

基于 ARIMA 预测 MFCC 特征的声纹同一性鉴定方法

王学光, 诸珺文, 张爱新

引用本文

王学光, 诸珺文, 张爱新. [基于 ARIMA 预测 MFCC 特征的声纹同一性鉴定方法](#)[J]. 计算机科学, 2022, 49(5): 92-97.

WANG Xue-guang, ZHU Jun-wen, ZHANG Ai-xin. [Identification Method of Voiceprint Identity Based on ARIMA Prediction of MFCC Features](#)[J]. Computer Science, 2022, 49(5): 92-97.

相似文章推荐 (请使用火狐或 IE 浏览器查看文章)

Similar articles recommended (Please use Firefox or IE to view the article)

[基于 MFCC 特征的声纹同一性鉴定方法](#)

Identification Method of Voiceprint Identity Based on MFCC Features

计算机科学, 2021, 48(12): 343-348. <https://doi.org/10.11896/jsjcx.210100038>

[基于 ARIMA 模型的 Ad-hoc 网络节点位置预测加权分簇算法](#)

ARIMA-based Weighted Clustering Algorithm for Prediction of Nodes' Location in Ad-hoc Network

计算机科学, 2012, 39(3): 47-50.

基于 ARIMA 预测 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 声纹识别技术的关键是从语音信号中提取具有说话人特征的语音特征参数。考虑到当下大多是运用鉴定人的经验对两段语音是否来源于同一人进行判定,在前期研究的基础上,结合 MFCC 特征,提出一种基于 ARIMA 预测的声纹同一性鉴定方法,以提高具有年份差距的检材与样本比对的准确率。此方法在 Mel 倒谱系数声纹同一性鉴定方法基础上,采用自回归综合移动平均季节序列作出线性最小均方估计,对声纹特征进行预测,改良了包含元音与响辅音的共振峰特性。实验证明,ARIMA 时间序列的预测结果很好,且使用 ARIMA 改良的基于 Mel 倒谱系数的文本无关同一性鉴定的准确率较高,相似度在 60% 以上。

关键词: ARIMA 预测; Mel 倒谱系数; MFCC 特征; 同一性鉴定

中图分类号 TP391

Identification Method of Voiceprint Identity Based on ARIMA Prediction of MFCC Features

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

¹ College of Criminal Justice, 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 The key of vocal pattern recognition technology is to extract the speech feature parameters with representative speaker characteristics from the speech signal. Considering that most of the contemporary determinations are made using the experience of the identifiers, combined with MFCC features, this paper proposes an ARIMA prediction-based vocal identity identification method on the basis of previous study to improve the accuracy of the comparison between the examination materials with year gaps and the samples. This method uses an autoregressive integrated moving average seasonal series based on the Mel inverse spectral coefficient vocalic identity identification method, makes linear least mean square estimation, and improves the resonance peak characteristics containing vowels and loud consonants. It is demonstrated that the prediction results of ARIMA time series are good, and the accuracy of text-independent identity identification based on Mel inverse spectral coefficients using the modified ARIMA is high, with a similarity of more than 60%.

Keywords ARIMA prediction, Mel cepstrum coefficient, MFCC characteristics, Identity identification

1 引言

随着电子信息技术的快速发展,对视听资料的分析已然成为司法鉴定与刑事侦查的重要手段。声纹识别作为身份验证的一种手段,已被广泛应用于实际生活中^[1-2]。据统计,在中国裁判文书网上以视听资料为关键字进行相应检索,可发现民事诉讼中涉及视听资料的案件有 9500 多起,而在刑事诉讼中则多达 27000 多件。因而,以语音内容作为案件的关键证据,其真实性及同一性的认定显得十分重要。在语音同一性鉴定方面,当今关于语音同一认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

音同一认定过程中的质量控制与标准化等方面^[3]。而传统语音智能识别方法对语音的识别精度较低,识别语音的错误率较高,识别效果并不理想。基于上述问题,不同学者提出不同的模型与算法来优化语音智能识别^[4]。Zhou 等提出一种基于 GMM-UBM 的声纹识别技术,该技术利用 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, LPCC, LPC 和 MFCC 结合动态参数的性能^[5]; French 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^[6]。但这些方法没有考虑检材与样本之间存在年份差距的情况,对于具有年份差距的检材与样本之间的比对,其识别效果并不

到稿日期:2021-04-07 返修日期:2021-07-01

基金项目:国家重点研发计划(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)

理想,而国际上也没有太多的研究策略来解决这个问题。

虽然人的声音随着时间的推移会发生改变,但同一人的个别音符的发声习惯一般不会发生改变。因此,对于有一定年限的检材与样本的同一性对比,有学者通过对元音、响辅音的部分共振峰特性进行对比来弥补年限差距。Morrison 提取了 27 位澳大利亚男性语音中的 5 个复合元音的共振峰特征,使用逻辑回归方法进行特征融合后,发现存在年限差距的两段语音的系统识别错误率大大降低^[7]。Rose 基于 297 位日本男性语音数据库提取了 5 个元音的 LPC 倒谱特征,特征融合后系统识别的准确性显著提高^[8]。Zhang 等从 60 位女性汉语语音数据库中选取了 /ei/, /ai/ 和 /iau/ 这 3 个元音的共振峰和基频特征进行融合,融合系统的识别性能虽不稳定但有一定的提升^[9]。近年来,国际上对法庭证据的检验和评价要求越来越高,以似然比为框架的法庭证据评价新范式也被越来越多地应用于各个国家的司法实践^[10]。因此,似然比作为反应真实性的一种指标,将其应用到存在年份差距的检材与样本的对比中也是同一性鉴定中常采用的一种策略。

上述研究方法仅仅是从元音、响辅音的部分共振峰特性而非整段语音来进行声纹同一性鉴定,而且采用概率论的策略来提高样本与检材的相似性,其本身具有一定的局限性。因此,根据人的声音随着时间的推移而发生变化的特性,对具有年份差距的样本声纹特征进行预测后再与检材进行对比的鉴定策略显得尤为重要。

2 基于 ARIMA 时间序列预测 Mel 频率倒谱系数模型

2.1 ARIMA 季节序列

在现有文献^[11-15]中,对有一定年限的检材与现存样本的同一性进行对比,文献中使用了不同的 DTW 算法对不同的

特征向量进行关键词识别。该方法主要适用于大量需要搜索的口语文档,以了解它们是否包含某些特定的单词^[16]。国外已经使用 MFCC 特征和 DTW 算法开发了多种语言的孤立词识别系统,如马拉地语^[17]、马拉雅拉姆语^[18]、孟加拉文^[19]和阿拉伯语^[20],以弥补这种年限差距造成的同一性认定的不稳定性。但此方法仅仅是通过关键字的识别来弥补具有年份差距的检材与样本之间的相似度不稳定性问题,并未拉平年份差距。

考虑到人的声音是按照时间顺序变换的,随着时间的变换,人的声音也会逐渐发生变化。因此本文认为,以不同时间采样到的声音构成时间序列,可以从所构成的这一事件序列过去的变化规律推断今后变化的可能性、变化趋势以及变化规律。由于人对一个完整的句型或词组的发音具有一定的规律,经过验证满足季节序列,所以此时间序列较为特殊,可采用自回归综合移动平均(Auto Regressive Integrated Moving Average, ARIMA)季节序列。

ARIMA 是一种基于统计的时间序列预测方法^[21]。许多研究人员提出 ARIMA 来预测不同领域的行为,如流感病毒(He 等)^[22]、SARS(Tao 等)^[23]。本文定义的平稳指宽平衡,其特性是序列的统计特性不随时间的平移而变化,即均值和协方差不随时间的平移而变化。除了存在失真、升沉和掩蔽外,人耳是最好的辨音“仪器”。因此,本文使用 Mel 倒谱系数对原始语音信息生成语音序列,以进行后续的 ARIMA 模型预测。

2.2 模型框架

基于 ARIMA 时间序列预测 Mel 频率倒谱系数模型的整体框架如图 1 所示,预测检材现在的 Mel 倒谱系数的计算过程详见第 2.3-2.5 节。

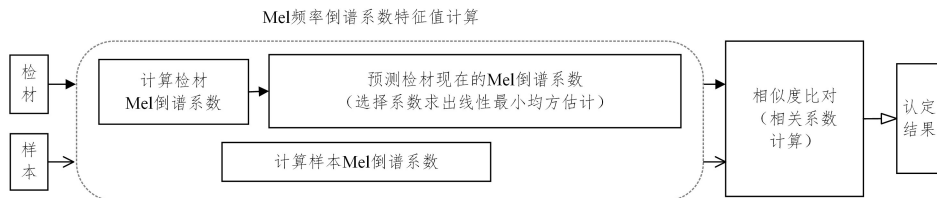


图 1 整体框架图

Fig. 1 Overall block diagram

2.3 选择系数的满足条件

设不同时间采样得到的 Mel 频率倒谱系数组成的序列 $\{X_t, t=0, \pm 1, \pm 2, \dots\}$ 是零均值平稳序列,从时间序列预报的角度引出偏相关函数。如果已知 $\{X_{t-1}, X_{t-2}, \dots, X_{t-k}\}$ 的值,要求对 X_t 做出预报。此时可以考虑由 $\{X_{t-1}, X_{t-2}, \dots, X_{t-k}\}$ 对 X_t 的线性最小均方估计,即选择系数 $\varphi_{k,1}, \varphi_{k,2}, \dots, \varphi_{k,k}$,使得

$$\min \delta = E[(X_t - \sum_{j=1}^k \varphi_{k,j} X_{t-j})^2] \quad (1)$$

将 δ 展开,得

$$\delta = \gamma_0 - 2 \sum_{j=1}^k \varphi_{k,j} \gamma_j + \sum_{j=1}^k \sum_{i=1}^k \varphi_{k,j} \varphi_{k,i} \gamma_{j-i} \quad (2)$$

$$\text{令 } \frac{\partial \delta}{\partial \varphi_{k,j}} = 0, j=1, 2, \dots, k, \text{ 得}$$

$$-\gamma_j + \sum_{i=1}^k \varphi_{k,i} \gamma_{j-i} = 0, j=1, 2, \dots, k \quad (3)$$

两端同除以 γ_0 并写成矩阵形式,可知 $\varphi_{k,j}$ 应满足下列线性方程组:

$$\begin{bmatrix} 1 & \rho_1 & \cdots & \rho_{k-1} \\ \rho_1 & 1 & \cdots & \rho_{k-2} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{k-1} & \rho_{k-2} & \cdots & 1 \end{bmatrix} \begin{bmatrix} \varphi_{k,1} \\ \varphi_{k,2} \\ \vdots \\ \varphi_{k,k} \end{bmatrix} = \begin{bmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_k \end{bmatrix} \quad (4)$$

2.4 非平稳序列的平稳化

经过一阶差分运算,原来的非平稳序列 X_t 转化为平稳序列 W_t 。如果一阶差分还不能使时间序列平稳化,还可以进行二阶差分、三阶差分,直至第 d 阶差分,最后将序列化为平稳序列。

一般地, d 阶差分的定义为:

$$\nabla^d X_t = (1-B)^d X_t \quad (5)$$

其中, ∇^d 为 d 阶差分算子, 有:

$$\begin{aligned} \nabla^d &\equiv (1-B)^d \\ &= 1 - \binom{d}{1}B + \binom{d}{2}B^2 + \dots + (-1)^{d-1} \binom{d}{d-1}B^{d-1} + \\ &\quad (-1)^d B^d \end{aligned} \quad (6)$$

设 $\{X_t, t=0, \pm 1, \pm 2, \dots\}$ 是非平稳序列。若存在正整数 d , 使得:

$$\nabla^d X_t = W_t \quad (7)$$

而 $\{W_t, t=0, \pm 1, \pm 2, \dots\}$ 是 $ARMA(p, q)$ 序列, 则称 X_t 是 $ARMA(p, d, q)$ 下序列。这时, X_t 满足

$$\varphi(B)\nabla^d X_t = \theta(B)\varepsilon_t \quad (8)$$

本文在建立 ARIMA 序列模型时, 使用 $d=2$ 的情形。先对 X_t 的样本 x_1, x_2, \dots, x_m 计算样本自相关函数与样本偏相关函数。如果是截尾的或是拖尾的(即被负指数控制的), 说明已服从 ARIMA 模型。若自相关函数与偏相关函数至少有 1 个不是截尾的或拖尾的, 说明 X_t 不是平稳的, 可以做 d 阶差分, 并求其样本自相关函数与样本偏相关函数, 再用上述方法讨论。这样, 直至判断 $\nabla^d X_t$ 是平稳序列未知。在实验中, 样本自相关函数或样本偏相关函数的图形虽然下降, 但下降很慢, 则这是非平稳序列, 应作差分运算。

2.5 ARIMA 时间序列的声纹特征预测

现在开始运用 ARIMA 序列模型对犯罪嫌疑人的声纹特征进行预报。

当 $d=2$ 时, $\nabla^2 \hat{X}_k(m) = \hat{W}_k(m)$, 即

$$\hat{X}_k(m) - 2\hat{X}_k(m-1) + \hat{X}_k(m-2) = \hat{W}_k(m) \quad (9)$$

复原 $\hat{X}_k(m)$, 可得

$$\hat{X}_k(m) = X_k + m(X_k - X_{k-1}) + \sum_{j=1}^m (m+1-j)\hat{W}_k(j) \quad (10)$$

得到的 ARIMA 序列如下:

ARIMA(2,0,2) Model (GaussianDistribution):

	Value	StandardError	TStatistic	PValue
Constant	0	0	NaN	NaN
AR{1}	0.16719	0.0175770	9.5124	1.8640×10^{-21}
AR{2}	0.17409	0.0071228	24.4410	6.2646×10^{-132}
MA{1}	-0.54918	0.0176950	-31.0360	1.7871×10^{-211}
MA{2}	-0.21499	0.0137340	-15.6530	3.1551×10^{-55}
Variance	32.36700	0.0294070	1100.6000	0

3 实验

3.1 案例及实验环境

在一起模拟案件中, 警方得到了 15 年前一起故意杀人案犯罪分子的录音, 需进行同一人鉴定, 以抓获凶手。警方怀疑录音的声音属于甲, 而乙、丙也有作案的嫌疑。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, 处理器为八核 Intel(R) Core(TM) i7-6700HQ CPU @ 2.60GHz, 内存为 32GB。

3.2 ARIMA 时间序列预测声纹特征的可行性验证

此案件是对不同时间、不同录音设备录制的语音进行同一性鉴定, 应根据 15 年前的录音片段对当年的声纹特征进行预测, 并完成相似度的对比。使用 ARIMA 时间序列对声纹特征进行预测, 得到的波形图和加直流偏移如图 2 所示。

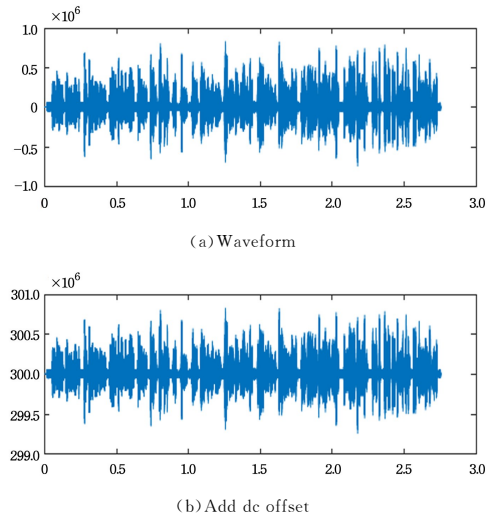


图 2 波形图和加直流偏置

Fig. 2 Waveform and DC bias

从图 2 中可以看出, 检材所选取的语音为饱和音轨, 且加直流偏移后语音呈均匀分布。

对语音加 300v 的直流偏移, 则它的波形就以 $y=5$ 为中心上下摆动, 产生向上偏移。将声纹波形信息放大, 如图 3 所示。

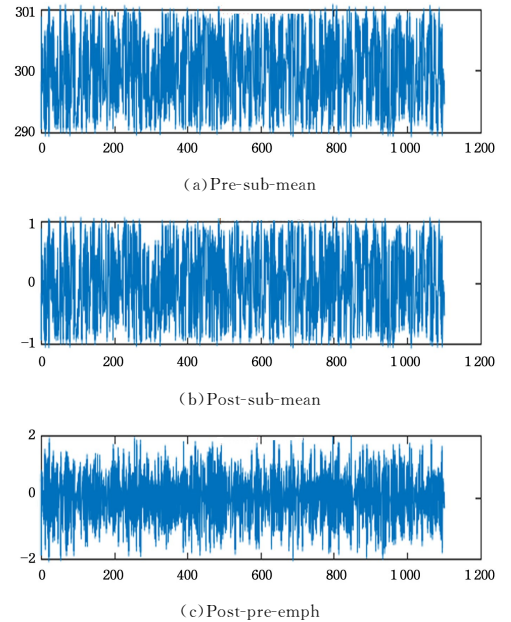


图 3 前均算子与后均算子

Fig. 3 Pre sub mean and post sub mean

从图 3 中可以看出, 其前均算子与后均算子分布于 $[299, 301]$ 和 $[-1, 1]$ 区间, 可进行 FFT 变换。

在进行 FFT 变换时, 帧和频率的关系如图 4 所示。从图中可以看出, 每一帧已趋于一定的稳定性。

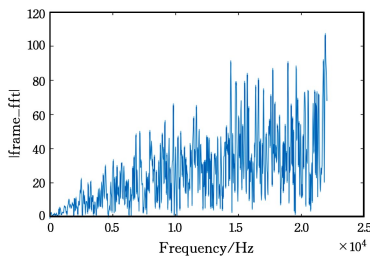


图 4 帧和频率的关系

Fig. 4 Relationship between frame and frequency

最后,经过 ARIMA 时间序列的预测,得到检材现在的 Mel 频率倒谱系数图,如图 5 所示。

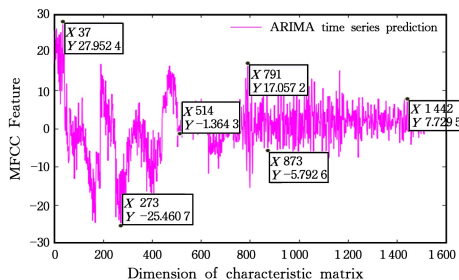


图 5 ARIMA 时间序列预测 MFCC 特征图

Fig. 5 MFCC feature map of ARIMA time series prediction

从图 5 中可以看出,检材预测的 MFCC 已随时间发生变化,但此变化符合人声音变化的特性。

3.3 ARIMA 时间序列预测声纹特征的准确性验证

将预测得到的 Mel 频率倒谱系数与甲样本的声纹特征进行相似度比对,发现相似度为 62.474%,并且通过观察发现,预测特征与甲样本特征大部分重合,可以基本认定为同一人,如图 6 所示。

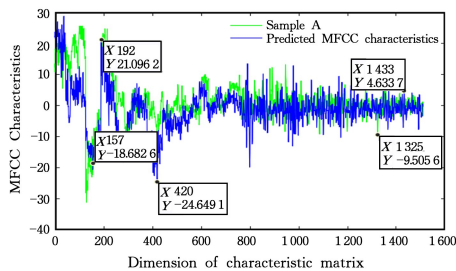


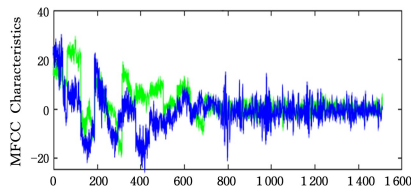
图 6 预测 MFCC 与甲声纹对比图

Fig. 6 Comparison of predicted MFCC and A's Sample

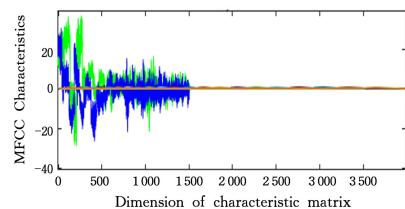
对图 6 的预测检材 MFCC 特征与甲样本声纹相应的共振峰信息进行多点采样比对。其中随机选取(157, -18.6826), (192, 21.0962), (420, -24.6491), (1325, -9.5056), (1433, 4.6337)5 个点,其比对信息如图 6 所示,此 5 个点的信息高度吻合。

将检材预测的 MFCC 特征与乙、丙的声纹特征进行比对,比对结果如图 7 所示。从图 7 比对的相似度可以看出,图 7(a)为预测 MFCC 与乙样本的比对结果,其相似度为 40.674%;图 7(b)为预测 MFCC 与丙样本的比对结果,其相似度为 13.312%。鉴定人员可以基本确定此段检材的声音不属于乙或丙。事实上,乙是甲的大学同学,而丙是甲的小学

同学,平时生活习惯趋同的人的外貌与声音有同一性的趋势。这和英格兰利物浦大学的托尼·利特尔学说也相符。



(a) Comparison of predicted MFCC characteristics and sample B the similarity was 40.674%



(b) Comparison of predicted MFCC characteristics and sample C the similarity was 13.312%

图 7 与其他人员的声纹对比图

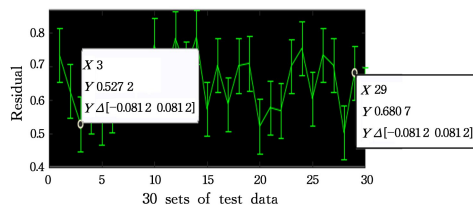
Fig. 7 Voiceprint comparison with other personnel

3.4 人群随机采样实验

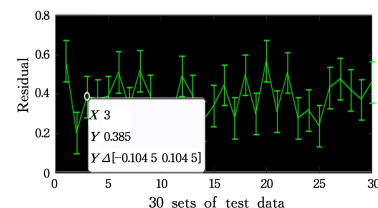
3.4.1 采样实验

对于随机人群的声纹预测,由于实验周期较长,而父子、母女之间的声纹在语气、清浊音、语调、情绪等方面具有事实上的遗传性,因此本文将子、女的声纹进行预测后与父、母的声纹进行相似度计算,以此对 ARIMA 预测的准确性进行评估。

对于本文提出的基于 ARIMA 时间序列的 MFCC 特征预测,本文在人群中随机采样了 30 组父子、母女的不可预测检材和 30 组非可预测的检材,并且对此 60 组同一人鉴定的相似度进行统计与归类。实验结果表明,30 组同一人鉴定的相似度平均值为 64.78%,方差为 8.12%;30 组非同一个人鉴定的相似度平均值为 38.25%,方差为 10.45%,实验结果如图 8 所示。



(a) Prediction and comparison results of the same person



(b) Comparison results of prediction by different people

图 8 人群随机采样预测实验的残差分析图

Fig. 8 Residual analysis chart of population random sampling prediction experiment

3.4.2 实验结果对比及分析

对于智能语音的识别, Huang 等提出一种基于线性预测倒谱系数(Linear Predictive Cepstral Coefficient, LPCC)的不同频段语音分析算法^[4], Zeng 等提出一种基于高斯混合模型(Gaussian Mixture Model, GMM)和 LPC-MFCC 联合特征参数的识别方法^[24]。考虑到 LPCC 与 LPC-MFCC 都是基于文本无关的语音识别特征, 因此基于此两种算法统计的文本无关性的语音相似度与所提算法的语音相似度结果可进行对比参考。Huang 等的算法中, 在开始识别后 10 分钟, 传统的 LPCC 文本无关识别准确率为 46%, HMM 与 LPCC 的融合模型文本无关识别准确率为 60%。Zeng 等的算法中, 其提出的基于 GMM 的 LPC-MFCC 联合特征参数的识别方法的文本无关识别准确率可达到 25%, 两者的实验结果都较为理想。相比较而言, 本文算法的语音识别相似度在 60% 以上, 可以看出, 本文提出的识别策略在其他原有研究的基础上得到了进一步完善。

分析原因可以看出, HMM 与 LPCC 的融合模型文本无关识别准确率较高, 是因为 HMM 是马尔可夫链的一种, 它的状态不能被直接观察到, 只能通过观测向量序列观察到, 每个观测向量都是通过某些概率密度分布表现为各种状态, 每一个观测向量是由一个具有相应概率密度分布的状态序列产生。使用 HMM 与 LPCC 的融合模型, 能更有力地描述出语音序列与内部隐藏状况的关系模式。基于 GMM 的 LPC-MFCC 联合特征参数的识别方法虽然具有模型体积小、转换耗时少、可本地化计算等优点, 但识别的准确度较低。考虑到在 MFCC 参数与 LPC 参数的融合过程中, 引入 MFCC 参数也会对 LPC 的合成阶段产生干扰, 导致合成语音存在些许噪声, 因此其准确率的提升相比 HMM 与 LPCC 的融合模型较小。而 ARIMA 时间序列模型之所以能在不同年代的检材与样本中的识别度较高, 是因为时间序列的获取是通过实验分析获得的, 同时语音序列的突兀点较少, 因此能极大地保证模型运算的准确性。此外, 非平稳序列经过差分运算之后能转换为平稳序列, 在消除其局部水平或者趋势之后, 能显示出一定的同质性, 而这一特性恰好与人的声纹特征随着时间的变化相符。考虑到不同人的声带结构、发音习惯等声纹特征随着时间的推移并不会趋于一定的同质性, 因此基于 ARIMA 时间序列运算后的检材并不会对样本的同一人识别产生干扰。基于这一特性, 采用 ARIMA 时间序列模型对不同年份差距的检材与样本之间的相似度进行比对, 可使对同一人与非同一个人的识别准确率得到有效的提升。

此外, 从人群随机采样预测实验的残差分析中可以发现, 父、子和母、女的声纹特征具有一定的遗传意义, 并且可作为样本以研究 MFCC 的预测特性。父、子和母、女两者的 Mel 频率倒谱系数特征的相似度高于人群中的陌生人, 但重合度并未高到足以混淆同一性结果。因此, 基于 Mel 倒谱系数的 ARIMA 模型仍具有良好的语音识别性。

结束语 当今声像资料鉴定多致力于对同一年份或年份相近的检材与样本的相似度进行对比。为弥补鉴定人存在

主观性与臆断性的不足, 科研人员多致力于改进 Mel 倒谱系数、LPCC 等算法, 并未关注跨越一定年度的语音同一性鉴定。考虑到人的声音特征随时间变化的特性, 本文参考了 ARIMA 时间序列的预测模型, 在 Mel 倒谱系数的基础上使用 ARIMA 时间序列进行预测, 对跨越一定年份的检材进行 ARIMA 时间序列预测, 并将预测得到的声纹特征与样本进行对比。实验结果表明, ARIMA 时间序列的预测结果很好, 且使用 ARIMA 时间序列改良的基于 Mel 倒谱系数的文本无关同一性鉴定的准确率较高。预测后的子、女与父、母的相似度与同一人鉴定的相似度有趋近的趋势, 在此两者进行同一人与非同一人鉴定的问题, 将是下一阶段的研究方向。

参考文献

- [1] JI M F, CHEN N. Research on vocal tract spectrum conversion based on GMM model and LPC-MFCC[J/OL]. Journal of East China University of Science and Technology. <https://doi.org/10.14135/j.cnki.1006-3080.20201209001>.
- [2] ZHANG X, KONG H F, WANG H Y, et al. Difference Analysis of Formant in Network Voice Identification[J]. Computer Applications and Software, 2019, 36(3): 187-191.
- [3] KANG J R, 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.
- [4] HUANG X B, ZHANG L, CAO L, et al. Intelligent speech recognition methods of different frequency bands based on LPCC[J]. Electronic Design Engineering, 2020, 28(2): 22-25.
- [5] 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.
- [6] FRENCH P, CAO H L, LEI Y J. A developmental history of forensic speaker comparison in the UK[J]. Evidence Science, 2019, 27(6): 730-740.
- [7] Morrison G S. Likelihood-ratio forensic voice comparison using parametric representations of the formant trajectories of diphthongs[J]. Journal of the Acoustical Society of America, 2009(4): 2387-2397.
- [8] ROSE P. Forensic voice Comparison with Japanese Vowels Acoustics—a likelihood ration-based approach using segmental cepstra[C]// Proceedings of the 17th International Congress of Phonetic Sciences, 2011: 1718-1721.
- [9] ZHANG C, ENZINGER E. Fusion of multiple formant-trajectory and fundamental-frequency-based forensic-voice-comparison systems: Chinese /ei/, /ai2/, and /iau1/[J]. Journal of the Acoustical Society of America, 2013(5): 3295-3300.
- [10] ZHANG C L. A new paradigm of forensic evidence evaluation [J]. Journal of People's Public Security University of China (Science and Technology), 2018(1): 25-30.
- [11] BAHI H, BENATI N. A new keyword spotting approach[C]// International Conference on Multimedia Computing and Systems. IEEE, 2009: 495-501.

- [12] ZHANG Y, GLASS J R. Unsupervised spoken keyword spotting via segmental DTW on Gaussian Posteriorgrams [C] // IEEE Workshop on Automatic Speech Recognition & Understanding, 2009. ASRU 2009. IEEE, 2010: 398-403.
- [13] BARAKAT M S, RITZ C H, STRILING D A. Keyword Spotting Based on Analysis of Template Matching Distances [C] // 2011 5th International Conference on Signal Processing and Communication Systems (ICSPCS). IEEE, 2011: 405-412.
- [14] KESHET J, GRANGIER D, BENGIO S. Discriminative keyword spotting [J]. *Speech Communication*, 2009, 51(4): 317-329.
- [15] ALEX J S R, VENKATESAN N. Spoken Utterance Detection Using Dynamic Time Warping Method Along With a Hashing Technique [J]. *International Journal of Engineering and Technology*, 2014, 6(2): 1100-1108.
- [16] SHETTY S, ACHARY K K. Audio Data Mining Using Multi-perceptron Artificial Neural Network [J]. *International Journal of Computer Science and Network Security*, 2008, 8(10): 224-229.
- [17] GANGONDA S S, MUKHERJI P. Speech Processing for Marathi Numeral Recognition using MFCC and DTW Features [J/OL]. *International Journal of Engineering research and Applications*. <https://www.semanticscholar.org/paper/Speech-Processing-for-Marathi-Numeral-Recognition-Gangonda-Mukherji/0439f6c236c87472e26b924faf5de915b9a87a26# citing-papers>.
- [18] DAS A, ACHARJEE P, TALUKDAR H. Isolated BODO Spoken Word Identification using Mel-frequency Cepstral Coefficients and K-means clustering [J]. *International Journal of Advanced Research in Computer Science and Software Engineering*, 2013, 3(11): 1501-1506.
- [19] ALI M A, HOSSAIN M, BHUIYAN M N. Automatic Speech Recognition Technique for Bangla Words [J]. *International Journal of Advanced Science & Technology*, 2013, 50: 51-60.
- [20] LIMKAR M, RAMARAO R, SAGVEKAR V. Isolated Digit Recognition Using MFCC and DTW [J]. *International Journal of Advanced Electrical & Electronics Engineering*, 2012, 1(1): 59-64.
- [21] ELSHEIKH A H, SABA A I, ELAZIZ M A, et al. Deep learning-based forecasting model for COVID-19 outbreak in Saudi Arabia [J]. *Process Safety and Environmental Protection*, 2021, 149: 223-233.
- [22] HE Z R, TAO H B. Epidemiology and ARIMA Model of Positive-Rate of Influenza Viruses among Children in Wuhan, China: A Nine-Year Retrospective Study [J]. *International Journal of Infectious Diseases Ijid Official Publication of the International Society for Infectious Diseases*, 2018, 74: 61-70.
- [23] WANG T, ZHOU Y, WANG L, et al. Using autoregressive integrated moving average model to predict the incidence of hemorrhagic fever with renal syndrome in Zibo, China, 2004-2014 [J]. *Japanese Journal of Infectious Diseases*, 2016, 69(4): 279-284.
- [24] ZENG X, ZHANG X W, SUN M, et al. Research on vocal tract spectrum conversion based on GMM model and LPC-MFCC [J]. *Technical Acoustics*, 2020, 39(4): 451-455.



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.

(责任编辑:柯颖)