

独立分量分析在说话人识别技术中的应用

邱作春¹, 曾庆宁²

(1. 盐城师范学院物理科学与电子技术学院, 江苏盐城 224002; 2. 桂林电子科技大学信息与通信学院, 广西桂林 541004)

摘要: 独立分量分析方法能够将线性混合信号进行分离, 得到统计独立的源信号, 能用于提取组合语音的特征基函数。倒谱矢量符合 ICA 变换的假设条件, 用 ICA 方法对 MFCC 特征进行转换得到 ICA 特征基, 继而用于说话人识别, 建立了一个基于独立分量分析的说话人识别系统。实验结果表明, 在噪声环境下此系统具有更高的识别率。

关键词: 独立分量分析; 说话人识别; 矢量量化; 高斯混合模型

中图分类号: TN912.3

文献标识码: A

文章编号: 1000-3630(2008)-06-0863-04

Application of independent component analysis to speaker recognition technique

QIU Zuo-chun¹, ZENG Qing-ning²

(1. School of Physical Science and Electronic Technology, Yancheng Teachers University, Yancheng 224002, Jiangsu;

2. School of Information and Communication, Guilin University of Electronic Technology, Guilin 541004, Guangxi, China)

Abstract: Independent Component Analysis (ICA) is a new method to separate the observed mixtures so as to estimate the independent components and can be used to extract the feature bases functions. The characteristic of the cepstrum vector agrees with the assumption of the ICA transformation. ICA bases is transformed from improved robust Mel Frequency Cepstrum Coefficients (MFCC) by ICA and then used in speaker recognition. A novel speaker recognition system based on ICA is presented. The results of experiments show that the proposed system has a higher recognition rate in noisy environments.

Key words: independent component analysis; speaker recognition; vector quantization; Gaussian mixture model

1 引言

说话人识别(Speaker Recognition)是一种重要的个人身份识别手段,在当今国家基础设施和高新技术服务领域,国家、公司、个人的机密、财产、信誉等需要越来越高的识别、支配与使用等安全措施,而说话人识别正是这样的一种安全技术。

目前,无噪声的说话人识别系统已达到较高的识别率,但是,当受到噪声的影响时,系统的识别率将急剧下降,甚至导致系统完全失效。然而,在许多

应用场合中,噪声又是不可避免的,比如有空调声的会议室、发动的汽车、行驶的轮船、轰鸣的工厂等。因此,研究如何提高说话人识别系统在这些噪声环境下的识别率,或者说如何提高说话人识别系统的稳健性,具有十分重要的意义^[1,2]。总结国内外研究成果,稳健说话人识别的方法主要有三种:前端处理、特征值处理和模型补偿。

本文首先利用基于麦克风阵的盲源分离算法对含噪语音进行前端处理,得到较为纯净的语音用于说话人识别,计算出 MFCC 特征后,用 ICA 对其进行转换,得到一种 ICA 基函数系数特征,用于说话人识别。

2 独立分量分析

将非高斯性作为 ICA 的一个准则,来判断信号

收稿日期:2008-04-10;修回日期:2008-07-23

基金项目:国家自然科学基金资助项目(60272038),广西自治区研究生教育创新计划项目。

作者简介:邱作春(1977-),男,湖北人,硕士研究生,主要研究方向为语音信号处理、盲信号处理。

通讯作者:邱作春,E-mail: qiuzech@163.com

的分离程度。而用负熵来度量信号的非高斯性,所以可以用负熵作为 ICA 的准则。定义负熵 $J(y)$ 为:

$$J(y) = H(y_{\text{gauss}}) - H(y) \quad (1)$$

其中 $H(y)$ 表示信息源 y 的熵, $H(y_{\text{gauss}})$ 表示与 y 具有相同协方差的高斯分布的熵。负熵总是非负的, 当且仅当 y 为高斯分布时为零。

用定义来计算负熵 $J(y)$ 复杂且不稳健, 这里采用非线性函数逼近概率密度函数的方法来逼近 $J(y)$ ^[3]。估计出 $J(y)$, 把它作为 ICA 的目标函数。

ICA 算法首先对观察数据去均值和白化。然后对白化输出信号 Z 进行变换, 使输出 $y_i(t)$ 的方差保持为 1, 同时使 $y_i(t)$ 尽可能互相独立。

用负熵作为准则^[4], 为了有好的收敛性能, 使用牛顿迭代算法。在每提取一个分量之后, 要从观察信号中消去该独立分量, 如此重复, 直到提取所有的独立分量。这里使用格拉姆-施密特 (Gram-Schmidt) 正交分解方法, 消去已提取的独立分量^[5]。

此算法的优点是: 由于采用牛顿迭代算法, 具有三阶的收敛速度, 收敛有保证, 迭代过程中无需引入调节步长等人为设置的参数, 因而更简单。

3 ICA 说话人特征提取

语音信号的倒谱矢量可以表示为一组语音特征基函数的加权组合^[6]:

$$c[n] = \sum_{i=1}^n a_i f_i[n] \quad (2)$$

其中 $c[n]$ 表示语音的倒谱矢量, $f_i[n]$ 表示特征基函数组。

每个可识别的说话人对应一个不同的特征基函数组, 对某个特定的输入语音, 只有与它对应的基函数组使其系数向量各分量之间的独立性最强, 即负熵最大。ICA 是一种将线性混合的信号进行分离, 得到统计独立的源信号的技术, 能用于提取组合语音的特征基。与式(2)对应的 ICA 模型为:

$$\mathbf{x} = \mathbf{A}\mathbf{s} \quad (3)$$

其中 \mathbf{x} 是语音帧的谱特征, 对应于 ICA 中的混合信号; \mathbf{A} 的每一列为一个语音特征基函数, 对应于 ICA 中的混合矩阵; \mathbf{s} 是未知的组成 \mathbf{x} 的基本成分, 对应于 ICA 中的源信号。

对倒谱特征进行转换的模型为:

$$\mathbf{H}\mathbf{x} = \mathbf{y} \quad (4)$$

\mathbf{y} 的列为具有独立性的系数向量, 文中用 ICAft 表示, 对应于 ICA 中的分离信号; \mathbf{H} 为 ICA 特征基函

数, 对应于 ICA 中的分离矩阵。 \mathbf{y} 和 \mathbf{H}^{-1} 分别为 \mathbf{s} 和 \mathbf{A} 的估计。基于语音 ICA 模型的特征 (下文中用 ICAft 表示) 提取过程如图 1 所示。

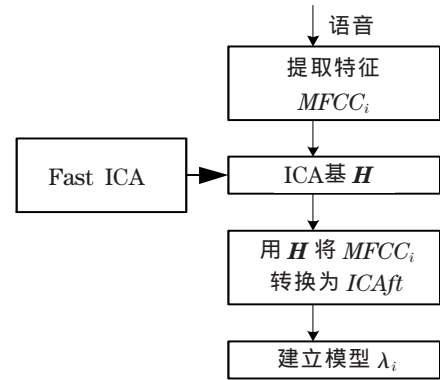


图 1 提取 ICA 特征基函数的流程图

Fig.1 Flowchart of extracting ICA feature bases functions

4 基于 ICA 的说话人识别系统

4.1 前端语音处理

训练过程可以离线进行, 可以录制相对纯净的语音用来训练, 前端语音处理主要针对测试阶段在线语音中的噪声消除和语音分离。环境噪声和其他说话人语音干扰是普遍存在的, 用混有干扰 (噪声和其他语音或其中之一) 的信号作为测试语音, 会使说话人识别系统的性能急剧下降, 所以非常必要采取措施, 对混有干扰的语音进行分离和消噪, 以得到较为纯净的语音用于说话人识别。这里采用基于自适应波束形成和 ICA 的消噪系统^[7]进行预处理。

4.2 训练阶段

训练过程可以离线进行, 训练的原理图如图 2 所示。

出于语音信号的短时平稳性, 在短时 (10ms~30ms) 内可认为声道形状、激励性质不变, 所以, 为

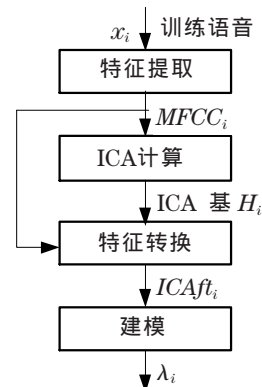


图 2 说话人识别系统的训练阶段

Fig2 Training phase of speaker recognition

了可以用平稳过程的方法来提取说话人特征，需要将语音分帧。文中取帧长为 16ms，帧移为 8ms。记语音帧序列为 $\{x_i(n) | n=1, \dots, T\}$ ， T 为语音帧总数。语音采样均为 16kHz，因此一个语音帧共有样本点 $16 \times 16 = 256$ 个。

说话人特征提取，就是从这 256 个语音样本点上提取能反映说话人个性特征的参数。采用 MFCC 特征，每帧取 16 个 MFCC 系数，则获得了一个 16 维的向量。因此经过特征提取，语音帧序列 $\{x_i(n) | n=1, \dots, T\}$ 就变成了特征向量序列，记为 $\{x_i(n) | n=1, \dots, T\}$ ，其中 $x_i(n)$ 为 16 维向量。

然后使用 ICA 将 MFCC 转换成 ICAft，这就是表示语音信号特征的 ICA 特征基函数系数，同时求取 ICA 基 H ，通过 GMM 或 VQ 方法构造说话人参考模型或码本。

4.3 测试阶段

说话人辨认实验系统的测试流程图如图 3 所示。用 y 表示说话人的语音，首先，从 y 中提取 MFCC 特征向量序列，然后用 ICA 基 H 转换特征产生新的 ICA 特征系数序列，最后是进行说话人识别。

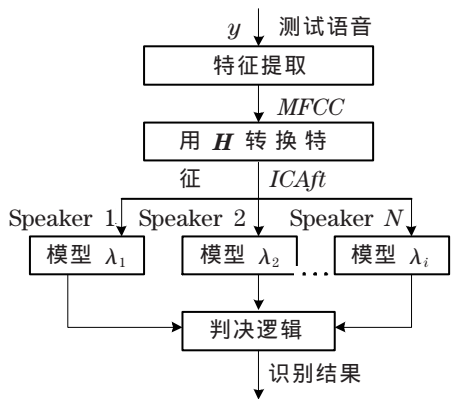


图 3 说话人识别的测试阶段

Fig.3 Test phase of speaker recognition

将使用 ICAft 特征和 GMM 模型的说话人识别系统称为 ICA-GMM 系统；将使用 ICAft 和 VQ 方法的说话人识别系统称为 ICA-VQ 系统。

5 实验与结果分析

实验使用我们自己制作的 6 通道语音库，包括 6 个女声和 12 个男声。录音时，语音源距离麦克风都为 40cm，多通道时采用间距为 4cm 的线阵，采样率为 16kHz。采集实验系统如图 4，其中 UA301 型 A/D 采集器采用美国新型 12 位 A/D 转换芯片，

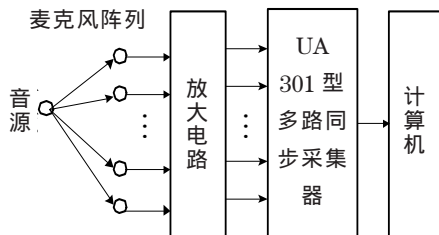


图 4 采集系统

Fig.4 Data acquisition system

可广泛应用于科学实验、工业测量控制领域。

实验中的训练语音长 30s，测试语音长 3s。说话人特征分别使用 MFCC 和 ICAft，含噪语音经过前端消噪处理，分别用 GMM、VQ、ICA-GMM、ICA-VQ 系统进行识别。每个人取 10 段音频，共 180 条测试语音，电脑配置为 P4 1.5G CPU, 256M RAM, Windows XP 操作系统，所得结果如表 1 所示。

表 1 MFCC 与 ICAft 特征的识别率比较

Table 1 Comparison of recognition rate between MFCC and ICAft

	MFCC		ICAft	
	GMM	VQ	ICA-GMM	ICA-VQ
不含噪声语音	100%	98.9%	100%	100%
含噪声语音 SNR=5dB	17.2%	11.1%	51.7%	85.6%
系统识别响应时间/s	5.1	5.3	7.3	5.5

从表 1 可以看出，在测试语音不含噪声时，各种方法都取得了很高的识别率。但是当有噪声干扰时，使用 MFCC 特征的 GMM 系统和 VQ 系统的识别性能急剧下降，使用 ICAft 特征的 GMM 系统的识别率也下降得非常厉害，而使用 ICAft 特征的 VQ 系统的识别率可以达到 85.56%。

对于同样使用 ICAft 特征的 ICA-GMM 和 ICA-VQ 系统，识别率却有很大的差别，分别为 51.67% 和 85.56%。这种结果与文献 [8] 中的分析是一致的，说明了本文在说话人识别中使用 ICA 方法是恰当的。

6 结 论

本文提出用 ICA 方法对 MFCC 特征进行转换，提取 ICAft 特征，继而用于说话人识别，这是因为倒谱矢量符合 ICA 变换的假设条件，盲源分离中的 ICA 方法是一种将线性混合信号进行分离，得到统计独立的源信号的技术，能用于提取组合语音的特征基函数。本系统在信噪比为 5dB 的噪声环境下，达到了较高的识别率，实时性较好。但

在识别率方面还有进一步的提升空间,这也是今后努力的方向。

参 考 文 献

- [1] Ozkurt T E, Akgul T. Robust Text-independent speaker identification using bispectrum slice. signal processing and communications applications conference[A]. 2004. Proceedings of the IEEE 12th[C]. 2004, 418-421.
- [2] Deng J, Zheng T F, Song Z J, et al. Using predictive differential power spectrum and subband mel-spectrum centroid for robust speaker recognition in stationary noises[A]. The 4th International Conference on Machine Learning and Cybernetics[C]. 2005, 8: 4846-4851.
- [3] Hyvarinen A. New approximations of differential entropy for independent component analysis and projection pursuit [J]. Advance in Neural Information Processing Systems, 1998, **10**: 273-279.
- [4] 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. 97-101.
YANG Fusheng, HONG Bo. The principle and application of independent component analysis[M]. Beijing: Tsinghua University publishing House, 2006, 97-101.
- [5] Hyvarinen A. Fast and robust fixed-point algorithm for independent component analysis[J]. IEEE Trans on Neural Networks, 1999, **10**(3): 626-634.
- [6] Jang Gil-Jin, Yun Seong-Jin, Oh Yung-Hwan. Feature vector transformation using independent component analysis and its application to speaker identification[J]. Euro Speech, 1999, **7**(6): 767-770.
- [7] 邱作春, 曾庆宁. 基于自适应波束形成和 ICA 的消噪系统[J]. 声学技术, 2008, **27**(1): 119-125.
QIU Zuochun, ZENG Qingning. Noise cancellation system based on adaptive beamforming and independent component analysis[J]. Technical Acoustics. 2008, **27**(1): 119-125.
- [8] 屈微, 刘贺平. 基于独立分量分析和矢量量化的说话人识别 [J]. 计算机应用, 2005, **125**(10): 2041-2043.
QU Wei, LIU Heping. Speaker recognition system based on ICA and VQ[J]. Computer Applications, 2005, **125**(10): 2041-2043.