背景

语言是人类进行信息传递和思想交流的最重要的工具，而语音则是语言的重要载体，是语言的声学表现，是人类进行信息交换的最常用、最有效的手段。随着科学技术的发展，为了使语音这个信息载体能够得到更加经济有效的利用以及更加高效准确的传输，并能服务于其它各个领域，语音信号处理技术由此产生并得到发展。语音信号处理技术的主要内容包括语音编码、语音合成、语音识别、说话人识别、语音增强等多个方面。其中语音增强技术是语音编码、语音识别等技术实现的前提和重要保证，在语音信号处理系统中占有非常重要的地位。

语音增强技术是随着 60 年代数字信号处理理论的成熟而发展起来的。在 70 年代人们对语音增强的研究达到高潮，而到 80 年代以后语音增强技术则进一步走向了实用化。

总体上看，语音增强包含了去除目标语音信号中的噪声和混响以及去除语音干扰两个方面。在语音去噪方面，可以充分利用噪声与语音信号在时域、频域以及统计特性等方面的不同，采用固定滤波器、自适应滤波器、相关性去噪、维纳滤波^[1,2]以及谱相减^[3]、波束形成^[4,5,6,7]等方法来消除噪声，提取纯净目标语音。而在消除语音干扰方面，由于干扰信号与目标信号同属于语音信号，其性能及其相似，常用的去噪方法对于消除语音干扰往往不再适用。此时需要考虑采用多语音分离方法来实现目标语音和干扰语音的分离，以达到提取目标语音，消除语音干扰的目的。盲源分离技术^[8]的诞生与发展为多语音分离方法提供了良好的基础和广阔的前景。

1.2 盲源分离技术发展概况

在多输入、多输出的信号传输模型中，输出信号是由输入信号经线性或非线性混合叠加而成，并能直接被观察者观测到。因此输出信号通常又被称为混合信号或观测信号，输入信号被称为源信号，由源信号到观测信号之间的传输通道被称为混合系统或混合矩阵。在源信号和混合系统未知的情况下，将源信号从观测信号中分离出来的过程称为盲源分离（Blind source separation, BSS）。本文后面所涉及的盲源分离内容均属于线性混合的情况。

盲源分离技术的研究是在二十世纪八十年中后期才开始起步的，是一项新兴的前沿技术。通过二十多年来许多学者的共同努力，盲源分离技术已取得了很大的发展。虽然目前盲源分离技术因发展时间较短还有待进一步的研究和完善，但是人们在对盲源分离的研究过程中已建立起了一系列的理论和算法，取得了初步的成果。

就语音信号而言,其盲源分离问题可描述为实际应用中的“鸡尾酒会”问题,即在嘈杂的语音环境下如何将某个人的语音从众多说话人的语音中分离出来。由于实际环境的复杂性不便于处理盲源分离问题,因此通过一些假设将实际环境条件进行简化,把从源信号到观测信号的混合过程抽象成相对简单的数学模型,以便于盲源分离问题的处理。一般说来,线性混合盲源分离的数学模型有三种,分别是瞬时混合模型^[9]、瞬时—延迟混合模型^[10]以及卷积混合模型^[11]。针对不同的混合模型,有各种不同的算法。当 N 个源信号通过一个 $M \times N$ 的混合矩阵 \mathbf{A} 得到 M 个观测信号时,需要根据源信号和观测信号的相对个数来确定分离的方法。当 $M \geq N$ 时,在源信号满足非高斯性和统计独立性的条件下,通过构造解混矩阵 \mathbf{W} ,满足 $\mathbf{W} = \mathbf{A}^{-1}$,且不考虑 \mathbf{W} 各行排列的顺序和各行的模值大小,可将源信号从观测信号中分离出来。此时分离得到的源信号的估计与原始的输入源信号相比,存在着幅值及排列顺序上的不确定性。其中当 $M = N$ 时, \mathbf{A} 为方阵,可直接通过解混矩阵 \mathbf{W} 分离源信号,当 $M > N$ 时, \mathbf{A} 为超定矩阵,可结合最小二乘法来求解源信号。当 $M < N$ 时,混合矩阵 \mathbf{A} 为欠定矩阵,由于欠定方程组解的不唯一性,使得源信号的分离变得较为复杂。通常采用非线性的方法来解决欠定混合问题。因此,解决一个具体的盲源分离问题,通常需要从其混合模型以及源信号与观测信号的相对个数两方面来考虑。

实际环境中的各种信号往往本身具有一些很好的特性,这些特性通常也是盲源分离问题中必不可少的前提条件。以语音信号为例,语音信号本身的短时平稳性、稀疏性、非高斯性以及不同语音源信号之间的统计独立性,是语音信号进行盲源分离的各种算法的理论前提。其中语音信号的非高斯性及统计独立性是解决混合矩阵为超定或方阵的情况下的盲源分离的必要前提,而语音信号的短时平稳性和稀疏性则是解决欠定混合情况下盲源分离的基础。

目前,盲源分离技术的发展主要经历了以下过程。在1986年,Herault和Jutten^[12]在非高斯、独立源信号线性方阵瞬时混合的分离中,最早使用了盲源分离方法。他们采用了递归人工神经网络来分离未知源信号,并在此基础上发展了自适应算法。1989年,Linsker^[13]从信息论的角度提出了基于源信号与观测信号之间互信息最大化的无监督学习算法。同年,Cardoso^[14]提出了利用四阶统计量分解独立信源及辨识混合矩阵的四阶盲辨识(Fourth Order Blind Identification, FOBI)算法。1990年,Gaeta和Lacoume等人^[15]提出了极大似然估计的方法。1994年,Comon^[16]提出了输入信号与输出信号之间的互信息量可以作为输出信号之间相互独立性的一种很好的度量方式,同时还论证了最大化输出信号的非高斯性与最小化输出信号之间的互信息等价,也就是说如果分离得到的源信号的非高斯性最大或它们之间的互信息最小,那么此时分离出的源信号将最接

近于原始的源信号（不考虑信号的幅值和排列顺序）。无论是非高斯性还是互信息量都可以作为信号统计独立性的度量尺度，因此独立分量分析（Independent Component Analysis, ICA）的概念由此正式提出。同年，Cichocki 等人^[17]提出了自然梯度法，提高了算法的收敛速度和有效性。1995 年，Bell 和 Sejnowski^[18]将 BSS 算法进一步发展，提出了 BS-Infomax 方法，该方法采用的是随机梯度法。随后，Cardoso^[19]证明了 BS-Infomax 算法和极大似然法之间的等价性。1997 年 Belouchrani 等人^[20]提出了二阶盲辨识（Second Order Blind Identification, SOBI）算法，同年，Hyvarinen 等人^[21]提出了基于四阶累积量的不动点算法（Fixed-Point Algorithm），而后在 1999 年又提出了基于负熵的不动点算法^[22]，并在两年后的著作中对该算法进行了进一步的简化^[23]。ICA 不动点算法以分离信号的非高斯性极大化作为源信号分离成功的判定条件，通过构造相应的梯度算法，将源信号从观测信号中提取出来。由于该算法稳定性好，收敛速度很快，因此又被称为“快速 ICA 算法”（Fast Independent Component Analysis, FastICA）。2000 年，Bingham 等人^[24]提出了基于复信号的快速 ICA 算法，用于频域盲分离。2001 年，Stone^[25]提出了基于时域相关性的算法，随后 Cheung 等人^[26]对该算法作了进一步的发展。2005 年，Tichavsky 等人^[27]对快速 ICA 算法又作了进一步的完善，以提高其精度和稳定性。

早期的盲源分离算法都集中于研究观测信号个数大于或等于源信号个数的情况，对于具有挑战性的观测信号个数小于源信号数目的欠定混合情况的研究，直到二十世纪末才逐步兴起。1998 年，Lewicki 和 Sejnowski^[28]提出了基于信号稀疏性的方法。信号稀疏性的引入，使得欠定盲源分离有了一系列非常有效的解决方法，并使得一般性的盲源分离问题有了一个全新的解决途径。2000 年，Jourjine^[29]首先提出了退化分离估计技术（Degenerate Unmixing Estimation Technique, DUET）算法，该方法能够用于解决瞬时混合模型以及瞬时一延迟混合模型下的欠定盲源分离问题。2004 年，Yilmaz 与 Rickard^[10]对 DUET 方法进行了有效的改进和拓展。2007 年，Thomas 及 Deville 等人^[30]提出了一种基于差分四阶累积量的 FastICA 算法，用于实现非平稳源信号之间的分离。而 Mitianoudis 等人^[31]则提出了采用拉普拉斯模型的批处理及在线算法来解决欠定盲分离问题。2008 年，Pedersen 和 Wang 等人^[32]提出了一种将 ICA 技术与时频掩蔽相结合的欠定盲分离算法，取得了较好的分离效果。同一年，Lathauwer 等人^[33]以张量分解为基础，提出了矩阵同时对角化的系统盲辨识方法，该方法结合了交替最小二乘法（Alternating Least Squares, ALS）及精确线搜索法（Exact Line Search, ELS）来进行实现。

近年来，对欠定情况的研究主要集中在盲源分离问题的解决这一方面；而对于非欠定情况的研究则主要集中在进一步提高源信号分离的精度以及加快算法的收敛性等方面。

1.3 本论文主要研究内容及章节安排

本文主要研究基于盲源分离的多语音分离语音增强方法，具体章节安排如下：

第1章主要介绍了语音增强的应用背景以及盲源分离技术的发展概况。

第2章简单介绍了语音信号处理的相关基础知识。

第3章研究了基于快速独立分量分析的盲源分离算法及其结合了语音信号特性的改进方法。

第4章研究了利用语音信号的稀疏性，采用聚类算法的欠定盲辨识方法。给出了一种采用迭代自组织数据分析技术的混合矩阵估计方法，并进一步给出了修正方法。

第5章研究了语音信号的欠定盲分离方法，提出了一种对语音信号进行逐层分离的方法，以进一步提高语音分离的质量。

最后，结论部分对本文研究的各种方法进行了总结和展望。

2 语音信号处理的基础知识

2.1 语音信号的基本特征

2.1.1 语音信号的时域特征

由于说话人在不同时刻的说话内容千变万化, 而且没有确定的规律性, 因此语音信号是时变的。但由于人类的发声器官的变化速率有限, 声道在短时间内(5-500ms)处于平稳状态, 因此语音信号具有短时平稳性。而这种短时平稳性, 也是语音处理中许多理论和算法的必要前提。由于人们在说话时, 各个音节或单词之间总会存在着时间上的间隔, 因此语音信号在时域上存在着有音段和无音段。通过有音段和无音段的检测, 可用去除带噪语音中平稳的噪声。此外, 无音段所占比例越大, 语音的稀疏性越好, 利用这一特点, 产生了许多基于语音稀疏性的增强算法。

2.1.2 语音信号的频域特征

语音信号的频谱能量主要集中在 300-3400Hz 范围内。语音本身由浊音和清音组成, 浊音含有语音信号的大部分能量, 其频谱分布主要集中在低频段中的基音频率及其各次谐波上, 呈现出明显的周期性; 清音则表现出随机性, 在频谱上类似于白噪声。

2.1.3 语音信号的统计特征

语音信号可以看作是一个遍历性随机过程的样本函数, 其统计特性可用其幅度的概率密度函数来描述。

对语音信号的统计特性的研究表明, 其幅度分布的概率密度函数可以用两种近似的表达式来较好地描述^[1], 其中一种是伽马 (Gamma) 分布

$$P(x) = \frac{\sqrt{k} e^{-k|x|}}{2\sqrt{\pi} \sqrt{|x|}}$$

式中 k 是一个常数, 与标准差 σ_x 有下列关系

$$k = \frac{\sqrt{3}}{2\sigma_x}$$

另一种是拉普拉斯分布 (Laplacian) 分布

$$P(x) = 0.5\alpha e^{-\alpha|x|}$$

式中, α 是一个由标准差 σ_x 决定的常数, 即

$$\alpha = \frac{\sqrt{2}}{\sigma_x}$$

相对说来,伽马分布对语音信号的幅度分布描述更为精确一点,而拉普拉斯分布的函数形式则更加简洁。

2.2 语音信号的传输模型

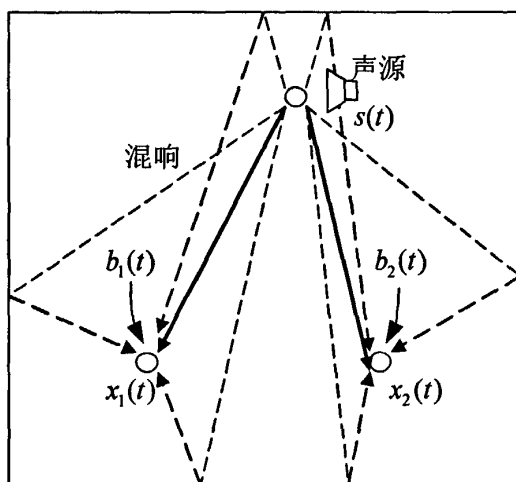


图 2.1 语音信号在实际环境中的传输模型

Fig. 2.1 The model of speech transmission in real world

实际环境中,麦克风阵列接收的信号可以用实际模型来描述^[34,35],如图 2.1 所示。其矢量形式表达如下

$$\mathbf{x}(t) = \mathbf{h}(t) * s(t) + \mathbf{b}(t)$$

这里, $\mathbf{x}(t)$ 为麦克风接收信号矢量, $\mathbf{h}(t)$ 是与麦克风阵列相关的房间传递函数向量, 算子 “*” 表示卷积操作, $s(t)$ 为声源信号, $\mathbf{b}(t)$ 为噪声干扰矢量。其频域的对应形式是

$$\mathbf{X}(\omega) = \mathbf{H}(\omega)S(\omega) + \mathbf{B}(\omega)$$

为了便于处理,需要对问题进行简化,通常将以上实际模型转换为理想情况下的瞬时一延迟模型,此时 N 个麦克风接收到的信号 $x_i(t)(i=1,2,\dots,N)$ 可以表示为

$$x_i(t) = \alpha_i s(t - \tau_i) + \beta_i n_i(t - \tau_i)$$

其中, $s(t)$ 为语音源信号, $n_i(t)$ 为噪声或干扰, α_i 是语音源信号传播的衰减因子, β_i 是相关噪声传播的衰减因子, τ_i 是语音源信号传播到第 i 个麦克风所需要的时间, τ_i' 是相关噪声传播到各个麦克风所需要的时间。

通过时延估计, 可以进一步将瞬时一延迟模型转化为瞬时模型:

$$x_i(t) = \alpha_i s(t) + \beta_i n_i(t)$$

式中, $n_i(t)$ 为第 i 个麦克风上的干扰, 且 $n_i(t)$ 和 $n_j(t)$ ($i \leq j \leq N, i \neq j$) 之间彼此不相关。

2.3 短时傅里叶变换

在对语音信号处理的研究中, 通常需要对语音进行时频分析, 而短时傅里叶变换 (Short Time Fourier Transform, STFT) 就是一种常用的时频分析工具^[36]。STFT 通过加窗处理将时域信号分帧, 并随着窗口的滑动对信号逐帧进行傅里叶变换。

STFT 的定义如下:

$$X(t, \Omega) = \int x(\tau) w(\tau - t) e^{-j\Omega\tau} d\tau$$

式中, $x(\tau)$ 为时域信号, $w(\tau)$ 为时域窗函数, $X(t, \Omega)$ 为 $x(\tau)$ 的时频分布。当窗函数 $w(\tau)$ 沿着 t 轴移动时, 它可以不断地截取一小段一小段的信号, 然后对每一小段的信号作傅里叶变换, 因此得到二维函数 $X(t, \Omega)$, 从而得到 $x(\tau)$ 的联合时频分布。尽管 $x(\tau)$ 是非平稳的, 但将其分段以后, 只要每段长度足够小, 每一段都是平稳的。

将信号离散化可得到离散的 STFT 如下:

$$X(n, e^{j\omega}) = \sum_{m=-\infty}^{\infty} x(m) w(m - n) e^{-j\omega m}$$

式中, $x(m)$ 为离散时域信号, $w(m)$ 为时域窗函数, n 为窗函数在时域的对称中心, 代表窗函数在时域轴上平移的位置。通常可选择 Hanning 窗或 Hamming 窗等常用窗函数, 窗口的长度要足以保证所截取每一帧的平稳性。

2.4 语音增强效果评价

语音增强效果的评定通常包括主观评定和客观评定两个方面。然而, 在某些情况下, 语音增强算法要同时满足这两个方面是很难的。比如, 有时候为了提高主观效果往往会造成客观指标一定程度的下降, 相反的情形也同样存在。此时, 就需要根据语音增强的实际需求来进行折中选择。

2.4.1 主观评价法

主观测评方法是对语音质量的综合评估，反映了人耳对语音听音效果的主观感觉，主要包括有：平均意见得分(Mean Opinion Score, MOS)、判断韵字测试(Diagnostic Rhyme Test, DRT)和判断满意度测试(Diagnostic Acceptability Measure, DAM)三种评测方法^[28]。其中，MOS 法通过测试者对所听到的语音效果进行打分。全体实验者的平均分就是所测语音质量的综合得分。DRT 方法让测试者从听到的具有相同韵母的文字中判断出所听到的音是哪一个字，全体实验者判断正确的百分比就是 DRT 得分。该方法反映了语音的清晰度和可懂度。DAM 是对语音质量的综合评估，它是在多种条件下对语音质量可接受程度的一种度量，同样采用百分比评分。由于各种原因，主观测评法每次得到的结果之间可能会存在着相对较大的波动，而参与测试者个体的差异对测评结果也有很大影响。因此，为了得到可靠的结果就需要大量的实验者参与测试，用来进行测试的语音材料也要足够丰富，并尽量保持相同的测试环境。

2.4.2 客观评价法

相比于主观评价法，客观评价法更加准确而具体。客观评价法往往通过给出一些具体的客观测量数据或者是增强后得到的信号波形，来给出语音增强效果的客观评定。客观评价法的结果并不能完全反映人对语音质量的感觉。

最常用的一种客观评价方法就是计算增强后的语音信号的信噪比 (Signal Noise Ratio, SNR)，其计算公式如下：

$$SNR(x, y) = 10 \log_{10} \left(\frac{\sum_{n=1}^N |x(t)|^2}{\sum_{n=1}^N |x(t) - y(t)|^2} \right)$$

式中， $x(t)$ 是未失真的理想信号，对应于原始纯净语音； $y(t)$ 是失真后的信号，对应于经过语音增强后的信号。

2.5 本章小结

本章主要介绍了语音信号的基本特征、语音的声学模型、短时傅里叶变换以及语音增强效果的评价方法，这些内容是语音增强的理论基础，是后续各章需要用到的基础知识。

3 基于快速 ICA 的语音信号盲源分离算法研究

3.1 基于 ICA 方法的盲信号混合模型

基于 ICA 方法的盲信号混合模型如下所示：

$$\mathbf{x} = \mathbf{A}\mathbf{s} \quad (3.1)$$

其中

$$\mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_N \end{bmatrix}, \quad \mathbf{s} = \begin{bmatrix} s_1 \\ \vdots \\ s_N \end{bmatrix}, \quad \mathbf{A} = \begin{bmatrix} a_{11} & \cdots & a_{1N} \\ \vdots & \ddots & \vdots \\ a_{N1} & \cdots & a_{NN} \end{bmatrix}$$

式中， \mathbf{x} 、 \mathbf{s} 和 \mathbf{A} 分别代表观测信号向量、源信号向量以及混合矩阵。其中 \mathbf{A} 为 $N \times N$ 的满秩矩阵， \mathbf{s} 的各分量 s_1, s_2, \dots, s_N 是相互独立的。只要能够找到一个 $N \times N$ 的解混矩阵 \mathbf{W} ，使得 $\mathbf{W} = \mathbf{A}^{-1}$ ，则由解混过程 $\mathbf{y} = \mathbf{W}\mathbf{x}$ 可得到 \mathbf{s} 的精确估计，且此时 $\mathbf{y} = \mathbf{s}$ ，其中 $\mathbf{y} = [y_1 \ y_2 \ \cdots \ y_N]^T$ 代表分离信号向量，即盲源分离问题求出的最终解，代表了对源信号的估计。然而，在盲源分离中， \mathbf{s} 和 \mathbf{A} 都是未知的，因此，仅仅利用已知的观测信号 \mathbf{x} 很难直接找到满足 $\mathbf{W} = \mathbf{A}^{-1}$ 的解混矩阵，此时，往往需要利用源信号 s_1, s_2, \dots, s_N 之间的统计独立性来求解。

3.2 基于负熵的优化判据

在基于 ICA 的盲源分离问题中，由混合过程 $\mathbf{x} = \mathbf{A}\mathbf{s}$ 及解混过程 $\mathbf{y} = \mathbf{W}\mathbf{x}$ 联立可得：

$$\mathbf{y} = \mathbf{W}\mathbf{A}\mathbf{s} = \mathbf{G}\mathbf{s} \quad (3.2)$$

式中 $\mathbf{G} = \mathbf{W}\mathbf{A}$ 表示从源信号到分离信号的增益矩阵。若设增益矩阵 \mathbf{G} 中的各个元素用 $G_{ij} (i, j = 1, 2, \dots, N)$ 来表示，则对于 \mathbf{y} 中的某一分量 y_i 有： $y_i = \sum_{j=1}^N G_{ij} s_j$ 。当 \mathbf{s} 中的各个独立源信号 s_1, s_2, \dots, s_N 都是非高斯分布（比如都是超高斯或亚高斯分布）时，由中心极限定理可知， y_i 比 s_j 更加接近高斯分布，也就是说 s_1, s_2, \dots, s_N 通过线性组合得到 y_1, y_2, \dots, y_N 后，非高斯性减弱了。在不考虑源信号的幅度和次序不定的情况下，当 \mathbf{G} 为单位矩阵时，每一个 y_i 由一个 s_j 的组成而不包含其它源信号，此时 \mathbf{y} 的非高斯性将达到最强，并且 \mathbf{y} 中的每一个分量正好对应于一个源信号。因此，可将最大化分离信号的非高斯性作为盲源分离算法设计的目标。

通常，采用负熵来作为信号非高斯性的度量，负熵的定义如下：

$$\begin{aligned}
J[p(y)] &= KL[p(y), p_G(y)] \\
&= \int p(y) \log \left[\frac{p(y)}{p_G(y)} \right] dy \\
&= H_G(y) - H(y) \geq 0
\end{aligned}$$

式中, $p(y)$ 和 $H(y)$ 分别表示随机变量 y 的概率密度函数及其熵, $p_G(y)$ 和 $H_G(y)$ 分别表示与 $p(y)$ 具有同样方差的高斯分布的概率密度函数及其熵, $KL[\cdot]$ 表示 KL 散度。当 y 的非高斯性最强时, 距离高斯分布达到最远, 此时它的负熵 $J[p(y)]$ 取得最大值。

为了便于计算, 将 $p(y)$ 的方差归一化之后, 可将负熵表示成非多项式函数的加权和:

$$J[p(y)] = H_G(y) - H(y) = \frac{1}{2} \sum_{i=1}^N E\{[G^{(i)}(y)]^2\}$$

其中, $G^{(i)}(y) (i=1, \dots, N)$ 为若干个非多项式函数, 且满足正交归一性和矩消失性。在对称的概率密度分布中负熵可简化为如下形式:

$$J[p(y)] = \{E[G^{(2)}(y)] - E[G^{(2)}(v)]\}^2 \quad (3.3)$$

式中, v 是与 y 具有同样方差的零均值高斯随机变量。

将负熵作为提取独立分量的目标函数, 令 \mathbf{w}_i 为分离第 i 个源信号的分离向量, \mathbf{w}_i^T 对应于解混矩阵 \mathbf{W} 的第 i 行, 并设 \mathbf{z} 为 \mathbf{x} 经过去均值和球化处理后所得到的白化信号, 则有关系式: $y_i = \mathbf{w}_i^T \mathbf{z}$ 。将分离信号 y_i 的负熵 $J[p(y_i)]$ 对 \mathbf{w}_i 求导

$$\Delta \mathbf{w}_i \propto \frac{\partial J[p(y_i)]}{\partial \mathbf{w}_i} = \gamma E[\mathbf{z} g(\mathbf{w}_i^T \mathbf{z})] \quad (3.4)$$

式中, $\gamma = E[G(y_i)] - E[G(v)]$, $g(\cdot)$ 是 $G(\cdot)$ 的导数。几种常用的 $G(\cdot)$ 、 $g(\cdot)$ 及 $g'(\cdot)$ 如表 3.1 所示。将 $J[p(y_i)]$ 取得极大值作为完成独立分量提取的优化判据, 可以此构造梯度算法, 对 \mathbf{w}_i 进行迭代, 并设定收敛条件, 当 $\Delta \mathbf{w}_i$ 趋近于 0 且 $\|\Delta \mathbf{w}_i\|$ 足够小, 即可完成对第 i 个独立分量的提取。

表 3.1 常用的非线性函数
Tab.3.1 Usually used nonlinear functions

$G(y)$	$g(y)$	$g'(y)$
$(1/a_1) \log \cosh a_1 y$	$\tanh a_1 y$	$a_1 [1 - \tanh^2(a_1 y)]$
$-e^{-\frac{y^2}{2}}$	$ye^{-\frac{y^2}{2}}$	$(1 - y^2)e^{-\frac{y^2}{2}}$
y^4	y^3	$3y^2$

3.3 快速 ICA 算法介绍

1999 年, 芬兰学者 Hyvarinen 等人提出了基于负熵的 ICA 固定点算法^[19]。该算法能够完成对独立分量的逐个提取, 并且具有很快的收敛速度, 因此又被称为快速 ICA 算法(Fast ICA, FICA)。

当每一次从观测信号中提取出一个独立分量时, 该算法又被称为一单元快速 ICA 算法^[18,19], 其迭代过程如下:

- (1) 将观测信号向量 \mathbf{x} 去均值并球化, 得到白化信号向量 \mathbf{z} ;
- (2) 初始化分离向量 \mathbf{w}_i ;
- (3) 对 \mathbf{w}_i 进行如下迭代:

$$\mathbf{w}_i^+ \leftarrow E[\mathbf{z}g(\mathbf{w}_i^T \mathbf{z})] - E[g'(\mathbf{w}_i^T \mathbf{z})]\mathbf{w}_i \quad (3.5)$$

式中, $g(\cdot)$ 及 $g'(\cdot)$ 所采用的非线性函数可从表 3.1 中选取, $E(\cdot)$ 所取的集总均值用信号时间序列的样本均值来代替。

$$(4) \text{ 归一化: } \mathbf{w}_i^+ \leftarrow \frac{\mathbf{w}_i^+}{\|\mathbf{w}_i^+\|}.$$

(5) 收敛性判定: 当 $\|\mathbf{w}_i^T \mathbf{w}_i^+ - 1\| < \varepsilon$ 时 (ε 为某个取值很小的正实数, 是判定是否收敛的条件), 说明 $\|\Delta \mathbf{w}_i\| = \|\mathbf{w}_i^+ - \mathbf{w}_i\|$ 已足够小, 可认为 $J[p(y_i)]$ 已达到极大值, 则算法结束; 否则令 $\mathbf{w}_i = \mathbf{w}_i^+$, 返回(3)继续迭代。

当需要求出所有的独立分量时, 只需多次使用一单元快速 ICA 算法, 并在分离向量 \mathbf{w}_i 的每一次迭代结束后进行正交化处理, 使得 \mathbf{w}_i 与前面已经求出的 $i-1$ 个分离向量 $\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_{i-1}$ 相互正交。此外, 也可采用对称快速 ICA 算法^[19], 使得提取各个独立分量的一单元快速 ICA 算法同时进行, 并在每一次迭代中对所有的分离向量 $\mathbf{w}_i (i=1, 2, \dots, N)$ 进行对称正交化, 迭代公式如下:

$$\mathbf{W}^+ \leftarrow g(\mathbf{W}\mathbf{z})\mathbf{z}^T - \text{diag}[g'(\mathbf{W}\mathbf{z})\mathbf{1}_N]\mathbf{W} \quad (3.6)$$

$$\mathbf{W} \leftarrow (\mathbf{W}^+ \mathbf{W}^{+T})^{-1/2} \mathbf{W}^+ \quad (3.7)$$

其中, \mathbf{W} 为 $N \times N$ 解混矩阵, $\mathbf{1}_N$ 为 $N \times 1$ 的元素全为 1 的列向量。

3.4 快速 ICA 的改进算法

3.4.1 快速 ICA 算法的误差分析

在基于 ICA 的盲源分离问题中, 核心任务是求解解混矩阵 \mathbf{W} 。在不考虑分离信号与源信号之间的幅度和次序关系的情况下, 当 $\mathbf{W} = \mathbf{A}^{-1}$ 时, 分离信号为源信号的精确恢复。此时, 增益矩阵 \mathbf{G} 为单位矩阵, \mathbf{G} 中非对角元上的元素值均为 0。而在实际运算中, 由于非线性函数 $g(\cdot)$ 的估计误差以及信号样本的有穷性, 源信号恢复的误差总会存在。当源信号为零均值单位方差的随机变量且长度一定时, 在每一次的分离结果中, 归一化增益矩阵 \mathbf{G} 中非对角元素的值将服从零均值的正态分布^[21], 其方差 $\text{var}[\mathbf{G}_{ij}](i \neq j)$ 反映了分离出的源信号之间相互干扰的平均程度, $\text{var}[\mathbf{G}_{ij}]$ 越小, 说明在分离出的第 i 个源信号中引入的第 j 个源信号的残留成分越少, 分离效果越好。

通常将克拉默-罗下限(Cramer-Rao lower bound, CRLB)^[37]作为 ICA 算法精度的衡量尺度, 它是理论上增益矩阵 \mathbf{G} 所能达到的方差下限, 其值为:

$$\text{CRLB}[\mathbf{G}_{ij}] = \frac{1}{N_s} \frac{\kappa_j}{\kappa_i \kappa_j - 1} \quad i \neq j \quad (3.8)$$

式中, $\kappa_i = E[\psi_i^2(x)]$, 而 $\psi_i(x) = -f'_i(x)/f_i(x)$, 其中 $\psi_i(x)$ 为评价函数, $f_i(x)$ 为第 i 个源信号 s_i 的概率密度函数, 这里假设 s_i 是零均值的, 且具有单位方差, N_s 为信号长度。

在一单元快速 ICA 算法和对称快速 ICA 算法中, 增益矩阵 \mathbf{G} 中各元素方差的理论值分别为^[21]:

$$\text{var}[\mathbf{G}_{ij}^{IU}] \approx \frac{1}{N_s} V_{ij}^{IU}, \quad i \neq j \quad (3.9)$$

$$\text{var}[\mathbf{G}_{ij}^{SYM}] \approx \frac{1}{N_s} V_{ij}^{SYM}, \quad i \neq j \quad (3.10)$$

其中

$$\begin{aligned} V_{ij}^{IU} &= \frac{\gamma_i}{\tau_i^2}, & V_{ij}^{SYM} &= \frac{\gamma_i + \gamma_j + \tau_j^2}{(\tau_i + \tau_j)^2} \\ \gamma_i &= \beta_i - \mu_i^2, & \tau_i &= \nu_i - \mu_i \\ \mu_i &= E[s_i g(s_i)], & \nu_i &= E[g'(s_i)], & \beta_i &= E[g^2(s_i)] \end{aligned} \quad (3.11)$$

当非线性函数 $g_i(\cdot)$ 取为源信号 s_i 的评价函数 $\psi_i(x)$ 时, ICA 的分离效果将达到最好。

3.4.2 提高快速 ICA 算法精度的改进算法

2006 年, Koldovsky 等人^[38]提出了一种有效的快速 ICA 算法(Efficient Fast ICA, EFICA), 用于提高快速 ICA 算法的精度。该算法实际上是对快速 ICA 算法的输出 $y_i(i=1, \dots, N)$ 做进一步的修正处理。该修正步骤分两步进行。

首先, 计算快速 ICA 算法的各个输出信号 $y_i(i=1, 2, \dots, N)$ 的四阶累积量, 由此判断为输出信号 y_i 选取的非线性函数 $g_i(\cdot)$ 是否足够接近于其评价函数 $\psi_i(x)$, 否则将为其重新选取更加合适的 $g_i(\cdot)$, 然后在已分离出的信号 y_i 的基础上, 作进一步的一单元快速 ICA 算法迭代, 直到收敛为止, 以得到更加精确的解混矩阵 $\mathbf{W}^+ = [\mathbf{w}_1^+, \mathbf{w}_2^+, \dots, \mathbf{w}_N^+]^T$ 。

然后, 对 $\mathbf{w}_1^+, \mathbf{w}_2^+, \dots, \mathbf{w}_N^+$ 作如下加权正交处理:

$$\mathbf{W}_i^+ = [c_{i1} \frac{\mathbf{w}_1^+}{\|\mathbf{w}_1^+\|}, c_{i2} \frac{\mathbf{w}_2^+}{\|\mathbf{w}_2^+\|}, \dots, c_{iN} \frac{\mathbf{w}_N^+}{\|\mathbf{w}_N^+\|}]^T \quad (3.12)$$

$$\mathbf{W}_i^+ \leftarrow (\mathbf{W}_i^+ \mathbf{W}_i^{+T})^{-1/2} \mathbf{W}_i^+ \quad (3.13)$$

其中

$$c_{ij} = \begin{cases} \frac{V_{ij}^{1U}}{V_{ji}^{1U} + 1}, & i \neq j \\ 1, & i = j \end{cases} \quad (3.14)$$

取 \mathbf{W}_i^+ 中的第 i 行作为第 i 个源信号的分离向量, 令 $i=1, 2, \dots, N$ 分别进行以上加权正交处理过程, 则可得到所有源信号的分离向量及最终输出。

通过以上修正处理以后, 增益矩阵 \mathbf{G} 的元素方差特性变为:

$$\text{var}[\mathbf{G}_{ij}^{EF}] \approx \frac{1}{N_s} \frac{V_{ij}^{1U}(V_{ji}^{1U} + 1)}{V_{ij}^{1U} + V_{ji}^{1U} + 1}, \quad i \neq j \quad (3.15)$$

当 $g_i = \psi_i, (i=1, \dots, N)$ 时, $\beta_i = v_i = \kappa_i, \mu_i = 1, V_{ij}^{1U} = \frac{1}{\kappa_i - 1}$, \mathbf{G} 的元素方差可化简为:

$$\text{var}[\mathbf{G}_{ij}^{EF}] \approx \frac{1}{N_s} \frac{\kappa_j}{\kappa_i \kappa_j - 1}, \quad i \neq j \quad (3.16)$$

从式(3.16)中可知, \mathbf{G} 的元素方差已约等于 CRLB, 其分离信号的误差已接近于理论最小值。

3.4.3 基于短时平稳性的快速 ICA 改进算法

在快速 ICA 算法中, 如果仅为每一个源信号 y_i 选取一个非线性函数 $g_i(\cdot)$ 来构造目标函数并进行优化迭代计算, 那么每一个长度为 N_s 的源信号 y_i 应该被视为某个未知随机变量的 N_s 次样本实现。也就是说 y_i 应是一个平稳随机过程, 在每一时刻都具有相同的统计特性。而事实上很多信号并不满足平稳特性。尽管如此, 快速 ICA 算法还是将这些信号视为平稳过程, 并用其时间均值代替集总均值来估计其统计特性。从实际应用来看, 这种方式对于分离非平稳信号还是非常有效的。

对于某些具有短时平稳性的信号, 比如语音信号, 为了充分利用其在长时间内统计特性变化, 在短时间内统计平稳的性质, 考虑将信号分成许多小段, 使其每一段都是平稳的, 然后为每一小段分别选取合适的非线性函数。将信号分段以后其混合模型表示如下:

$$\mathbf{x}^{(I)} = \mathbf{A}\mathbf{s}^{(I)}, \quad I=1, 2, \dots, J \quad (3.17)$$

式中, $\mathbf{x}^{(I)}$ 和 $\mathbf{s}^{(I)}$ 分别为观测信号与源信号的第 I 小段, J 表示信号被分成的总的段数。

2009 年, Koldovsky 等人^[39]利用对信号分段选择非线性函数的思想, 设计了一种分段有效快速 ICA 算法(Block Efficient Fast ICA, BEFICA)。该算法的实现前提是信号应具有分段平稳性。

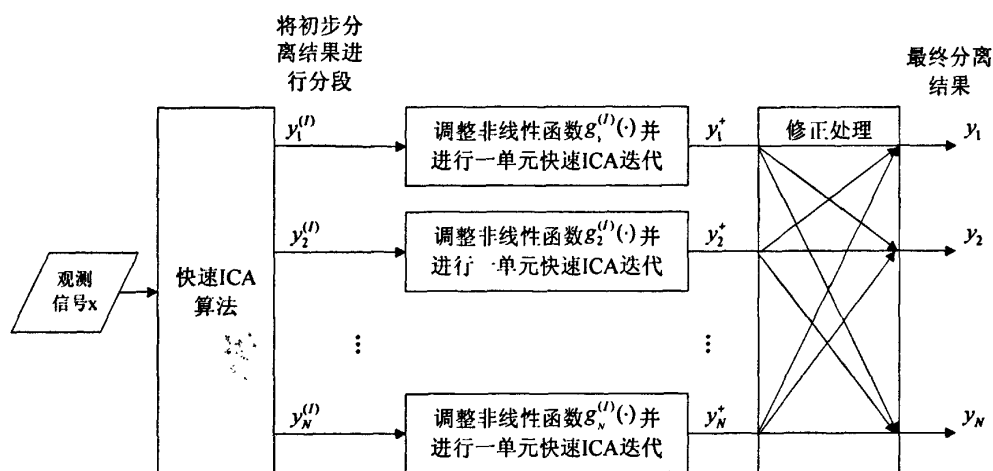


图 3.1 分段有效快速 ICA 算法框图

Fig.3.1 Flow of Block Efficient FastICA

算法流程图如图 3.1 所示。算法步骤如下:

(1) 首先采用快速 ICA 算法从观测信号 \mathbf{x} 中得到初始分离矩阵

$$\mathbf{W}(0) = [\mathbf{w}_1(0), \mathbf{w}_2(0), \dots, \mathbf{w}_N(0)]^T$$

以及分离信号的初始结果

$$y_i(0) = \mathbf{w}_i(0)^T \mathbf{z}, (i=1, 2, \dots, N)$$

(2) 将初始结果 $y_i(0)$ 等分成 J 小段 $y_i^{(l)} (l=1, 2, \dots, J)$, 使每一段信号都具有平稳性;

(3) 为 $y_i^{(l)} (l=1, 2, \dots, J)$ 选取合适的非线性函数 $G_i^{(l)}(\cdot)$ 和 $g_i^{(l)}(\cdot)$, 其选取方法采用 Pham 等人^[40]所介绍的参数估计方法, 由此可以得到提取源信号 s_i 相应的目标函数为:

$$c(\mathbf{w}_i) = \lambda_i^{(1)} E[G_i^{(1)}(\mathbf{w}_i^T \mathbf{z}^{(1)})] + \dots + \lambda_i^{(J)} E[G_i^{(J)}(\mathbf{w}_i^T \mathbf{z}^{(J)})] \quad (3.18)$$

式中, $\lambda_i^{(1)}, \dots, \lambda_i^{(J)}$ 为权重系数, 且 $\lambda_i^{(1)} = \lambda_i^{(2)} = \dots = \lambda_i^{(J)} = \frac{1}{J}$ 。由此可以由步骤(1)中得到的 $\mathbf{w}_i(0)$ 作为初值, 对 \mathbf{w}_i 进行进一步的分段一单元快速 ICA 算法迭代, 迭代公式如下:

$$\begin{aligned} \mathbf{w}_i^* \leftarrow & \lambda_i^{(1)} \left(E[\mathbf{z}^{(1)} g_i^{(1)}(\mathbf{w}_i^T \mathbf{z}^{(1)})] - \mathbf{w}_i E[g_i^{(1)'}(\mathbf{w}_i^T \mathbf{z}^{(1)})] \right) + \dots \\ & \dots \lambda_i^{(J)} \left(E[\mathbf{z}^{(J)} g_i^{(J)}(\mathbf{w}_i^T \mathbf{z}^{(J)})] - \mathbf{w}_i E[g_i^{(J)'}(\mathbf{w}_i^T \mathbf{z}^{(J)})] \right) \end{aligned} \quad (3.19)$$

当迭代收敛以后, 可得到优化了的分离向量 \mathbf{w}_i , 并能提取出精度更高的分离信号 $y_i^* = \mathbf{w}_i^T \mathbf{z}$ 。当所有的分段一单元快速 ICA 迭代结束以后, 便得到了优化后的解混矩阵 $\mathbf{W} = [\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_N]^T$ 。此时增益矩阵 \mathbf{G} 的元素方差特性为:

$$\text{var}[G_y^{BIU}] \approx \frac{1}{N_s} V_y^{BIU}, \quad i \neq j \quad (3.20)$$

其中

$$\begin{aligned} V_y^{BIU} &= \frac{\overline{\beta_i} - \overline{\mu_i}^2}{\overline{\tau_i}^2}, \quad i \neq j \\ \overline{\beta_i} &= \frac{1}{J} \sum_{l=1}^J (\lambda_i^{(l)})^2 \beta_i^{(l)}, \quad \overline{\mu_i} = \frac{1}{J} \sum_{l=1}^J \lambda_i^{(l)} \mu_i^{(l)}, \quad \overline{\nu_i} = \frac{1}{J} \sum_{l=1}^J \lambda_i^{(l)} \nu_i^{(l)}, \quad \overline{\tau_i} = \overline{\nu_i} - \overline{\mu_i} \\ \beta_i^{(l)} &= E[g_i^{2(l)}(s_i^{(l)})], \quad \nu_i^{(l)} = E[g_i^{(l)'}(s_i^{(l)})], \quad \mu_i^{(l)} = E[s_i^{(l)} g_i^{(l)}(s_i^{(l)})] \end{aligned} \quad (3.21)$$

(4) 对 \mathbf{W} 进行最后的修正处理:

$$\mathbf{W}_i^* \leftarrow [c_{i1} \mathbf{w}_1, \dots, c_{iN} \mathbf{w}_N]^T \quad (3.22)$$

$$\mathbf{W}_i \leftarrow (\mathbf{W}_i^* \mathbf{W}_i^{*T})^{-1/2} \mathbf{W}_i^* \quad (3.23)$$

其中

$$c_{ij} = \begin{cases} \frac{V_{ij}^{BIU}}{V_{ji}^{BIU} + 1}, & i \neq j \\ 1 & i = j \end{cases} \quad (3.24)$$

取 \mathbf{W}_i 的第 i 行 \mathbf{w}_i^T 作为提取第 i 个源信号的最终分离向量, 得到最终分离结果为

$$y_i = \mathbf{w}_i^T \mathbf{z}, \quad (i=1, 2, \dots, N)$$

通过对分段信号进行修正处理后, 最终增益矩阵 \mathbf{G} 的元素方差特性变为:

$$\text{var}[\mathbf{G}_{ij}^{BEF}] \approx \frac{1}{N_s} \frac{V_{ij}^{BIU} (V_{ji}^{BIU} + 1)}{V_{ji}^{BIU} + V_{ji}^{BIU} + 1}, \quad i \neq j \quad (3.25)$$

当 $g_i^{(l)} = \psi_i^{(l)}$, $(i=1, 2, \dots, N; l=1, 2, \dots, J)$ 时, 有 $\beta_i^{(l)} = \nu_i^{(l)} = \kappa_i^{(l)}$, $\mu_i^{(l)} = 1$ 及 $\tau_i^{(l)} = \gamma_i^{(l)} = \kappa_i^{(l)} - 1$, 则 $\bar{\beta}_i = \bar{\nu}_i = \bar{\kappa}_i, \bar{\mu}_i = 1$, 可得到 $V_{ij}^{BIU} = \frac{1}{\bar{\kappa}_i - 1}$, 因此 \mathbf{G} 的元素方差可化

简为:

$$\text{var}[\mathbf{G}_{ij}^{BEF}] \approx \frac{1}{N_s} \frac{\bar{\kappa}_j}{\bar{\kappa}_i \bar{\kappa}_j - 1}, \quad i \neq j \quad (3.26)$$

由式(3.26)可知, \mathbf{G} 的元素方差已约等于 CRLB, 其分离信号的误差已接近于理论最小值。

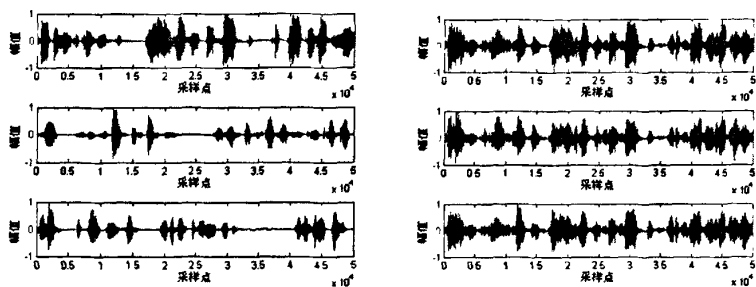
3.5 计算机仿真实验结果

本文选用 3 路经 PCM 采样后得到的纯净语音信号作为源信号, 采样率为 16kHz, 如图 3.2(a)所示。将源信号经过线性混合以后得到观测信号如图 3.2(b)所示。其中, 混合矩阵为:

$$\mathbf{A} = \begin{bmatrix} 1.23 & 0.97 & 1.01 \\ 0.65 & 0.73 & 1.18 \\ 0.71 & 0.92 & 0.56 \end{bmatrix}$$

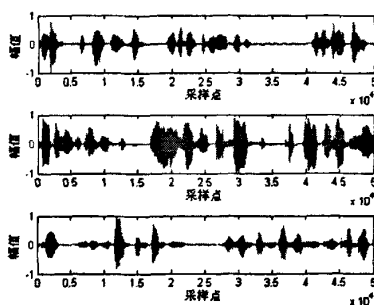
分别采用对称快速 ICA 算法、改进的有效快速 ICA 算法以及改进的分段有效 ICA 算法从混合信号中提取源信号, 得到的分离信号结果分别如图 3.2(c)、3.2(d)、3.2(e)所示。在不考虑源信号的分离顺序和幅度的情况下, 从时域波形来看, 每一种算法都有效地分离出了源信号。由式(3.10)、(3.15)及 (3.25)分别计算出对称快速 ICA 算法、改进的有效快速 ICA 算法以及改进的分段有效 ICA 算法各自的干扰信号比矩阵 \mathbf{ISR}_{sym} 、

ISR_{efica} 、 ISR_{befica} 。其中 ISR_{sym} 、 ISR_{efica} 、 ISR_{befica} 非对角元上的元素分别为 $\text{var}[G_y^{SYM}]$ 、 $\text{var}[G_y^{EF}]$ 、 $\text{var}[G_y^{BEF}]$ ，非对角元的值越小分离效果越好。

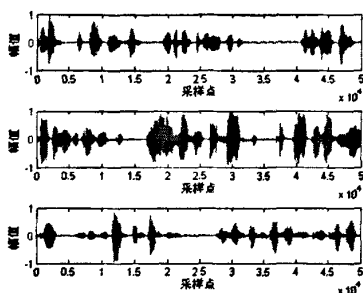


(a)源信号
(a)The sources

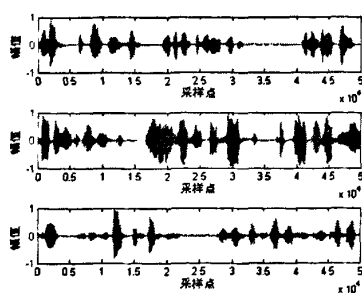
(b)混合信号
(b)The mixed signals



(c)对称快速 ICA 算法分离信号
(c)The signals separated by symmetric fast ICA Algorithm



(d)改进的有效快速 ICA 算法分离信号
(d) The signals separated by
efficient fast ICA Algorithm



(e)改进的分段有效 ICA 算法分离信号
(e) The signals separated by
block efficient fast ICA Algorithm

图 3.2 不同快速 ICA 算法的分离结果

Fig.3.2 The result of different fast ICA algorithm

$$\mathbf{ISR}_{sym} = \begin{bmatrix} 0 & 0.1140 & 0.1234 \\ 0.0903 & 0 & 0.1030 \\ 0.0764 & 0.0791 & 0 \end{bmatrix} \times 10^{-4}, \quad \mathbf{ISR}_{efica} = \begin{bmatrix} 0 & 0.0359 & 0.0360 \\ 0.0255 & 0 & 0.0256 \\ 0.1043 & 0.1043 & 0 \end{bmatrix} \times 10^{-5}$$

$$\mathbf{ISR}_{befica} = \begin{bmatrix} 0 & 0.0850 & 0.0850 \\ 0.2634 & 0 & 0.2634 \\ 0.0419 & 0.0419 & 0 \end{bmatrix} \times 10^{-7}$$

从实验结果可以看出, 采用改进的有效快速 ICA 算法, 其精度比传统的对称快速 ICA 算法更好。而结合了语音信号短时平稳性, 对信号分段选取非线性函数构造目标函数, 并进行迭代和修正的分段有效快速 ICA 算法, 又进一步提升了信号分离的质量, 更好地抑制了在分离结果中源信号之间的相互干扰。

3.6 本章小结

本章讨论了快速 ICA 算法的原理及其算法精度, 研究并实现了在此基础上改进的两种有效的快速 ICA 算法。理论分析及实验结果表明了改进算法的有效性, 并验证了改进算法提高了原有算法的信号分离精度。

4 基于语音稀疏性的欠定盲辨识方法

4.1 稀疏性信号欠定混合模型的特征分析

4.1.1 信号欠定混合的数学模型

假设源信号的个数为 N ，观测信号的个数为 M ，且 $M < N$ ，则时域信号欠定混合的数学模型为：

$$\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t) \quad (4.1)$$

其中

$$\begin{aligned} \mathbf{s}(t) &= [s_1(t), s_2(t), \dots, s_N(t)]^T \\ \mathbf{x}(t) &= [x_1(t), x_2(t), \dots, x_M(t)]^T \\ \mathbf{A} = [\mathbf{a}_1 \quad \mathbf{a}_2 \quad \dots \quad \mathbf{a}_N] &= \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1N} \\ a_{21} & a_{22} & \dots & a_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ a_{M1} & a_{M2} & \dots & a_{MN} \end{bmatrix} \\ \mathbf{a}_i &= [a_{1i} \quad a_{2i} \quad \dots \quad a_{Mi}]^T, \quad i=1, 2, \dots, N \end{aligned}$$

式中， $\mathbf{s}(t)$ 为时域源信号向量， $\mathbf{x}(t)$ 为时域观测信号向量， \mathbf{A} 为 $M \times N$ 的混合矩阵，其元素 a_{ij} 代表了第 j 个源信号 $s_j(t)$ 在第 i 个观测信号 $x_i(t)$ 中的幅值增益大小，因此可考虑 a_{ij} 均为正实数。由于在盲源分离问题中，分离出的源信号存在着幅度上的模糊性，因此每个源信号对应的混合系数向量 $\mathbf{a}_i (i=1, 2, \dots, N)$ 的模值也存在着不确定性。这里不妨设 $\|\mathbf{a}_i\| = 1$ ，即未知混合矩阵 \mathbf{A} 中的各个列向量都是归一化了的。

利用短时傅里叶变换将(4.1)式变换到时频域如下：

$$\mathbf{X}(\tau, \omega) = \mathbf{A}\mathbf{S}(\tau, \omega) \quad (4.2)$$

其中

$$\begin{aligned} \mathbf{X}(\tau, \omega) &= [X_1(\tau, \omega) \quad X_2(\tau, \omega) \quad \dots \quad X_M(\tau, \omega)]^T \\ \mathbf{S}(\tau, \omega) &= [S_1(\tau, \omega) \quad S_2(\tau, \omega) \quad \dots \quad S_N(\tau, \omega)]^T \\ X_i(\tau, \omega) &= STFT[x_i(t)], \quad i=1, 2, \dots, M \\ S_i(\tau, \omega) &= STFT[s_i(t)], \quad i=1, 2, \dots, N \end{aligned}$$

语音信号在时频域的稀疏性比时域更好，因此欠定混合问题常常在时频域中解决。

4.1.2 信号稀疏性的特征及其利用

若源信号向量 $S(\tau, \omega)$ 在时频域上的稀疏性达到理想状态, 即信号严格满足正交不重叠条件, 则在时频域上某一点 (τ, ω) 处只有一个源信号的值不为零, 其余源信号的值均为零, 在这种情况下, 式(4.2)可以简化为如下形式:

$$\mathbf{X}(\tau, \omega) = \mathbf{a}_i S_i(\tau, \omega), \quad i = 1, 2, \dots, N \quad (4.3)$$

写成向量形式为:

$$[X_1(\tau, \omega) \ X_2(\tau, \omega) \ \dots \ X_M(\tau, \omega)]^T = \mathbf{a}_i S_i(\tau, \omega), \quad i = 1, 2, \dots, N \quad (4.4)$$

考虑 \mathbf{a}_i 的各分量均为正实数的情况, 对式(4.4)等式两边向量的各分量先取模, 再进行归一化如下, 即

$$\begin{aligned} [|X_1(\tau, \omega)| \ |X_2(\tau, \omega)| \ \dots \ |X_M(\tau, \omega)|]^T &= \mathbf{a}_i |S_i(\tau, \omega)| \\ \frac{[|X_1(\tau, \omega)| \ |X_2(\tau, \omega)| \ \dots \ |X_M(\tau, \omega)|]^T}{\| [|X_1(\tau, \omega)| \ |X_2(\tau, \omega)| \ \dots \ |X_M(\tau, \omega)|]^T \|} &= \frac{\mathbf{a}_i |S_i(\tau, \omega)|}{\| \mathbf{a}_i \| |S_i(\tau, \omega)|} \end{aligned} \quad (4.5)$$

令

$$\begin{aligned} \mathbf{X}_{abs}(\tau, \omega) &= [|X_1(\tau, \omega)| \ |X_2(\tau, \omega)| \ \dots \ |X_M(\tau, \omega)|]^T \\ \mathbf{X}_{abs_norm}(\tau, \omega) &= \frac{\mathbf{X}_{abs}(\tau, \omega)}{\| \mathbf{X}_{abs}(\tau, \omega) \|} \end{aligned} \quad (4.6)$$

联立式(4.5)和式(4.6)两式, 并考虑到 $\| \mathbf{a}_i \| = 1$, 则可以得到:

$$\mathbf{X}_{abs_norm}(\tau, \omega) = \mathbf{a}_i, \quad i = 1, 2, \dots, N \quad (4.7)$$

由式(4.7)可知, 将观测信号向量先进行分量取模, 再归一化得到的向量 $\mathbf{X}_{abs_norm}(\tau, \omega)$, 是 M 维的实数单位向量, 其值分布在 M 维的单位超球面上, 并聚集成 N 个点, 每个点的值对应于一个混合矩阵向量 \mathbf{a}_i 。而实际应用中的信号由于不能达到理想的稀疏性条件, 使得 $\mathbf{X}_{abs_norm}(\tau, \omega)$ 的值在单位超球面上的分布比较分散, 不再重叠于 \mathbf{a}_i 对应的 N 个中心点上。但只要信号本身的稀疏性达到一定程度, $\mathbf{X}_{abs_norm}(\tau, \omega)$ 的取值分布还是能够表现出较为明显的聚类特性。在后面的内容中, 本文将以语音信号为实际例子来说明这种聚类特性, 并利用这一特点, 来进行混合矩阵的估计。

4.2 语音混合信号向量在空间中的投影特性

语音信号在时频域中表现出了良好的稀疏性, 具体表现为: 在含有语音信号能量分布的某个时频点上, 通常只有一个源信号的能量占绝对主导成分, 而其它源信号的能量

相对太小,可忽略不计。因此,可以认为一个时频点上仅有一个语音源信号存在,而没有其它源信号的成分。由于语音信号中大多数的时频点都满足以上特性,因此由语音信号线性混合得到的混合信号向量在向量空间中将呈现出良好的聚类特征。以4路源信号经线性混合得到3路混合信号的情况为例,本文通过实验得到时频域中混合信号向量在空间中的分布如图4.1(a)所示(这里实际上只显示了混合信号数据实部的分布,其虚部具有同样的分布特性)。

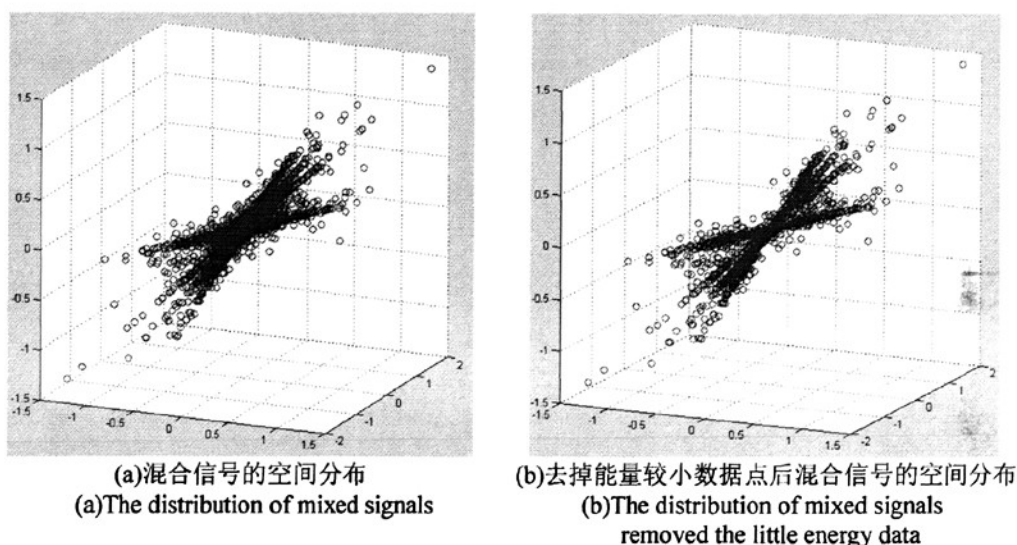


图4.1 混合信号向量的空间分布
Fig.4.1 The distribution of mixed signals

如图4.1(a)可见,混合信号的数据点集中分布在一些直线上,事实上每条直线方向正好对应于一个源信号。将能量较小的数据点去掉,保留含有语音主要能量的绝对值较大的数据点,其在空间分布中聚集而成的直线显得更加明显,如图4.1(b)所示。按照式(4.5)~(4.7)将这些数据点取模,并进行归一化,使其投影在单位球面上。得到数据点在球面上的分布情况如图4.2所示。

在图4.2中,通过选择不同的阈值 δ (这里分别取 $\delta = 0.02$ 、 $\delta = 0.1$ 以及 $\delta = 0.2$),去掉混合信号中模值小于 δ 倍混合信号最大模值的点,得到混合信号不同的分布情况分别如图4.2(a)~4.2(c)所示。从图4.2中可以看出,含有语音主要能量的数据点在单位球面上的分布中出现了一些聚类中心。球面上每个聚类中心的值可近似认为是某个源信号的归一化混合系数向量,分别对应于混合矩阵中的某一行,只要找到并计算出所有聚类中心的值,便可得到混合矩阵的估计。

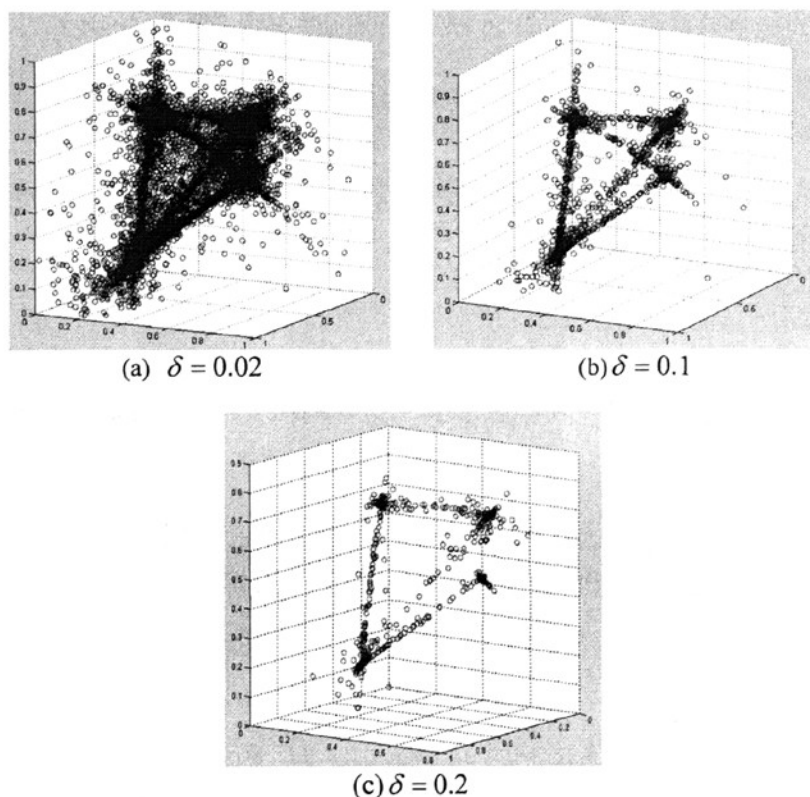


图 4.2 取不同阈值时混合信号数据点在单位球面上的分布

Fig. 4.2 The distribution of the mixed signals on the unit sphere with different threshold

4.3 源信号个数已知情况下的盲辨识

针对混合信号数据点在单位球面上的分布特性，可以采用聚类方法来估计混合矩阵。在现有的多种盲辨识算法^[10, 31, 32, 41-44]中，通常需要预先知道源信号的个数，然后再进行混合矩阵估计。其中，k-means 聚类算法^[45]及其修正算法^[43, 44]是效果较好的经典方法。

k-means 算法的流程如下：

(1) 假设要把已知样本集聚成 K 个类 S_1, S_2, \dots, S_K 。需要人为决定 K 个初始聚类中心 $Z_1(1), Z_2(1), \dots, Z_K(1)$ 。

(2) 在第 k 次叠代中，样本集 $\{Z\}$ 用如下方法分类：

对所有 $i=1, 2, \dots, K$, $i \neq j$

若 $\|Z - Z_j(k)\| < \|Z - Z_i(k)\|$ ，则 $Z \in S_j(k)$

(3) 令由(2)得到的 $S_j(k)$ 的新的类中心为 $Z_j(k+1)$

$$\text{令 } J_j = \sum_{Z \in S_j(k)} \|Z - Z_j(k+1)\|^2 \text{ 最小, } j=1,2,\dots,K$$

$$\text{则 } Z_j(k+1) = \frac{1}{N_j} \sum_{Z \in S_j(k)} Z, \quad N_j \text{ 为 } S_j(k) \text{ 中样本的个数}$$

(4) 对于所有的 $j=1,2,\dots,K$, 若 $Z_j(k+1) = Z_j(k)$, 则终止; 否则转到步骤(2)继续执行分类。

为了执行 k-means 算法, 需要预先知道聚类中心的个数, 并初始化聚类中心。转化到混合矩阵的估计问题中, 也就是需要预先知道源信号的个数, 并且需要对混合矩阵各列作一个粗略的估值以作为算法迭代的初始值。可以通过肉眼观察样本点在空间的分布, 将观察到的各个聚类中心的大致位置作为初值的选取。k-means 算法的优点在于计算量小, 收敛速度快, 精度较高, 缺陷在于需要事先知道分类的正确个数, 且算法的收敛性对初值的选取依赖性较强, 初值选取不当将得不到正确的分类结果。

4.4 源信号个数未知情况下的盲辨识

当源信号个数未知时, 以 k-means 算法为代表的各种要求源信号个数已知的盲辨识方法不再适用, 如何估计未知源信号的个数也是当前研究的一个热点。通过肉眼观察混合信号的数据点在空间中的分布情况可以大致确定源信号的个数, 但这种方法只有在数据点不太多, 数据点之间的相互重叠情况不太严重的情况下才适用。一旦在数据点的分布范围内出现了大范围的重叠情况, 肉眼将无法辨识出聚类中心, 即对于样本点分布的密度信息人眼是难以感知的。并且采用肉眼辨识的方法很不方便, 是非智能化的, 应该设计出相应的自适应算法来估计源信号的个数。

4.4.1 ISODATA 算法用于混合矩阵的估计

将混合信号的数据点作为聚类问题中的样本点, 那么对于源信号个数的估计也就是对样本点正确分类数的估计, 对混合矩阵的估计也就是对聚类中心的估计。针对以上问题, 本文给出了采用迭代自组织数据分析技术算法 (Iterative Self Organizing Data Analysis Techniques Algorithm, ISODATA)^[46]来进行源信号个数未知情况下的混合矩阵估计的方法。采用 ISODATA 算法可以自适应的估计出聚类中心的个数并同时求解出分类结果, 混合矩阵的估值由分类结果中的各个聚类中心直接得到。

ISODATA 算法的思想主要体现在三个方面, 一是直接取消样本点太少的类, 并将其样本点分配给其它类, 二是对于类内离散程度太大的类要分裂, 三是对于相互距离太

近的要合并。因此, ISODATA 算法不需要知道原始类的个数, 只需根据算法自身每一步迭代得到的中间分类结果, 依据所得分类的质量好坏, 自适应的调整分类状态, 直到得出符合预定条件的分类结果。

ISODATA 算法流程如下:

(1) 控制参数的设定与聚类中心的初始化

在 ISODATA 算法中, 首先需要设定 7 个控制参数, 分别是

- ① C : 预期的大致分类数
- ② N_c : 初始聚类中心个数
- ③ θ_n : 每一类允许的最少样本数
- ④ θ_s : 类内各分量分布的距离标准差上界
- ⑤ θ_d : 两类中心间的最小距离下界
- ⑥ L : 每次迭代中可以合并的类的最多对数
- ⑦ I : 允许的最多迭代次数

同时, 需要选定初始聚类中心, 可直接从待分类的样本点集 $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N\}$ 中选取 N_c 个样本点作为初始聚类中心 $\mathbf{z}_j (j=1, 2, \dots, N_c)$, 对应于每个聚类中心, 存在相应的类 $S_j (j=1, 2, \dots, N_c)$

(2) 按最小距离原则将样本点集中的每个样本点 $\mathbf{x}_i (i=1, 2, \dots, N)$ 分到某一类 $S_l (l=1, 2, \dots, N_c)$ 中。即若 $d_{il} = \min_j \|\mathbf{x}_i - \mathbf{z}_j\|$, $(i=1, 2, \dots, N)$, 则 $\mathbf{x}_i \in S_l$ 。式中 d_{il} 表示样本点 \mathbf{x}_i 和类中心 \mathbf{z}_l 之间的距离。

(3) 依据 θ_n 判断合并。如果类 S_j 中样本数 $n_j < \theta_n$, 则取消该类的中心 \mathbf{z}_j , 且置聚类中心个数 $N_c = N_c - 1$, 然后转至步骤(2)

(4) 计算分类后的各类中心、类内距离及总体平均距离:

① 计算各类中心:

$$\mathbf{z}_j = \frac{1}{n_j} \sum_{\mathbf{x}_i \in S_j} \mathbf{x}_i, \quad (j=1, 2, \dots, N_c)$$

② 计算各类中样本点到类中心的平均距离:

$$\bar{d}_j = \frac{1}{n_j} \sum_{\mathbf{x}_i \in S_j} \|\mathbf{x}_i - \mathbf{z}_j\|, \quad (j=1, 2, \dots, N_c)$$

③ 计算各个样本点到其类内中心的总体平均距离:

$$\bar{d} = \frac{1}{N} \sum_{j=1}^{N_c} n_j \bar{d}_j$$

(5) 依据迭代次数 I_p 以及 N_c 判断停止、分裂或合并

① 若 $I_p \geq I$ ，则置 $\theta_d = 0$ ，转到步骤(9)；否则转②作进一步的判断

② 若 $N_c \leq \frac{C}{2}$ ，则直接转到步骤(6)（分裂处理）；否则转③作进一步的判断

③ 若 $N_c \geq 2C$ ，则转至步骤(9)（合并处理）；否则转④作进一步的判断

④ 若 $\frac{C}{2} < N_c < 2C$ ，当迭代次数是奇数时，转至步骤(6)；迭代次数是偶数时，转至

步骤(9)

(6) 计算各类类内距离的标准差矢量

$$\sigma_j = (\sigma_{1j}, \sigma_{2j}, \dots, \sigma_{N_d j})', \quad (j=1, 2, \dots, N_c)$$

$$\sigma_{kj} = \left[\frac{1}{n_j} \sum_{x_i \in S_j} (x_{ki} - z_{kj})^2 \right]^{\frac{1}{2}}, \quad (k=1, 2, \dots, N_d; j=1, 2, \dots, N_c)$$

式中， k 为分量编号， j 为类的编号， N_d 为样本点维数， x_{ki} 是样本点 \mathbf{x}_i 的第 k 个分量， z_{kj} 是 \mathbf{z}_j 的第 k 个分量。

(7) 对每一个聚类，求出类内距离标准差矢量 σ_j 中的最大分量 $\sigma_{j\max}$

$$\sigma_{j\max} = \max_k [\sigma_{kj}]$$

(8) 在 $\{\sigma_{j\max}\}$ 中，对任一 $\sigma_{j\max}$ ，若有 $\sigma_{j\max} > \theta_s$ ，同时又满足以下两个条件之一：

① $\bar{d}_j > \bar{d}$ 和 $n_j > 2(\theta_n + 1)$

② $N_c \leq \frac{C}{2}$

则将类 S_j 分裂为两个聚类，得到两个新的聚类中心 \mathbf{z}_j^+ 和 \mathbf{z}_j^- ，且令 $N_c = N_c + 1$ 。 \mathbf{z}_j^+ 和 \mathbf{z}_j^- 只是在 \mathbf{z}_j 中相应于 $\sigma_{j\max}$ 的分量分别加上和减去 $\mu\sigma_{j\max}$ ，而其它分量保持不变，其中 $0 < \mu \leq 1$ ， μ 的取值应该使 \mathbf{z}_j^+ 和 \mathbf{z}_j^- 仍在 S_j 的类域空间中，且其它类 $S_i (i \neq j)$ 中的样本点到 \mathbf{z}_j^+ 和 \mathbf{z}_j^- 的距离较远，而原类 S_j 中的样本点到 \mathbf{z}_j^+ 和 \mathbf{z}_j^- 的距离较近。分裂以后， $I_p = I_p + 1$ ，然后转至步骤(2)

否则，转至步骤(9)

(9) 计算各类的聚类中心之间的距离

$$D_{ij} = \|\mathbf{z}_i - \mathbf{z}_j\|, \quad (i=1,2,\dots,N_c-1; j=i+1,\dots,N_c)$$

(10) 将小于 θ_d 的 D_{ij} 按递增顺序排列, 取前面 L 个, $D_{i_1j_1} < D_{i_2j_2} < \dots < D_{i_Lj_L}$ 。从最小的 D_{ij} 开始, 将相应的两个类合并。如果合并前两个类的中心为 \mathbf{z}_i 和 \mathbf{z}_j , 则合并后的聚类中心为:

$$\mathbf{z}_l = \frac{1}{n_i + n_j} [n_i \mathbf{z}_i + n_j \mathbf{z}_j]$$

$N_c = N_c - N_b$, 其中 N_b 为合并中被消去的类数与合并中生成的类数之差。在一次迭代中, 一个类最多被合并一次。

(11) 如果迭代次数 I_p 已达 I 次或过程已收敛, 则算法结束。否则 $I_p = I_p + 1$ 并转至步骤(2); 若需调整参数, 则转至步骤(1)。

以上是 ISODATA 算法的整个过程, 对于如何判定算法是否已经收敛从而停止迭代还未明确说明。通过分析可知, 当算法趋于收敛时, 一方面分类的个数将趋于稳定, 即分裂与合并过程都不再执行, 另一方面分类结果也将趋于稳定, 分给同一个类的样本点不再变化。本文通过实验测试, 认为只要满足以下两个条件:

(1) 第 $I_p + 1$ 迭代后的分类结果与第 I_p 次相同

(2) 第 $I_p - 1$ 次和第 I_p 次的迭代中, 类的分裂或合并过程均没有执行

则可判定算法已经收敛, 在第 $I_p + 1$ 次迭代完成后, 停止迭代并得到最终分类结果。

4.4.2 在混合矩阵估计中 ISODATA 算法的控制参数设定

在 ISODATA 的算法中, 初始控制参数的选择很重要, 因具体样本点在空间中分布情况的不同而设置不同的参数值。根据本文所涉及的样本点的分布特点, 并结合各个控制参数在算法中所起到的作用, 这里对各个参数的设置说明如下:

① C : 预期的大致分类数。该参数主要给出一个期望得到的分类个数的大致范围。

当分类个数小于 $\frac{C}{2}$ 时, 算法会优先考虑对已有类进行分裂处理; 当分类个数大于 $2C$ 时, 则优先考虑合并处理。事实上, 究竟能否进行分裂或合并还需要结合其它参数来共同判定。当实际聚类个数不在预期范围内时, 最终的分类结果类数往往以实际情况为准。

② N_c : 初始聚类中心个数。本文设定聚类中心个数的初值为 1, 即开始假设只有一个类。当算法开始以后, 会根据各个控制参数的判定情况, 对每次迭代得到的中间分类结果进行分裂或合并处理, 使之逐步趋近于正确的分类个数。

③ θ_n : 每一类允许的最少样本数。当类中样本点太少时,则需要取消该类,并将类中样本点分配给其它类。本文根据参与聚类的混合信号数据的样本点数,为 θ_n 选择合适的取值。当去除混合信号中能量较小的数据点所设定的阈值 δ 较小时,参与聚类的样本数会增加,此时需要将 θ_n 的值调整得更大。反之, θ_n 的取值应当更小。

④ θ_s : 类内各分量分布的距离标准差上界。这是算法中非常重要的一个参数,它的取值为类内样本点的离散程度提供了一个上限,一旦超过该上限,则需要将该类分裂成两个新的类,并重新分配样本点。类内样本点各分量分布的距离标准差大于 θ_s 是分裂过程能够进行的必要条件,因此 θ_s 的取值是否合理,决定了分类是否能够有效进行。既然要进行聚类,那么一般说来实际类的个数应当不止一个。当初始聚类个数 N_c 取值为1时,为了得到正确的分类结果,分裂过程必然要进行。如果 θ_s 取值太大,将可能导致原本属于不同类的样本点无法分裂成多个类,使得最终分类数少于实际类数。反之,若 θ_s 取值太小,则可能导致本该属于同一个类的样本点被分裂成多个类,使得最终分类数多于实际的类数。

⑤ θ_d : 两类中心间的最小距离下界。这是合并条件控制参数,当两个类中心的距离小于 θ_d 时,需要合并。为了使得合并过程有实际意义,避免刚合并后的新类因过于离散而马上被分裂,选择 θ_d 取值小于分裂过程控制参数 θ_s 比较合适。

⑥ L : 每次迭代中可以合并的类的最多对数。这里取 $L=2$,即每次合并过程中只将符合合并条件的相互间距离最小的两个类合并。

⑦ I : 允许的最多迭代次数。经过分裂和合并过程使聚类个数达到稳定之后,算法在对样本点的分类与迭代中将很快收敛,这里取 $I=30$ 已足够了。

在算法执行结束以后,将得到混合信号分布于单位球面上的样本点的分类结果。各个聚类中心的取值即可作为混合矩阵各个列向量的估计值。

4.4.3 对聚类进行修正的改进算法

由于混合信号样本点本身的分布不够集中,除了较为密集地分布在各个聚类中心周围,还存在大量稀疏而离散地分布于单位球面上远离各个聚类中心位置的点,本文将这些样本点称之为“野点”。大量野点的存在会降低聚类算法的质量,其影响表现在两个方面:一方面是影响聚类算法的稳定性,使得算法可能不收敛或者直接得到错误的聚类结果;另一方面是影响聚类算法的精度,导致聚类中心的取值与混合矩阵各列的真实值产生较大偏差。本文针对聚类算法在以上两个方面可能存在的问题,分别给出了改进措施,即采用预处理来加强算法稳定性,采用后置处理来提高算法精度。

为了提高聚类算法的稳定性,在算法开始前先对所有的样本点进行预处理,去除野点。当一个样本点附近分布的样本点很少时,说明该样本点位于距离聚类中心较远的稀疏区域,可认为该样本点是野点,将其删除。具体处理步骤如下:

- (1) 设定参数 ρ 和 N_p
- (2) 遍历所有的样本点,并统计在每一个样本点的半径为 ρ 的邻域内分布的样本点个数。
- (3) 将邻域内分布样本点个数少于 N_p 的样本点从样本集中删除,使聚类算法在保留下来的样本点集中进行。

在实际应用中,如果样本点的分布集中性较好,野点较少,也可不用预处理,直接对原始样本点进行聚类。

为了提高聚类算法的精度,在聚类结束后采用逐个去除类中最大偏差点的方法来进行聚类中心的修正。所谓类中最大偏差点,就是指类中距离聚类中心最远的样本点。每去除一个最大偏差点,将使得聚类更加集中,聚类效果更好。具体算法如下:

- (1) 设定偏差控制阈值 ε , 开始迭代;
- (2) 计算类 S_i 中各个样本点 $\{\mathbf{x}_{i1}, \mathbf{x}_{i2}, \dots, \mathbf{x}_{iN_i}\}$ 到聚类中心 \mathbf{z}_i 的距离, N_i 表示类 S_i 中样本点个数。找出满足距离关系 $d_{im} = \max_j \|\mathbf{x}_{ij} - \mathbf{z}_i\|, (j=1, 2, \dots, N_i)$ 的最大偏差点 \mathbf{x}_{im} , 如果 \mathbf{x}_{im} 到 \mathbf{z}_i 的距离 $d_{im} < \varepsilon$, 则停止迭代。否则转步骤(3)执行;
- (3) 从类 S_i 中删除该最大偏差点 \mathbf{x}_{im} , 并令 $N_i = N_i - 1$, 然后用类中剩余样本点更新类 S_i , 并重新计算类中心 \mathbf{z}_i 。然后转到步骤(2)执行。

通过以上修正,将使聚类中心的取值进一步逼近归一化混合矩阵列向量的真实值。

4.5 计算机仿真实验结果

本文选取由 PCM 采样得到的一组不同说话人的语音信号作为语音库,信号的采样率为 16kHz,采样点数为 50000,实验中的源信号将从该语音库中选取。

选取 4 路不同的源信号经线性混合得到 3 路混合信号,混合矩阵如下:

$$\mathbf{A} = \begin{bmatrix} 0.5070 & 0.5883 & 0.8835 & 0.3328 \\ 0.6881 & 0.3138 & 0.4016 & 0.6239 \\ 0.5191 & 0.7452 & 0.2410 & 0.7071 \end{bmatrix}$$

将混合信号变换到时频域,STFT 取 Hamming 窗,窗长 512 点,帧间交叠 256 点,FFT 点数 1024。时频域的 3 路混合信号构成了一个由三维矢量样本点组成的样本集。取

$\delta = 0.2$ ，去掉该样本集中模值小于 0.2 倍最大样本点模值的样本点，将保留下来的样本点经对称点取反处理，投影到三维空间中的单位球面的一侧上，如图 4.3(a)所示。

设对混合矩阵 \mathbf{A} 的估计结果为 $\hat{\mathbf{A}}_\varepsilon$ ，其百分比误差定义为：

$$e_\varepsilon = \left| \mathbf{A} - \hat{\mathbf{A}} \right| ./ \left| \mathbf{A} \right|$$

式中，符号 $./$ 表示矩阵对应元素相除，下标 ε 代表估计结果偏差控制阈值。

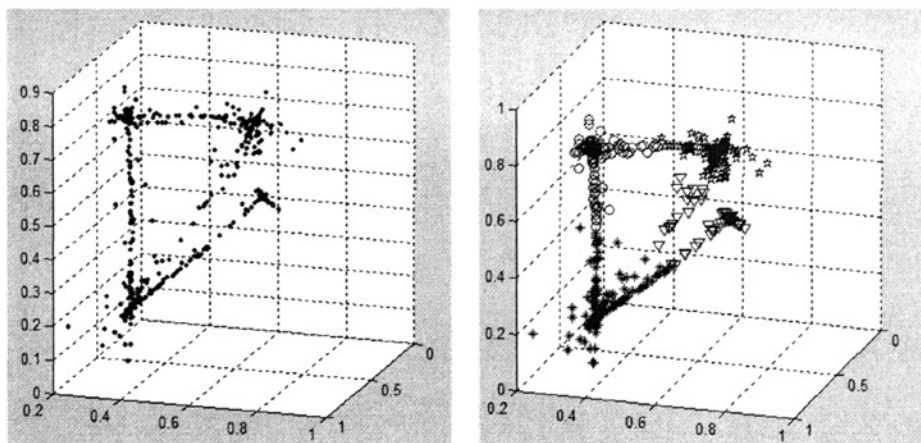
采用 ISODATA 算法直接对球面上的样本点进行聚类，设定控制参数为 $C = 4$ ， $N_c = 1$ ， $\theta_n = 120$ ， $\theta_s = 0.08$ ， $\theta_d = 0.06$ ， $L = 1$ ， $I = 30$ 。聚类结果如图 4.3(b)所示，属于同一个类的数据点具有相同的符号形状。直接把聚类得到各个类中心作为对混合矩阵的估计，此时相当于取 $\varepsilon = \infty$ ，得到结果如下：

$$\hat{\mathbf{A}}_\infty = \begin{bmatrix} 0.5208 & 0.5974 & 0.8640 & 0.3516 \\ 0.6765 & 0.3283 & 0.4152 & 0.6180 \\ 0.5207 & 0.7317 & 0.2847 & 0.7032 \end{bmatrix} e_\infty = \begin{bmatrix} 2.72\% & 1.54\% & 2.21\% & 5.66\% \\ 1.69\% & 4.63\% & 3.38\% & 0.95\% \\ 0.31\% & 1.82\% & 18.15\% & 0.55\% \end{bmatrix}$$

对聚类结果进行逐步去除最大偏差点的修正，取偏差控制阈值 ε 分别为 0.05、0.02 和 0.01，得到聚类结果分别如图 4.3(c)、图 4.3(d) 及图 4.3(e)所示。相应的混合矩阵估计结果及百分比误差如下（在实际运算中估计得到的混合矩阵列向量的次序是不确定的，这里为了分析混合矩阵的估计质量，将估计矩阵中各个列向量的次序作了调整，使之与原混合矩阵的列向量次序一一对应）：

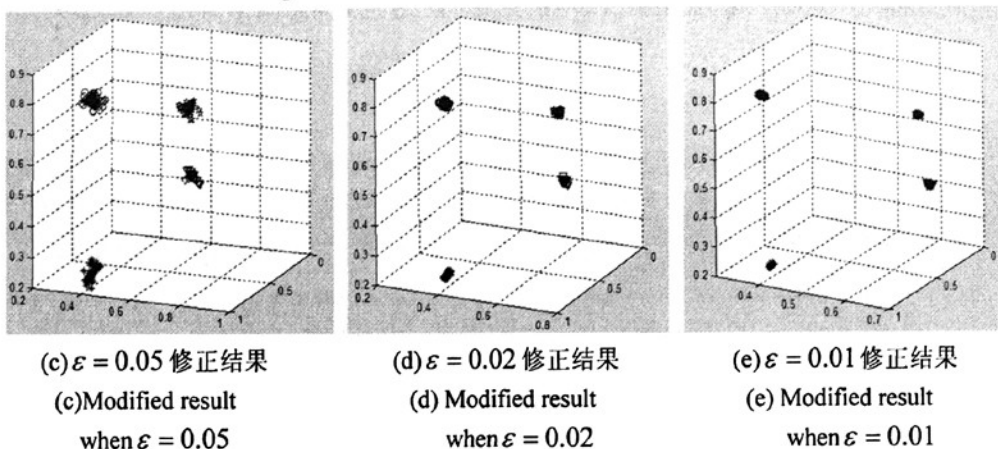
$$\begin{aligned} \hat{\mathbf{A}}_{0.05} &= \begin{bmatrix} 0.5071 & 0.5875 & 0.8824 & 0.3353 \\ 0.6884 & 0.3138 & 0.4018 & 0.6228 \\ 0.5186 & 0.7459 & 0.2448 & 0.7069 \end{bmatrix} e_{0.05} = \begin{bmatrix} 0.02\% & 0.14\% & 0.13\% & 0.75\% \\ 0.04\% & 0.00\% & 0.06\% & 0.17\% \\ 0.09\% & 0.09\% & 1.57\% & 0.03\% \end{bmatrix} \\ \hat{\mathbf{A}}_{0.02} &= \begin{bmatrix} 0.5072 & 0.5886 & 0.8834 & 0.3337 \\ 0.6879 & 0.3140 & 0.4011 & 0.6232 \\ 0.5192 & 0.7450 & 0.2422 & 0.7073 \end{bmatrix} e_{0.02} = \begin{bmatrix} 0.03\% & 0.05\% & 0.01\% & 0.29\% \\ 0.03\% & 0.06\% & 0.12\% & 0.12\% \\ 0.03\% & 0.04\% & 0.50\% & 0.03\% \end{bmatrix} \\ \hat{\mathbf{A}}_{0.01} &= \begin{bmatrix} 0.5075 & 0.5881 & 0.8829 & 0.3338 \\ 0.6875 & 0.3142 & 0.4021 & 0.6232 \\ 0.5193 & 0.7453 & 0.2426 & 0.7072 \end{bmatrix} e_{0.01} = \begin{bmatrix} 0.10\% & 0.05\% & 0.07\% & 0.31\% \\ 0.08\% & 0.12\% & 0.11\% & 0.11\% \\ 0.08\% & 0.01\% & 0.68\% & 0.02\% \end{bmatrix} \end{aligned}$$

由结果可知，选取足够小的 ε 值，可以提高混合矩阵的估计精度。



(a)混合信号数据在单位球面上的分布
(a)Mixed signal data distribution
on the unit sphere

(b)ISODATA 算法结果
(b)Result of ISODATA algorithm



(c) $\epsilon = 0.05$ 修正结果
(c)Modified result
when $\epsilon = 0.05$

(d) $\epsilon = 0.02$ 修正结果
(d) Modified result
when $\epsilon = 0.02$

(e) $\epsilon = 0.01$ 修正结果
(e) Modified result
when $\epsilon = 0.01$

图4.3 聚类结果

Fig.4.3 The result of clustering

以上实验中，由于混合信号样本点的分布特性较好，所以未经预处理而直接采用 ISODATA 算法取得的聚类结果较好，估计得到的混合矩阵精度较高。但是，当混合矩阵的某两个列向量距离太近时，为了分辨出距离较近的两个类，需要将类内标准差上界 θ_s 的值取得足够小，否则较大的 θ_s 取值可能会导致类的欠分裂。但大量野点的存在使得同一个类中样本点的分布较为分散，需要 θ_s 的取值足够大，否则较小的 θ_s 取值可能会导致类的过度分裂。因此，为了进一步提高算法稳定性，在接下来的实验中加入预处理过程。

将4路源信号经线性混合得到3路混合信号，混合矩阵如下：

$$\mathbf{A} = \begin{bmatrix} 0.7151 & 0.6970 & 0.3625 & 0.6376 \\ 0.3283 & 0.4378 & 0.3381 & 0.7335 \\ 0.6171 & 0.5679 & 0.8685 & 0.2354 \end{bmatrix}$$

混合信号样本点的分布如图4.4(a)所示。设定预处理参数 $\rho = 0.05$, $N_p = 20$, 去除较小幅值点的阈值参数 $\delta = 0.25$, ISODATA算法控制参数 $C = 4$, $N_c = 1$, $\theta_n = 20$, $\theta_s = 0.05$, $\theta_d = 0.04$, $L = 1$, $I = 30$, 聚类结果偏差控制阈值 $\varepsilon = 0.01$ 。ISODATA算法结果以及修正的聚类结果分别如图4.4(b)~4.4(d)所示, 得到的混合矩阵最终估计结果如下:

$$\hat{\mathbf{A}} = \begin{bmatrix} 0.7151 & 0.6972 & 0.3640 & 0.6380 \\ 0.3276 & 0.4367 & 0.3386 & 0.7327 \\ 0.6175 & 0.5685 & 0.8677 & 0.2367 \end{bmatrix}$$

这里采用估计矩阵与混合矩阵对应列向量之间的平均偏差距离来衡量估计效果的好坏, 其定义如下:

$$d_e = \frac{1}{N} \sum_{i=1}^N \|\hat{\mathbf{a}}_i - \mathbf{a}_i\| \quad (4.8)$$

式中, \mathbf{a}_i 和 $\hat{\mathbf{a}}_i$ 分别为混合矩阵 \mathbf{A} 与估计矩阵 $\hat{\mathbf{A}}$ 的列向量, N 为矩阵列数。由式(4.8)计算得到本次实验结果所得估计矩阵与混合矩阵对应列向量之间的平均偏差距离为 $d_e = 0.0014$ 。

为了进一步测试经过预处理及聚类结果修正的ISODATA算法的性能, 进行10次随机实验。每次实验选取不同的4路源信号混合得到3路观测信号, 每次实验的混合矩阵随机产生, 其矩阵元素值服从区间(0.2, 1)上的均匀分布, 并对各列进行模值归一化。设定预处理参数 $\rho = 0.02$, $N_p = 20$, 去除较小幅值点的阈值参数 $\delta = 0.25$, ISODATA算法控制参数 $C = 4$, $N_c = 1$, $\theta_n = 20$, $\theta_s = 0.02$, $\theta_d = 0.01$, $L = 1$, $I = 30$, 以式(4.8)定义的平均偏差距离 $d_e(\varepsilon)$ 来衡量估计效果, 其中 $d_e(\varepsilon)$ 表示取不同偏差控制阈值 ε 得到的 d_e 值。实验结果如表4.1所示。

由实验结果可以看出, 混合矩阵的估计结果是非常有效的, 且 ε 取值越小, 估计值与真实值之间的偏差越小, 精度越高。

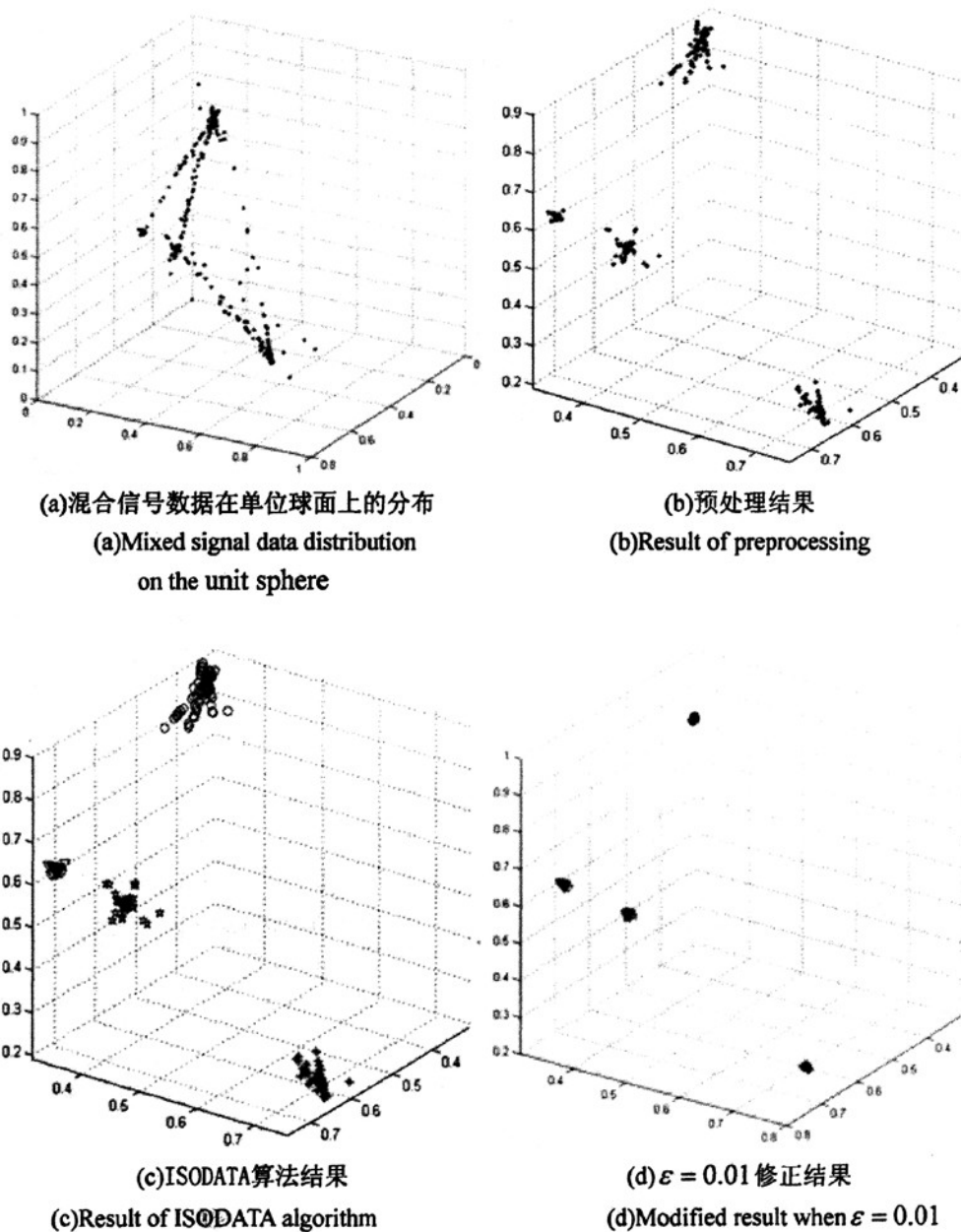


图4.4 加入预处理后的实验结果

Fig.4.4 The result of experiment with preprocessing

表 4.1 估值与真实值之间的偏差距离

Tab.4.1 Error distance between estimate value and real value

实验次数	$d_e(\infty)$	$d_e(0.02)$	$d_e(0.01)$
1	0.0009	0.0007	0.0006
2	0.0014	0.0013	0.0011
3	0.0047	0.0012	0.0007
4	0.0022	0.0016	0.0017
5	0.0018	0.0012	0.0006
6	0.0014	0.0014	0.0006
7	0.0007	0.0007	0.0007
8	0.0009	0.0009	0.0006
9	0.0018	0.0017	0.0012
10	0.0035	0.0007	0.0008
平均值	0.0019	0.0011	0.0009

4.6 本章小结

本章研究了基于聚类的欠定盲辨识方法，并采用ISODATA算法，对语音信号的欠定混合系统进行了有效的估计。该方法将所有混合信号的数据样本初始化为一个类，根据设定的控制参数自适应地对类进行分裂与合并，不需要预先知道源信号的精确个数，且该方法不依赖初值的选取，可随机选择某些样本点作为类中心的初始值。因此具有很大的灵活性。此外，本章还提出了对聚类结果的修正算法。通过消除野点的预处理方法，提高了聚类算法的稳定性。通过逐步去除类中最大偏差点并更新聚类中心的方法，提高了混合矩阵估计结果的精度。实验结果说明了该方法的有效性。

5 基于逐层分离的欠定语音盲分离方法

设源信号的个数为 N ，混合信号的个数为 M ， $M < N$ 。则重写欠定瞬时混合模型在频域的表达式如式 (5.1) 所示：

$$\mathbf{X}(\tau, \omega) = \mathbf{A} \mathbf{S}(\tau, \omega) \quad (5.1)$$

其中

$$\mathbf{X}(\tau, \omega) = \begin{bmatrix} X_1(\tau, \omega) \\ X_2(\tau, \omega) \\ \vdots \\ X_M(\tau, \omega) \end{bmatrix}, \mathbf{S}(\tau, \omega) = \begin{bmatrix} S_1(\tau, \omega) \\ S_2(\tau, \omega) \\ \vdots \\ S_N(\tau, \omega) \end{bmatrix}, \mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1N} \\ a_{21} & a_{22} & \cdots & a_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ a_{M1} & a_{M2} & \cdots & a_{MN} \end{bmatrix}$$

这里， $\mathbf{X}(\tau, \omega)$ 为混合信号向量， $\mathbf{S}(\tau, \omega)$ 为源信号向量， \mathbf{A} 为混合矩阵， \mathbf{A} 中任取 M 个列向量组成的向量组都是一个线性无关向量组。 \mathbf{A} 中的元素 a_{ij} 代表了源信号的相对幅值大小，通常取值为正实数。

5.1 语音信号的时频特性分析

5.1.1 语音信号的近似正交不重叠性质

如果源信号在时频域上两两互不重叠，则称源信号严格满足正交不重叠 (W-Disjoint Orthogonality, W-DO) ^[10] 条件，由此可在时频域中构造与各个源信号对应的二值模板，将源信号从混合信号中提取分离出来。因此，对于满足 W-DO 条件的各种类型的源信号，其欠定盲源分离问题将能得到有效解决。

Yilmaz 和 Rickard 通过研究发现，语音信号在时频域上能够近似满足 W-DO 条件^[47]。

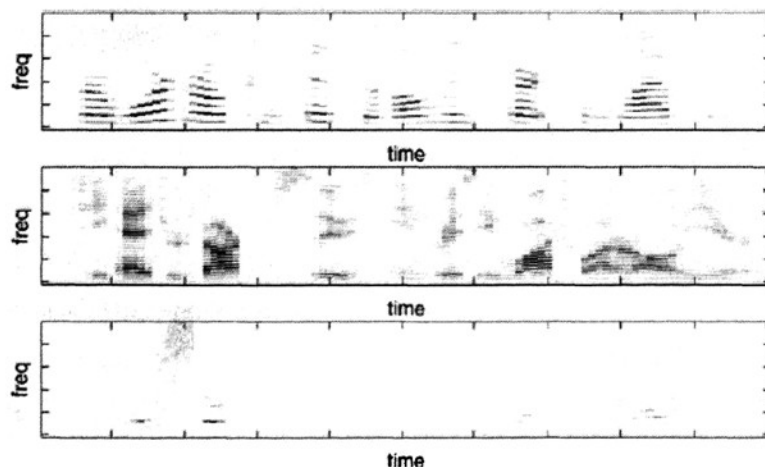


图 5.1 语音信号满足 W-DO 条件的性能

Fig. 5.1 The quality of speech satisfying the W-DO condition

从图 5.1 中上面和中间两图分别表示两个不同语音信号 $s_1(t)$ 和 $s_2(t)$ 在时频域的能量分布 $|S_1(\tau, \omega)|$ 、 $|S_2(\tau, \omega)|$ ，下图则表示 $|S_1(\tau, \omega)S_2(\tau, \omega)|$ 的能量分布。可以看出， $|S_1(\tau, \omega)S_2(\tau, \omega)|$ 所含能量成分远远小于 $|S_1(\tau, \omega)|$ 和 $|S_2(\tau, \omega)|$ 各自所含能量。通过进一步的研究分析发现，含有 s_1 的 90% 能量的时频点上只含有 s_2 中 1.1% 的能量，而含有 s_2 的 90% 能量的时频点上只含有 s_1 中 0.6% 的能量^[10]。由此说明语音信号具有近似满足 W-DO 条件的特性，因此可以考虑利用该特性来解决语音信号的欠定盲源分离问题。但解决效果的好坏，与其对 W-DO 条件的满足程度密切相关。

5.1.2 语音信号正交性能好坏对语音分离效果的影响

如果源信号在时频域上两两互不重叠，则称源信号满足 W-DO 条件。由此可在时频域中构造与各个源信号对应的二值掩蔽模板，将源信号从混合信号中提取分离出来。而语音信号并非严格满足 W-DO 条件，那么在时频域中的某些时频点上，语音信号之间必定存在着重叠，由此构造模板直接提取语音源信号必将引入语音间的相互干扰。

将源信号在时频域上的分布区域分成两部分，即非重叠的区域和存在重叠的区域。在非重叠区域中，源信号严格满足 W-DO 条件，理论上而言，可以通过构造二值模板从分布于该区域上的混合信号中精确提取出各个源信号；在重叠区域中，源信号不满足 W-DO 条件，无法直接从混合信号中精确提取出源信号。因此源信号之间相互干扰的强弱由源信号本身对 W-DO 条件满足的程度决定。对 W-DO 条件满足得越好，则各个源信号的能量在非重叠区域中分布得越多，源信号之间的干扰越小，分离效果越好；反之，源信号在重叠区域中的能量分布越多，信号之间的干扰越大，分离效果越差。

5.2 提取源信号的逐层分离法

5.2.1 按源信号重叠个数对时频域进行分层

本文根据源信号在时频域的重叠情况对时频点进行划分，以层数表示，同一层上每一个时频点重叠的源信号个数相同，且重叠源信号的个数越多，层数越高。本文假设，第 1 层上，每一个时频点只含有 1 个源信号；第 2 层上，每一个时频点是 2 个源信号叠加；以此类推，第 M 层上，每一个时频点是 M 个源信号的叠加。在同一层时频点上，根据所重叠的源信号的不同，又可将同一层时频点分为不同的类，同一类时频点是相同源信号的叠加。以第 $m(1 \leq m \leq M-1)$ 层时频点为例，可将其分为 C_N^m 类，每一类时频点上重叠了 m 个相同组合的源信号。

5.2.2 语音源信号的逐层分离

在明确了时频域划分的思路以后，需要设计一个对源信号的重叠情况进行划分和标记的方法，以便于后续处理中对源信号进行分离。由于各个时频点上有哪些源信号的重叠事先未知，因此，需要利用观测信号及混合矩阵的信息来对时频点进行划分。

时频点划分的整个过程，是从寻找第1层时频点开始，由低层向高层逐次推进，分层获得的。而找出第 m 层中各类时频点的分布，需要以前面 $m-1$ 层的划分结果为基础。

下面以寻找第1层的时频点为例，来说明如何找到各层的时频点及其分类方法。基本思路是采用了消源找零（消去源信号找出零点）的方法，来判定时频点所在的层数，并根据源信号的重叠情况对其进行分类。

首先，在式(5.1)两边分别左乘一个 $M \times M$ 阶的变换矩阵 $\mathbf{L}^{(i)}$ ， $1 \leq i \leq N$ ，从原始的混合信号中消去某个源信号 $S_i(\tau, \omega)$ ，即

$$\mathbf{L}^{(i)}\mathbf{X}(\tau, \omega) = \mathbf{L}^{(i)}\mathbf{A}\mathbf{S}(\tau, \omega) \quad (5.2)$$

若令 $\mathbf{X}^{(i)}(\tau, \omega) = \mathbf{L}^{(i)}\mathbf{X}(\tau, \omega)$ ， $\mathbf{A}^{(i)} = \mathbf{L}^{(i)}\mathbf{A}$ ，则可以得到式(5.3)

$$\mathbf{X}^{(i)}(\tau, \omega) = \mathbf{A}^{(i)}\mathbf{S}(\tau, \omega) \quad (5.3)$$

从式(5.2)可知，在混合信号 $X_2^{(i)}(\tau, \omega), \dots, X_M^{(i)}(\tau, \omega)$ 中，源信号 $S_i(\tau, \omega)$ 的成分已被消去，那么， $X_2^{(i)}(\tau, \omega), \dots, X_M^{(i)}(\tau, \omega)$ 中新增的零值点可以认为是只含有源信号 $S_i(\tau, \omega)$ 的时频点。由此可直接根据零值点构造二值掩蔽模板 $Mask^{(i)}(\tau, \omega)$ 如下

$$Mask^{(i)}(\tau, \omega) = \begin{cases} 1, & (\tau, \omega) \in \left(\bigcap_{m=2}^M \{(\tau, \omega) | X_m^{(i)}(\tau, \omega) = 0\} \right) \\ 0, & otherwise \end{cases} \quad (5.4)$$

由于对应于二值掩蔽模板 $Mask^{(i)}(\tau, \omega)$ 中值为1的时频点 (τ, ω) ，只含有第 i 个源信号，因此，在时频域中，可以用该模板将源信号 $S_i(\tau, \omega)$ 与其它源信号没有重叠的部分从混合信号中提取出来。由(5.1)式可知

$$X_1(\tau, \omega) \cdot Mask^{(i)}(\tau, \omega) = a_{1i} S_i(\tau, \omega) \cdot Mask^{(i)}(\tau, \omega)$$

用 $S_i^{(1)}(\tau, \omega)$ 表示由第1层时频点分离出的第 i 个信号，则可解得第1层时频点上各源信号为：

$$S_i^{(1)}(\tau, \omega) = S_i(\tau, \omega) \cdot Mask^{(i)}(\tau, \omega) = \frac{1}{a_{1i}} X_1(\tau, \omega) \cdot Mask^{(i)}(\tau, \omega), \quad i = 1, 2, \dots, N \quad (5.5)$$

第1层分离完成以后，为了避免重复分离，需要标记已分离的点。因此，构造二值屏蔽模板 $Mask_1(\tau, \omega)$ 如式(5.6)所示

$$Mask_1(\tau, \omega) = \begin{cases} 1, & (\tau, \omega) \in \left(\bigcup_{i=1}^N \{(\tau, \omega) | Mask^i(\tau, \omega) = 1\} \right) \\ 0, & otherwise \end{cases} \quad (5.6)$$

模板 $Mask_1(\tau, \omega)$ 中, 值为 1 的点 (τ, ω) 只含有一个源信号, 并且这些源信号已经被分离出来, 在以后各层的时频点及源信号的分离过程中, 这些点将不再参与。至此, 第 1 层时频点的寻找以及第 1 层时频点上源信号的分离全过程结束。

接下来进一步介绍第 2 层时频点的寻找方法及这些时频点上所叠加源信号的分离过程。这里需要以第 1 层处理所得的一些计算结果为基础, 采用类似的方式来做进一步处理。

首先, 将第 1 层处理后得到的混合信号向量 $\mathbf{X}^{(i)}$ 及混合矩阵 $\mathbf{A}^{(i)}$ 保留下来, 这里 $i=1, 2, \dots, N$ 。在式(5.3)两边同乘以变换矩阵 $\mathbf{L}^{(ij)}$:

$$\mathbf{L}^{(ij)} \mathbf{X}^{(i)}(\tau, \omega) = \mathbf{L}^{(ij)} \mathbf{A}^{(i)} \mathbf{S}(\tau, \omega)$$

若令 $\mathbf{X}^{(ij)}(\tau, \omega) = \mathbf{L}^{(ij)} \mathbf{X}^{(i)}(\tau, \omega)$, $\mathbf{A}^{(ij)} = \mathbf{L}^{(ij)} \mathbf{A}^{(i)}$, 则可得到:

$$\mathbf{X}^{(ij)}(\tau, \omega) = \mathbf{A}^{(ij)} \mathbf{S}(\tau, \omega) \quad (5.7)$$

式中上标 (ij) 表示寻找只存在第 i 个和第 j 个源信号重叠的时频点。从式(5.7)中可知混合信号 $X_3^{(ij)}(\tau, \omega), \dots, X_M^{(ij)}(\tau, \omega)$ 中源信号 $S_i(\tau, \omega)$ 和 $S_j(\tau, \omega)$ 已经被消去, 在 $X_3^{(ij)}(\tau, \omega), \dots, X_M^{(ij)}(\tau, \omega)$ 中新增的零值点可以认为是只含有源信号 $S_i(\tau, \omega)$ 和 $S_j(\tau, \omega)$ 的成分叠加的时频点。由此可直接根据新增零值点并结合上一层处理所得屏蔽模板 $Mask_1(\tau, \omega)$ 构造二值掩蔽模板 $Mask^{(ij)}(\tau, \omega)$ 如下:

$$Mask^{(ij)}(\tau, \omega) = \begin{cases} 1, & (\tau, \omega) \in \left(\bigcap_{m=3}^M \{(\tau, \omega) | X_m^{(ij)}(\tau, \omega) = 0\} \right) \cap \{(\tau, \omega) | Mask_1(\tau, \omega) = 0\} \\ 0, & otherwise \end{cases} \quad (5.8)$$

对应于二值掩蔽模板 $Mask^{(ij)}(\tau, \omega)$ 中值为 1 的时频点 (τ, ω) 上, 只含第 i 个源信号和第 j 个源信号叠加的成分。因此可用该模板将只含源信号 $S_i(\tau, \omega)$ 和 $S_j(\tau, \omega)$ 叠加的成分从混合信号中提取出来, 由式(5.1)可知:

$$\begin{bmatrix} S_i(\tau, \omega) \cdot Mask^{(ij)}(\tau, \omega) \\ S_j(\tau, \omega) \cdot Mask^{(ij)}(\tau, \omega) \end{bmatrix} = \begin{bmatrix} a_{1j} & a_{1i} \\ a_{2j} & a_{2i} \end{bmatrix}^{-1} \begin{bmatrix} X_1(\tau, \omega) \cdot Mask^{(ij)}(\tau, \omega) \\ X_2(\tau, \omega) \cdot Mask^{(ij)}(\tau, \omega) \end{bmatrix} \quad (5.9)$$

将 i, j 的取值遍历 $1, 2, \dots, N$, 则可找出所有只含源信号 i 与源信号 j 相叠加的时频点。为了避免 i 和 j 在组合上的重复取值, 规定 i 和 j 在每次取值时满足 $i < j$ 。用 $S_i^{(2)}(\tau, \omega)$

表示由第 2 层时频点分离出的第 i 个信号，这样可求出第 2 层时频点上源信号的分离结果如下：

$$S_i^{(2)}(\tau, \omega) = \sum_{n=1}^{i-1} S_i(\tau, \omega) \cdot Mask^{(n,i)}(\tau, \omega) + \sum_{n=i+1}^N S_i(\tau, \omega) \cdot Mask^{(i,n)}(\tau, \omega), \quad i = 1, 2, \dots, N \quad (5.10)$$

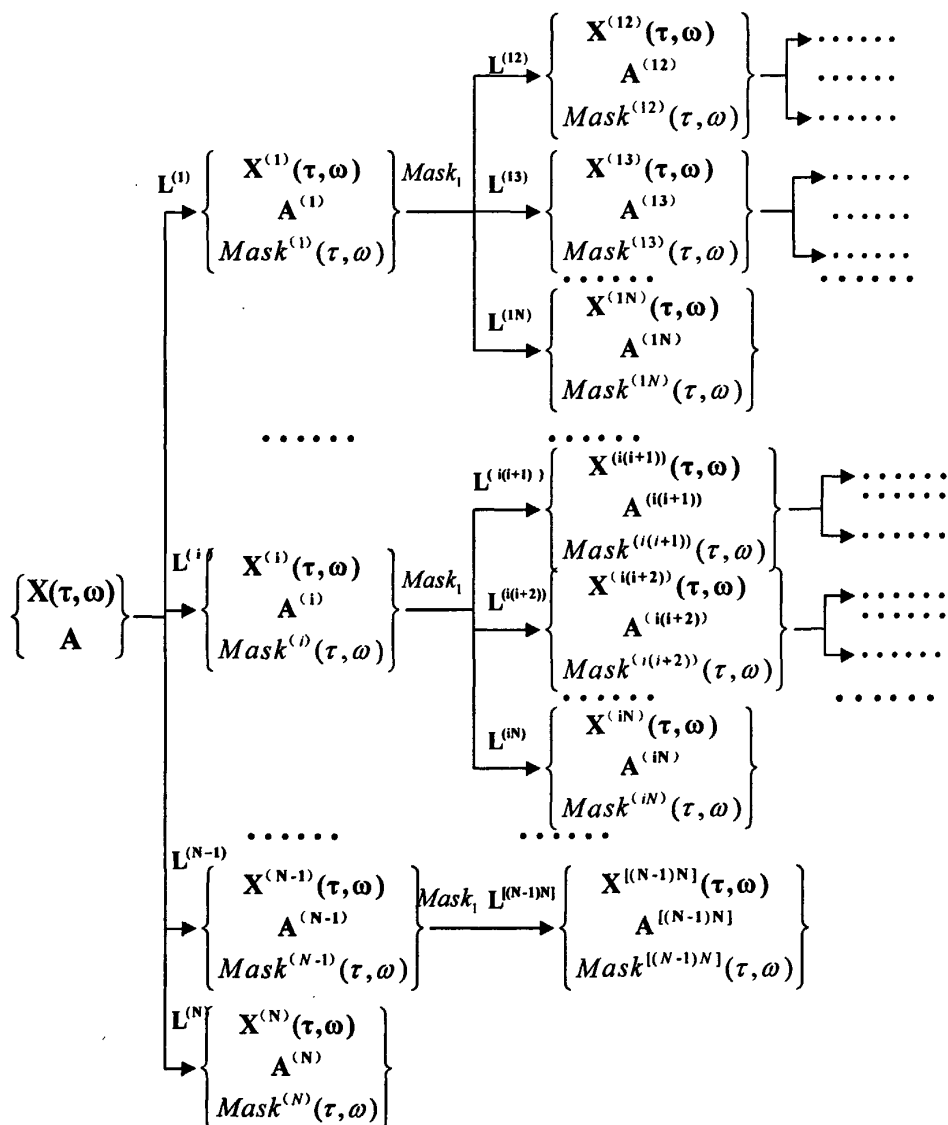


图 5.2 源信号逐层分离的流程图

Fig.5.2 Flow of stepwise separation

第 2 层的分离处理完成以后, 接下来还需要构造屏蔽二值模板 $Mask_2(\tau, \omega)$:

$$Mask_2(\tau, \omega) = \begin{cases} 1, & (\tau, \omega) \in \bigcup_{i=1}^{N-1} \bigcup_{j=2}^N \{(\tau, \omega) | Mask^{(ij)}(\tau, \omega) = 1\} \cup \{(\tau, \omega) | Mask_1(\tau, \omega) = 1\} \\ & (i < j) \\ 0, & otherwise \end{cases} \quad (5.11)$$

$Mask_2(\tau, \omega)$ 的值为 1 的点 (τ, ω) 上所叠加的源信号已被分离出来, 因此在接下来寻找更高层数的时频点及源信号的分离过程将在 $Mask_2(\tau, \omega)$ 的值为 0 的点 (τ, ω) 上进行。

接下来仿照第 1、2 层的处理方式, 继续分离更高层时频点上的源信号, 整个流程如图 5.2 所示。

当算法进行到第 k 层时, 已知 M 为观测信号个数, N 为源信号个数, 则 $k < M < N$ 。 n_1, \dots, n_k 对应着被消去的源信号 $S_{n_1}(\tau, \omega), \dots, S_{n_k}(\tau, \omega)$, 且有如下关系:

$$n_1 < n_2 < \dots < n_k \quad (5.12)$$

其中

$$1 < k < M, \quad n_1, \dots, n_k \in \{1, 2, \dots, N\}$$

混合矩阵 $\mathbf{A}^{(n_1 \dots n_k)}$ 代表消去了源信号 $S_{n_1}(\tau, \omega), \dots, S_{n_k}(\tau, \omega)$ 之后的混合矩阵, 其对应的矩阵元素表达方式如下:

$$\mathbf{A}^{(n_1 \dots n_k)} = [a_{ij}^{(n_1 \dots n_k)}], (i = 1, 2, \dots, M; j = 1, 2, \dots, N) \quad (5.13)$$

消去源信号 $S_{n_1}(\tau, \omega), \dots, S_{n_k}(\tau, \omega)$ 的变换矩阵 $\mathbf{L}^{(n_1 \dots n_k)}$ 定义如下:

$$\mathbf{L}^{(n_1 \dots n_k)} = \begin{bmatrix} l_{11} & & & & \\ & \ddots & & & \\ & & l_{kk} & & \\ & & -\frac{a_{(k+1)n_k}^{(n_1 \dots n_{k-1})}}{a_{kn_k}^{(n_1 \dots n_{k-1})}} l_{(k+1)(k+1)} & & \\ & & \vdots & \ddots & \\ & & -\frac{a_{Mn_k}^{(n_1 \dots n_{k-1})}}{a_{kn_k}^{(n_1 \dots n_{k-1})}} & \dots & l_{MM} \end{bmatrix} \quad (5.14)$$

其中

$$l_{ii} = 1, (i = 1, 2, \dots, M)$$

第k层的混合矩阵 $\mathbf{A}^{(n_1 \cdots n_k)}$ 是由第k-1层的混合矩阵 $\mathbf{A}^{(n_1 \cdots n_{k-1})}$ 经变换矩阵 $\mathbf{L}^{(n_1 \cdots n_k)}$ 得到, 其关系如下:

$$\mathbf{A}^{(n_1 \cdots n_k)} = \mathbf{L}^{(n_1 \cdots n_k)} \mathbf{A}^{(n_1 \cdots n_{k-1})} \quad (5.15)$$

第k层的混合信号向量表达式如下:

$$\mathbf{X}^{(n_1 \cdots n_k)}(\tau, \omega) = \begin{bmatrix} X_1(\tau, \omega) \\ X_2^{(n_1)}(\tau, \omega) \\ \vdots \\ X_k^{(n_1 \cdots n_{k-1})}(\tau, \omega) \\ X_{k+1}^{(n_1 \cdots n_k)}(\tau, \omega) \\ \vdots \\ X_M^{(n_1 \cdots n_k)}(\tau, \omega) \end{bmatrix} \quad (5.16)$$

与前面类似, $\mathbf{X}^{(n_1 \cdots n_k)}(\tau, \omega)$ 表示消去了源信号 $S_{n_1}(\tau, \omega), \dots, S_{n_k}(\tau, \omega)$ 的混合信号向量, 其分量 $X_1(\tau, \omega)$ 包含了所有源信号的成分, $X_2^{(n_1)}(\tau, \omega)$ 中消去了源信号 $S_{n_1}(\tau, \omega)$, $X_3^{(n_1 n_2)}(\tau, \omega)$ 中消去了源信号 $S_{n_1}(\tau, \omega)$ 和 $S_{n_2}(\tau, \omega)$, 以此类推, 分量 $X_k^{(n_1 \cdots n_{k-1})}(\tau, \omega)$ 中消去了源信号 $S_{n_1}(\tau, \omega), \dots, S_{n_{k-1}}(\tau, \omega)$, 而从 $X_{k+1}^{(n_1 \cdots n_k)}(\tau, \omega)$ 开始一直到 $X_M^{(n_1 \cdots n_k)}(\tau, \omega)$ 的分量中则消去了相同的源信号 $S_{n_1}(\tau, \omega), \dots, S_{n_k}(\tau, \omega)$ 。

第k层的混合信号向量 $\mathbf{X}^{(n_1 \cdots n_k)}(\tau, \omega)$ 与第k-1层的观测信号向量 $\mathbf{X}^{(n_1 \cdots n_{k-1})}(\tau, \omega)$ 之间的关系如下:

$$\mathbf{X}^{(n_1 \cdots n_k)}(\tau, \omega) = \mathbf{L}^{(n_1 \cdots n_k)} \mathbf{X}^{(n_1 \cdots n_{k-1})}(\tau, \omega) \quad (5.17)$$

由向量 $\mathbf{X}^{(n_1 \cdots n_k)}(\tau, \omega)$ 中的第k+1个分量到第M个分量 $X_i^{(n_1 \cdots n_k)}(\tau, \omega), (i = k+1, \dots, M)$ 构造模板如下:

$$\begin{aligned} & \text{Mask}^{(n_1 \cdots n_k)}(\tau, \omega) \\ &= \begin{cases} 1, & (\tau, \omega) \in \bigcup_{i=k+1}^M \{(\tau, \omega) | X_i^{(n_1 \cdots n_k)}(\tau, \omega) = 0\} \cap \{(\tau, \omega) | \text{Mask}_{k-1}(\tau, \omega) = 0\} \\ 0, & \text{otherwise} \end{cases} \end{aligned} \quad (5.18)$$

$\text{Mask}^{(n_1 \cdots n_k)}(\tau, \omega)$ 用于提取源信号 $S_{n_1}(\tau, \omega), \dots, S_{n_k}(\tau, \omega)$ 在时频域上的重叠区域, 由式(5.1)可知:

$$\begin{bmatrix} X_1(\tau, \omega) \cdot \text{Mask}^{(n_1 \cdots n_k)}(\tau, \omega) \\ \vdots \\ X_k(\tau, \omega) \cdot \text{Mask}^{(n_1 \cdots n_k)}(\tau, \omega) \end{bmatrix} = \begin{bmatrix} a_{1n_1} & \cdots & a_{1n_k} \\ \vdots & \ddots & \vdots \\ a_{kn_1} & \cdots & a_{kn_k} \end{bmatrix} \begin{bmatrix} S_{n_1}^{(n_1 \cdots n_k)}(\tau, \omega) \cdot \text{Mask}^{(n_1 \cdots n_k)}(\tau, \omega) \\ \vdots \\ S_{n_k}^{(n_1 \cdots n_k)}(\tau, \omega) \cdot \text{Mask}^{(n_1 \cdots n_k)}(\tau, \omega) \end{bmatrix}$$

解得：

$$\begin{bmatrix} S_{n_1}^{(n_1 \cdots n_k)}(\tau, \omega) \cdot \text{Mask}^{(n_1 \cdots n_k)}(\tau, \omega) \\ \vdots \\ S_{n_k}^{(n_1 \cdots n_k)}(\tau, \omega) \cdot \text{Mask}^{(n_1 \cdots n_k)}(\tau, \omega) \end{bmatrix} = \begin{bmatrix} a_{1n_1} & \cdots & a_{1n_k} \\ \vdots & \ddots & \vdots \\ a_{kn_1} & \cdots & a_{kn_k} \end{bmatrix}^{-1} \begin{bmatrix} X_1(\tau, \omega) \cdot \text{Mask}^{(n_1 \cdots n_k)}(\tau, \omega) \\ \vdots \\ X_k(\tau, \omega) \cdot \text{Mask}^{(n_1 \cdots n_k)}(\tau, \omega) \end{bmatrix} \quad (5.19)$$

将 n_1, \dots, n_k 的取值遍历 $1, 2, \dots, N$ ，则可找出所有只含源信号 n_1, \dots, n_k 相叠加的时频点。用 $S_i^{(k)}(\tau, \omega)$ 表示由第 k 层时频点分离出的第 i 个信号，这样可求出第 k 层时频点上各个源信号的分离结果如下：

$$S_i^{(k)}(\tau, \omega) = \sum_{n_1=1}^{N-k+1} \cdots \sum_{n_k=k}^N S_i(\tau, \omega) \cdot \text{Mask}^{(n_1 \cdots n_k)}(\tau, \omega), \quad i = 1, 2, \dots, N \quad (5.20)$$

第 k 层的分离处理完成以后，构造屏蔽模板：

$$\text{Mask}_k(\tau, \omega) = \begin{cases} 1, & (\tau, \omega) \in \bigcup_{n_1=1}^{N-k+1} \cdots \bigcup_{n_k=k}^N \{(\tau, \omega) | \text{Mask}^{(n_1 \cdots n_k)}(\tau, \omega) = 1\} \cup \{(\tau, \omega) | \text{Mask}_{k-1}(\tau, \omega) = 1\} \\ 0, & \text{otherwise} \end{cases} \quad (5.21)$$

$\text{Mask}_k(\tau, \omega)$ 用于屏蔽前 k 层已经处理过的点，保证第 $k+1$ 层待处理的点是前面没有处理过的。

假设处理到第 P ($0 < P < M$) 层算法流程结束，将同一个源信号在各层时频点上所得分离结果相加，得到各个源信号最终的分离结果如下：

$$S_i(\tau, \omega) = \sum_{k=1}^P S_i^{(k)}(\tau, \omega), \quad i = 1, 2, \dots, N \quad (5.22)$$

整个算法流程采用分层处理的方式，将源信号逐次分离。处理的层数可灵活控制，层数越多源信号最后分离的结果越精细，但相对的计算量也越大。该方法最多只能处理到 $M-1$ 层，即可以提取出存在 $M-1$ 个源信号叠加的时频点。就语音信号而言，随着处理层数的增加，能够提取出的时频点在急剧减少，其所含能量也变得微乎其微，因此根据实际精度要求，选择处理到足够多的层数即可终止算法，得到源信号的分离结果。

最后通过短时傅里叶变换，得到最终恢复的时域信号如下：

$$s_i(t) = \text{ISTFT}[S_i(\tau, \omega)], \quad i = 1, 2, \dots, N \quad (5.23)$$

5.3 计算机仿真实验结果

从语音库中选取采样率为 16kHz 的 4 路不同说话人的数字语音信号作为源信号，在时域将其进行线性混合。混合矩阵如下：

$$\mathbf{A} = \begin{bmatrix} 1.23 & 0.97 & 1.01 & 0.71 \\ 0.65 & 0.73 & 1.18 & 1.25 \\ 0.71 & 0.92 & 0.56 & 0.83 \end{bmatrix}$$

将得到的混合信号经过STFT变换到频域，STFT的时域信号帧长取为512点，帧间取256点交叠，频域分辨率取为1024点。

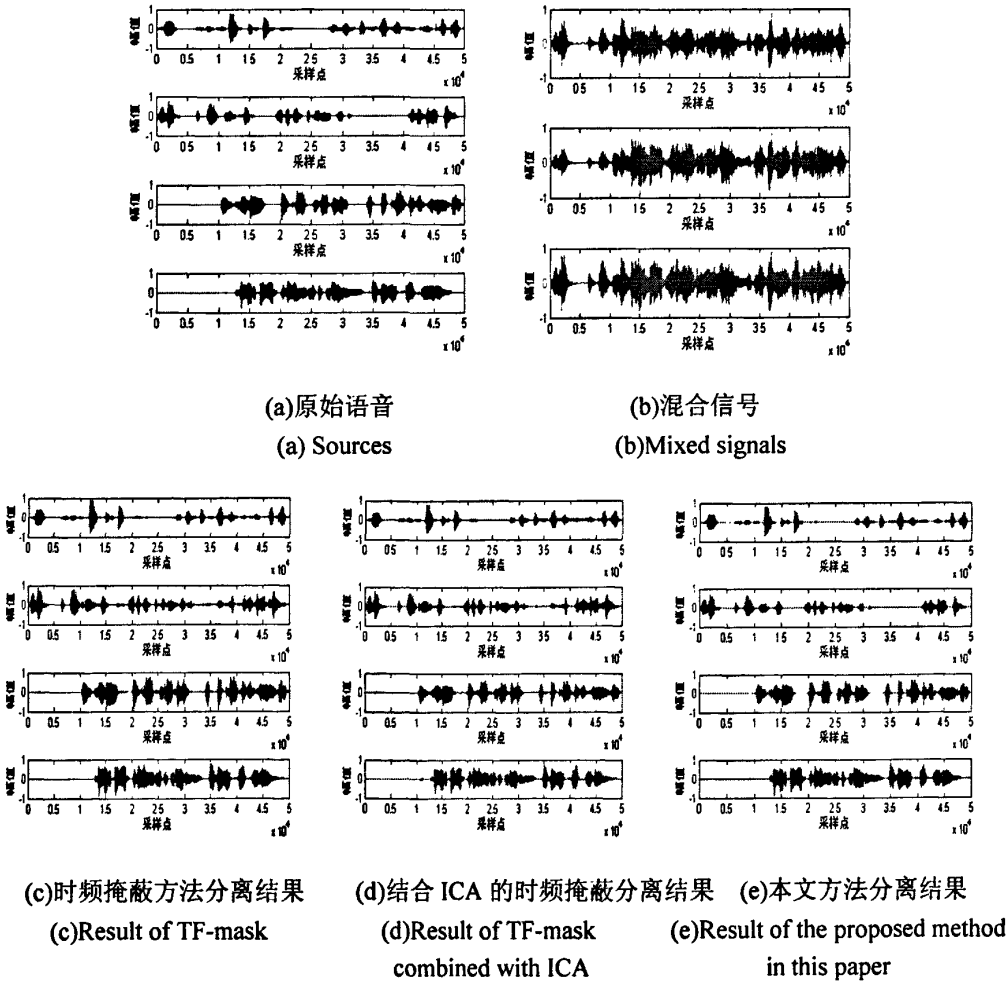


图5.3 欠定盲源分离结果

Fig.5.3 Result of underdetermined separation

在时频域中,分别采用时频掩蔽方法^[10]、结合ICA的时频掩蔽法^[26]以及本文方法从已知混合信号中分离出源信号。其中时频掩蔽算法所需的输入混合信号从已知混合信号中选取,本文方法所需的混合矩阵则是通过ISODATA聚类算法从混合信号中获得。三种方法分离出的源信号的时域波形图如图5.3所示。图5.3(a)为原始语音时域波形,图5.3(b)为混合信号时域波形。图5.3(c)、图5.3(d)、图5.3(e)依次为时频掩蔽算法、结合ICA的时频掩蔽算法以及本文算法所得语音分离的结果。从时域波形上看,本文方法所得分离结果更加接近原始语音。

另外,本文利用输出信噪比从客观评价角度来进一步衡量上述三种方法的分离效果,每一路分离信号的输出信噪比的计算由式(5.24)所示:

$$SNR_i = 10 \log_{10} \left(\frac{\sum_{n=1}^{N_s} |s_i(n)|^2}{\sum_{n=1}^{N_s} |s_i(n) - \hat{s}_i(n)|^2} \right) \quad i=1,2,\dots,N \quad (5.24)$$

式(5.24)中, $s_i(n)$ 为第 i 个原始语音信号, $\hat{s}_i(n)$ 为盲源分离输出得到的第 i 个原始语音信号的分离结果。 N_s 为信号的采样点数, N 为源信号个数。为了避免幅度不确定带来的影响,这里 $s_i(n)$ 和 $\hat{s}_i(n)$ 在进行 SNR_i 计算前都经过了平均功率归一化处理。上述三种方法得到分离结果的输出信噪比如表5.1所示。

从表5.1可以看出,采用本文方法可以使得分离出来的语音源信号的信噪比得到进一步的提升。对分离出的语音信号进行主观听音测试也表明,本文方法所得分离语音中所含的其它语音源信号的干扰更小,语音更为纯净,听音效果更好。

表 5.1 输出信噪比
Tab.5.1 SNR of output signal

输出信号 $SNR_i(dB)$	$\hat{s}_1(n)$	$\hat{s}_2(n)$	$\hat{s}_3(n)$	$\hat{s}_4(n)$	输出信号 平均SNR
时频掩蔽 方法	10.3540	6.4335	9.8311	12.4676	9.7716
结合ICA的 时频掩蔽法	11.7118	7.0303	8.4051	8.8002	8.9869
本文方法	11.2415	13.4128	13.1783	15.5361	13.3422

为了进一步测试算法的性能，下面进行10次实验测试，每次实验选取不同的4路源信号经线性混合得到3路混合信号，混合矩阵随机生成，其元素值的取值范围为0.2~1.0，且服从均匀分布。混合矩阵生成后需进行列向量模值归一化处理。每次实验由不同方法得到的输出信号平均信噪比值SNR(dB)如表5.2所示。

表 5.2 平均输出信噪比
Tab.5.2 Average SNR of output signal

实验次数	时频掩蔽方法	结合ICA的时频掩蔽法	本文方法
1	9.8439	9.9151	14.0200
2	8.7079	8.6349	10.9192
3	8.2987	8.1365	11.6749
4	9.0482	9.3876	11.2832
5	10.7404	10.4767	14.2920
6	7.7892	9.4108	10.4168
7	9.0954	9.0945	11.4366
8	7.5548	8.2635	9.4695
9	8.3300	7.4210	12.3754
10	7.9947	8.8715	11.2167

从表5.2可以看出，经过多次实验，本文方法都能进一步改善分离信号的信噪比。在每次实验中，通过听音测试也发现，本文方法所提取的源信号更为纯净，干扰更小。

5.4 本章小结

本章研究了语音信号欠定混合的分离方法。根据混合信号在时频域上的分布情况，本文提出从混合信号中逐个消除源信号，以产生的零值点来找到混合信号在各个时频点上源信号的叠加情况，并构造多层二值掩蔽模板，进而以线性方式分离求解各个源信号。从理论推导来看，该方法尽最大可能地找出了时频域上混合信号中各个源信号的分布及叠加情况，提高了分离的有效性。通过实验结果也发现，本文方法所得源信号分离结果无论是在客观输出信噪比还是主观评价方面，比之原有的时频掩蔽方法的分离效果，都有了进一步的改进。

结 论

本文研究了基于盲源分离技术的多语音分离语音增强方法。在实际环境中,当多个语音源信号从不同方位传递到麦克风阵列接收器时,接收到的语音信号将产生线性混叠,即每一个语音源信号中都混入了其它语音源信号的干扰。此时接收到的语音混合信号不仅在听觉上难以辨识其信息内容,而且在后续的语音编码或语音识别等语音处理环节中,这种混合信号也变得难以利用。针对上述问题,采用盲源分离方法可以较好地消除语音干扰,将纯净的原始语音从混合信号中提取分离出来。目前,人们对于盲源分离技术的研究已取得了很大进展,尽管如此,盲源分离技术在理论上还存在着许多不完善的地方,因此在实际应用中其局限性还比较大,有待于进一步的发展和完善。根据从麦克风阵列接收到的观测信号(即混合信号)与源信号相对个数的不同,盲源分离的解决方法又可分为非欠定盲源分离和欠定盲源分离两个方面。本文结合语音信号的特性,对这两方面进行了研究和方法改进。

本文工作的第一部分针对非欠定盲源分离问题,研究了快速 ICA 的算法及其改进方法。通过文中分析可知,由于信号样本的有穷性使得信号本身的统计特性无法通过集总均值得到,因此基于信号统计特征的快速 ICA 算法在分离源信号时必将产生误差,而该误差尚未达到理论上的极小值。通过对快速 ICA 算法的结果进行修正而得到的改进有效快速 ICA 算法可以使得信号分离误差得以减小,并接近于理论极小值。而结合了语音信号短时平稳性的分段有效快速 ICA 算法则充分利用了信号本身的时变统计特性,进一步提高了源信号分离的精度。本文通过实验仿真,验证了上述改进方法的有效性。

本文工作的第二部分研究了欠定盲源分离中的混合矩阵盲辨识问题。在欠定盲辨识中,利用信号的稀疏性,采用聚类算法来进行混合矩阵的估计是一类较为常用的方法。而以往用于矩阵估计的 k-means 聚类及其改进方法需要预先知道源信号的个数。本文给出了一种采用 ISODATA 算法来进行混合矩阵估计的方法,该方法不需要预先知道源信号的个数,且对于迭代初值的选取不敏感,通过设定合适的控制参数,能够自适应的进行聚类并求出混合矩阵。本文还提出了对混合信号样本点进行预处理的方法以及逐步去除最大偏差点的聚类结果修正方法,改善了 ISODATA 算法在盲辨识聚类中的稳定性,并进一步提高了对混合矩阵的估计精度。

本文工作的第三部分研究了语音信号欠定混合的分离问题。在欠定问题的解决方法中,比较经典的时频掩蔽方法利用源信号在时频域所具有的近似正交不重叠特性,以极大似然方法构造二值掩蔽模板,直接从观测信号中提取出源信号。但语音信号本身不满足严格正交不重叠条件,会使得分离出的源信号之间存在着较为明显的相互干扰和音乐

噪声。本文针对这一问题提出了从混合信号中逐步消去各个源信号，并由各混合信号中逐次产生的零值点来构造多层二值掩蔽模板的方法，将源信号进行逐层分离。该方法有效抑制了分离出的语音源信号之间的相互干扰及音乐噪声，进一步改进了语音源信号恢复的质量。

由于时间及个人水平有限，本文所研究的改进方法以及提出的新算法还存在着较大的局限性，有待于进一步的改进。目前盲源分离技术在源信号非欠定瞬时混合情况下的研究已取得了很大的进展，并已逐步应用于实际工作中。而对于更具挑战性的欠定混合以及卷积混合情况下的盲源分离技术的研究目前还处于起步阶段，其理论体系还很不完善，需要不断深入研究和改进，是盲源分离技术今后的发展方向之一。此外，为了更加有效地应用于实际环境中，研究低信噪比以及非线性混合情况下的盲源分离问题也是今后的一大发展趋势。语音分离及语音增强技术也将随着盲源分离技术的不断完善而得到更加有效地发展。

参 考 文 献

- [1] 胡航. 语音信号处理[M]. 哈尔滨:哈尔滨工业大学出版社, 2005.
- [2] 赵力. 语音信号处理[M]. 北京:机械工业出版社, 2003.
- [3] Cho J, Crishnamurthy A. Speech enhancement using microphone array in moving vehicle environment[C]. Proceedings of Intelligent Vehicles Symposium, Columbus, USA, 2003: 366-371.
- [4] Frost O L. An algorithm for linearly constrained adaptive array processing[J]. Proceedings of the IEEE, 1972, 60(8): 926-935.
- [5] Griffiths L J, Jim C W. An alternative approach to linearly constrained adaptive beamforming[J]. IEEE Transactions on Antennas and Propagation, 1982, 30(1): 27-34.
- [6] Hoshuyama O, Sugiyama A, Hirano A. A robust adaptive beamformer for microphone arrays with a blocking matrix using constrained adaptive filters[J]. IEEE Transactions on Signal Processing, 1999, 47(10): 2677-2684.
- [7] Gannot S, Cohen I. Speech enhancement based on the general transfer function GSC and postfiltering[J]. IEEE Transactions on Speech and Audio Processing, 2004, 12(6): 561-571.
- [8] 马建仓, 牛奕龙, 陈海洋. 盲信号处理[M]. 北京:国防工业出版社, 2006.
- [9] Hyvarinen A. Survey on independent component analysis[J]. Neural Computing Surveys, 1999, 2: 94-128.
- [10] Yilmazand O, Rickard S. Blind separation of speech mixtures via time-frequency masking[J]. IEEE Transactions on Signal Processing, 2004, 52(7): 1830 - 1847.
- [11] 杨福生, 洪波. 独立分量分析的原理与应用[M]. 北京: 清华大学出版社, 2006.
- [12] Herault J, Jutten C. Space or time adaptive signal processing by neural network models[C]. Proceedings of the Conference on Neural Networks in Computing, Snowbird, Utah, USA, 1986:206-211.
- [13] Linsker R. An Application of the principle of maximum information preservation to linear systems[C]. Neural Information Processing Systems Conference, Denver, USA, 1989: 186-194.
- [14] Cardoso J F. Source separation using higher order moments[C]. International Conference on Acoustics, Speech, and Signal Processing, Glasgow, UK, 1989: 2109-2112.
- [15] Gaeta M, Lacoume J L. Source separation without prior knowledge: the maximum likelihood solution[C]. Fifth European Signal Processing Conference, Barcelona, Spain, 1990: 621-624.

- [16]Comon P. Independent component analysis, A new concept[J]. Signal Processing, 1994, 36: 287-314.
- [17]Cichocki A, Unbehauen R, Moszczynski L, et al. A new on-line adaptive learning algorithm for blind separation of source signal[C]. International Symposium on Artificial Neural Networks, Tainan, Taiwan, 1994: 406-411.
- [18]Bell A J, Sejnowski T J. An information maximization approach to blind separation and blind deconvolution[J]. Neural Computation, 1995, 7: 1129 - 1159.
- [19]Cardoso J F. Infomax and maximum likelihood for blind source separation[J]. IEEE Signal Processing Letters, 1997, 4(4): 112-114.
- [20]Belouchrani A, Karim A M, Cardoso J F, et al. A blind source separation technique using second-order statistics[J]. IEEE Transactions on Signal Processing, 1997, 45(2): 434-444.
- [21]Hyvarinen A, Oja E. A fast fixed-point algorithm for independent component analysis[J]. Neural Comput, 1997, 9(7): 1483-1492.
- [22]Hyvarinen A. Fast and robust fixed-point algorithms for independent component analysis[J]. IEEE Transactions on Neural Network, 1999, 10(3): 626-634.
- [23]Hyvarinen A, Karhunen J, Oja E. Independent component analysis[M]. New York: John Wiley and Sons, 2001.
- [24]Bingham E, Hyvarinen A. A fast fixed-point algorithm for independent component analysis of complex Valued Signals[J]. Neural System, 2000, 10(1): 1-8.
- [25]Stone J V. Blind source separation using temporal predictability[J]. Neural Computation, 2001, 13(7): 1559-1574.
- [26]Cheung Y M, Liu H L. A new approach to blind source separation with global optimal property[C]. Proceedings of the IASTED International Conference on Neural Networks and Computational Intelligence, Grindelwald, Switzerland , 2004: 137-141.
- [27]Tichavsky P, Koldovsky Z, Oja E. Asymptotic performance of the FastICA algorithm for independent component analysis and its improvements[C]. IEEE Statistical Signal Processing Workshop, Bordeaux, France, 2005: 1084-1089.
- [28]Lewicki M S, Sejnowski T J. Learning overcomplete representations[J]. Neural Computation, 1998, 12: 337 - 365.
- [29]Jourjine A, Rickard S, Yilmaz O. Blind separation of disjoint orthogonal signals: demixing N sources from 2 mixtures[C]. IEEE International Conference on Acoustics, Speech, and Signal Processing, Istanbul, Turkey, 2000: 2985 - 2988.
- [30]Thomas J, Deville Y, Hosseini S. Differential fast fixed-point algorithms for underdetermined instantaneous or convolutive partial blind source separation[J], IEEE Transactions on Signal Processing, 2007, 55(7): 3717-3729.

- [31]Mitianoudis N, Stathaki T. Batch and online underdetermined source separation using laplacian mixture models[J]. IEEE Transactions on Audio, Speech and Language Processing, 2007, 15(6): 1818-1832.
- [32]Pedersen M S, Wang D L, Larsen J, Kjems U. Two-microphone separation of speech mixtures[J], IEEE Transactions on Neural Networks, 2008, 19(3): 475-492.
- [33]Lathauwer L D, Castaing J. Blind identification of underdetermined mixtures by simultaneous matrix diagonalization[J], IEEE Transactions on Signal Processing, 2008, 56(3): 1096-1105.
- [34]刘小华. 语音分离和语音增强方法研究[D]: (硕士学位论文). 大连:大连理工大学, 2007.
- [35]马晓红. 传声器阵列语音增强中关键技术的研究[D]: (博士学位论文). 大连:大连理工大学, 2006.
- [36]胡广书. 数字信号处理: 理论、算法与实现[M]. 北京:清华大学出版社, 2003.
- [37]Tichavsky P, Koldovsky Z, Oja E. Performance analysis of the FastICA algorithm and cramer-rao bounds for linear independent component analysis[J]. IEEE Transactions on Signal Processing, 2006, 54(4): 1189 - 1203.
- [38]Koldovsky Z, Tichavsky P, Oja E. Efficient variant of algorithm FastICA for independent component analysis attaining the cramer-rao lower bound[J]. IEEE Transactions on Neural Networks, 2006, 17(5): 1265-1277.
- [39]Koldovsky Z, Malek J, Tichavsky P, et al. Blind separation of piecewise stationary nongaussian sources[J]. Signal Processing, 2009, 89: 2570-2584.
- [40]Pham D T, Garat P, Blind separation of mixture of independent sources through a quasi-maximum likelihood approach[J], IEEE Transactions on Signal Processing, 1997, 45(7): 1712 - 1725.
- [41]Zibulevsky M, Kisilev P, Zeevi Y Y, et al. Blind source separation via multinode sparse representation[C]. Advances in Neural Information Processing Systems, Vancouver, Canada, 2002: 1049 - 1056.
- [42]Zibulevsky M, Zeevi Y Y. Extraction of a source from multichannel data using sparse decomposition[J]. Neurocomputing, 2002, 49: 163-173.
- [43]O'Grady P D, Pearlmutter B A. Hard-LOST: Modified k-means for oriented lines[C]. Irish Signals and Systems Conference, Belfast, Northern Ireland, 2004: 247-252.
- [44]O'Grady P D, Pearlmutter B A. Soft-LOST: EM on a mixture of oriented lines[C]. Fifth International Conference on Independent Component Analysis, Granada, Spain, 2004: 430-436.
- [45]武森, 高学东, 巴斯蒂安 M. 高维稀疏聚类知识发现[M]. 北京:冶金工业出版社, 2003.
- [46]钟珞. 模式识别[M]. 武汉:武汉大学出版社, 2006.

- [47]Rickard S, Yilmaz O. On the approximate W-disjoint orthogonality of speech[C]. International Conference on Acoustics, Speech, and Signal Processing, Orlando, USA, 2002: 529-532.

攻读硕士学位期间发表学术论文情况

- 1 杨捷, 马晓红. 欠定语音信号分离的逐层分离法. 大连理工大学研究生网络学刊.
主办单位: 大连理工大学. 已录用(本硕士学位论文第五章)

致 谢

本论文是在导师马晓红教授的悉心指导和亲切关怀下完成的。在整个研究生学习阶段，马老师在学业上一直对我悉心教导、严格要求，不仅让我学到了许多专业知识，更让我在学习的过程中充分锻炼了个人的学习能力和做事的毅力与耐心。马老师严谨务实、认真负责的治学和科研态度对我的学习和工作都产生了深远影响，让我终身受益。从论文的选题、研究到撰写、修改，再到最终完成，整个过程马老师都严格把关、耐心指导，使我能够从中不断的学习和提高。在此，谨向马老师致以崇高的敬意和衷心的感谢！

感谢两年多以来和我一同进入教研室学习和工作的王立新、刘蕾、秦艳、高文婷几位同学，感谢你们的陪伴和帮助。感谢贾冬梅、柳永娟、朱妹丽、甄会、巩俊以及所有的师弟师妹们在我研究工作期间给教研室带来的欢快与和谐的生活气氛，让我在此留下了美好的回忆。

感谢游月辉、江辉及张德峰在研究生期间对我学习及生活中的帮助。

感谢何顺春在我刚进入研究生阶段时所给予的各方面的帮助。

感谢一北两年多以来在我的研究生生涯中一直给予我的关心、理解和支持，感谢你对我无时无刻的体贴、安慰和鼓励。你在这两年多的时间里为我所默默承受和付出的一切，我将铭记于心。

感谢我的爸爸，感谢您把我养育成人，感谢您在我的学业上所给予的正确引导，让我树立了坚定的信念，并一直坚持到今天。我知道您总是孜孜不倦地忙碌着，默默无闻地奉献着，不求任何回报。我只希望我的每一次进步都能够让您高兴，让您欣慰。

感谢我的妈妈，感谢您的养育之恩，感谢您在我的成长与学习生涯中对我无微不至的关怀和照顾。您总是不断地开导我、鼓励我，包容我的一切并坚定地支持着我。感谢您在我研究生期间所给予的最大支持。

感谢我所有的家人，在我最困难的时候，为我撑起了一片蓝天，使我能够在研究生期间顺利度过困难时期，并安心完成学业，祝愿您们一生都幸福平安。

最后感谢所有关心我和帮助过我的人。