文章编号：1000-5013(2015)05-0517-05　　**doi**：10.11830/ISSN.1000-5013.2015.05.0517

声纹识别在开放仪器管理中的应用

赖丽旻[1]，洪青阳[2]

（1. 厦门大学 环境与生态学院，福建 厦门 361005；
2. 厦门大学 信息科学与技术学院，福建 厦门 361005）

摘要：在现有的仪器工作站中加入基于 GMM-HMM 算法的声纹识别系统，利用声纹的唯一性识别仪器用户，实现开放仪器的无人监管. 应用该声纹识别系统，进行语音动态口令测试. 结果表明：该系统语音动态口令的错误接受率低于 1%，可有效地防范录音冒充，保证系统的可靠性.

关键词：声纹识别；仪器管理；身份认证；高斯混合模型；隐马尔可夫模型
中图分类号：TP 393　　**文献标志码**：A

为了提高科研经费的使用效益，贵重仪器一般对外开放，共享使用. 由于时间和精力限制，仪器管理员很难对仪器监管到位，机时统计不真实，仪器故障率高. 为了规范化管理仪器，降低仪器的故障率，需要在仪器周边安装摄像头进行监控. 但外加设备成本较高，且受限于摄像头的安装位置，往往难以拍摄到所需的画面. 因此，需要发展一种能有效识别仪器使用者，并记录仪器使用机时和使用状况的管理系统. 传统的方式是通过账号识别仪器使用者，但账号容易被人借用，存在较大的管理漏洞[1]. 为确保身份的唯一性，更有效的方式是采用生物特征识别技术. 声纹识别也称说话人识别[2-4]，由于每个人的声带震动频率不同，声道结构不同，再加上发音习惯不同，组合形成了各具一色的声纹特征. 不同人说同样的话，对应的语谱图也会不一样. 因此，可用来比对两段语音的同一性，即是否来自同一人. 声纹采集方便、硬件成本低、用户容易接受，因此，得到越来越多的应用. 本文将声纹识别技术应用到仪器管理中，并创造性地采用语音动态口令，达到防录音冒充的效果.

1　基于声纹识别的仪器管理系统

大部分贵重仪器是通过计算机上的工作站控制，在计算机上加入声纹识别系统，控制仪器软件的开启，以达到只有通过审核的人才能使用仪器的目的. 用户无需任何其他设备，直接采用电脑麦克风录音，进行声纹采集. 系统结构图，如图 1 所示.

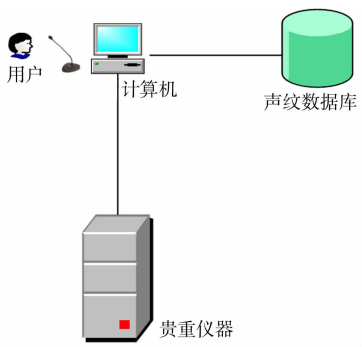

图 1　系统结构图
Fig. 1　System structure diagram

利用声纹的唯一性确认仪器用户身份，实现无人监管. 电脑麦克风可设置比较高的采样率，并可持续录音，使送到验证服务器的声纹信息最大限度地不失真，这样声纹验证更可靠. 对于部分没有连接计算机的仪器，可通过增加声纹识别模块，控制仪器电源的开关，从而达到控制仪器使用的目的. 基于声纹识别技术的共享仪器平台管理系统，具体包括以下 5 个步骤.

步骤 1　声纹登记. 用户通过仪器培训后，在仪器管理员监督和指导下，通过麦克风录音，朗读计算机屏幕上的文字，进行声纹特征值的采集. 达到有效时长后，提示用户录音结束，系统检测语音合格后，登记该声纹模型，屏幕显示声纹登记成功.

收稿日期：2015-04-09
通信作者：洪青阳(1977-)，男，副教授，博士，主要从事语音识别和说话人识别的研究. E-mail：qyhong@xmu.edu.cn.
基金项目：国家自然科学基金资助项目(61105026)

步骤 2 用户开启仪器工作站时,自动启动声纹验证程序.用户通过麦克风朗读屏幕上的文字,达到有效时长后,提示用户录音结束.

步骤 3 系统判断用户声纹是否与登记声纹模型一致,识别用户身份是否为授权用户.

步骤 4 已授权用户,仪器可正常启动,用户正常使用仪器,后台记录用户信息和统计机时.

步骤 5 若用户为非授权用户,仪器则不能正常启动,用户无法使用该仪器.用户可联系仪器管理员,告知存在的问题.

2 基于 GMM-HMM 算法的声纹识别系统

2.1 基本原理

声纹识别是一个模式识别过程,其基本原理如图 2 所示.首先对目标说话人的语音特征提取;然后进行声纹建模,验证语音也要经过特征提取,才能进行声纹比对;声纹比对得分与事先设定的阈值比对,最后得到验证结果.图 2 是一个典型的模式识别过程,关键是声纹特征要与语音信号建立一一对应的关系.如果语音信号包含噪声等杂音,则还需进行降噪等前端处理.后端模型用来刻画声纹的统计分布,比较通用的是采用高斯混合模型(Gaussian mixture model,GMM)[5-6].

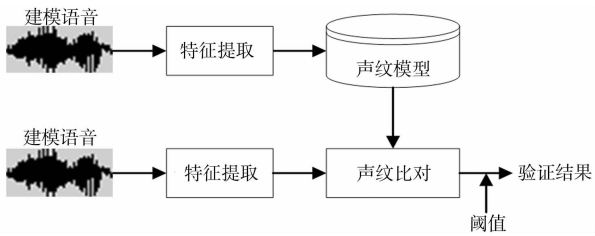

图 2 声纹识别基本原理

Fig. 2 Principle of voiceprint recognition

GMM 通过若干个高斯概率密度函数的线性组合逼近任意分布,从而模拟出各种形式的语音特征分布,以区分不同的说话人.GMM 能很好地刻画参数空间中训练数据的空间分布及其特征,并且具有简单高效的特点,已广泛应用于与文本无关的声纹识别系统 .

为解决录音冒充问题,进一步结合隐马尔可夫模型(hidden Markov model,HMM)[7],采用一种语音动态口令的建模和验证方法[8],把声纹识别和语音识别技术更好地融合在一起,使得身份认证系统更加可靠.

2.2 声纹建模过程

系统依据说话人的训练语音,进行语音预处理,提取说话人特征,并通过相应的建模算法,生成声纹模型.声纹动态口令系统训练模型所需要的语音是 N 段文本内容不同的短语音,一般取 3 至 5 段.训练过程,如图 3 所示.用户录完的语音,将被训练成与该用户相关的声纹模型(包括说话人模型和语音模型).其中,说话人模型为 GMM 模型,采用最大后验概率(MAP)方法[6],由全局背景模型(UBM)自适应而来.具体实现时,只需要自适应均值参数,即

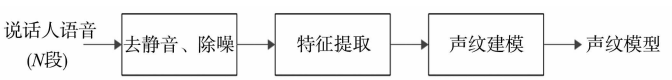

图 3 声纹建模过程

Fig. 3 Modeling process of voiceprint

$$\hat{\mu}_i = \beta E_i(x) + (1-\beta)\mu_i. \tag{1}$$

式(1)中:i 是 UBM 所包含的每个高斯函数对应的索引;$E_i(x)$为自适应数据 x 的均值期望;μ_i 为原始 UBM 的均值;$\hat{\mu}_i$ 为自适应后得到的均值;β 为调节系数.

语音模型采用隐马尔可夫模型.基于 HMM 的通用语音识别器,也可实现自适应训练,变成与目标说话人相关的特定识别器,如图 4 所示.

图 4 HMM 自适应训练

Fig. 4 Adaptation of HMM

Hong 等[8]采用单音子(monophone)模型,没有考虑音素的上下文关联,一定程度上会导致识别率的下降.而文中进一步采用了三音子(triphone)模型,使声韵母之间的关联性也能得到建模.经过重新组合[7-8],采用的三音子模型(不考虑 yi 和 yao)包括 sil,+i_one,s-i_one,sp,s+an,s-an,_w+u,_w-u,q+i,q-i,b+a,b-a,l+ing,l-ing,j+iou,j-iou,_e+er,_e-er,l+iou,l-iou.

2.3　声纹验证过程

在验证阶段，声纹系统根据说话人的语音，判决说话人是否为其所申明的身份(identity claimed). 这个阶段有 2 个输入信息，即说话人的语音和其所申明的身份信息. 首先，系统对语音进行预处理；然后，提取声纹特征，将其与对应的声纹模型进行模式匹配；最后，判决这段语音是否属于该说话人.

在文中方法里，声纹验证过程是个融合的过程. 输入语音经特征提取后，分别进行基于 HMM 的语音识别和基于 GMM 的声纹确认，得到相应的语音识别得分和声纹确认得分. 基于 HMM 的语音识别，是根据提示文本，产生相应的受限语法. 如数字串“43825769”，其对应的受限语法如下

$digit1＝si；

$digit2＝san；

$digit3＝ba；

$digit4＝er；

$digit5＝wu；

$digit6＝qi；

$digit7＝liu；

$digit8＝jiu；

(SENT-START [$digit1] [$digit2] [$digit3] [$digit4] [$digit5] [$digit6] [$digit7] [$digit8] SENT-END)

其中：$digit1 表示第一个数字；si 对应数字 4；括号里的 SENT-START 是句子的开头；SENT-END 是句子的结尾；[$digit1] [$digit2] [$digit3] [$digit4] [$digit5] [$digit6] [$digit7] [$digit8]表示只能识别为 8 个数字.

基于以上受限语法，采用 Viterbi 解码算法[7]，就可得到语音识别得分. 由于受限语法是与提示文本关联的，也就是相当于为文本内容建立了对应的语言模型. 如果用户故意说别的数字串，或用录音设备录制回放别的数字串，则正确识别到的数字个数就很少，识别得分也会很低. 因此，该方法可起到内容鉴别的作用，有效避免录音冒充.

系统融合得分计算，表达为

$$S_{\mathrm{F}}=\frac{1}{1+\exp(-(S_{\mathrm{ASR}}/2+\alpha\times S_{\mathrm{VPR}}))}. \tag{2}$$

式(2)中：S_{F} 为系统融合得分；S_{ASR} 为基于 HMM 的语音识别得分；S_{VPR} 为 GMM 的声纹确认得分；α 是调节系数，可根据实际应用调节.

声纹验证过程，如图 5 所示. 由图 5 可知：系统融合得分将与预设阈值比对，超过阈值则表示接受通过，未超过则予以拒绝. 阈值可根据实际应用做调整.

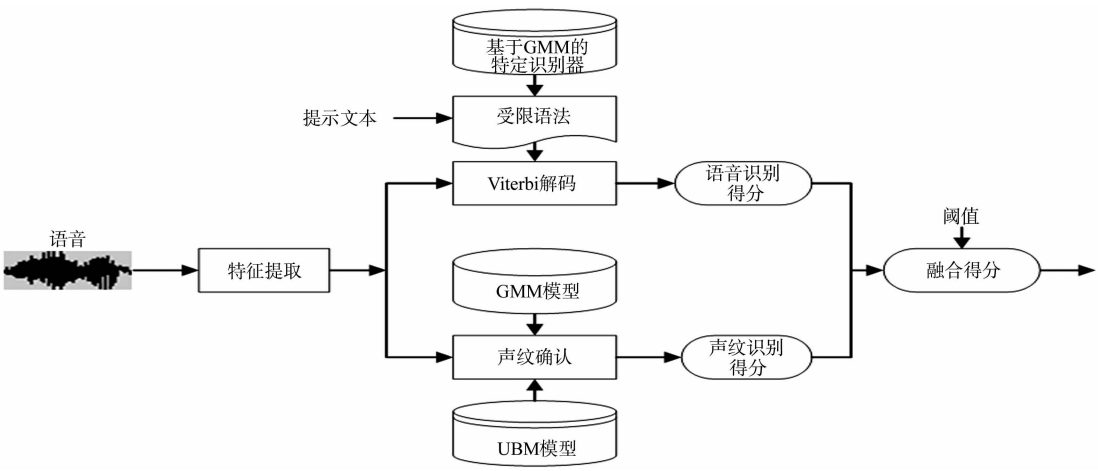

图 5　声纹验证过程

Fig. 5　Verification process of voiceprint

3 结果与分析

进行了两组语音动态口令实验.一组在办公室进行声纹的登记和测试,采集对象以年轻人为主.说话人与麦克风之间的距离在0.3～1 m之间,以说话人感觉舒适为度.采样率为8 K,量化位数为16 bit.样本总共20人,每人录音20句以上,随机抽取16句作为登记,其他剩下的作为本人认证测试,不同人之间进行交叉测试.测试结果,如表1所示.表1中:R_{FR}表示错误拒绝率,即本人认证被拒绝的比例;R_{FA}表示错误接受率,即他人冒充通过的比例.

表1 语音动态口令的测试结果

Tab.1 Experimental results of speech dynamic password

测试方式	总数	成功次数	错误次数	错误比例/%
本人认证	157	153	4	$R_{FR}=2.55$
他人冒充	9 063	9 006(拒绝)	57(通过)	$R_{FA}=0.63$

从表1可以看出:R_{FR}为2.55%,即本人通过率为97.45%,说明本文系统对真实用户通过率较高,已可满足应用需求;R_{FA}为0.63%,即他人冒充通过的可能性低于1%,说明文中系统具有很强的防冒充能力,能有效地保证贵重仪器的安全管理.有文献[9]报道基于指纹识别的开放式仪器管理系统,R_{FR}为2.50%,R_{FA}为1.11%.

第2组实验数据是在比较复杂的环境下采集的.采集环境可能在办公室、马路边、商场、家里等地方,以模拟各种噪声背景.样本总共30人,每个人用智能手机采集8个随机数字,登记语音5遍,验证语音3遍以上.采样率为16 K,量化位数为16 bit.本人测试149次,冒充测试7 305次.实验结果采用DET曲线[10]绘制,如图6所示.图6中:R_{FA}为错误接受率;R_{FR}为错误拒绝率.图6中:曲线越靠近零点表示识别效果越好;曲线与对角线的交叉点是等错误率(R_{EE},即R_{FA}与R_{FR}相等的地方).由图6可知:三音子模型明显优于单音子模型,三音子的R_{EE}约为1%.

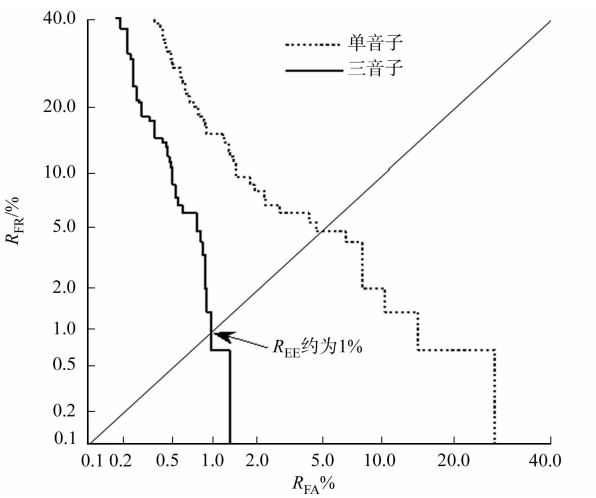

图6 声纹验证结果

Fig.6 Verification results of voiceprint

与文献[9]方法相比,在本人通过率相差不大的情况下,文中方法的他人冒充通过率更低.考虑到指纹识别的开放式仪器管理系统需要部署指纹采集仪,成本较高,因此,文中方法具有较高的性价比.

文中方法将基于传统模型GMM和HMM的声纹识别技术有机地结合起来,应用到实际系统中,实现内容＋身份的识别,而不是简单的GMM身份识别.尤其采用了8个数字随机动态口令,非法用户无法通过录音冒充通过,有效地提高了仪器管理的安全性.

在实际应用中,声纹采集时,操作是否规范直接影响声纹识别效果.因此,需要仪器管理员在现场指导.这样,一方面提高声纹采集样本的质量;另一方面,从源头防止冒充他人使用仪器的可能.

4 结束语

在贵重仪器现有的工作站系统内加入声纹识别部分,通过声纹识别判定仪器使用者的身份[11],并从后台记录仪器使用机时,有利于仪器的规范化管理,防止仪器使用者漏登记机时.通过测试发现,语音动态口令的效果很好,错误接受率低于1%,可有效防范冒充,保证了系统的可靠性.

参考文献:

[1] 王云平.国外大学实验室管理及其对国内开放实验室的启示[J].实验技术与管理,2010,27(3):149-151.

[2] HONG Q Y,KWONG S. Discriminative training for speaker identification based on maximum model distance algorithm[C]// IEEE International Conference on Acoustics, Speech, and Signal Processing. Montreal:IEEE Press, 2004:25-28.
[3] 张彩红,洪青阳,陈燕. 基于 GMM-UBM 的说话人确认系统的研究[J]. 心智与计算,2007,1(4):420-425.
[4] 陈燕,洪青阳,张彩虹. 声纹识别在司法身份鉴定中的应用[J]. 心智与计算,2008,2(1):1-7.
[5] REYNOLDS D A. Speaker identification and verification using Gaussian mixture speaker models[J]. Speech Communication,1995,17(1/2):91-108.
[6] REYNOLDS D A,QUATIERI T F,DUNN R B. Speaker verification using adapted Gaussian mixture models[J]. Digital Signal Processing,2000,10(1/2/3):19-41.
[7] 韩纪庆,张磊,郑铁然. 语音信号处理[M]. 北京:清华大学出版社,2004:200-213,239-241.
[8] HONG Qing-yang,WANG Sheng,LIU Zhi-jian. A robust speaker-adaptive and text-prompted speaker verification system[J]. Lecture Notes in Computer Science,2014,8833:385-393.
[9] 卢畅. 基于指纹检测识别的开放式实验室管理系统研究与设计[J]. 实验室研究与探索,2013,32(12):211-215.
[10] DODDINGTON G R,PRZYBOCKI M A,MARTIN A F,et al. The NIST speaker recognition evaluation: Overview, methodology, systems, results, perspective[J]. Speech Communication,2000,31(2/3):225-254.
[11] DEHAK N,KENNY P,DEHAK R,et al. Front-end factor analysis for speaker verification[J]. IEEE Transactions on Audio, Speech, and Language Processing,2011,19(4):788-798.

Application of Voiceprint Recognition in Opening Instrument Management

LAI Li-min[1], HONG Qing-yang[2]

(1. College of Environment and Ecology, Xiamen University, Xiamen 361005, China;
2. School of Information Science and Technology, Xiamen University, Xiamen 361005, China)

Abstract: In order to have a standard management of opening instruments, we apply the voiceprint recognition system based on GMM-HMM algorithm into the instrument workstation. The uniqueness of voiceprint is utilized to verify the user identity, which realizes unmanned supervision of opening instrument. Based on the voiceprint recognition system, we conducted the experiment of dynamic password speaker verification. The results showed that the false acceptance rate of dynamic password version was lower than 1%, which could avoid the danger of recording playback and assure the system′s reliability.

Keywords: voiceprint recognition; instrument management; identity authentication; Gaussian mixture model; hidden Markov model

（**责任编辑**：黄晓楠　　**英文审校**：吴逢铁）