

音频取证若干关键技术研究进展

包永强 梁瑞宇 丛 榭 高冲红 王青云

(南京工程学院通信工程学院, 南京, 211167)

摘要: 介绍了音频取证领域的最新研究进展、音频真实性的研究状况。对音频取证研究领域的历史进行了回顾,探讨了音频取证的分类,构建了音频取证框架。对音频取证的若干个关键技术进行了总结,包括音频主动取证技术、基于电网频率特征的音频篡改技术、无电网频率成分下的音频篡改检测技术、录音设备的特征参数、模式识别、数据库建设情况以及录音场合识别等。最后对音频取证技术进行了总结和展望。

关键词: 音频取证; 取证水印; 音频篡改检测; 录音设备识别; 录音场合识别

中图分类号: TP391 **文献标志码:** A

Research Progress on Key Technologies of Audio Forensics

Bao Yongqiang, Liang Ruiyu, Cong Yun, Gao Chonghong, Wang Qinyun

(School of Communication Engineering, Nanjing Institute of Technology, Nanjing, 211167, China)

Abstract: The latest research progress in audio forensics is introduced with the emphasis on audio authenticity. First, the history of audio forensics research is reviewed. The classification of audio forensics is discussed. Then, the framework of audio forensics is designed. Several key technologies of audio forensics are summarized including audio active forensics technology, audio tamper technology based on electrical network frequency (ENF), audio tamper detection technology with different sampling rates and audio tamper detection technology with the same sampling rates under the passive power grid frequency components, the characteristic parameters of recording equipment, pattern recognition, situation of database construction, recording environment identification and so on. Finally, the prospective of audio forensics technology is presented.

Key words: audio forensics; forensic watermarking; audio tamper detection; recording equipment identification; recording environment identification

引 言

随着音频信号处理技术的飞速发展^[1],对音频内容进行修改变的十分便捷。对音频信号合法性、真实性和关联性的研究即音频取证成为法庭取证中迫切需要解决的问题。当前的录音主要为手机和录音笔所录制,其中手机录音占的比重很大。中国互联网络信息中心的数据显示,截止到2015年10月,中

国手机用户超过 13 亿,智能手机市场快速发展,厂商数量虽然在 100 家以下,但在售的智能手机数量已超过 1 200 款,其中 4 G 机型发展迅猛,达到 845 款。这些数据直接伴随了一个现象的出现:手机录音功能的便捷使用直接导致了音频证据中大量私录音频的出现。音频编辑种类繁多功能强大。常见的有 Syntrillium 公司的 Cool Edit, Cakewalk 公司的 Cakewalk 软件和德国 STEINBERG 公司的 Steinberg Nuendo, MP3 Splitter 等软件。这些编辑软件能够实现录制、混合、编辑和控制音频,可以随心所欲地编辑出想要的内容。针对以上情况,人们开展了对音频取证的研究^[2],可分为 3 类:

(1) 主动取证技术。2005 年以前,人们主要关心如何保证合法音频文件不被篡改或伪造。因此对音频主动取证的研究主要集中于将脆弱音频水印技术嵌入到合法音频文件中,用于鉴别音频文件的真实性和完整性^[3-5]。这一阶段虽然有取证语音增强^[6]、取证说话人识别^[7]等方面的研究,但也是将语音增强和说话人识别的最新研究成果与音频取证需求简单结合,目的是提高音频取证的工作效率。

(2) 音频真实性取证技术。法庭在收到音频证据的时候,需要对当前音频进行真实性判断,即判断是否经过了篡改。音频真实性取证技术一般可分为两种:(a) 基于电网频率特性的真实性取证技术。2002 年以来,大量私录音频作为音频证据合法地出现在法庭证据中。因此,这一时期主要面临的问题是如何鉴别私录音频是否经过了篡改,即私录音频的真实性取证问题。脆弱水印嵌入技术由于无法嵌入到私录音频中,导致对主动取证技术的研究慢慢退出主流研究视野。2005 年,Grigoras 发现音频中蕴含了电网频率成分(Electric network frequency, ENF)^[8],可以利用电网频率成分特征来确定音频的真实性和录音时间。由于供电部门电网频率成分不断波动,可以通过提取音频信号中的电网频率变化曲线与供电部门的电网频率数据库中数据进行比对,来确定录音时间和是否进行过篡改^[9]。利用电网频率特性进行音频篡改检测和录音时间检测成为了这一时期最热门的研究方向,这一技术的应用有一个限制,环境声音中必须有电网频率成分。(b) 基于环境特征的真实性取证技术。2009 年以后,由于 4G, WLAN 等宽带技术的发展,人们对音频音质的需求不断提高,语音信号处理技术特别是滤波技术得到不断应用。音频中可能并不包含有电网频率成分,在一些场合中,为了进一步提高带宽使用效率,基站在处理手机通话中的无声段时候,可能并不发送真实的无声段信号,而是根据背景特征向用户输送一段舒适的噪声,针对电网频率的研究开始回落。此时研究开始转向利用噪声帧功率谱^[10]、不同采样率拼接带来的线性插值^[11]和统计特征分析^[12]等方面。

(3) 音频来源的识别技术。证据的合法性、真实性和关联性是构成证据的 3 个要素。因此,在确定录音的真实性基础上,需要对音频的录音设备、录音场合、录音时间、涉及的说话人和采用的编码方式等进行研究。

随着人们法制观念的不断增强、大量手机用户的普及及音频信号处理技术的发展,私录音频成为了法庭认可的证据。音频取证融合了语音信号处理、信息安全和数据分析等,成为了越来越多科学家关注的热点,并取得了一系列的研究成果。具体主要包括音频主动取证、音频真实性取证、取证说话人识别、录音设备识别及录音场合识别等。

1 音频取证基本框架

通过分析文献发现,近 20 年的研究主要还是围绕 1958 年美国麦科夫案中确立的准则^[2]:(1) 识别出正确的录音设备;(2) 确定音频文件是否进行了篡改;(3) 音频文件存放是否安全;(4) 正确识别说话人;(5) 音频中的会话未被诱导或强迫。根据麦科夫案准则和目前的研究进展,音频取证技术可以分为两类:音频主动取证技术和音频被动取证技术(盲取证)。前者是在确定该音频为原始录音情况下,存储或传输过程中如何防止音频证据被调换、篡改采用的方法,音频水印技术是音频主动取证研究的主要技术。后者针对待鉴别音频证据,包括如何确定该音频证据的录音时间、录音场合、采用的录音设备、对应说话人和是否被篡改过。

2 音频主动取证技术

音频主动取证技术^[3-5,13-18]是通过在合法音频中嵌入隐秘信息来达到防篡改的目的。主要用于合法音频的版权保护,解决合法音频因存储、传输带来的安全隐患,验证音频文件的完整性或真实性。音频主动取证一般采用数字水印的方式,如图 2 所示。

音频主动取证嵌入的水印应当满足以下特性:(1)不可感知性。数字水印的嵌入应当不改变音频的听觉质量,不能被

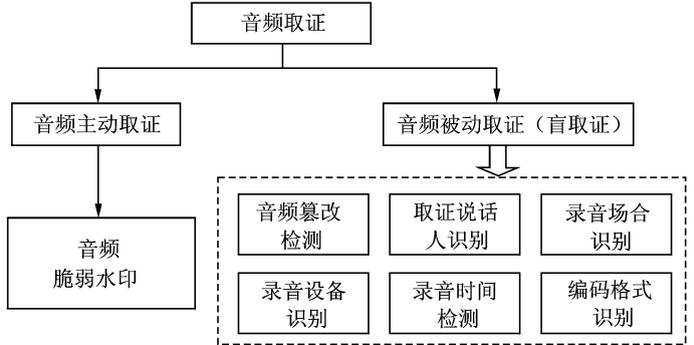


图 1 音频取证基本框架

Fig. 1 Framework of audio forensics

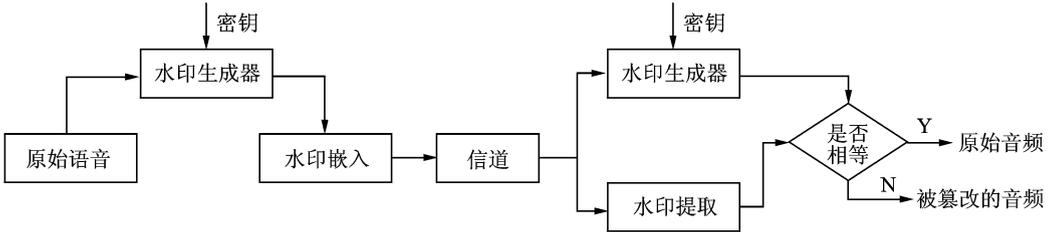


图 2 音频主动取证模型

Fig. 2 Audio active forensics model

感知;(2)脆弱性。音频主动取证嵌入的水印应当脆弱,任何改动都会导致水印不能够正常提取,从而能够抵抗恶意的插入、删除和替换等攻击;(3)安全性。音频信号中嵌入的水印必须不能被伪造,水印算法完全依赖于密钥;(4)盲检测。音频水印的提取必须不依赖于原始音频信号;(5)嵌入容量。满足音频真实性取证的基本要求所需的比特数越少越好。

2005 年 Petrovic 采用伪随机码产生易脆水印来验证音频文件的完整性^[3],2007 年 Serap 针对带限高级音频编码(Advanced audio coding, AAC)音频包络嵌入水印^[4,5],目前国内外对音频主动取证技术集中在变换域的嵌入水印的方法,包括单变换域和多变换域嵌入法。单变换域嵌入法将音频信号进行分帧,针对每帧音频信号进行变换,水印信息就加载在变换域系数上,如离散余弦变换(Discrete cosine transform, DCT)变换^[13]、离散小波变换(Discrete wavelet transform, DWT)变换^[14]、复倒谱变换(Complex cepstrum transform, CCT)变换^[15]和奇异值分解(Singular value decomposition, SVD)^[16]等。多变换域嵌入法一般首先对音频信号进行离散小波变换,然后采用单变换域方法进行处理,如离散小波-离散余弦法(Discrete wavelet transform-discrete cosine transform, DWT-DCT)、离散小波-奇异值分解法(Discrete wavelet transform-singular value decomposition, DWT-SVD)和离散小波-复倒谱法(Discrete wavelet transform-complex cepstrum transform, DWT-CCT)。随着音频编辑软件功能的越来越强大和法庭证据中大量私录音频的出现,音频主动取证技术已不能完全满足法庭取证的要求。因此需要研究音频被动取证技术,即音频真实性取证技术。

3 音频真实性取证

3.1 基于电网频率特征的音频真实性取证技术

由于供电和负载的相对变化,电网频率随着时间的变化存在着不大于±0.5 Hz 的波动。2005 年罗马尼亚的 Grigoras, C 博士^[8]在 3 个相距 200 km 以上的城市采用声卡进行录音,采用 DCLive/Forensics

V5.03 音频分析软件进行分析发现:这 3 个地方的录音显示电网频率特征具有相同的变化规律。他第 1 次提出了可利用这个特点进行录音时间检测,并提出了录音时间识别模型(见图 3)。同时也可用于音频篡改检测,能够有效地解决以往技术无法解决的音频数据被部分抹去或拼接问题。

2008 年丹麦林格比丹麦声学研究院的 Brixen^[19] 通过实验进一步证实了这一点。这一发现使得以往篡改技术难以克服相同采样率篡改问题的解决成为可能。Daniel^[20,21] 提出了一种电网频率和相位的高精度测量算法,可用于篡改音频导致的相位跳变检测。电网频率特征的提取成为了音频篡改领域的研究热点^[22,23]。目前采用 ENF 对音频真实性取证的研究主要集中在电网频率数据库构建和电网频率特征提取。

由于目前还缺乏权威的 ENF 数据库,ENF 曲线变化缓慢,一些录音设备中的语音信号处理模块可能滤除了 ENF 特征,部分手持录音设备采用电池供电等原因,导致对 ENF 用于音频真实性取证的研究转入低谷。

3.2 无电网频率成分下的音频篡改检测技术

如果音频信号中不包含电网频率特征,音频篡改检测就变得更加困难,文献[10]在分数余弦变换域上通过分析参数的不连续性以判决该音频是否发生了篡改,文献[11]研究了子带谱平滑度与音频篡改的关系,文献[24]研究了高阶谱相关性与音频篡改的关系,由于音频篡改位置通常发生在具有类噪声谱特征的寂音段,单纯从语音频谱角度研究音频篡改存在虚检和漏检现象,必须兼顾噪声帧和语音帧。

对于不同采样率或不同比特率的两段音频进行拼接处理时,针对不同比特率或不同采样率,可以采用重采样篡改检测方法,其原理是针对采样率提高时导致的线性插值现象进行检测。线性插值会使部分抽样点成为周围样点的线性组合,而且这种线性组合是周期性存在的。因此可以通过 EM、子带平滑度检测和奇异值分解统计检测等算法来检测样点的关联性从而判断是否发生了篡改。

对于相同采样率或同一段音频信号用音频编辑软件如 Cooledit, DC Live Forensics 等进行编辑处理时,就不会产生插值处理,不跟原始音频对比,无法从时频域上进行区分。而且当前一般的录音设备通常为电池供电,录音证据中通常也不包括市电频率成分,因此只能基于音频本身存在着语音或噪声的连续特性来进行判决^[10]。

Suwan 等人^[25] 将集成学习引入音频篡改检测中,通过对多个分类器输出结果进行综合考虑,其识别性能优于支持向量机(Support vector machine, SVM)、径向基函数(Radial basis function, RBF)及概率神经网络(Probabilistic neural network, PNN)等技术。Malik 等人将混响时间作为表征不同录音环境的特征参数^[26],用于音频编辑痕迹检测。混响时间是重要的声学参数,是室内声波的复杂传播环境的特征参数之一。混响时间可以通过最大似然估计算法获得^[27]。

4 录音设备识别

在法庭音频证据提供过程中,经常有声称采用某设备录制,但目前无更有效的方法验证这一点,往往会导致对方不认可这一证据。一般的录音设备的录音模块包括前端调理部分、模数转换部分、降噪算法部分和压缩算法部分,对于大部分品牌录音设备或手机厂商来讲,从专利保护角度出发,品牌录音设备厂商一般会采用不同的模拟电路和数字信号处理算法来实现录音功能。这些技术的不同就会导致录音信号中蕴含了区别于其他录音设备的个性特征。

录音设备特征数据库的构建具有十分重要的意义和广阔的应用前景,但也存在着许多关键技术瓶

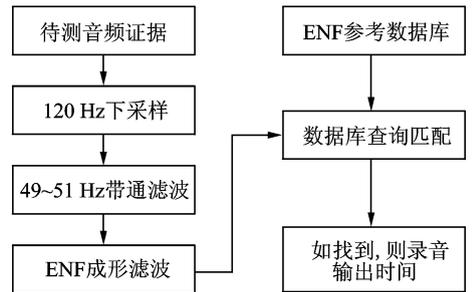


图 3 基于 ENF 技术的录音时间识别
Fig. 3 Recording time recognition based on ENF technology

颈,制约着在音频取证中的实际应用,突出的表现在录音设备特征表征、录音设备特征与其他音频特征之间的分离和录音设备识别模型等方面。目前的录音设备特征一般为声压频率响应曲线、有效频率范围和频谱参数等。其中声压频率响应曲线和有效频率范围特征很难从实际的音频信号中获得,需要在专业的消声室内采用纯音信号测量法等办法对录音设备进行测量,对声学环境要求很高;频谱参数在实际过程中应用较多,但该参数也蕴含了丰富音频特征,如语音特征、说话人特征和环境特征等,并非专门的录音设备特征,如何将录音设备特征从这些音频特征中分离开来是录音设备取证需要着重研究的课题。

4.1 录音设备特征方面

录音设备识别中目前一般采用梅尔频率倒谱参数(Mel-frequency cepstrum coefficient, MFCC)、噪声功率谱等作为特征参数。2007年以来,研究人员分别采用了时频域混合参数^[28]、接近寂音段的音频帧的傅里叶参数^[29]、幅度谱距离和相位谱的互相关、寂音段的功率谱密度函数^[30]、MFCC倒谱参数^[31]、随机谱特征^[32]、MFCC倒谱参数和归一化功率例谱参数(Power-normalized cepstral coefficients, PNCC)组成的混合参数作为录音设备的特征参数^[33],取得了较好的效果;文献^[34]研究了音频中的基音、共振峰和MFCC倒谱参数与录音设备之间的关联。

无论是傅里叶参数、共振峰、随机谱特征还是MFCC等倒谱参数,这些特征参数是包括语音识别、说话人识别在内的语音类模式识别的通用特征参数,它们蕴含了语音、语义、信道、噪声和录音设备等信息,并非专门用于录音设备识别的特征参数。如在设备识别中效果较好,使用最多的为MFCC倒谱参数,其根据人耳的听觉特性设计,常用于说话人识别,其低维部分一般反映说话人的语义特征,高维部分一般反映说话人的个性特征,如果不做处理,直接用作录音设备特征参数,势必会影响录音设备识别正确率的提高;如果寂音段的噪声平稳,功率谱、幅度谱等参数能够较大程度地表征录音设备特征,但如果时长不够或受到其他语音或噪声干扰时,来自语音和干扰的功率谱时变特性会严重影响录音设备的特征,就无法进行有效识别。

从目前的研究情况看,专门针对录音设备的特征参数的研究还非常少。实际上,不同品牌的录音设备由于专利、版权等原因,采用的放大电路、滤波电路、语音采集、降噪算法和压缩算法都有可能不同,这些不同就蕴含了录音设备的个性特征。因此录音设备特征提取需要解决以下问题:从底层硬件和算法角度出发,寻找电路和算法与表征录音设备特征之间的关联,从现有的语音类参数中提取录音设备个性特征和寻找具有较强抗噪声和各种干扰能力的特征参数。

4.2 录音设备识别模型

录音设备识别中一般采用回归分析、贝叶斯分类器、决策树、SVM及GMM等识别方法,其中表现较好、使用较多的是SVM模型。Kraetzer等^[28]采用K-均值和Bayes分类算法相结合对在11个房间内分别采用4种麦克风录制的音频信号进行录音设备分类研究,第1次提出了录音设备识别的模型,取得了一定的效果,录音设备识别率达到75.99%。Ling等^[33]采用了说话人识别中常用的高斯混合模型(Gaussian mixture model, GMM)作为录音设备识别模型,作为单状态隐马尔可夫模型(Hidden Markov model, HMM),GMM方法是说话人识别中主流识别方法,但该方法依赖于训练数据,模型参数设置比较困难,需要根据实验情况进行调整,这导致了算法在不同的问题或数据库上的通用性不高;而且GMM在小样本情况下表现不佳,也存在着过学习、局部极小等问题,在输入的特征数据不具有线性可分度情况下,区分性能大幅下降。Cemal^[35]等从手机录音信号中提取了手机的个性特征,并采用了SVM作为识别模型,对14种不同手机进行了识别,获得了96%的识别率,SVM在小样本情况下能够取得较好的性能,具有很高的分类精度、很强的学习能力和很好的泛化性能,被广泛应用于模式分类等领域。但在SVM分类器的具体设计中,参数选择对于分类准确率的影响很大,而且也存在着数据量大时分类能力低、速度慢等缺点。就目前的研究而言,无论是SVM,GMM还是回归分析等模型,皆为现有的语音识别或说话人识别中表现较好的模型直接应用,参数设置和模型设计并没有专门针对录音设备特点进行改进。录音设备特征往往与语义、说话人及信道特征混合在一起,呈复杂的分布特性。

4.3 数据库建设方面

目前公开报道的标准录音设备数据库主要如下:1997年,林肯实验室的 Reynolds 等^[36]采用9部手机和1个高质量麦克风对来自 TIMIT 数据库的384个说话人的音频进行再录音,得到了 HTMIT 数据库;同时,又采用这些录音设备对53个说话人进行录音,得到了 LLHDB 数据库,直到2014年,大量的研究都基于这两个数据库。随着移动通信技术的到来,新型品牌和型号的手机不断涌现,这两种数据库所涉及的录音设备已经淘汰。针对这种情况,Panagakos 等^[37,38]采用14部不同的手机录制了24个说话人的3360句话,重新构建了新的录音设备数据库。2014年,Constantine 构建了 MOBIPHONE 数据库^[39],该数据库采用了21部不同型号的手机录制了来自 TIMIT 数据库的12名男女的说话。

随着4G时代的到来,市场上的手机品牌和型号不断增加,这些数据库所涉及的手机品牌型号不仅数量偏少,没有或者很少涉及智能手机、其他录音设备以及占据市场较大份额的国产手机;所录制的说话人语言不包括汉语也未兼顾各种方言,录音对象没有覆盖各个年龄段,录音场合比较单一。

5 录音场合识别

音频的录音环境识别是指通过从音频中提取蕴含环境特征的参数来进行模式识别,从而判断出待测音频的录音环境。由于音频的录音环境多样且复杂,不少录音环境呈现出时变快衰落特性,而且环境特征参数覆盖频段宽,通常与说话人特征、录音设备特征和噪声混在一起,很难进行分离,因此对于录音环境识别的研究报道不多。Mubarak 等^[40]采用过零率特征和 MPEG-7 底层特征(音频波形、功率、谱包络和谱质心等)作为特征参数,采用 K-最近邻算法作为分类器,对包括餐馆、办公室、喷泉池边、自助餐厅、购物中心、会客室和走廊在内的多个录音场合进行识别,取得了平均为40%的识别率。Hong Zhao 等^[41]从混响和背景噪声中提取环境特征参数,采用 SVM 作为分类器,针对2240个压缩和非压缩音频进行识别,实验中采用了2个说话人、4种不同种类的麦克风和9种录音环境,取得了94%以上的识别率。

6 结束语

本文讨论了音频取证若干个方向的研究进展。首先将音频取证研究分成主动取证技术、真实性取证技术和音频来源性取证技术3类,综合近年来的研究,构建了音频取证框架;然后分别讨论了音频主动取证技术的基本概念、模型和研究进展;介绍了基于电网频率特征的音频篡改技术、无电网频率成分下的不同采样率的音频篡改检测技术和无电网频率成分下的相同采样率篡改检测的国内外发展情况;分析了录音设备的特征参数、模式识别和数据库建设的国内外进展;最后对录音场合识别进行了介绍。国内外虽然对音频取证进行了研究,但只是简单的将原来在语音识别和说话人识别中行之有效的技术应用到音频取证中,还没有深入探讨录音设备特征、录音环境特征与其他声学特征(如说话人特征、语音特征)的分离问题,也没有找到通用的篡改特征的提取方法。故应根据从硬件电路、语音信号处理算法角度分析国内市场上常见的品牌型号的手机和录音笔的内在个性机理,寻找录音设备特征抽取和识别一体化模型;从录音设备和场合角度研究音频篡改特征的提取方法,从而推动音频证据的真实性、来源性取证的不断发展。

参考文献:

- [1] 孙林慧,杨震. 语音压缩感知研究进展与展望[J]. 数据采集与处理,2015,30(2):275-288.
Sun Linhui, Yang Zhen. Compressed speech sensing for research progress and prospect[J]. Journal of Data Acquisition and Processing, 2015, 30(2): 275-288.
- [2] Maher R. Audio forensic examination[J]. Signal Processing Magazine, IEEE, 2009, 26(2): 84-94.
- [3] Petrovic R. Digital watermarks for audio integrity verification[C]//7th International Conference on Telecommunications in Modern Satellite, Cable and Broadcasting Services, Serbia and Montenegro. Yugoslavia: IEEE Computer Society Press, 2005: 215-220.
- [4] Kirbiz S, Lemma A N, Celik M U, et al. Decode-time forensic watermarking of AAC bitstreams[J]. IEEE Transactions on In-

formation Forensics and Security, 2007, 2(4): 683-696.

- [5] Kirbiz S, Celik M, Lemma A, et al. Forensic watermarking during AAC playback[C]//IEEE International Conference on Multimedia and Expo. Beijing, China; IEEE Computer Society Press, 2007: 1111-1114.
- [6] Singh L, Sridharan S. Speech enhancement for forensic applications using dynamic time warping and wavelet packet analysis [C]//TENCON '97, IEEE Region 10 Annual Conference. Brisbane, Australia; IEEE Computer Society Press, 1997(2): 475-478.
- [7] Ortega-Garcia J, Cruz-Llanas S, Gonzalez-Rodriguez J. Speech variability in automatic speaker recognition systems for forensic purposes[C]//IEEE 33rd Annual 1999 International Carnahan Conference on Security Technology. Madrid, Spain; IEEE Computer Society Press, 1999: 327-331.
- [8] Grigoras C. Digital audio recording analysis: The electric network frequency (ENF) criterion[J]. *Int J Speech Language Law*, 2005, 12(1): 63-76.
- [9] Rodriguez D P N, Apolinario J A, Biscainho L W P. Audio authenticity: Detecting ENF discontinuity with high precision phase analysis[J]. *IEEE Transactions on Information Forensics and Security*, 2010, 5(3): 534-543.
- [10] 钟巍, 孔祥维, 尤新刚, 等. 基于分数倒谱变换的取证语音拼接特征提取与分析[J]. *数据采集与处理*, 2014, 29(2): 248-253. Zhong Wei, Kong Xiangwei, You Xinggang, et al. Splicing feature extraction and analysis based on fractional cepstrum transform in voice forensics[J]. *Journal of Data Acquisition and Processing*, 2014, 29(2): 248-253.
- [11] 丁琦, 平西建. 针对语音变换的语音篡改检测[J]. *数据采集与处理*, 2012, 27(1): 57-62. Ding Yi, Ping Xijian. Speech tampering detection for voice transformation[J]. *Journal of Data Acquisition and Processing*, 2012, 27(1): 57-62.
- [12] 谢玲, 范明泉. 基于时域统计特征的音频内容取证新算法[J]. *电讯技术*, 2013, 11: 1476-1481. Xie ling, Fan Mingquan. A novel audio content forensic scheme based on time domain statistical characteristic[J]. *Telecommunications Technology*, 2013, 11: 1476-1481.
- [13] Gulbis M, Muller E, Steinebach M. Audio integrity protection and falsification estimation by embedding multiple watermarks [C]//International Conference on Intelligent Information Hiding and Multimedia Signal Processing, IHH-MSP '06. Pasadena, CA, United States; IEEE Circuits and Systems Society Press, 2006: 469-472.
- [14] Wu Shaoquan, Huang Jiwu, Huang Daren. Self-synchronized audio watermark in DWT domain[C]//2004 IEEE International Symposium on Circuits and Systems. Vancouver, BC, Canada; IEEE Circuits and Systems Society Press, 2004: 712-715.
- [15] Lee Sang Kwang, Ho Yo Sung. Digital audio watermarking in the cepstrum domain[J]. *IEEE Transaction on Consumer Electronics Press*, 2000, 46(3): 203-207.
- [16] Liu J, Niu X, Kong W. Image watermark based on singular value decomposition[C]//Proc of the Int Conf Intelligent Information Hiding and Multimedia Signal Processing. Daejeon, Korea; Springer Verlag Press, 2006: 457-460.
- [17] Li D, Kim J W. Audio forensic marking using quantization in DWT-SVD domain[C]//13th International Conference on Advanced Communication Technology (ICACT). Gangwon-Do, Korea; IEEE Communication Society Press, 2011: 988-991.
- [18] Park J, Seo Y, Yoo W Y, et al. Audio forensic marking for music streaming service[C]//2013 International Conference on ICT Convergence (ICTC). Jeju Island, Korea; IEEE Computer Society Press, 2013: 817-818.
- [19] Brixen E B. ENF—Quantification of the magnetic field[C]//Proc Audio Engineering Society 33rd Conf, Audio Forensics—Theory and Practice. Denver, CO, United States; Audio Engineering Society Press, 2008: 1-6.
- [20] Daniel P N, Apolinario J A, Biscainho L W P. Audio authenticity: Detecting ENF discontinuity with high precision phase analysis[J]. *IEEE Transactions on Information Forensics and Security Press*, 2010, 5(3): 534-543.
- [21] Nicolalde D P, Apolinario J A. Evaluating digital audio authenticity with spectral distances and ENF phase change[C]//IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP 2009. Taipei, China; IEEE Signal Processing Society Press, 2009: 1417-1420.
- [22] Gerazov B, Kokolanski Z, Arsov G, et al. Tracking of electrical network frequency for the purpose of forensic audio authentication[C]//13th International Conference on Optimization of Electrical and Electronic Equipment (OPTIM). Brasov, Romania; Transilvania University of Brasov Press, 2012: 1164-1169.
- [23] Nicolalde-Rodriguez D P, Apolinario J A, Biscainho L W P. Audio authenticity based on the discontinuity of ENF higher harmonics[C]//Proceedings of the 21st European Signal Processing Conference (EUSIPCO). Marrakech, Morocco; European Signal Processing Conference Press, 2013: 1-5.
- [24] 高阳, 黄征, 徐彻, 等. 基于高阶频谱分析的音频篡改鉴定[J]. *信息安全与通信保密*, 2008(2): 94-96. Gao Yang, Huang Zheng, Xu Ch et al. Audio tampering identification based on higher order spectral analysis [J]. *Information Security and Communication Security*, 2008(2): 94-96.

- [25] Suwan T, Jaiyen S, Wiangsr R. Edited audio detection using ensemble learning[C]//7th International Conference on Knowledge and Smart Technology (KST). Chonburi, Thailand; IEEE Press,2015;71- 74.
- [26] Malik H, Farid H. Audio forensic from acoustic reverberation[C]//IEEE International Conference on Acoustic Speech and Signal Processing-ICASSP 2010. Dallas, TX, United States;IEEE Signal Processing Society Press,2010;1710-1713.
- [27] Amelia C, Tudor C, Cristian N, et al. Analysis of reverberation time blind estimation used in audio forensics[C]//2014 11th International Symposium on Electronics and Telecommunications, ISETC 2014. [S. l.];IEEE Press,2014;281-284.
- [28] Kraetzer C,Oermann A,Dittmann J,et al. Digital audio forensics: A first practical evaluation on microphone and environment classification[C]// Proc 9th Workshop on Multimedia and Security. Dallas, TX, USA;IEEE Press, 2007;178-181.
- [29] Robert B,Christian K,Jana D. Microphone classification using Fourier coefficients[C]//LNCS. Berlin Heidelberg;Springer-Verlag, 2009;235-246.
- [30] Pandey V, Verma V K, Khanna N. Cell-phone identification from audio recordings using PSD of speech-free regions[C]// 2014 IEEE Students' Conference on Electrical, Electronics and Computer Science (SCEECS). Bhopal, India;IEEE Computer Society Press,2014; 1-6.
- [31] Kotropoulos C, Samaras S. Mobile phone identification using recorded speech signals[C]//2014 19th International Conference on Digital Signal Processing (DSP). Hong Kong, China;IEEE Press, 2014; 586-591.
- [32] Aggarwal R, Singh S, Roul A K. Cellphone identification using noise estimates from recorded audio[C]//2014 International Conference on Communications and Signal Processing (ICCSPP). Melmaruvathur, Tamil Nadu, India;IEEE Press,2014; 1218-1222.
- [33] Zou Ling, Yang Jichen, Huang Tangsen. Automatic cell phone recognition from speech recordings[C]//2014 IEEE China Summit & International Conference on Signal and Information Processing. Xi'an, China;IEEE Press,2014 ;621-625.
- [34] 达钊,李倩,郭霞生,等.不同录音系统对声纹检测的影响[J].南京大学学报:自然科学,2011,47(2):201-206.
Da zhao,Li qian,Guo Xiasheng,et al. The influence of different recording system for detecting voiceprint[J]. Journal of Nanjing University:Natural Science,2011,47(2):201-206.
- [35] Hanilci C, Ertas F, Ertas T, et al. Recognition of brand and models of cell-phones from recorded speech signals[J]. IEEE Transactions on Information Forensics and Security,2012,7(2): 625-634.
- [36] Reynolds D A. HTIMIT and LLHDB: Speech corpora for the study of handset transducer effects[C] //Proc 1997 IEEE Int Conf Acoustics,Speech, and Signal Processing. Munich, Germany;IEEE Press, 1997;1535-1538.
- [37] Panagakis Y, Kotropoulos C. Automatic telephone handset identification by sparse representation of random spectral features [C]//Proc 14th ACM Multimedia and Security Workshop. Coventry, UK;IEEE Press,2012;91-95.
- [38] Panagakis Y, Kotropoulos C. Telephone handset identification by feature selection and sparse representations[C]// Proc 2012 IEEE Int Workshop Information Forensics and Security. Tenerife, Spain;IEEE Press, 2012;73-78.
- [39] Constantine L K. Source phone identification using sketches of features[J]. IET Biom,2014,3(2):75-83.
- [40] AlQahntani M O, Muhammad G, Alotaibi Y A. Environment sound recognition using zero crossing features and MPEG-7[C]// 5th International Conference on Digital Information Management (ICDIM). Thunder Bay,Ontario,Canada;IEEE Computer Society Press,2010;502-506.
- [41] Hong Z, Malik H. Audio forensics using acoustic environment traces[C]//2012 IEEE Statistical Signal Processing Workshop (SSP). Ann Arbor, MI, United States;IEEE Computer Society Press,2012;373-376.

作者简介:



包永强(1973-),男,博士,教授,研究方向:语音信号处理及音频取证,E-mail: jybyq@163.com。



梁瑞宇(1978-),男,博士,副教授,研究方向:语音信号处理,E-mail: lly1711@163.com。



丛韞(1995-),女,本科生,研究方向:语音信号处理。



高冲红(1995-),女,本科生,研究方向:语音信号处理。



王青云(1972-),女,博士,教授,研究方向:语音信号处理。

