

声纹识别在虚拟仪器平台的实现

王会清^{1,2}, 张 涛², 周 帆²

(1. 武汉工程大学智能机器人湖北省重点实验室, 湖北 武汉 430074;

2. 武汉工程大学计算机科学与工程学院, 湖北 武汉 430074)

摘 要:采用 LabVIEW 软件和声卡构建一个声纹识别系统. 通过对音频信号的采集及处理, 分析说话人的声音, 提取出语音的频率时间序列作为特征值, 建立模型库. 识别时, 由系统对现场输入的语音采样、提取特征值, 并与注册样本模型进行匹配, 根据其结果作出是否一致的判断. 系统测试结果表明, 用提取文本语音频率时间序列进行身份认证是一种能满足实际要求的可行方法. 该方法可以区分不同的说话人, 能有效地克服环境噪声带来的影响, 并且在做说话人确认测试时达到较高的识别率. 在采用麦克进行语音录入与匹配时, 学习、识别时间短, 声纹确认准确, 声纹确认的漏识率小于 1%. 具有无区域限制, 简单方便, 成本低等特点, 可以为声纹识别系统的应用提供一个安全可靠的技术支持.

关键词:虚拟仪器; 声纹识别; 特征提取; 说话人确认

中图分类号:TP277

文献标识码:A

doi:10.3969/j.issn.1674-2869.2012.12.014

0 引 言

声纹识别系统是对人的声音进行特征识别和确认, 分辨不同个体的装置, 是信息安全的重要组成部分. 在网络时代, 电子商务、网络银行等网络上的双方交互, 利用声纹识别技术不需要采用加密算法和建立严密的安全认证体系, 不需本人到场, 只需在网络服务器上安装声纹识别软件, 用户的客户机上备有话筒即可. 这种识别方法不会出现遗忘、被盗现象, 满足“随身携带”的特点, 很难进行伪造, 是确保信息安全的好方法. 并且, 其应用价值还可体现在诸如门禁系统, 用语音控制门锁的自动开启; 电话银行可用语音控制转账密码; 以及情报监听、各种声音的鉴别系统等方面.

以声卡为数据采集卡, LabVIEW(虚拟仪器)软件为开发平台, 进行语音采集、分析与识别研究. 系统首先采集鉴别体发出的语音样本, 通过学习建立该个体的应用模型. 一旦该模型被调入系统, 系统就会对待测语音进行识别, 并提示匹配程度.

1 声纹识别系统的设计原理

声纹识别技术与语音的生理特性及行为特性密切相关. 说话者嗓音体现其生理学和行为特征, 可以运用语言学模式进行分析, 根据说话人语音波形中反映其生理和行为特征的语音参数来识别

说话人的身份. 声纹识别本质上是一种模式识别的过程, 是对个性特征的认识, 所以特征提取非常重要. 每个人的声音都有自己的频率、声调以及断续特征, 经过学习和鉴别, 分析并处理说话人语音信号波形, 在确定的频率段提取频率时间序列, 为每个人构造一个具有指纹特征的数字化文件, 以特征模型的形式存储在计算机中. 应用时系统采集待识别的目标声音数据, 经过相同的分析和处理, 与存档的样本声音文件进行精确匹配, 从而得出二者是否一致的判断^[1].

声纹识别系统的主要任务包含: 语音信号的采集与处理、声纹特征提取和构建模型、模型匹配与识别, 其工作过程如图 1 所示, 分为建模和识别两个分支过程. 首先建立鉴别对象的声纹模型文件. 采集和分析语音信号, 经过滤波降噪等必要的处理后, 进行特征提取, 即从声音数据中选取唯一表现说话人的有效且稳定可靠的特征, 将其提取后生成用户声纹模型文件存放在计算机中. 在声纹识别过程中最主要的内容是特征提取和模式匹配. 特征提取就是从声音中选取唯一表现说话人身份的有效且稳定可靠的特征; 模式匹配就是把提取的特征与建模存档文件做相似性匹配. 识别时, 现场向系统输入特定语音, 提取其特征, 同预先存储的特征模式进行相似性匹配, 根据匹配结果确定其真假.

收稿日期: 2012-11-14

作者简介: 王会清(1957-), 男, 辽宁沈阳人, 副教授, 硕士. 研究方向: 计算机测控和数字信号处理.

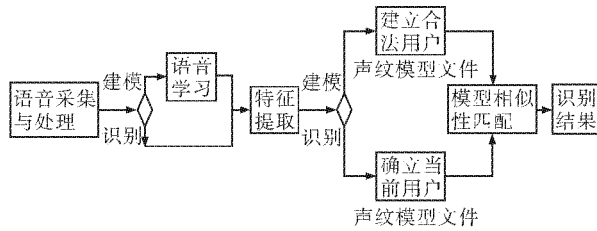


图 1 声纹识别系统工作流程

Fig. 1 The voiceprint recognition systems workflow

从图 1 可知系统的设计与实现分二步进行:一是建模步骤,包括语音的学习与注册阶段.考虑到用户声纹模型应该代表该用户的常态语音特征,建模步骤中系统多次、重复从注册语音信号中提取声纹特征序列,取平均构建说话人的特征模型;二是识别步骤,包括语音特征提取和声纹确认阶段.对待认证人的语音信号进行特征提取后,与学习阶段的模板进行比较,通过判断测试语音与所声称说话人的模型之间的相似度是否大于所规定的判决阈值作为识别的结果.无论是建模还是识别,都需要首先对输入的原始语音进行预处理,并进行特征提取.注册语音录入和待检语音录入两个子系统由同一个硬件及软件机构完成.本文的声纹识别系统是文本声纹认证,要求说话人提供指定的语句或词语,即注册语音,进行学习,认证时必须输入与注册语音相同的内容.学习时,每个说话人重复一定次数的发音,然后检测并分析每次发声的语音段,以提取特征,并利用动态时间规整技术,在时间上对齐特征序列且多次平均,形成每个说话人的参考模型.识别时,对语音信号进行特征分析,然后计算与参考模型的差距,与设定的阈值比较,若高于阈值则拒绝判决,低于阈值则接受判决.

2 系统程序设计与实现

2.1 语音信号的采集与处理

数据采集过程分为三步:初始化配置声卡、采样、释放声卡.首先对声卡进行设置,用 LabVIEW 自带的 Sound Input config(声音输入设置)来配置声卡采样所需的各个参数.将声卡的采样率设为 11 025 Hz,采样位数为 8 位,采样方式为单通道,并设置 131 072 字节大小的缓冲区,以保证声卡与 CPU(中央处理器)协调工作.然后进入样本声音录入程序^[2].

声卡对外部信号的采样在起始部分会有一些不稳定的数据,因此忽略开始的一段数据.在前面板设置了一个启动按钮,没有按动启动按钮时,程序处于等待循环阶段.当用户单击按钮时,开始录

制声音,根据用户设置的声音格式从声卡获得数据、存储以及转到信号分析处理模块.在程序框图设计中,为避免系统对声音的错误记录,安排了有效音频等待环节,只有当系统检测到的语音信号幅值超过预先设定值 8(经多次实验确定)后,开始对声音信号进行采集.该子程序将采集的信号作为一个波形,通过“提取单频信息”模块提取信号主频率的幅值,当信号的主频率幅值大于 8,则进入信号处理环节,即将有效的波形信号用带通滤波器滤波后,对波形作快速傅立叶变换,提取出主频率,否则继续等待.启动前后的等待环节均采用 While 循环.将单击启动按钮作为启动前循环停止的条件.在单击按钮之前,系统处于等待录制命令的状态.当单击事件发生,进入后面的等待循环.程序每循环一次,声卡采集的信号形成一个波形,再用“提取单频信息”模块提取波形的主频率的幅值进行判断,当幅值高于设定值 8 时,退出 While 循环,进入录制环节.

2.2 声纹建模子系统

语音录制与建模子程序主要是实现系统用户注册声音样本功能,即用户向系统输入自己的语音,经过提取声纹的主要特征,然后作为唯一的“密钥”存档.由于每个人说话时,内容不同,语音特征频率也不同,因此,当有语音输入时,对语音信号进行连续 10 次采样.每循环一次,将声卡采集的信号用“提取单频信息”模块提取波形主频率的幅值^[3-4].

其核心是通过录入的语音按时间顺序进行采集,提取出每一段信号的主频率,组成一个频率时间序列.这个频率时间序列表征了用户语音的特征信息.该特征信息不仅和语音的频率有关,还与语音的内容有关.用户需要说一句特定的话,并且用特定的语速,才能被系统正确识别.这样,就将声音样本信号变换成特征矢量文件.然后根据样本信号的特征矢量的分布状态,建立声纹识别模型.

该子程序通过 SI Read(声音读入子)模块读取声音输入,输入的语音信号捆绑为一个波形后,按时间先后创建波形,通过滤波器和窗模块对信号进行处理,窗是用于时间信号的时域窗,选为 Hanning(汉字窗).滤波后的信号分两路:一路是时间信号作为时域波形的输入数组.通过连线数据至 FFT(快速傅立叶变换)模块的时间信号输入端,计算时间信号的平均 FFT 频谱,输出的结果为平均 FFT 谱的幅度范围,至前面板的波形图显示;另一路的信号作为提取部分信号模块的输入信号,将动态数据类型转换成数组数据类型,其输

出转换为波形数据一维数组送给提取单频信息模块. 实现在输入的时域波形信号中, 查找幅值最高的单频, 返回单频的频率、幅值和相位. 在该模块的输出端检测到的单频的频率, 就是提取的单频信息, 输出给前面板的样本声音频率显示控件.

作为提取信号特征模块的输入信号, 查找幅值最高的单频, 或在指定频域内搜索, 返回单频的频率、幅值和相位. 输入信号可以是实数或复数、单个通道或多通道. 其时间信号输入为时域波形, 检测到的频率是检测到的单频的频率, 以赫兹为单位. 检测到的幅度是检测到的单频的幅度, 以 V_p 为单位. 该子 VI(虚拟仪器子程序)用于在 While 循环内部连续处理单通道或多通道. 提取单频信息的详细信息可通过下式表示实数单频信号:

$$x(n) = A * \cos(\frac{2\pi fn}{F_s} + \phi)$$

式中, A 、 f 和 ϕ 分别是单频信号的是幅值、频率和相位, F_s 是输入波形信号的采样率.

复数单频信号可通过下式表示:

$$x(n) = A * \exp\{j(\frac{2\pi fn}{F_s} + \phi)\} =$$

$$A * \{\cos(\frac{2\pi fn}{F_s} + \phi) + j * \sin(\frac{2\pi fn}{F_s} + \phi)\}$$

式中, A 、 f 和 ϕ 分别是单频信号的是幅值、频率和相位, F_s 是输入波形信号的采样率.

对于实数信号, 频率范围 = $(0, F_s/2)$. 对于复数信号, 频率范围 = $(-F_s/2, F_s/2)$.

程序设计同样采用 While 循环对输入的语音信号连续进行 10 次采集, 每次采集前都要进入等待模式, 以防止声音的错误记录. 当有语音信号输入这一环节时, 就进行一次信号采集, 对于采集的波形进行带通滤波, 滤掉过低和过高的频率成分, 获取有效的语音频率成分. 系统将滤波后的波形进行快速傅立叶变换, 提取出主频率. 然后等待下一次输入, 直到完成 10 次信号的采集. 在这一过程中, 用户输入的声音信号被转换成一个频率时间序列. 滤波器的参数设置为带通滤波器. 考虑到对系统产生干扰的噪声频率均较低, 集中在几十赫兹, 因而低频截止频率设定为 100 Hz, 远大于工频频率, 基本可以将低频噪声部分滤除, 改善采集信号的质量. 高频截止频率设定为 2 000 Hz, 因为大多数人的语音主要成分在 2 000 Hz 以下, 通过对高频的滤波以避免系统对主频率的误判. 此外, 为了提高滤波器的选频特性, 选择有限长冲激响应滤波器.

2.3 语音检测与识别子系统

用户登录系统时的用户名是调取库存语音特征文件的句柄, 系统按用户名检索到该用户的库

存密钥. 密钥读取模块使用文件读取控件读出注册的序列值. 在输入待测语音步骤用户按下启动按钮, 输入自己的声音. 待检声音被采集后得到采样数据, 系统通过声音数据前端处理, 提取其特征矢量. 产生的频率时间序列与读取密钥文件得到的序列被送入检测模块, 进行匹配计算. 模型匹配计算过程如下: 使用取整模块对密钥语音和待检声音两个时间频率序列取整数, 并对两个取整后的时间频率序列做减法, 得到一个差值序列. 这一差值序列体现了两组声音样品的匹配程度. 然后对差值序列取绝对值, 使每一个元素为正数, 与设定的阈值序列做比较. 最后将比较结果用 LED(发光二极管)指示灯阵列显示出来. 若差值序列的某一元素大于阈值, 则说明该位置不匹配, 在 LED 灯阵列中对应的灯上显示关闭状态; 反之, 差值序列的某一元素小于或等于阈值, 对应 LED 灯显示打开状态. 10 个灯分别对应显示 10 个频率的匹配结果. 当差值序列所有元素均在阈值范围之内时, 则用另一个 LED 灯显示 100% 匹配. 本文的阈值序列值取 50, 它体现了系统容许的匹配误差范围, 其确定方法采用预定初始值, 然后根据测试数据进行多次调整取得^[5-7].

3 测试结果

在基于 LabVIEW 的语音识别系统的前面板的下方有 10 个 LED 灯分别显示频率时间序列 10 个元素的匹配情况, 其旁边的另外一个 LED 灯显示全部频率的匹配结果. 当频率时间序列全部匹配时, 绿灯亮.

例如, 在系统程序测试中, 选择某密钥语句, 录入的样本密钥序列为: 302 Hz、220 Hz、396 Hz、243 Hz、533 Hz、748 Hz、1 175 Hz、988 Hz、1 200 Hz、920 Hz, 输入的待检声音也采用与密钥相同的语音内容, 测试结果为: 其中有 9 个频率特征匹配绿灯亮, 显示为匹配, 基本正确地反映了注册的语音样本密钥. 反复测试的结果, 基本大同小异, 并多次出现完全匹配, 绿灯全部亮的结果. 当说出不同的语音内容则没有绿灯亮. 一般来说, 若有 80% 左右的绿灯亮, 表示基本匹配. 所以, 本系统基本可以正确对密码声音进行匹配. 作者还选定其它十余个密钥语句分别测试, 匹配率都在 80% 以上.

请 10 个不同年纪和性别的参试者, 每个人自己选择一个密钥语句进行测试, 全部准确识别, 没有出现漏识案例.

他人伪冒待测人, 套用其密钥语句、模仿其声调测试误识率. 请两个男声和两个女声分别仿冒

另外的一个男声和女声进行测试,以及男女声相互仿冒的测试,都没有匹配率大于 40% 的情况,即误识率为零. 由于测试误识的样本数量太少,语音特征的代表性不够广泛,还不能得出系统的误识率为零的结论. 还没有对更复杂的语音对象进行测试研究,例如,没有对声音相似的兄弟、姊妹等等的测试及误识率数据.

整个用户测试案例超过 100 余次,仅有一个匹配率低于 8 个绿灯. 以 80% 以上的绿灯表示基本匹配,全部测试结果表明:声纹确认的漏识率小于 1%.

笔者用 MP3(音频器件及软件格式)反复向声卡输入同一个测试语句,经过 60 余次测试,识别匹配率都为 9 个以上绿灯. 表明识别系统运行稳定,声源的重复性是影响系统漏识率的主要因素.

4 结 语

通过对声纹识别的研究,设计了基于 LabVIEW 和声卡的声纹识别系统. 实现了对说话人的语音采集与处理,以及进行声纹特征的提取、建模和匹配等功能. 系统测试结果表明,用提取文本语音频率时间序列进行身份认证是一种能满足实际要求的可行方法,该方法可以区分不同的说话人,能有效地克服环境噪声带来的影响,并且在做说话人确认测试时达到较高的识别率. 在采用麦克进行语音录入与匹配时,学习、识别时间短,

声纹确认的漏识率小于 1%,认证准确、安全可靠、具有无区域限制、简单方便、成本低等特点.

但还存在一些实际问题有待进一步解决,如说话人的确定,即使是同一个人的语音,也会随着音量、语速以及身体状况带来的音质变化而变化,进而影响说话人的模型精确匹配. 此外,不同设备之间的语音采样率、压缩性、传输率等也会影响特征提取和识别结果. 人的语音复杂、多样,地域、方言、年龄、身体状态等等都影响识别结果,误识问题更是一个高安全系统的重要指标,有待使用更多、更广泛的样本检验系统的误识率.

参考文献:

- [1] 赵力. 语音信号处理[M]. 北京:机械工业出版社,2011.
- [2] 王会清,程勇. 家庭安防系统中声音信号的小波分析与降噪[J]. 武汉工程大学学报,2011,33(10):96-100.
- [3] National Instruments. LabVIEW User Manual[M]. Texas:National Instruments, 2005.
- [4] 黄松岭,吴静. 虚拟仪器设计基础教程[M]. 北京:清华大学出版社,2008.
- [5] 王会清,程勇. 基于 LabVIEW 的软件许可证系统设计[J]. 武汉工程大学学报,2011,33(4):81-84.
- [6] 谈宏华,潘正春,腾达. 基于 LabVIEW 的液压站监控系统[J]. 武汉工程大学学报,2010,32(12):94-97.
- [7] 李振国,宋吉江,李月然. 基于虚拟仪器的声音识别系统设计[J]. 山东理工大学学报:自然科学版,2011,25(1):101-103.

On implementation of voiceprint recognition in virtual instrument platform

WANG Hui-qing^{1,2}, ZHANG Tao², ZHOU Fan²

(1. Hubei Province Key Laboratory of Intelligent Robot, Wuhan 430074, China;

2. School of Computer Science and Engineering, Wuhan Institute of Technology, Wuhan 430074, China)

Abstract: A voiceprint recognition system was built using LabVIEW and sound card. Through the acquisition and processing of the audio signal, including analyzing the speaker's voice and extracting voice frequency-time sequence as the characteristic value, the model library was created. During the authentication, the on-site input voice was sampled by the system and extracted the characteristic value, and matched with the registered sample model, thus a consistency judgment was made based on the result. System test results show that the identity verification by extracting text-to-speech frequency-time series is a feasible method to meet the actual requirements. The method can distinguish between different speakers, effectively overcome the impact of environmental noise, and achieve higher recognition rate in the speaker verification test. Using the microphone for voice input and matching, learning and recognition time is short, confirmation by the voiceprint is accurate, and the ratio of missing recognition is less than 1%. With the advantages of no regional restrictions, simplicity and convenience, and low cost, the system can provide a safe and reliable technical support for the application of the voiceprint recognition system.

Key words: LabVIEW; voiceprint recognition; characteristics extraction; speaker verification

本文编辑:陈小平