

文章编号: 1002-8684(2007)02-0045-02

声纹识别技术及其应用

· 综述 ·

杨 阳, 陈永明

(厦门大学 机电工程系, 福建 厦门 361005)

【摘 要】声纹是生物特征的一种,对于人体来说,声纹是长期稳定的特征信号,通过声纹鉴别技术可以区分不同个体。介绍了声纹、声纹识别的概念和原理;指出了声纹识别技术的应用范围和前景。

【关键词】声纹; 声纹识别; 说话人辨认; 说话人确认

【中图分类号】TN912.34

【文献标识码】A

Voice Identification Technology and Its Application

YANG Yang, CHEN Yong-ming

(Department of Mechanical and Electrical Engineering, Xiamen University, Xiamen Fujian 361005, China)

【Abstract】Voiceprint which is a steady signal for every person in long time is one kind of biologic characteristics. Different people can be distinguished by the voice identification technology. The concept and principles of the voiceprint and voice identification are introduced. Finally, the future and range of the application on the voiceprint identification are discussed.

【Key words】voiceprint; voice identification; speaker identification; speaker verification

1 引言

声纹(voiceprint)是指通过专用的电声转换仪器(声谱仪、语图仪等)将声波特征绘制成的波谱图形,它是各种声学特征图谱的集合。“voiceprint”在辞典中的注释为:“用仪器描录的因人而异的声波纹”。声纹是人体的一张“身份证”,是长期稳定的特征信号。

声纹识别(voice identification)是把未知人语音材料(检材)与已知人语音材料(样本)分别通过电声转换仪器绘成声纹图谱,再根据图谱上的语音声学特征进行比较和综合分析,以得出两者是否同一的判断过程。

声纹识别有着十分广阔的应用前景,在世界范围内正广泛应用于金融、证券、社保、公安、军队及其他民用安全认证等领域。目前,中国市场尚属启动阶段,其发展空间更为广阔。

2 声纹识别的内涵

声纹识别广义上分为语音识别和说话人识别两种^[1]。语音识别是根据说话人的发音辨认其所说的语音、音节、单词或单句,这就要排除不同说话人的个人特色,找出代表各个语音单位的共性特征。说话人识别是根据语音来辨认说话人,而并不考虑声音的内容和意义,这就需要分离出每个个体的特性。目前,普遍意

义上声纹识别的概念是指说话人识别。

说话人识别包括说话人辨认(speaker identification)和说话人确认(speaker verification)两个方面。说话人辨认是一对多的分析过程,即判断出某段语音是若干人中哪一个所说,主要应用于刑侦破案、罪犯跟踪、国防监听、个性化应用等。说话人确认是一对一的确定过程,即确认某段语音是否属于指定的某人,主要应用于证券交易、银行交易、个人计算机声控锁、汽车声控锁、身份证、信用卡等。识别的核心是预先录入声音样本,并提取每个样本独一无二的特征,建立特征数据库,使用时将待检声音与数据库中的特征进行匹配,通过分析计算,实现说话人识别^[2]。

3 声纹识别的原理

3.1 特征提取

特征提取是提取声音中能够反映个体信息的基本特征,这些基本特征必须能够准确、有效地区分不同的发声个体,且对于同一个体,这些基本特征应具有稳定性。

目前的声纹识别系统主要依靠较低层次的声学特征进行识别。这些声学特征主要有以下几个方面:
(1) 语音信息通过滤波器组输出,以合适的速率对输出

进行抽样得到谱包络特征参数; (2) 基于发声器官如声门、声道和鼻腔的生理结构提取的特征参数, 如基音轮廓、共振峰频率带宽及其轨迹等; (3) 以线性预测导出的特征参数, 如线性预测系数、自相关系数、反射系数等; (4) 模拟人耳对声音频率感知的特性而得到的听觉特性参数, 如 Mel 倒谱系数、感知线性预测等。

随着对声纹识别应用范围的不断扩大, 以及对系统准确性要求的不断提高, 只考虑较低层次的声学特征并不能满足要求, 这就需要同时考虑高层次的特征信息, 如语速、语法、韵律、语种、方言、特性发音、特性词、信道(声音信号获取的渠道)等。对于这些高层次的信息, 最关键的问题是选择, 同时要针对具体情况来决定。例如, 对于信道这一特征来说, 在刑侦破案方面, 就希望不采用, 即希望信道对识别不产生影响, 从而使得录音等间接手段获得的声音能够成为帮助破案的证据; 而在银行交易中, 就希望采用, 也就是希望信道对识别产生影响, 这样才能剔除录音等恶意行为带来的危害。因此, 在声纹识别过程中, 必须根据实际情况, 安排不同特征参量的组合, 以提高实际系统的性能, 当各组合参量间相关性不大时, 会得到更好的识别效果。

3.2 模式匹配

声纹识别技术的关键在于对各种声学特征参数进行处理, 并确定模式匹配方法, 主要的模式匹配方法为:

(1) 概率统计方法: 声音信息在短时内较为平稳, 通过对稳态特征如基音、声门增益、低阶反射系数的统计分析, 可利用均值、方差等统计量和概率密度函数进行分类判决。这种方法不用对特征参数在时域上进行规整, 适合与文本无关的声纹识别。

(2) 动态时间规整方法: 说话人信息不仅有稳定因素(发声器官的结构和发声习惯等), 也有时变因素(语速、语调、重音和韵律等), 将识别模板与参考模板进行时间对比, 按照某种距离测定得出两模板间的相似程度。

(3) 矢量量化方法: 把每个人的特定文本编成码本, 识别时将测试文本按此码本进行编码, 以量化产生的失真度作为判决标准, 具有识别精度高、判断速度快的特点。

(4) 隐马尔可夫模型方法: 隐马尔可夫模型是基于转移概率和传输概率的随机模型, 它把语音看成由可观察到的符号序列组成的随机过程, 符号序列则是发声系统状态序列的输出。识别时, 为每个发音个体建立

发声模型, 通过训练得到状态转移概率矩阵和符号输出概率矩阵。识别时计算未知语音在状态转移过程中的最大概率, 根据最大概率对应的模型进行判决。这种方法不需要时间规整, 可节约判决时的计算时间和存储量, 目前已被广泛应用; 缺点是训练时计算量较大。

(5) 人工神经网络方法: 人工神经网络在某种程度上模拟了生物的感知特性, 是一种分布式并行处理结构的网络模型, 具有自组织和自学习能力、很强的复杂分类边界区分能力以及对不完全信息的鲁棒性, 其性能近似理想的分类器; 缺点是训练时间长, 动态时间规整能力弱, 网络规模随说话人数目增加时可能大到难以训练的程度。

4 声纹识别技术的发展、现状及前景

身体在讲话时使用的器官——舌、牙齿、喉头、肺、鼻腔, 在尺寸和形态方面每个人的差异很大, 这就使每个人的声音特征具有与其他人不同的唯一性和一定时期内不变的稳定性。虽然模仿他人的声音在一般的人耳听起来可能极其相似, 但如果采用声纹识别技术进行识别, 就能显示出巨大的差异, 因此, 无论是多么高明、相似的声音模仿都可通过声纹识别技术辨别。

声纹识别技术是在 20 世纪中期在美国提出的, 最早进行此项技术研究的是美国贝尔实验室的劳伦斯·克斯特, 他对一百多名健康人的上万个声纹图进行了分析鉴定, 准确率达到 99.65%。中国的声纹识别技术起步较晚, 在 20 世纪 90 年代才开始进行正式的研究, 目前进行相关研究的有北京大学、清华大学、中国科学院声学研究所以及一些政法部门^[3]。

现在, 在世界范围内, 声纹识别技术已应用于如金融、证券、刑侦、以及其他民用安全认证系统等诸多领域。尤其是在安全认证方面, 声音不涉及隐私问题, 相关设备造价低廉, 利用声纹进行身份识别是自然又经济的方法, 使用者接受程度相对较高。例如银行、证券系统的密码可用声音来代替, 即利用声纹技术, 使声音转化为密钥, 这样人们不需要记住复杂的密码, 也不需随身携带钥匙、智能卡之类的东西。另外, 在电话勒索一类的案件中, 最容易获得的证据就是电话录音, 通过声纹识别技术, 就能根据电话录音获得线索, 缩短破案周期。虽然现在大多数国家还没有把声音列入法庭审判的有效证据范围, 但是声音样本在刑侦和司法上的作用已经日渐受到重视。

(下转第 50 页)

5 结论

针对盲信号分离的原理和特点,采用基于延时估计的语音盲分离算法,有效地实现了对卷积混迭语音信号的盲分离,并通过实验证明笔者所提方法不仅具有更小的运算复杂度和更快的执行速度,大大降低了硬件的开销,很大程度上增强了算法的实时性,还大大改善了盲分离的分离效果。但在强烈噪声和反射的情况下,算法的性能将大大下降,因此,在这些情况下的盲分离需要进一步改进。

参考文献

- [1] TAN Li-li, WEI Gang. Blind signal separation of convolution mixture signals via minimum mutual information (MMI) method[J]. Journal of China Institute of Communications, 1999, 20(10): 49-55.
- [2] TAN Li-li, WEI Gang. Multi-input and multi-output (MIMO) blind deconvolution via maximum entropy (ME) method[J]. Acta Electronica Sinica, 2000, 28(1): 114-116.
- [3] ARAKI S, MUKAI R, MAKINGO S, et al. The fundamental limitation of frequency domain blind source separation for convolutive mixtures of speech[J]. IEEE Tran. on Speech and Audio Processing, 2003, 11(2): 109-116.

(上接第46页)

总之,声纹识别技术是一种方便、快捷、安全的识别技术,随着各个研究领域(如材料、通信、计算机、生命科学等)的不断发展,声纹识别技术也在飞速发展,其可靠性和准确性将不断提高,可以预见,在不久的将来,声纹识别技术会在更多领域内获得更广泛的应用,使人类的生活更方便、安全。当今社会,人们的工作和生活越来越多地依赖于网络,而网络在带给人们方便和快捷享受的同时,也存在巨大的安全隐患,黑客们可以通过网络侵入他人的计算机系统,获取资料、信息甚至金钱,在日益盛行的电子购物、网上炒股等电子商务活动中,资金的安全问题更加重要,因此,如果能够利用声音作为本地或远程登录系统密码,那么即使再高明的黑客也只能望洋兴叹,可有效防止密码被盗、黑客攻击等恶性事件。此外,随着声纹识别技术准确性的提高,以及录音设备(如录音笔、MP3)的更新换代,新型录音设备携带方便、存储容量大、使用简单的特点,使人们在很多场合下可以容易地获取声音样本。通过声纹识别技术,声音样本在公安机关侦破案件、司法部门做出判决等方面都将起到关键性作用。相信各国也会

- [4] MITIANOUDIS N, MICHEAL E D. Audio source separation of convolutive mixture[J]. IEEE Tran. on Speech and Audio Processing, 2003, 11(5): 489-497.
- [5] MURATA N, IKEDA S, ZIEHE A. An approach to blind source separation based on temporal structure of speech signals[J]. Neurocomputing, 2001, 41(1-4): 1-24.
- [6] 殷斌,何培宇, SOMMEN P C W. 一种有效的语音盲信号分离简化混合模型[J]. 电子学报, 2002, 30(10): 1438-1440.
- [7] BELOUHRANI A, ABED-MERAIM K, CARDOSO J F, et al. A blind source separation technique using second-order statistics[J]. IEEE Tran. on Singal Processing, 1997, 45(2): 434-444.
- [8] BENESTY J. Time-delay estimation via linear interpolation and cross correlation[J]. IEEE Tran. on Speech and Audio Processing, 2004, 12(5): 509-519.

作者简介

刘燕华, 硕士研究生, 主要研究方向为语音信号处理;

方勇, 教授, 博士生导师, 主要研究方向为盲信号处理、通信信号处理和智能信息系统。

[责任编辑] 潘浩然

[收稿日期] 2006-10-08

逐渐将声音证据列入法定证据范围,中国已经在1996年修订的刑事诉讼法将视听资料规定为第七种证据,这极大推动了声像物证技术的发展。

5 结语

声纹和指纹、面部特征、DNA、气味、虹膜、视网膜纹理等一样,都属于生物特征,将这些特征数字化后可识别地球上的每一个人。但就目前的技术水平和各方因素的限制,每一种特征识别都存在缺陷,因此,包括声纹识别在内的特征识别手段的应用还并不普遍。在目前的技术条件下,如果能够将某几种特征识别结合起来使用,如声纹和指纹识别相结合,优势互补,将会大大提高应用空间和应用效果。

参考文献

- [1] 朱民雄, 闻新, 黄健群, 等. 计算机语音技术[M]. 修订版. 北京: 北京航空航天大学出版社, 2002.
- [2] 蔡耿平, 黄顺珍, 徐志鸿, 等. 声纹识别系统[J]. 深圳大学学报: 理工版, 2002, 19(2): 78-80.
- [3] 宋森, 李敞阳. 浅谈声纹检验鉴定技术[J]. 北京人民警察学院学报, 2005(2): 35-36.

[责任编辑] 潘浩然

[收稿日期] 2006-10-16