



研究与开发

基于逆梅尔对数频谱系数的回放语音检测算法

林朗, 王让定, 严迪群, 李璨
(宁波大学, 浙江 宁波 315211)

摘要: 高保真录音设备和回放设备的普及化及便携化, 给说话人识别系统的抗回放语音攻击带来了严峻挑战。通过语谱图分析原始语音和回放语音在高频区的差异, 有针对性地将语音信号在求取 Mel (梅尔) 倒谱系数过程中的 Mel 滤波器组逆置, 并将 DCT 前的 Mel 对数频谱系数作为算法的特征。最后, 利用支持向量机作为分类器对待测语音进行判别。实验结果表明, 此算法能够有效地检测回放语音。另外, 将此算法加载到 GMM-UBM 说话人识别系统后, 显著地提升了系统的抗回放语音攻击能力。

关键词: 说话人识别; 回放语音检测; 梅尔对数频谱; 逆梅尔滤波器组

中图分类号: TN912.3

文献标识码: A

doi: 10.11959/j.issn.1000-0801.2018020

A playback speech detection algorithm based on log inverse Mel-frequency spectral coefficient

LIN Lang, WANG Rangding, YAN Diquan, LI Can
Ningbo University, Ningbo 315211, China

Abstract: The popularity and portability of high-fidelity audio recording equipment and playback equipment poses a serious challenge for speaker recognition systems against playback attacks. Based on the differences between the original speech and the playback speech in high frequency region, the algorithm reversed the Mel-filter bank in Mel-frequency cepstral coefficient (MFCC) calculation, and the coefficients before the DCT were used as the features of the algorithm. SVM was utilized as the classifier. Experimental results show that this algorithm can effectively detect the playback speech. In addition, the algorithm is integrated into the GMM-UBM speaker recognition system, which significantly improves the systems' capability of resisting the playback attack.

Key words: speaker recognition, playback speech detection, log Mel-frequency spectrum, inverse Mel-filter group

收稿日期: 2017-06-11; 修回日期: 2017-12-07

基金项目: 国家自然科学基金资助项目 (No.61672302, No.61300055); 浙江省自然科学基金资助项目 (No.LZ15F020002, No.LY17F020010); 宁波大学科研基金资助项目 (No.XKXL1405, No.XKXL1420, No.XKXL1509, No.XKXL1503); 宁波大学王宽诚幸福基金资助项目

Foundation Items: The National Natural Science Foundation of China (No. 61672302, No.61300055), The Natural Science Foundation of Zhejiang Province of China (No.LZ15F020002, No.LY17F020010), The Scientific Research Foundation of Ningbo University (No.XKXL1405, No.XKXL1420, No.XKXL1509, No. XKXL1503), K.C. Wong Magna Fund in Ningbo University

1 引言

说话人识别技术^[1]以其自身独特的优势,诸如语音获取方便、使用者接受度高、说话人系统算法复杂度低等优点,在司法、金融、生活等领域得到了广泛应用。同时,说话人识别技术的安全性也成为亟待解决的问题。如何防止仿冒语音的攻击已成为研究的重点。

根据攻击手段的不同,仿冒语音主要分为两类:一是模仿特定说话人的声音进行攻击,称为说话人仿冒攻击;二是通过专业的设备和技术仿冒说话人的声音^[2,3](如合成语音、拼接语音、回放语音等)。对于第一种攻击方式,现有的说话人识别技术已经能够有效地检测。而对于第二种攻击方式,目前还没有比较成熟有效的手段能够完全检测出来。在实际场景中,由于合成、拼接语音需要比较专业的技术支持,而回放语音相对来说操作简单,便于仿冒,已经成为不法分子最善用的攻击手段^[4]。回放语音的产生过程如图1所示。

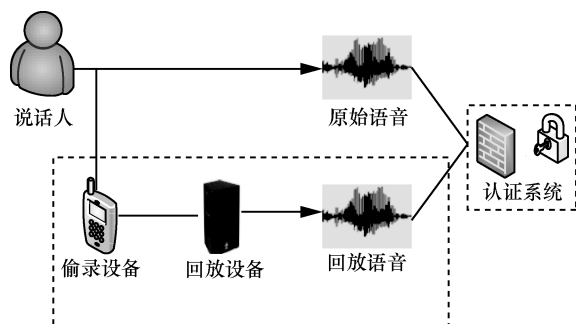


图1 回放语音产生过程

由图1可知,回放语音是真实地来源于说话人本人的声音,相较于其他仿冒语音来说,对说话人识别系统威胁更大。与原始语音相比,回放语音多经历了偷录设备的录制和回放设备的播放等过程,因此会不可避免地引入设备噪声和设备编码、解码的失真以及环境噪声等,使得回放语音和原始语音产生了细微的差异。目前,针对回放语音检测问题,主要分为以下两类:一种是基于语音随机性的检测

算法,如Shang等人^[5,6]利用语音产生的随机性,提出了一种检测待测语音和合法语音在峰值图上的相似度的算法,此方法只能够应用于文本相关的声纹认证系统;另一种是基于语音信道的检测算法,如张利鹏等人^[7]通过探究回放语音产生的机理,对语音的静音段信道进行建模,提出了一种基于语音静音段信道差异的回放语音检测算法,该算法由于静音段幅度很小,容易受到噪声的污染,因此很难建立精确的信道模型。王志峰等人^[8]分析了回放语音产生过程中由不同设备引入的信道噪声,提出了一种基于信道模式噪声的录音回放检测算法。该方法解决了文本相关问题,但实验过程中只涉及一种录音设备和回放设备,存在实验设备单一的问题。

通过分析原始语音和回放语音的语谱图发现,在高频区域原始语音和回放语音有显著的差异。而传统的MFCC(Mel-frequency cepstral coefficient, Mel倒谱系数)特征提取使用的Mel滤波器由于低频分辨率高、高频分辨率低的特点,使得原始语音和回放语音在高频区的差异性削弱,从而不利于人们对回放语音的检测。其次在分析MFCC特征参数时,经过实验发现,使用MFCC提取过程中去离散余弦变换(discrete cosine transform, DCT)前的梅尔对数频谱系数(log Mel-frequency spectral coefficient, MFSC)在回放语音的检测上有更好的检测效果。

基于上述分析,本文提出了一种基于逆Mel滤波器的梅尔对数频谱(log inverse Mel-frequency spectral coefficient, I-MFSC)的算法,逆Mel滤波器的设计是由Mel滤波器逆置得到的,表现为高频区域频谱分辨率高、低频区域分辨率低的特点,这样使得高频区回放语音和原始语音的差异性会更加显著地显示出来。实验结果表明,在回放语音的检测上I-MFSC有较好的检测效果。将本文提出的回放语音检测方法加载到目前主流的GMM-UBM说话人识别系统中后,系统安全性能有了显著的提高。



2 高频区回放语音和原始语音差异分析

尽管现有的偷录和回放设备都有着良好性能，能够做到较小的失真录制和高保真的回放。但原始语音和回放语音仍会存在一定的差别。本文从语谱图^[9]着手，考究了原始语音和回放语音在频谱上的差异性，如图2和图3所示，实验语音是由Aigo R6620采集的一段4 s的语音，语音内容为“芝麻开门，我是土豪，千里共婵娟”。其中图(a)代表原始语音的语谱图。图(b)、图(c)、图(d)分别对应偷录设备为iPhone6、Mi4以及Sony PX440的语谱图。图2使用的回放设备为Huawei AM08，而图3对应的回放设备为Philips DTM3115。

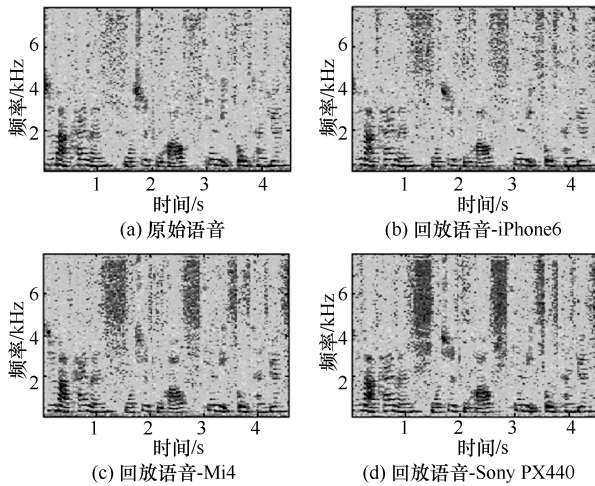


图2 原始录制和回放语音的语谱图(回放设备: Huawei AM08)

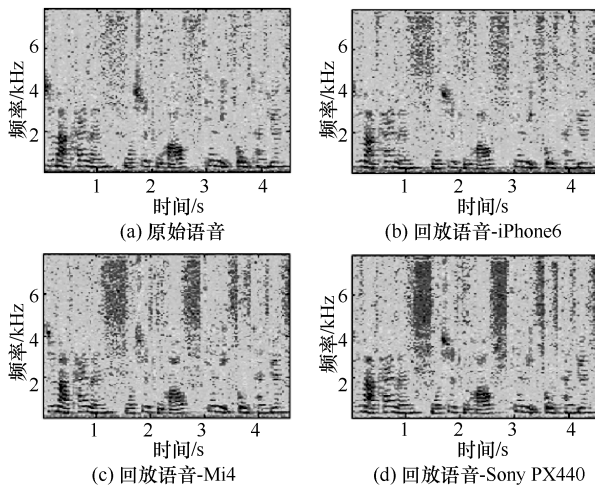


图3 原始录制和回放语音的语谱图(回放设备: Philips DTM3115)

从图2和图3中可以看出，与图(a)的原始语音语谱图相比，3种偷录设备高频区域(6~8 kHz)的频率值均小于原始语音，其中图(c)、图(d)显示出了更大的差异性。这表明Mi4以及Sony PX440两种设备失真度更高。此外在4 kHz左右，原始语音和iPhone6以及Sony PX440语谱图频谱过渡相对平滑，而Mi4分界线较为明显，产生了跳变现象，探究后发现这和Mi4设备自身固有的设备性能有关。

3 基于I-MFSC回放语音检测算法

3.1 MFCC特征提取过程

在说话人识别中，MFCC因能够较好地模拟人耳听觉系统的感知能力被广泛应用。图4为MFCC特征的提取过程。

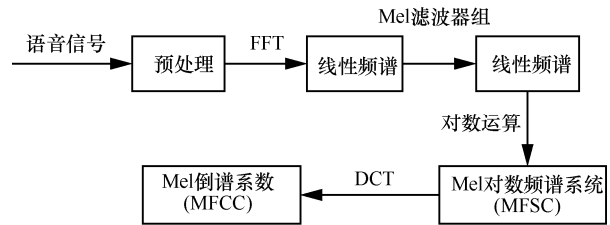


图4 MFCC特征的提取过程

首先对语音信号 $x(n)$ 进行预处理，包括预加重、分帧和加窗，得到分帧后的语音信号 $x_i(n)$ ，其中下标 i 表示分帧后的第 i 帧。然后对每帧语音信号 $x_i(n)$ 进行FFT得到各帧的线性频谱 $X_i(k)$ ，即：

$$X_i(k) = \sum_{n=0}^{N-1} x_i(n) e^{-j2\pi k n / N}, 0 \leq k \leq N \quad (1)$$

其中， N 表示傅里叶变换的点数；将得到的线性频谱 $X_i(k)$ 经由Mel滤波器进行滤波生成Mel频谱，然后再对Mel频谱计算对数能量得到对数频谱 $S_i(m)$ ，即：

$$S_i(m) = \ln \left(\sum_{k=0}^{N-1} |X_i(k)|^2 H_m(k) \right), 0 \leq m \leq M \quad (2)$$

其中， M 表示滤波器个数， $m=1,2,\dots,M$ ，这里 M

通常取 27~40。最后经离散余弦变换 (DCT) 得到 L 阶的 MFCC:

$$C(n) = \sum_{m=0}^{N-1} S_i(m) \cos\left(\frac{\pi n(m-0.5)}{M}\right), n=1, 2, \dots, L \quad (3)$$

其中, $C(n)$ 为 L 阶的 MFCC, L 通常取 12~16。

3.2 基于逆 Mel 滤波器组的 MFSC

传统的 Mel 滤波器组, 在低频段带宽较窄, 频谱分辨率相对较高。而高频段带宽较宽, 频谱分辨率相对较低。因此在高频区域较宽的滤波器平滑了高频信息, 削弱了不同频率带之间的差异, 使得高频区域的部分信息丢失。而由第 2.1 节的分析可知, 回放语音和原始语音在高频区域存在明显的差异, 因此为了利用两种语音在高频区域的差异性, 本文采用一种逆 Mel 滤波器组^[10-12]的设计来提取本文特征参数。逆 Mel 滤波器组的设计如图 5 所示。

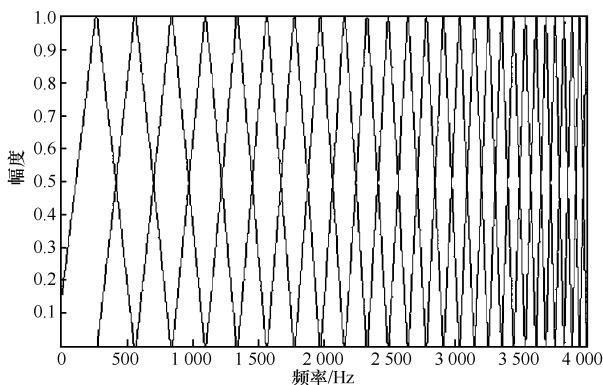


图 5 逆 Mel 滤波器组的设计

逆 Mel 滤波器组是基于逆 Mel 刻度变换得到的, 逆 Mel 刻度变换的物理频率和 Mel 频率的对应关系为:

$$\begin{cases} f_{\text{I-Mel}} = -2595 \times \lg\left(\frac{f_{\text{max}} - f}{700}\right) + b \\ b = 2595 \times \lg\left(\frac{f_{\text{max}} - 0}{700}\right) \end{cases} \quad (4)$$

其中, $f_{\text{I-Mel}}$ 表示逆 Mel 频率, f_{max} 表示语音信号最大频率, f 表示语音信号的物理频率。逆 Mel 和 Mel 变换关系如图 6 所示。

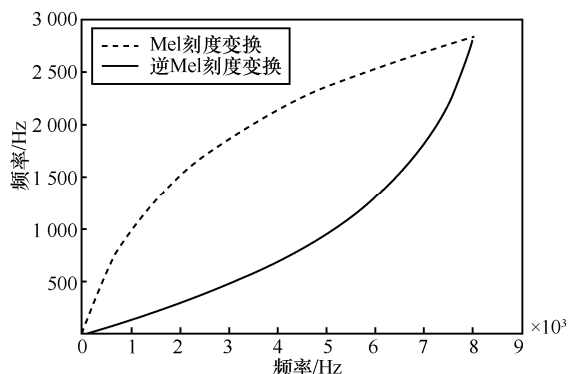


图 6 逆 Mel 和 Mel 变换关系

此外, 在 MFCC 特征提取时, 会将得到的 Mel 频谱取对数后再进行 DCT, 得到最后的 MFCC 特征参数。这里的 DCT 有两种作用^[13]: 一是利用 DCT 后较强的能量集中特性, 能够对数据进行压缩; 二是由于 MFCC 提取过程中使用的 Mel 滤波器之间是有重叠的, 因此能量值之间具有很强的相关性, 使用 DCT 可以达到去相关的目的。而 Mohamed 等人^[14]的研究表明, 相较于 MFCC 特征, MFSC 具有更高的相关性和维度, 更适合于机器学习的建模方式。因此将 DCT 前的 MFSC 作为参数特征, 在滤波器的设计上, 采用逆 Mel 滤波器组, 新得到的参数定义为 I-MFSC。

3.3 特征构造与选择

I-MFSC 的特征提取过程和 MFCC 相似, 在第 3.1 节 MFCC 提取的基础上, 将 Mel 滤波器组换成逆 Mel 滤波器组, 然后将去 DCT 前的对数 Mel 频谱 $S_i(m)$ 作为本文算法的最后特征参数, 最终得到本文的特征参数 I-MFSC。

为了更好地说明 I-MFSC 特征的性能, 分别对 Mel 倒谱系数 (MFCC)、Mel 对数频谱系数 (MFSC)、逆 Mel 倒谱系数 (I-MFCC) 和逆 Mel 对数频谱系数 (I-MFSC) 在均值上做了对比分析, 图 7 分别为这 4 种特征系数在均值上的差异。

由图 7 可以看出, 虽然 MFCC 和 I-MFCC 在某些特征维度上也体现了差异性, 可以作为检测回放语音的特征, 但相较于 MFSC 与 I-MFSC 两种, 后者在原始语音和回放语音的检测上体现出

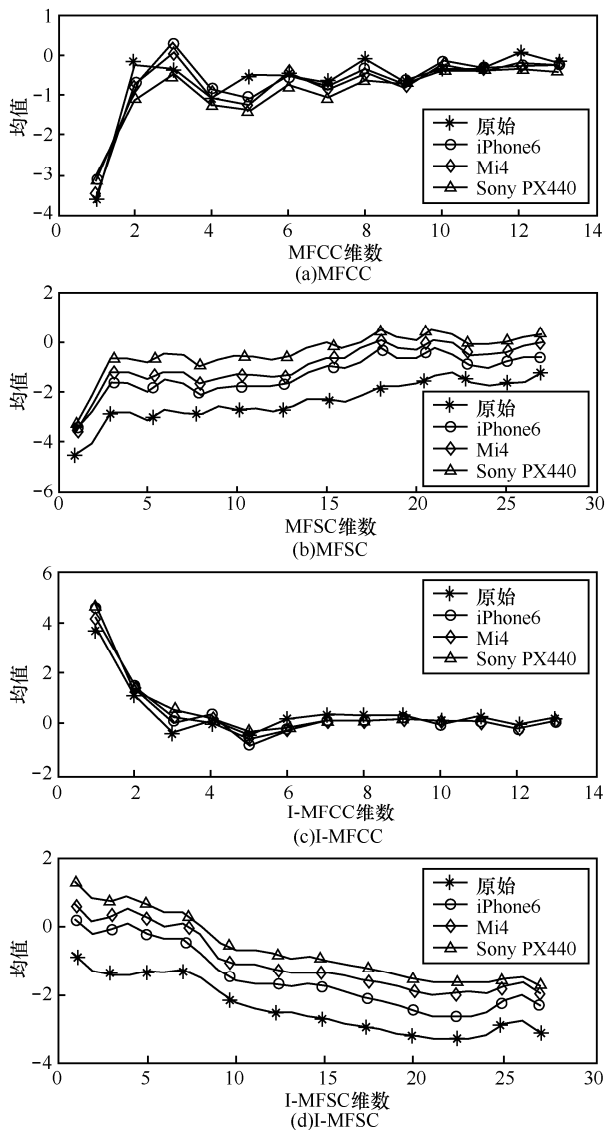


图7 4种特征均值分布

了更好的性能。而尽管 MFSC 特征与 I-MFSC 特征都具有很好的检测性能,但是第 4 节后续的实验表明,本文提出的 I-MFSC 特征参数略优于 MFSC 特征参数。

3.4 回放语音检测算法

实验分类器的选择是 LibSVM^[15],它是一种有监

督的机器学习方法。在分类训练和测试时需要为每一条数据设置标签,这里将原始语音设置为正样本,标签设置为“Y”,回放语音设置为负样本,标签设置为“N”,LibSVM 分类器其他参数均使用默认参数。

将提取的矩阵特征(MFCC、MFSC、I-MFCC、I-MFSC)按帧求取均值,得到每一条语音的均值特征,并将每一条语音特征设置标签。在训练阶段,使用训练集训练模型参数。在测试时,根据训练阶段得到的模型参数对测试数据进行判别归类,然后根据分类结果和测试数据的已知标签比较,得出回放语音检测的准确率,具体流程如图8所示。

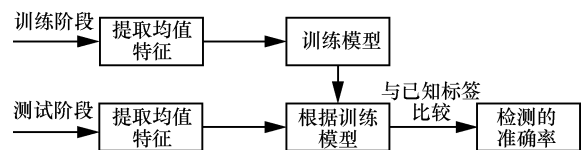


图8 回放语音检测流程

4 实验仿真及分析

4.1 实验设置

为了说明本文算法的有效性和适用性,分别构建了原始语音库和回放语音库。数据库的具体设置如下:语料库来源于 863 语料库^[16];人员分布为 10 男 6 女;考虑到现实场景中说话人验证系统中的语音采集模块通常采用 16 kHz 的采样率,所以在实验中使用了与其采样率相同的设备 Aigo R6620;而偷录和回放设备则选择了多款常见的高保真设备,设备的具体详情见表 1。

实验库的构建环境为安静办公室,具体录音流程为:说话人按照语料库语料录音,并使用采集设备进行语音采集,通常将采集设备采集到的语音称

表1 设备的具体详情

类别	原始录制设备		偷录设备		回放设备	
	Aigo R6620	iPhone6	Mi4	Sony PX440	Huawei AM08	Philips DTM3115
语音格式	wav	m4a	mp3	mp3	—	—
参数	16 kHz	44.1 kHz	44.1 kHz	44.1 kHz	—	—
	16 bit/s	64 kbit/s	128 kbit/s	192 kbit/s		

为原始语音。与此同时,使用3种偷录设备同时录制说话人语音,并在同样的环境下,将偷录设备录制的语音经音响回放,并使用采集设备录制该回放的语音,将此种条件下采集的语音称为回放语音。数据集共有实验样本15 000个,其中原始语音样本2 400个,两种回放设备对应3种偷录设备共计有12 600个回放语音样本,所有样本详情见表2。

4.2 实验结果与分析

4.2.1 不同特征的实验结果与分析

本节主要检测MFCC、I-MFCC、MFSC以及I-MFSC这4种特征对回放语音检测性能影响。实验所用样本详情见表2。表3为不同特征的检测结果,TPR为真阳性率(true positive rate),FPR为假阳性率(false positive rate),ACC为正确检测率。

由表3可知,两种Mel对数频谱系数MFSC和

I-MFSC的性能显著优于MFCC和I-MFCC。尽管MFSC对回放语音的检测率也有很优越的性能,但最好的检测结果仍然是I-MFSC,因此本文后续实验所使用的特征选择I-MFSC作为最后的特征参数。

4.2.2 不同设备之间的交叉实验对检测率的影响

现实场景中由于偷录设备的多种多样,回放语音检测算法能否对多种偷录设备和回放设备都有较好的检测结果是衡量算法有效性的关键所在。本节实验目的是检测本文提出的算法在不同设备之间交叉是否依然有较好结果。实验采用的样本见表2,实验采用的特征是本文提出的I-MFSC特征,具体检测结果见表4。

由实验结果可知,在不涉及交叉设备时,每种设备的检测率均达到了100%。当iPhone6和Sony PX440作为训练样本时,虽然检测效果有所

表2 原始语音和回放语音样本详情

语音	原始录制设备	回放设备	偷录设备	样本数/个
原始语音	Aigo R6620	—	—	2 400
回放语音	Aigo R6620	Huawei AM08	iPhone6、Mi4、Sony PX440	6 300
		Philips DTM3115	iPhone6、Mi4、Sony PX440	6 300

表3 不同特征的检测结果

特征	Philips DTM3115			Huawei AM08			两种设备的交叉		
	FPR	TPR	ACC	FPR	TPR	ACC	FPR	TPR	ACC
MFCC	99.60%	1.30%	99.58%	96.90%	7.00%	96.92%	96.70%	16.90%	96.67%
I-MFCC	99.90%	0.20%	99.92%	98.2%	3.70%	98.16%	97.30%	14.00%	97.29%
MFSC	100%	0	100%	99.30%	0.20%	99.33%	99.70%	0.30%	99.67%
I-MFSC	100%	0	100%	100%	0	100%	99.90%	0.90%	99.86%

表4 不同设备之间的交叉实验检测结果

	回放设备	偷录设备	测试集					
			Huawei AM08			Philips DTM3115		
			iPhone	Mi	Sony	iPhone	Mi	Sony
训练集	Huawei AM08	iPhone	100%	100%	100%	100%	100%	99.86%
		Mi	98.43%	100%	96.79%	97.86%	100%	82%
		Sony	100%	100%	100%	99.07%	99.50%	99.79%
	Philips DTM3115	iPhone	99.85%	100%	92.79%	100%	100%	99.85%
		Mi	96.70%	99%	77.29%	99.93%	100%	72.79%
		Sony	100%	99.14%	100%	100%	97.04%	100%



表 5 噪声环境下实验检测结果

算法	纯净条件	30 dB 噪声	25 dB 噪声	20 dB 噪声	15 dB 噪声
MFCC	96.67%	95.66%	90.89%	87.74%	85.71%
I-MFCC	97.29%	96.57%	95.52%	90.89%	88.81%
MFSC	99.67%	98.62%	98.23%	97.57%	96.57%
I-MFSC	99.86%	99.35%	98.95%	98.21%	97.43%

下降,但检测效果依旧可观;当 Mi4 作为训练集样本时,检测效果不甚理想。由第 2.1 节的语谱图分析结论可知,这是因为受到 Mi4 设备本身特性的影响,对高频区有些许的抑制作用;但综合来看,本文提出的 I-MFSC 特征在设备交叉的实验检测中具有良好的检测性能。

4.2.3 噪声环境下几种特征的检测效果

现实场景中偷录环境的多样性和复杂性也是影响回放语音检测性能的重要因素。考虑到声纹识别系统的实际应用场景,过高的噪声环境在声纹认证领域意义不大,且噪声环境特别大时,声纹系统会直接拒绝待测语音进入系统,因此本实验将回放语音和原始语音加上 15~30 dB 的高斯白噪声以检测本算法的顽健性。检测结果见表 5。

由表 5 可以看出,噪声的引入对回放语音检测有一定的影响,当信噪比从 30 dB 降到 15 dB 时,回放语音的检测率总体呈下降趋势。但相对而言,本文提出的算法 I-MFSC 相较于其他 3 种特征,在噪音环境下表现出更好的性能。

4.2.4 对比试验

为检验本算法的有效性和适用性,将本文的算法以单独的模块形式加载到 GMM-UBM 说话人识别系统^[17]中,加载回放语音检测模块的说话人识别系统如图 9 所示。

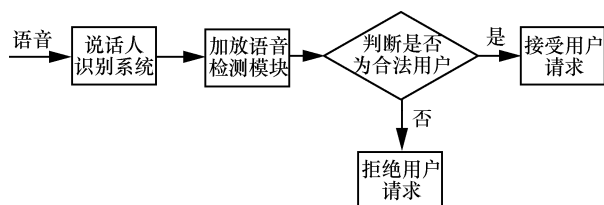


图 9 加载回放语音检测模块的说话人识别系统

语音进入说话人识别系统后,说话人识别系统对语音进行第一次判别,当说话人识别系统判决为非法用户时,系统会直接拒绝该语音,不再进入回放语音检测模块。只有当说话人识别系统判决该语音来自于合法用户后,才会进行回放语音检测模块;如果回放语音检测模块判决为原始语音,则系统接受该用户请求,否则拒绝该用户请求。

此外,将本文算法同参考文献[4]和参考文献[5]提出的算法进行比较。参考文献[4]的算法采用短时能量法提取静音,并用谱减法进行去噪。采用 12 维 MFCC 和一阶差分 Δ MFCC 作为特征参数。而参考文献[5]的算法则采用高通滤波器进行去噪,提取 6 个统计特征及 6 阶 Legendre 多项式系数(共 12 维)作为信道模式噪声的特征参数。检测检测结果见表 6,其中 ACC 表示检测的正确率, EER 表示将检测模块加载到 GMM-UBM 系统的等错误概率。

表 6 对比实验检测结果

算法	ACC	EER
参考文献[4]算法	75.42%	25.45%
参考文献[5]算法	83.23%	19.09%
本文算法	99.86%	5.90%

由表 6 可知,本文提出的算法相较于参考文献[4]和参考文献[5]在性能上有很大的提升。最后将几种方法以单独的模块加载到说话人识别系统,加载回放语音检测模块前后的等错误概率如图 10 所示。由图 10 可以看出,加载到 GMM-UBM 系统后,本文提出的算法更能有效地提高说话人识别系统对回放语音的抵抗能力。

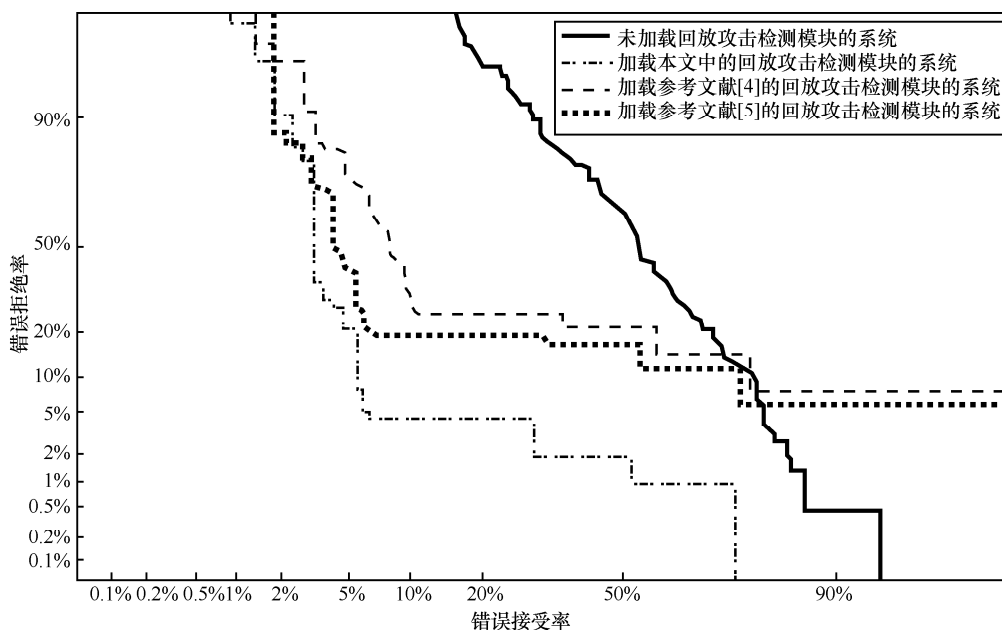


图 10 加载回放语音检测模块前后的等错误概率

5 结束语

本文利用原始语音在回放语音在高频区的差异,通过逆 Mel 滤波器和提取去 DCT 前的 MFSC 特征,提出了一种基于高频区 I-MFSC 特征的回放语音检测算法。该算法能够有效地弥补现有算法中设备单一的问题。通过实验表明,本文算法能够有效地检测回放语音和原始语音,将本文算法加载到 GMM-UBM 说话人识别系统时,对说话人识别的性能有了很大的提高。现实场景中,偷录设备和回放设备种类繁多,高保真的设备层出不穷,因此今后的工作将进一步探索回放语音产生的机理以及回放语音和原始语音产生差异性的具体因素。

参考文献:

- [1] ZHU D, MA B, LI H. Speaker verification with feature-space MAPLR parameters[J]. IEEE Transactions on Audio Speech & Language Processing, 2011, 19(3): 505-515.
- [2] 易克初, 胡征. 一种应用矢量量化的语音合成新方法[J]. 电信科学, 1987(11): 1-6.
YI K C, HU Z. A new speech synthesis method using vector quantization[J]. Telecommunications Science, 1987(11): 1-6.

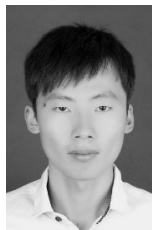
- [3] 郭弘. 录音证据的真实性检验与研究[J]. 电信科学, 2010, 26(Z2): 56-60.
GUO H. Authenticity verification and research of recording evidence[J]. Telecommunications Science, 2010, 26(Z2): 56-60.
- [4] 李臻, 王让定, 严迪群, 等. 基于相位谱的翻录语音攻击检测算法[J]. 电信科学, 2017, 33(8): 145-154.
LI C, WANG R D, YAN D Q, et al. Detection algorithm of riprap voice attack based on phase spectrum[J]. Telecommunications Science, 2017, 33(8): 145-154.
- [5] SHANG W, STEVENSON M. A playback attack detector for speaker verification systems[C]//IEEE International Symposium on Communications Control and Signal Processing (ISCCSP), March 12-14, 2008, St Julians, Malta. Piscataway: IEEE Press, 2008: 1144-1149.
- [6] SHANG W, STEVENSON M. Score normalization in playback attack detection[C]//IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP), March 14-19, 2010, Dallas, USA. Piscataway: IEEE Press, 2010: 1678-1681.
- [7] 张利鹏, 曹攀, 徐明星. 防止假冒者闯入说话人识别系统[J]. 清华大学学报(自然科学版), 2008, 48(S1): 699-703.
ZHANG L P, CAO J, XU M X. Prevention of impostors entering speaker recognition systems[J]. Journal of Tsinghua University (Science and Technology), 2008, 48(S1): 699-703.
- [8] 王志峰, 贺前华, 张雪源, 等. 基于模式噪声的录音回放攻击检测[J]. 华南理工大学学报, 2011, 39(10): 7-12.
WANG Z F, HE Q H, ZHANG X Y, et al. Channel pattern noise based playback detection algorithm speaker recognition[J]. Journal of South China University of Technology (Natural Science Edition), 2011, 39(10): 7-12.



- [9] 李富强, 万红, 黄俊杰. 基于MATLAB的语谱图显示与分析[J]. 微计算机信息, 2005(20): 172-174.
LI F Q, WAN H, HUANG J J. The display and analysis of sonogram based on MATLAB[J]. Control & Automation, 2005(20): 172-174.
- [10] BURILLO P, BUSTINCE H. Entropy on intuitionistic fuzzy sets and on interval-valued fuzzy sets[J]. Fuzzy Sets & Systems, 1996, 78(3): 305-316.
- [11] 项要杰, 杨俊安, 李晋徽, 等. 一种适用于说话人识别的改进 Mel 滤波器[J]. 计算机工程, 2013(11): 214-217.
XIANG Y J, YANG J A, LI J H, et al. An improved Mel-frequency filter for speaker recognition[J]. Computer Engineering, 2013(11): 214-217.
- [12] 陶佰睿, 郭琴, 苗凤娟, 等. 基于改进 Mel 滤波器组的声纹特征提取 SoC 设计[J]. 微电子学, 2015(6): 785-788.
TAO B R, GUO Q, MIAO F J, et al. SoC design of voiceprint features extraction based on improved Mel filter banks[J]. Microelectronics, 2015(6): 785-788.
- [13] 胡永刚, 吴翊, 王洪志, 等. 高维数据降维的 DCT 变换[J]. 计算机工程与应用, 2006(32): 21-23.
HU Y G, WU Y, WANG H Z, et al. Discrete cosine transform in data dimensionality reduction[J]. Computer Engineering and Applications, 2006(32): 21-23.
- [14] MOHAMED A. Deep neural network acoustic models for ASR[J]. Doctoral, 2014.
- [15] CHANG C C, LIN C J. LIBSVM: a library for support vector machines[J]. ACM Transactions on Intelligent Systems & Technology, 2012, 2(3): 1-27.
- [16] 王天庆, 李爱军. 连续汉语语音识别语料库的设计[C]//第六届全国现代语音学学术会议论文集, 2003年10月1日, 天津, 中国. [出版地不详: 出版者不详], 2003: 1-4.
WANG T Q, LI A J. The design of the continuous Chinese speech recognition corpus[C]//The Sixth National Conference on Modern Phonetics Learning, Oct 1, 2003, Tianjin, China. [S.l.:s.n.], 2003: 1-4.
- [17] CHAKROBORTY S, ROY A, SAHA G. Improved closed set

text-independent speaker identification by combining MFCC with evidence from flipped filter banks[J]. International Journal of Signal Processing, 2007, 4(2): 114-122.

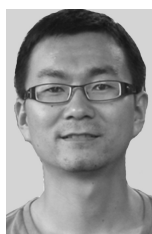
[作者简介]



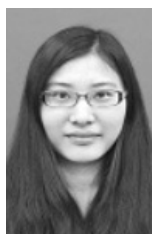
林朗 (1994-), 男, 宁波大学信息科学与工程学院硕士生, 主要研究方向为多媒体通信与信息安全等。



王让定 (1962-), 男, 博士, 宁波大学信息科学与工程学院教授、博士生导师, 主要研究方向为多媒体通信与取证、信息隐藏与隐写分析、智能抄表及传感网络技术。



严迪群 (1979-), 男, 博士, 宁波大学信息科学与工程学院副教授、硕士生导师, 主要研究方向为多媒体通信、信息安全、基于深度学习的数字语音取证等。



李臻 (1992-), 女, 宁波大学信息科学与工程学院硕士生, 主要研究方向为多媒体通信与信息安全等。