

基于 DTW 模型补偿的伪装语音说话人识别研究

李燕萍, 陶定元, 林 乐

(南京邮电大学 通信与信息工程学院, 江苏 南京 210003)

摘 要: 语音变声器及各种手机变声软件的出现, 在提供了极其方便且丰富的娱乐交互体验的同时, 也给语音通信带来了新的安全问题。由于其产生的电子伪装语音掩盖了语音本身的个性特征, 对现有的说话人识别技术来说是一种挑战, 且一旦被犯罪分子利用, 后果将十分严重。因此, 伪装语音说话人识别的研究成为当下的研究热点。提出一种针对电子伪装的说话人识别方法。对于由手机变声软件产生的电子伪装语音, 提取该语音的梅尔倒谱系数 (Mel Frequency Cepstral Coefficients, MFCC) 作为特征参数, 通过动态时间规整 (Dynamic Time Warping, DTW) 模型进行伪装程度鉴定, 再利用矢量量化 (Vector Quantization, VQ) 模型进行说话人识别, 从而设计了 DTW 与 VQ 相结合的电子伪装语音说话人识别系统。实验结果表明: 该系统能够有效解决 VQ 说话人识别系统对电子伪装语音识别率过低的问题, 识别效果得到了明显改善。

关键词: 电子伪装语音; 梅尔倒谱系数; 说话人识别; 动态时间规整; 矢量量化

中图分类号: TP302

文献标识码: A

文章编号: 1673-629X(2017)01-0093-04

doi:10.3969/j.issn.1673-629X.2017.01.021

Study on Electronic Disguised Voice Speaker Recognition Based on DTW Model Compensation

LI Yan-ping, TAO Ding-yuan, LIN Le

(College of Telecommunications & Information Engineering, Nanjing University of Posts and Telecommunications, Nanjing 210003, China)

Abstract: The appearance of voice changer and various voice software of mobile phone provides a very convenient and rich entertainment interaction experience, and at the same time, also gives voice communication new security issues. Electronic disguised voice produced masks the personality characteristics of voice itself, so the existing speaker recognition technology is a challenge, and once they are used by criminals, the consequences will be severe. Therefore, disguised voice speaker recognition is becoming a research hotspots. In view of electronic disguised voice produced by cell phone voice software, Mel Frequency Cepstral Coefficients (MFCC) are extracted as the characteristic parameters of voice signals, identifying the disguised degree of it by DTW model and carrying out speaker recognition by VQ to design a speaker recognition system of electronic disguised voice. The experimental results show that the system can efficiently solve the problem that VQ has a poor recognition rate for electronic disguised voices, and improve the performance obviously.

Key words: electronic disguised voice; MFCC; speaker recognition; DTW; VQ

0 引 言

近年来, 手机变声软件的流行, 在丰富人们业余生活的同时, 也给犯罪分子进行违法犯罪活动提供了新的途径^[1-3]。犯罪分子通过手机变声软件产生的电子伪装语音能掩盖自身语音, 从而躲避公安机关的侦查, 给此类案件的侦破增加了不少阻力^[1,4]。

语音作为人与人之间交流的基本方式之一, 也是重要的生物特征之一。目前在说话人识别领域, 作为

表征个体之间差异的特征参数主要有 MFCC 和线性预测谱系数 (Linear Prediction Cepstrum Coefficient, LPCC)。其中 MFCC 是基于听觉特性, LPCC 是基于声道特性^[5]。文中选取 MFCC 作为语音特征参数。

手机变声软件主要通过改变原始语音的音调, 产生电子伪装语音。随着伪装程度的加深, 说话人的原始语音与伪装处理后的语音差异增大^[6-7]。目前常用的 VQ 说话人识别模型对电子伪装语音的识别率低

收稿日期: 2015-08-25

修回日期: 2015-12-23

网络出版时间: 2017-01-04

基金项目: 国家自然科学基金资助项目 (61401227); 江苏省博士后基金 (1402067B); 智能语音技术公安部重点实验室 2014 年度开放课题 (2014ISTKFKT02)

作者简介: 李燕萍 (1983-), 女, 博士, 副教授, 研究方向为说话人识别、语音转换; 陶定元 (1989-), 男, 硕士研究生, 研究方向为说话人识别。

网络出版地址: <http://www.cnki.net/kcms/detail/61.1450.TP.20170104.1017.016.html>

下,无法完成识别此类语音的任务。在这种情况下,文中提出一种适用于识别电子伪装语音的新模型——DTW 与 VQ 相结合的模型,并将两者结合之后对 VQ 识别系统的性能进行分析,最后通过实验完成对该系统性能的测试。提高对电子伪装语音的识别率有助于与手机变声软件相关的违法犯罪案件,提高对犯罪嫌疑人身体的辨识度,从而为公安机关侦破此类案件提供帮助。

1 基于 DTW 与 VQ 的识别模型

1.1 电子伪装语音伪装程度的量化

在进行电子伪装语音识别模型研究之前,需要对伪装程度概念进行量化处理。文中的电子伪装语音由名为“高保真录音变声器”的手机变声软件产生,该软件主要通过改变音调来伪装原始语音。音调改变分为正向与负向两种,正向即提高原始语音的音调,改变幅度为 1,负向即降低原始语音的音调,改变幅度同样为 1。伪装程度可用符号加改变量表示。例如,一段语音音调提高了 9 个幅度,其伪装程度可用+9 表示。通过测试发现,经过该软件处理后,伪装程度高于+11 以及低于-11 的电子伪装语音语义基本丧失,即无法通过人耳辨别出此段语音的内容。据此将伪装程度分为从-11 至+11 的 22 个伪装级别,这与电子伪装语音的半音分类^[8-9]类似。

1.2 DTW 匹配模型

动态时间规整 (DTW) 是一种基于时间规整与距离测度的非线性规整技术^[10]。模板中已存在的语音称为参考模板,用于测试的语音称为测试模板。动态时间规整需要寻找一个时间规整函数 $m = \omega(n)$,使得测试模板的时间轴 n 非线性映射到参考模板时间轴 m 上,函数 ω 应满足:

$$D = \min \sum_{n=1}^N d[T(n), R(\omega(n))] \quad (1)$$

其中, $T(n)$ 为测试模板第 n 帧的特征参数; $d[T(n), R(\omega(n))]$ 与参考模板第 m 帧的特征参数 $R(m)$ 之间的欧氏距离测度; M 、 N 为参考模板与测试模板的长度; D 为测试模板矢量与参考模板矢量之间的最佳匹配路径。

但是动态规划计算量较大,所以采用 DTW 改进型路径^[11],改进后的匹配路径算法为:

$$D(n, m) = d(n, m) + \min[D(n-1, m), D(n-1, m-1), D(n-1, m+1)] \quad (2)$$

其中, $d(n, m)$ 是 $d[T(n), R(\omega(n))]$ 的简写。

利用 DTW 算法可实现对电子伪装语音伪装程度的鉴定,该理论基于假设:伪装程度相同或相似的语音

更容易匹配。其过程为:将一段待测语音与系统参考模板中的某个说话人的多段伪装语音进行匹配,可得到一个伪装程度的最佳估计值,若模板中有 N 个人的多段伪装语音,则得到 N 个伪装程度估计值,再取其平均值,由于伪装程度是整数值,所以结果需要进行四舍五入处理,最终结果作为该语音的伪装程度估计值。

1.3 VQ 识别模型

在说话人识别领域,矢量量化(VQ)是一种重要的信号压缩和识别方法^[12-13],而 VQ 码本的设计对 VQ 有着重要的影响,一个拥有 M 个说话人集合的系统需要为每一个人建立码本 Y_1, Y_2, \dots, Y_M 。目前,生成码本最常用的方法是 LBG 算法^[14],对训练矢量集合以及某种迭代算法生成更符合训练语音特征的码本。在识别时,提取待识别语音的特征矢量序列 X_1, X_2, \dots, X_N ,并用已生成的码本对特征矢量序列依次进行矢量量化^[15],并计算平均量化误差,公式为:

$$D_i = \frac{1}{M} \sum_{n=1}^M \min_{1 \leq m \leq M} [d(X_n, Y_m^i)] \quad (3)$$

其中, Y_m^i 是第 m 个码本的第 i 个码字; $d(X_n, Y_m^i)$ 表示待测特征矢量与码本矢量之间的距离,采用欧氏距离测度。

最终平均矢量量化误差 D_i 最小值所对应的第 i 个说话人即为系统的识别结果。

在电子伪装语音伪装程度已知的情况下,对 VQ 识别模型进行补偿,调整训练语音的伪装程度使其与测试语音相同,完成说话人识别向电子伪装语音说话人识别的过渡。

1.4 DTW 与 VQ 相结合的模型

通过 DTW 模型鉴定伪装程度,再通过 VQ 模型进行识别,完成对电子伪装语音的说话人识别,其系统框图如图 1 所示。

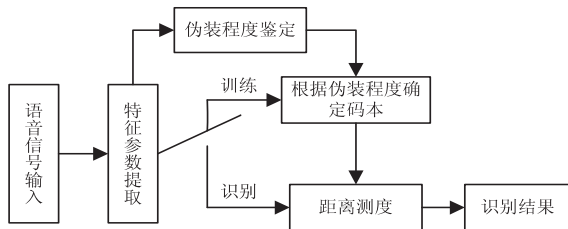


图 1 DTW 与 VQ 相结合的模型框图

2 实验分析

实验所用硬件为 PC 并配备普通声卡,软件为 Matlab 开发平台,录音环境为普通机房。

有 15 位男生和 15 位女生共计 30 人参与录音,每人采集从伪装程度-11 到+11 的 22 段语音,共计 660 段语音,作为 DTW 的参考模板语音,同时也是 VQ 模型的训练语音。语音长度为 20 s 左右,内容为一段描

述性语句,由于内容较长,故不在此赘述。30 位参与者录制测试语音,语音内容选为“不许报警,不许让别人知道,否则你的孩子就没命了”,长度为 5 s 左右,经过伪装处理,得到 660 段语音。

对实验语音进行端点检测,得到有效语音段。之后进行预加重、分帧(帧长 256,帧移 128)、加窗(汉明窗)处理,提取 20 维的 MFCC 参数。选取一段语音为例,语音内容为:“今天是 5 月 21 号,天气很好万里无云”,提取该段语音的 MFCC,如图 2 所示。经过电子伪装处理(伪装程度为+11)之后再提取其 MFCC,如图 3 所示。

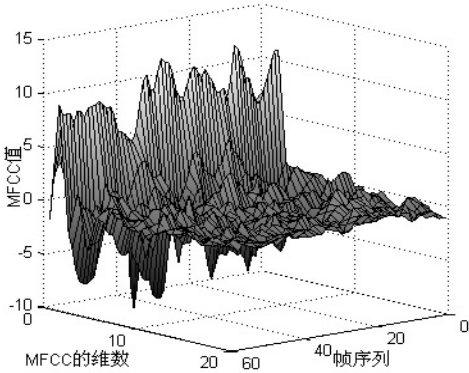


图2 提取一段正常语音的 MFCC

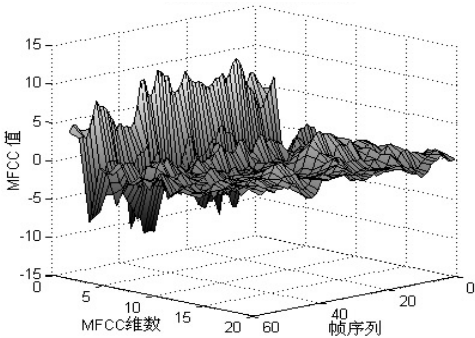


图3 提取伪装语音的 MFCC

通过对比图 2 与图 3 可知,正常语音经过电子伪装之后,特征参数 MFCC 会发生明显的改变。

实验根据测试语音和训练语音是否经过伪装处理分为四个部分:

- (1)测试语音与训练语音均为 30 人的正常语音,各计 30 段。
- (2)测试语音是 30 人的伪装语音(660 段),训练语音是 30 人的正常语音(30 段)。
- (3)测试语音与训练语音均为 30 人的伪装语音,各计 660 段且伪装程度未知。
- (4)测试语音与训练语音均为 30 人的伪装语音,各计 660 段且伪装程度已知。

实验部分(4)中,在电子伪装语音识别之前通过 DTW 模型进行伪装程度鉴定,使测试语音与训练语音的伪装程度已知。

DTW 模型对于伪装程度的鉴定效果如图 4 和图 5 所示。

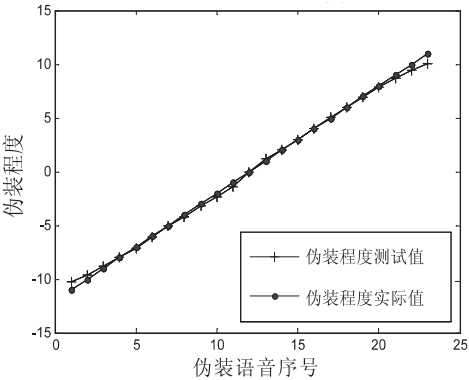


图4 测试语音伪装程度鉴定值

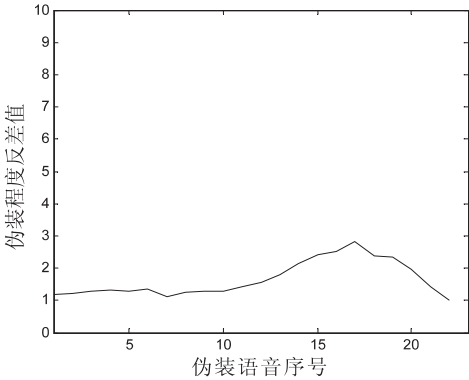


图5 测试语音伪装程度鉴定值方差

由图 4 可知,30 人的电子伪装语音经过 DTW 匹配之后的伪装程度均值曲线与实际伪装程度曲线比较接近,说明匹配效果良好。由图 5 可知,30 人的电子伪装语音经过 DTW 匹配之后伪装程度方差值在 0~3 之间,说明伪装程度鉴定值的浮动较小,DTW 匹配模型较为稳定。经过鉴定之后 30 人的伪装语音的伪装程度可视为已知。

实验识别阶段采用 VQ 识别模型。码本长度分别选择 16,32,64。四部分实验中 VQ 模型的识别率如表 1 所示。

表1 四部分实验 VQ 系统识别率汇总表

识别率	16	32	64
实验部分(1)	0.833 3	0.833 3	0.866 7
实验部分(2)	0.165 2	0.169 7	0.190 9
实验部分(3)	0.642 4	0.666 7	0.701 5
实验部分(4)	0.759 1	0.780 3	0.812 1

由表 1 可知,经过 DTW 模型匹配之后,VQ 模型识别效果与实验部分(2)相比,按照不同码本识别率分别提高了 59.39%,61.06%和 62.12%;与实验部分(3)相比,按照不同码本识别率分别提高了 11.67%,11.36%,11.06%;与实验部分(1)相比,按照不同码本识别率分别降低了 7.42%,5.30%和 5.46%,说明

DTW 与 VQ 相结合的模型在电子伪装语音存在的情况下,识别性能有很大提升,识别效果明显改善。但在说话人识别领域,该模型的识别效果并不理想,后续的研究可以通过使用改进后的伪装鉴定模型或者选取更为有效的特征参数等方法来进一步提高系统的性能。

3 结束语

电子伪装语音的存在,使得基于 VQ 模型的说话人识别性能降低,识别效果变得不理想。文中利用 DTW 模型匹配出测试语音的伪装程度,再将 VQ 模型训练语音的伪装程度调整至与测试语音同一伪装程度层面,实现对该模型的补偿,使其性能得到明显改善。实验结果表明:经过补偿之后的 VQ 模型对电子伪装语音的识别性能显著提升,识别效果良好。

参考文献:

- [1] Neustein A, Patil H A. Forensic speaker recognition: law enforcement and counter - terrorism [M]. [s. l.]: Springer, 2011.
- [2] 张翠玲,谭铁军,刘 昇. 伪装语音的自动话者识别研究 [J]. 刑事技术,2007(2):18-21.
- [3] 张翠玲. 伪装语音的声学研究 [D]. 天津:南开大学,2005.
- [4] 张桂清,金怡珠,刘红伟,等. 电子伪装语音的变声规律研究 [J]. 证据科学,2010,18(4):503-509.

(上接第 92 页)

加密时,采用了结构化加密的思想,只对 OLAP 系统中部分维表数据以及事实表中重要的敏感数据进行加密,用于连接维表与事实表间的主键不予加密,从而保证了进行 OLAP 分析时,数据加/解密时间开销不会影响 OLAP 系统的查询分析性能。通过实践证明,这种混合的加/解密方案,既保证了 OLAP 系统中数据的安全性,又不影响 OLAP 系统的查询分析性能。该研究方案在同类系统中有一定的推广应用价值。

参考文献:

- [1] 陈京民. 数据仓库与数据挖掘技术 [M]. 北京:电子工业出版社,2003:12-14.
- [2] 李雄飞,杜钦生. 数据仓库与数据挖掘 [M]. 北京:机械工业出版社,2013:18-19.
- [3] 王丽珍,周丽华. 数据仓库与数据挖掘原理及应用 [M]. 北

- [5] 余建潮,张瑞林. 基于 MFCC 和 LPCC 的说话人识别 [J]. 计算机工程与设计,2009,30(5):1189-1191.
- [6] Tan T J. The effect of voice disguise on automatic speaker recognition [C]//Proceedings of 3rd international congress on image and signal processing. Yantai:IEEE,2010:3538-3541.
- [7] Zhang C, Tan T. Voice disguise and automatic speaker recognition [J]. Forensic Sci. Int., 2008,175(2-3):118-122.
- [8] Wu H J, Wang Y, Huang J W. Blind detection of electronic disguised voice [C]//Proceedings of IEEE international conference on acoustics, speech and signal processing. Vancouver, BC:IEEE,2013:3013-3017.
- [9] Wu H J, Wang Y, Huang J W. Identification of electronic disguised voices [J]. IEEE Transactions on Information Forensics And Security, 2014,9(3):489-500.
- [10] 文 翰,黄国顺. 语音识别中 DTW 算法改进研究 [J]. 微计算机信息,2010,26(7-1):195-197.
- [11] 刘长明,任一峰. 语音识别中 DTW 特征匹配的改进算法研究 [J]. 中北大学学报:自然科学版,2006,27(1):37-40.
- [12] 丁艳伟,戴玉刚. 基于 VQ 的说话人识别系统 [J]. 电脑知识与技术,2008,4(5):1181-1183.
- [13] 赵 力. 语音信号处理 [M]. 北京:机械工业出版社,2003.
- [14] 孔勇平. 矢量量化 LBG 算法的研究 [J]. 硅谷,2008(6):39-40.
- [15] 王 伟,邓辉文. 基于 MFCC 参数和 VQ 的说话人识别系统 [J]. 仪器仪表学报,2006,27:2253-2255.

京:科学出版社,2005.

- [4] 于醒兵. 数据库结构加密理论与技术 [D]. 秦皇岛:燕山大学,2006.
- [5] 李 华. OLAP 技术在生产线质量控制决策系统中的应用 [D]. 贵州:贵州大学,2012.
- [6] 武 彤,程 辉. 基于决策树算法的电视机故障维修模型设计 [J]. 计算机技术与发展,2014,24(5):150-152.
- [7] 漆 媛,武 彤. 基于生产线质量控制系统的 OLAP 安全性研究 [J]. 计算机技术与发展,2014,24(9):179-182.
- [8] 汤姆森. OLAP 解决方案-创建多维信息系统 [M]. 朱建秋,译. 第 2 版. 北京:电子工业出版社,2004:8-10.
- [9] NIST. Advanced Encryption Standard (AES) [M]. [s. l.]: Federal Information Processing Standards Publication, 2001.
- [10] JavaScript 实现 SHA-1 加密算法的方法 [EB/OL]. 2014. <http://www.jb51.net/article/62035.htm>
- [11] 武 彤. 电视机生产线质量控制决策系统 OLAP 模型设计 [J]. 微计算机信息,2012(11):283-284.