

基于 MFCC 特征的声纹同一性鉴定方法

王学光¹ 诸珺文¹ 张爱新²

1 华东政法大学刑事法学院 上海 200052

2 上海交通大学网络空间安全学院 上海 200240

摘要 声纹作为当代司法鉴定技术发展的产物,在现代声像资料鉴定中发挥了至关重要的作用。传统的声纹分析方法是基于声音处理工具进行手工分析的,考虑到其具有严格的文本相关性以及比对的臆断性的缺点,其作为证据鉴定意见的证明力有待加强。文中提出了一种基于 Mel 频率倒谱系数的同一性鉴定方法,即提取并量化包含原始声音的共振峰及其时间轴信息的包络作为声纹特征进行同一性比对。此方法改进了传统 Mel 频率倒谱系数的不足,提取共振峰的突变并将元音与响辅音的转变特性加入声纹特征,以提高其识别度。实验证明,此方法在检材与样本无关的情况下,同一性鉴定的准确率达到 85%,方差控制在 9% 左右,具有良好的同一性识别;而在非同一性鉴定中,该方法也能在结合人工分析的情况下给出较准确的结果。

关键词: MFCC 特征; Mel 频率倒谱系数; 同一性鉴定; 证明力补强

中图法分类号 TP391

Identification Method of Voiceprint Identity Based on MFCC Features

WANG Xue-guang¹, ZHU Jun-wen¹ and ZHANG Ai-xin²

1 Criminal Justice College, East China University of Political Science and Law, Shanghai 200052, China

2 School of Cyber Science and Engineering, Shanghai Jiao Tong University, Shanghai 200240, China

Abstract As a product of the development of modern forensic technology, voiceprint plays an important role in modern audio-visual identification. The traditional voiceprint analysis method is based on the sound processing tools for manual analysis. Considering the shortcomings of strict text relevance and conjecture of comparison, its evidential power as evidence appraisal opinion needs to be strengthened. In this paper, a method of identification based on Mel frequency cepstrum coefficient is proposed, which is to extract and quantify the envelope containing the original sound formant and its time axis information as voiceprint features for identity comparison. This method improves the shortcomings of traditional Mel frequency cepstrum coefficient, which extracts the mutation of formant, and adds the transformation characteristics of vowels and consonants into voiceprint features to improve the correctness of recognition. Experiments show that the accuracy of identification is 85% and the variance is about 9% when the test material is independent of the sample text. Therefore, it has good identifiability for the same person identification of voiceprint. In the case of non same person identification of voiceprint, it proves to be far more accurate in combination with traditional manual analysis.

Keywords MFCC characteristics, Mel frequency cepstrum coefficient, Identity identification, Proof force reinforcement

1 引言

随着当今信息技术的快速发展,信息犯罪已然成为新兴犯罪的一种趋势。美国学者米尔建曾指出伴随着过去 50 年惊人的科学进步,越来越多的重要事实只能通过高科技手段进行调查。证据法学者赵毅宇认为用信息电子技术取代电子技术的概念更加合适,而何家弘定义电子数据为借助电子技术或者电子设备而形成的一切证据^[1]。现今视听资料已正式纳入我国诉讼法,成为法官判决和处理案件的有效依据。而关于语音同一性鉴定的专业论述大多集中在听觉分析方法、语音学及声学特征分析、自动识别技术以及语音同一性鉴定

过程中的质量控制与标准化等方面^[2]。在语音鉴定中,我国司法鉴定中心大多通过鉴定人的听觉语音分析来识别语谱图中的语音声学特征,如过零率曲线、辅音浊化现象、音渡特征、音强曲线、基频曲线、音节间过渡特征,以及分析辅音、噪音起始时间宽带语谱图的共振峰等特性来进行同一性鉴定。可以说,鉴定人的主观意志占比较大,因而该方法在鉴定行业中饱受诟病。

鉴于上述困境,学者们提出了不同的声纹特征提取与鉴定的方法。Zhou 等提出了一种基于 GMM-UBM 的声纹识别技术,利用 MATLAB 实现了与文本无关的声纹识别系统,从说话人识别应用角度,比较了主流静态特征参数 MFCC,

到稿日期:2021-01-06 返修日期:2021-04-23

基金项目:国家重点研发计划项目(2017YFB0802103)

This work was supported by the National Key R & D Program of China(2017YFB0802103).

通信作者:王学光(wangxueguang@ecupl.edu.cn)

LPCC, LPC 和结合了动态参数的 MFCC 的性能^[3];Li 提出了一种关于应用秩和检验法进行声纹鉴定的综合评判方法,采用数理统计中“非参数检验”的秩和检验方法进行综合评价,解决了身份识别的一些问题^[4];Peter 等提出建立不同语言的听觉特征集,并在瑞典语和芬兰语两种语言中进行了初步尝试^[5]。虽然上述方法在某些特定条件下能达到较好的鉴定效果,但是并未提供一套实际可行的通用方法。因此,本文在之前学者研究的基础上提出了一种通用的同一性鉴定方法,而这也是现今研究的主要目标之一。

考虑到语音具有严格的文本相关性以及比对的臆断性的缺点,本文提出了一种基于 Mel 频率倒谱系数的同一性鉴定方法,并对其性能进行了测试与分析。

2 Mel 倒谱系数

声纹分析提供了一种相对简单和客观的方法来观察和检查元音(单元音或双元音)的产生,通常是通过波形和频谱图进行分析^[6]。2010 年,英国英格兰与威尔士相关法院为对声纹进行统计分析,着手开发了自动语音识别系统(automatic speech recognition)。其原理是将检材与样本转换为基于 MFCC 的一个统计学模型,对其声纹特性进行统计分析并进行简化样本记录。在比对过程中,技术人员还将检材与样本转换后的统计学信息与现有罪犯的声纹数据库进行比对,以快速鉴定其音频信息来源。但由于检材时长的有限性以及系统无法自动选择合适的饱和音轨作为样本,因此该软件的开发陷入了困境。

Mel 倒谱系数是 Mermelstein 等在 1980 年根据人类听觉特性提出的一种语音信号特征,它与频率呈非线性关系。随着对语音信号研究的深入,Mel 倒谱系数在国际上已被证明是语音识别与检测应用中最成功的特征描述之一^[7]。所有语音类的研究基本上都是在 Mel 频率倒谱系数的基础上进行的。在声音处理中,Mel 倒谱系数是一种基于线性余弦变换的日志功率谱的非线性 Mel 刻度频率,体现了功率谱的声音特性。Mel 倒谱系数考虑到了人耳识别的特征,模拟了声音在耳内器官传播的过程,在确保外部声音衰减的同时引发阻塞效应^[8]。本文选择 Mel 倒谱系数是考虑到掩蔽效应的原理,即低音容易掩蔽高音,高音难以掩蔽低音,低频声波在内耳基膜上的行波距离大于高频声波。而且 Mel 倒谱系数能很好地提升高频的声音,防止其被低频声音掩蔽,这一特性与人耳识别的特性相符。

传统的 Mel 倒谱系数计算的是包含共振峰以及时间轴信息的包络,理论上其能在文本无关的情况下计算每个人的声纹相似度。但考虑到检材的有限性以及饱和音轨选择的困难程度,在文本无关的情况下不同人的声纹特征对于同一人鉴定的准确率较低。因传统的 Mel 倒谱系数仍存在着一定的局限性,所以有必要对传统的 Mel 倒谱系数进行改进,以实现在文本无关的情况下精准识别同一人的声纹。

3 一种改进的基于 MFCC 特征的声纹同一性鉴定方法

3.1 Mel 频率倒谱系数特征提取及改进

对于提取 Mel 频率倒谱系数特征,首先进行预加重,以

消除声带和嘴唇的影响;其次为避免信号的过渡变化进行分帧、加窗;接着进行加帧与快速傅里叶变化,并将语音通过三角带通滤波;最后进行对数变换与离散余弦变换,以将低频的包络信息与高频的频谱细节分离,从而求得 Mel 频率倒谱系数特征。本文鉴定方法的整体框架如图 1 所示。

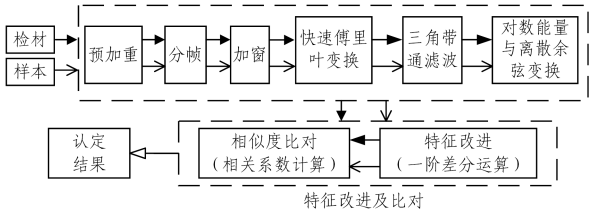


图 1 整体框架图

Fig. 1 Overall block diagram

3.1.1 预加重

预加重的目的是消除语音产生过程中声带和嘴唇的影响,通过将语音导入高通滤波器,从而补偿语音系统抑制的语音信号的高频部分。

$$H(z)=1-\mu z^{-1}$$
(1)

其中,μ取 0.98,以得到语音信号中的高频部分,突出高频共振峰。

3.1.2 分帧

考虑到音频信号的非平稳性,本文使用的音频处理技术基于概率模型,因此信号应是平稳的,以防止均值方差和其他统计数据失去意义。在信号分析过程中,必须将信号分割成包含原始特征的小帧^[9]。本文将音频信号分为多个帧,假设每个帧是稳定的,且以 20~30ms 为一帧,最后设置一定的比例重叠率。重叠是为了避免信号之间的过渡变化,如图 2 所示。



图 2 分帧与重叠示意图

Fig. 2 Schematic diagram of framing and overlapping

3.1.3 加窗

为了增加帧左右两端的连续性,每个帧应乘以 Hamming 窗口^[10]。假设成帧后的信号是 S(n),n=0,1,⋯,N,其中 n 是帧大小,将其乘以 Hamming 窗口后得到:

$$S'(n)=S(n)\times W(n)$$
(2)

而 W(n)的形式如下:

$$W(n,a)=(1-a)-a\times\cos\left(\frac{2\pi n}{N-1}\right),0\leq n\leq N-1$$
(3)

3.1.4 快速傅里叶变换

共振峰可以用来量化语音,以便更好地理解语音的清晰度,尤其是元音。根据语音声学的规律,共振峰频率通常与声道长度成反比,并且可以从穿过声音信号的稳态部分的信号声谱横截面求得^[11-13]。共振峰不仅仅是一段语音的音质决定因素,更是人说话发声部位的物理结构的描述。为了同时获取共振峰的信息以及时间轴信息,需要拟制一条包含所有共振峰点的平滑曲线并对其进行量化,本文借用模拟电路中

的一个术语“包络”来描述此平滑曲线。

由于包络信息处于语音的低频,而频谱的细节处于语音的高频,为了得到语音的包络,需要将音频信息从空间域转换到频率域并进行相关的数学运算,而以此得到的加窗、分帧后的每一帧则代表语音的能量。在语谱图中,越亮的地方代表能量越大,而传统的人工分析就是通过直接观察共振峰和转变信息来分析语音是否来自同一人。

快速傅里叶变换(FFT)是离散傅里叶变换(DFT)的快速算法,其计算的是离散傅里叶变换的序列(DFT),或其倒数(IDFT)^[14]。傅里叶分析将信号从其原始域(通常是时间或空间)转换为频域,反之亦然。DFT 是将一系列值分解为不同频率的分量而获得的。而语音信号的功率谱由语音信号功率谱的平方得到。语音信号的 DFT 如下:

$$X_{\omega}(k)=\sum_{n=0}^{N-1}x(n)e^{-j2\pi kn/N},0\leq k<N \tag{4}$$

3.1.5 三角带通滤波

将得到的语音频率域信息通过 Mel 滤波器组,即仿照人耳的特性将语音频率域变换到 Mel 域,以模拟鉴定人在听语音信息时,语音从外部进入人耳的过程,使其更符合人耳的听觉特性。考虑到三角带通滤波器的频率响应曲线最平滑,因此本文选择三角带通滤波进行运算。

大多数的 Mel 频率倒谱系数以标准赫兹计算,这主要是因为人体在一般情况下无法产生超出常规的声波^[15]。过滤器组是相互重叠的带通滤波器。基于 Mel 标度,过滤器组在频率 1 kHz 以下是线性的,在频率 1 kHz 以上呈对数关系^[16]。将快速傅里叶变换得到的结果输入三角带通滤波中,并且定义相应频率为:

$$H_m(k)=\begin{cases} 0, & k < f(m-1) \\ \frac{2(k-f(m-1))}{(f(m+1)-f(m-1))(f(m)-f(m-1))}, & f(m-1)\leq k\leq f(m) \\ \frac{2(f(m+1)-k)}{(f(m+1)-f(m-1))(f(m)-f(m-1))}, & f(m)\leq k\leq f(m+1) \\ 0, & k\geq f(m+1) \end{cases} \tag{5}$$

其中, $\sum_{m=0}^{M-1}H_m(k)=1$ 。

3.1.6 对数能量与离散余弦变换

为了将低频的包络信息与高频的频谱细节分离,假定声道是一个线性时不变系统,将原先的卷性信号通过快速傅里叶变换转变成乘性信号,然后通过取对数法将乘性信号转化为加性信号,最后使用低通滤波器组过滤掉高频的频率细节信号,并进行逆运算恢复,从而得到包络量化信息。

为了减少变换过程中的冗余数据,降低运算成本,本文选用离散余弦变换来完成这步操作。由于每帧的音量亦是语音的特征之一,因此应加上其对数能量来扩充 Mel 倒谱系数矩阵并进行后续计算。其子带能量的定义为:

$$s(m)=\ln(\sum_{k=0}^{N-1}|X_{\omega}(k)|^2H_m(k)),0\leq m\leq M \tag{6}$$

采用离散余弦变换(DCT)对 MFCC 进行单帧计数^[17],接着将对数能量进行离散余弦变换,得到 Mel 频率倒谱系数:

$$C(n)=\sum_{m=0}^{N-1}s(m)\cos\left(\frac{\pi n(m-0.5)}{M}\right),n=1,2,\cdots,L \tag{7}$$

通过加入时间导数得到基本稳定参数,可以提高语音识别系统的精度^[18]。

3.1.7 特征改进

经过以往的实验验证,纯粹的 Mel 倒谱系数准确率较低,并且纯粹的 Mel 倒谱系数只能反映共振峰以及其横轴上的时间信息,无法反映共振峰的变化规律。而且对于时长固定且有限的共振峰信息,由于 Mel 倒谱系数的有限性会导致部分样本与检材的相似度趋于一定的共同区间。因此,为了弥补 Mel 频率倒谱系数的有限性,一种策略是同时增加样本与检材的有效时长。但是,检材的有限性导致其无法通过“扩充”证据的方法进行检材填充。另一种策略是在现有的检材与样本的基础上增加特征采样点,以提高声纹信息的识别度^[19-21]。

因此,本文在传统的 Mel 频率倒谱系数的基础上,加入 Mel 频率倒谱系数的一阶差分,以求解得出共振峰的转变特征,即元音与响辅音之间的声纹变换特征。具体原理如下:

$$\Delta^n y_x=\Delta(\Delta^{n-1}y_x)=\Delta^{n-1}y_{x+1}-\Delta^{n-1}y_x \tag{8}$$

将 Mel 倒谱系数融入相似度计算公式,有:

$$d_t=\begin{cases} C_{t+1}-C_t, & t < k \\ \frac{\sum_{k=1}^K k(C_{t+k}-C_{t-k})}{\sqrt{2\sum_{k=1}^K k^2}}, & \text{others} \\ C_t-C_{t-1}, & t\geq Q-K \end{cases} \tag{9}$$

通过分析纯粹的 Mel 频率倒谱系数求得的相似度与包含一阶差分系数的 Mel 频率倒谱系数的相似度的关联与信息,可以基本确定检材和样本是否属于同来源的语音。

3.2 基于 Mel 频率倒谱系数的同一性鉴定

通过特征提取推导可以得出 Mel 频率倒谱系数矩阵。过长的音频得到的 Mel 频率倒谱系数矩阵较大,不利于两段录音特征的相似度比较。因此,首先对音频进行降噪处理,并将样本分割成 200 ms 等长的录音段;然后求多段录音的 Mel 频率倒谱系数并求出其平均值,得到样本的声纹特征;接着针对音频的特点,逐一计算样本特征与检材特征的相关性系数。相关性系数公式如下:

$$r(X,Y)=\frac{Cov(X,Y)}{\sqrt{Var[X]Var[Y]}} \tag{10}$$

其中, $Cov(X,Y)$ 为 X 与 Y 的协方差, $Var[X]$ 为 X 的方差, $Var[Y]$ 为 Y 的方差。

最后对得到的多个相关性系数求平均值,从而得到两段录音的相似度。此相似度是同一性鉴定的依据。

4 仿真实验

4.1 案例及实验环境

对于使用 Mel 频率倒谱系数进行同一性鉴定的方法,本文采用“先整体,后局部”的方法进行计算,即先将样本与检材进行整体比对,再引入一阶差分对 Mel 频率倒谱系数矩阵进行扩充,最后融入相似度计算公式进行比较。在一起模拟案件中,警方取证到某关键检材,并且获得犯罪嫌疑人甲、乙、

丙、丁 4 人的录音进行同一性鉴定。实验环境:本文所采用的操作系统是 Windows 10 专业版,软件是 MATLAB R2019a,处理器为 8 核 Intel(R) Core(TM) i7-6700HQ CPU @ 2.60GHz,内存为 32GB。

4.2 同一性验证

为了验证 Mel 倒谱系数在文本无关性的情况下也能得到良好的同一人比对结果,甲、乙、丙、丁 4 人采集样本时的语音内容与检材所录音的内容均完全不同,并在此基础上进行同一人鉴定分析。为了对语音进行分段处理,首先对语音进行加窗,Hamming 加窗的情况如图 3 所示。原始的语音信号分帧后与 Hamming 函数相乘,产生周期性信息,以体现要分析的重点信息。这是用来补偿在人类声音产生机制中被强调的谈话信号的高频部分^[15]。

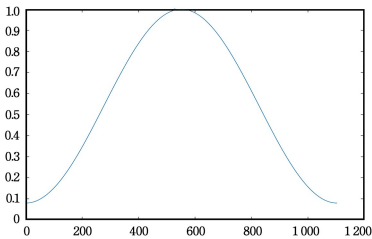


图 3 Hamming 窗函数
Fig. 3 Hamming window function

在进行加窗以后,对 4 份样本与 1 份检材进行传统的 Mel 倒谱系数运算,并且对样本进行相似度计算,运算结果如图 4 所示。

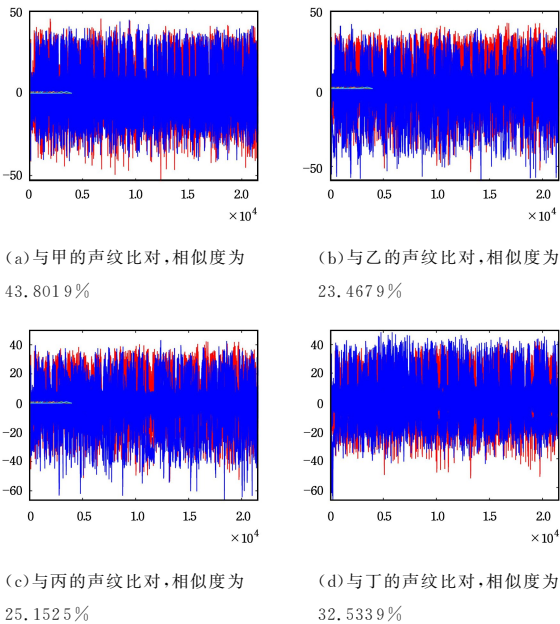


图 4 传统的 Mel 倒谱系数比对结果
Fig. 4 Comparison results of traditional Mel cepstrum coefficients

通过比对结果发现,检材与乙、丙的声纹相似度相对较低,而检材与甲、丁的声纹相似度相对较高。因此,需引入 Mel 倒谱系数进行一阶差分,求出语音共振峰的变换特征,并且分析两者之间的关联。改进的 Mel 倒谱系数比对结果如图 5 所示。

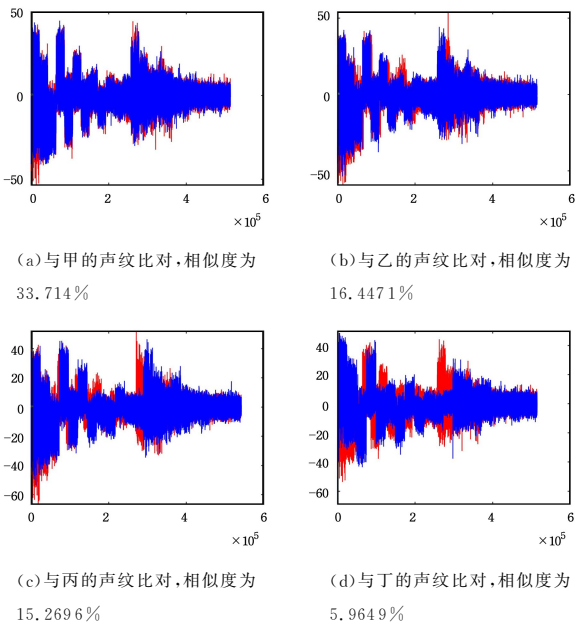


图 5 改进的 Mel 倒谱系数比对结果
Fig. 5 Comparison results of improved Mel cepstrum coefficients

通过总体分析后发现,与使用纯粹的 Mel 频率倒谱系数方法求得的甲与检材的声纹相似度相比,使用改进后的 Mel 频率倒谱系数方法求得的甲与检材的声纹相似度下降幅度最小,下降了 10.0879%,为 33.714%,基本可以认定为同一人;乙、丙的声纹在改进的 Mel 倒谱系数比对结果中,相似度只有 15%左右,基本可以认定为非同一人;而丁的声纹虽然在传统的 Mel 倒谱系数比对结果中相似度为 32.5339%,但是经过改进的 Mel 倒谱系数,发现丁的声纹与检材的声纹相似度仅为 5.9649%,下降幅度为 26.569%,因此基本上排除为同一人的可能性。

将甲的声纹比对图放大,以进一步对甲的同一性相似度进行精确比对。将检材中的声纹分割为 200 ms 等长的样本后求其平均值,并使用滑动窗口求相似度,结果如图 6、图 7 所示。

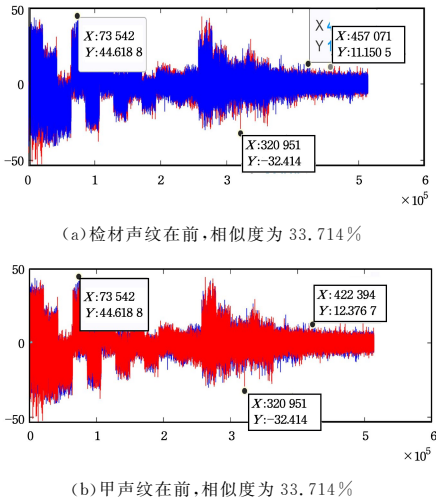


图 6 甲与检材的整体放大对比图(相似度为 33.714%)
Fig. 6 Enlarge the contrast map between the whole a and the test material(similarity of 33.714%)

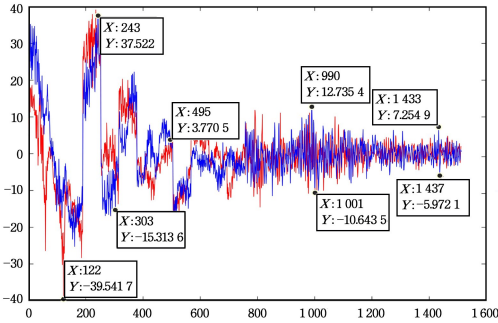


图 7 放大后的细节对比图(相似度为 70.907%)
Fig. 7 Enlarged detail comparison diagram(similarity of 70.907%)

对甲与检材的声纹的共振峰以及其转变信息进行多点采样对比,并将部分结果显示在图 6、图 7 中。从图 6 显示的信息可以看出,(73 542, 44.6188),(320 951, -32.414)和(422 394, 12.3767) 3 个点高度吻合。同理从图 7 中显示的信息可以看出,(122, -39.5417),(243, 37.522),(303, -15.3136),(495, 3.7705),(990, 12.7354),(1 001, -10.6435),(1433, 7.2549)和(1437, -5.9721) 8 个点高度吻合。因此可以看出,甲声纹与检材声纹高度相似,可以基本认为甲与检材是同一人。

4.3 人群随机采样实验

对于本文提出的基于 Mel 倒谱系数的文本无关性的同一人鉴定理论,我们在人群中随机采样了 30 组同一人的检材和 30 组非同一个人的检材,并且对此 60 组同一人鉴定的相似度进行统计与归类,实验结果如图 8 所示。

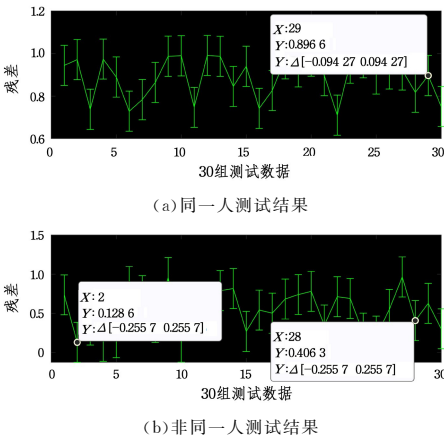


图 8 人群随机采样实验的残差分析
Fig. 8 Residual analysis of random sampling experiment

由图 8 可知,30 组同一人鉴定的相似度平均值为 88.71%,方差为 9.43%,可以看出,所提方法对于同一人测试的比对结果准确率较高,测试结果与预期值并未偏离太多,说明所提方法具有良好的同一人识别性。30 组非同一人鉴定的相似度平均值为 52.28%,方差为 25.57%,可以看出,所提方法对于非同一人声纹鉴定,不同人之间的声纹相似度略高,且方差较大,对于非同一个人的鉴定存在一定的误差,此时需要在人工分析的情况下结合改进的 Mel 倒谱系数,再次进行非同一人结果的判断。

结束语 声纹识别技术的关键是从语音信号中提取具有代表说话人特征的语音特征参数。考虑到当下鉴定人大多运

用自己的从业经验对两段语音是否为同源进行判定,本文提出了一种基于 MFCC 特征的声纹同一性鉴定方法,改进了 Mel 倒谱系数的不足,在声纹特征中加入共振峰的突变以及元音与响辅音的转变特性,以提高辨识度。实验证明,对于同一人的声纹鉴定,本文方法具有很好的识别度;而对于非同一个人的声纹鉴定,该方法则需结合人工审阅的方式来弥补其较大的方差,这也是我们下一步要深入研究的方向。

参 考 文 献

[1] ZHAO Y Y. The confusion and solution of audio-visual materials, electronic data and electronic evidence— from the perspective of information electronic technology[J]. Journal of South China University of Technology (Social Science Edition), 2020 (2):1-10.

[2] KANG J T, WANG L, WANG X D, et al. A Review on Researches of Forensic Phonetics and Acoustics in 2017[J]. Forensic Science and Technology, 2018, 43(3):179-186.

[3] ZHOU Y Y, KONG Q. Research of Feature Parameters in Voiceprint Recognition Technology Based on GMM-UBM[J]. Computer Technology and Development, 2020(5):1-11.

[4] LI W P. A study on the application of rank-sum test for comprehensive evaluation of acoustic pattern identification[J]. Journal of Criminal Investigation Police University of China, 1994(3): 62-64.

[5] PETER F, CAO H L, LEI Y J. A developmental history of forensic speaker comparison in the UK[J]. Evidence Science, 2019, 27(6):730-740.

[6] WANG Z N, CHEN Y, WU M H, et al. Acoustic Analysis of Mandarin Chinese Vowels Produced by Young Adults[J]. Rehabilitation Medicine, 2020, 30(3):183-191.

[7] YU Y S. Research on Pornographic Audio Detection Algorithm Using MFCC Features and Vector Quantization[D]. Lanzhou: Lanzhou University, 2015.

[8] BRUMMUND M K, SGARD F, PETIT Y, et al. Three-dimensional finite element modeling of the human external ear: Simulation study of the bone conduction occlusion effecta[J]. Journal of the Acoustical Society of America, 2014, 135(3):1433-1444.

[9] WU J Q, YU J J. An improved arithmetic of MFCC in speech recognition system[C]//2011 International Conference on Electronics, Communications and Control (ICECC). IEEE, 2011:719-722.

[10] HIDAYAT R, BEJO A, SUMARYONO S, et al. Denoising Speech for MFCC Feature Extraction Using Wavelet Transformation in Speech Recognition System[C]//2018 10th International Conference on Information Technology and Electrical Engineering (ICITEE). 2018:280-284.

[11] PETERSON G E, BARNEY H L. Control methods used in a study of the vowels[J]. J. Acoust. Soc. Am. , 1952, 24(2):175-184.

[12] PETERSON G E. Parameters of vowel quality [J]. J. of Speech & Hear. Res. , 1961, 4(1):10-29.

[13] STRANGE W. Evolving theories of vowel perception[J]. J. Acoust. Soc. Am. , 1989, 85(5):2081-2087.

[14] VIJAYAN A, MATHAI B M, VALSALAN K, et al. Throat microphone speech recognition using mfcc[C]//2017 International Conference on Networks & Advances in Computational Technologies(NetACT). 2017;392-395.

[15] YANG J B, XING Y L, CAO T Y, et al. Research on Speaker-Independent Speech Recognition Feature Based on Mellin Transform and Mel Frequency Analysis[J]. Pattern Recognition and Artificial Intelligence, 2020, 18(3):350-353.

[16] MESEGUER N A. Speech analysis for automatic speech recognition[D]. Trondheim: Norwegian University of Science and Technology, 2009.

[17] WINURSITO A, HIDAYAT R, BEJO A. Improvement of MFCC feature extraction accuracy using PCA in Indonesian speech recognition[C]//2018 International Conference on Information and Communications Technology(ICOIACT). 2018;379-383.

[18] RIGAZIO L, JUNQUA J, WELLEKENS C. Fundamentals of Speech Recognition[J]. AT&T, 1993;507.

[19] SENTHILDEVI K A, CHANDRA E. Keyword spotting system

for Tamil isolated words using Multidimensional MFCC and DTW algorithm[C]// International Conference on Communications & Signal Processing. IEEE, 2015;550-554.

[20] MAHESHA P, VINOD D S. LP-Hillbert transform based MFCC for effective discrimination of stuttering dysfluencies[C]// International Conference on Wireless Communications. 2017; 2561-2565.

[21] LI Q, YANG Y, LAN T, et al. MSP-MFCC; Energy-Efficient MFCC Feature Extraction Method with Mixed-Signal Processing Architecture for Wearable Speech Recognition Applications [J]. IEEE Access, 2020(8):48720-48730.



WANG Xue-guang, born in 1975, Ph.D, professor, is a member of China Computer Federation. His main research interests include computer networks, big data application and electronic data.