

基于欺骗性语音声纹特征分析研究

屈李晶 余琳* 王子曦

湖南警察学院

DOI:10.12238/acair.v2i2.7394

[摘要] 随着各种声音处理技术快速发展,利用欺骗性语音实施的电信诈骗案件也日益增多。为了改善欺骗性语音存在的涉及类型单一、不稳定且耗时较长及未拓展至快速认定欺骗性语音等问题,本研究通过对欺骗性语音进行声纹特征分析以实现快速分辨语音真实性。采用迅捷音频转换器和智能声纹鉴定工作站等软件进行语音处理并制作样本,最终进行听觉和频谱检验分析。

[关键词] 欺骗性语音; 声纹特征; 频谱分析

中图分类号: H11 文献标识码: A

Research on the analysis of deceptive speech voiceprint features

Lijing Qu Lin Yu* Zixi Wang

Hunan Police Academy

[Abstract] With the rapid development of various sound processing technologies, the number of telecommunications fraud cases using deceptive speech is also increasing. In order to improve the problems of single type, unstable, time-consuming, and not expanding to quickly identify deceptive speech, this study conducted voiceprint feature analysis on deceptive speech to achieve rapid recognition of speech authenticity. We use software such as the Swift Audio Converter and Intelligent Voiceprint Identification Workstation for speech processing and sample production, and ultimately conduct auditory and spectral analysis.

[Key words] deceptive speech; voiceprint features; spectral analysis

引言

近年来电信诈骗案件发案率呈“指数级”增长,音频证据重要性由此凸显,也带来系列问题。当前,有许多研究在部分欺骗性语音特征和研究方法上取得一定进展,但存在涉及类型单一、不稳定且耗时较长及未拓展至快速认定欺骗性语音等问题,这导致对其造成的潜在危害未得到应有警惕和防范。本研究采用文字、图表形式对转换语音、合成语音与原始语音的声纹频谱在共振峰频率、基频曲线等方面进行探讨。总结研究发现并提出对未来研究的展望,旨在为欺骗性语音相关领域提供新研究视角和理论支持。

1 欺骗性语音原理概述

欺骗性语音是指利用伪装手段制作的非原始且带有欺骗目的性的语音^[1],可大致分为语音转换(Voice Conversion, VC),语音合成(Speech Synthesis, SS),重播语音。

1.1 语音合成。其中语音合成的目标是将文本转换为可听的语音音频,使计算机能够“说话”。它涉及多个步骤,包括文本预处理、声学模型训练、参数生成和声音波形合成等。语音合成系统通过学习和模拟人类语音产生的过程,以达到生成自然语音的效果^[2]。

1.2 语音转换。语音转换指将一个人(原说话人)的声音个性化特征(如频谱、韵律等)通过“修改变换”,使之听起来像另外一个

人(目标说话人)的声音,同时保持说话内容信息不变^[3]。广义上把改变语音中说话人个性特征的语音处理技术统称为语音转换。广义的语音转换可分为非特定人语音转换和特定人语音转换。

2 实验方法

2.1 实验环境与器材。选取较为安静的场地(如带有隔音效果的面积较小的房间),使用SONY PCM-D10款录音笔进行录制。

2.2 样本选择。语音转换、语音合成录制语料如下:

(1)“我是xxx,最近有件急事,着急用钱,需要找你借钱,你放心我肯定没有骗你,给我支付宝转账,告诉我你收到的验证码”。(2)“我是公安局的,你已经涉嫌违法犯罪,已经在网上被通缉了,请你来一趟派出所,加一下我的微信或者qq号”。(3)“不要挂我的电话,你是本人吗,为什么不接我电话啊,为什么不早点告诉我呢,这个绝对不是诈骗”。(4)“找个隐蔽的地方,我这边已经成功帮你贷好款了,为了安全和保密不要和任何人说,支付宝还是微信?”(5)“我们这边是有银行担保的,你需要开通网上银行,提供一下你的银行卡号,我会把你的钱转到安全账户”(6)“明天天气如何,晚自习还去吗,怎么还是这个样子,我已经不想再和你说话了”。

2.3 数据收集与处理。(1)建立欺骗性语音数据集。语音样本的年级主要分布在20级、21级、22级,男女比例约为7:3;共采集

了1310条语音样本,其中包括1000条转换语音、160条合成语音和150条重放语音。关于语音转换及语音合成原始语音的录制,为保证原始语音的录制质量,避免过多的噪声对后续实验产生干扰,实验人应选取较为安静的场地,并且统一要求以距离录音笔10cm的标准,要求志愿者阅读指定录制语料(注:为避免出现录制语音效果不佳对频谱造成影响的情况,志愿者每条语音内容连续阅读三遍,实验人员选取最佳语音进行实验),实验负责人应做好语音数据的储存,并统一以“原始.语音种类.录制内容编号”命名分类。(2)语音分析软件。转换器,在语音转换里可称为“变声器”,是一种能改变声音的设备或软件,通过音频和数字信号处理算法实现声音的变换,即通过对输入的声音信号的系列处理和变换,达到改变声音的效果。在本实验中,采集完相应数据后使用pr软件进行降噪处理,再对语音转换的语音材料利用迅捷音频转化器进行转换,语音合成的语音材料利用剪辑中的“克隆音色”功能进行合成,得到的相应语音以“实验.语音种类.录制内容编号”命名分类,并同样做好数据储存。(3)智能声纹鉴定工作站(FIAS)。智能声纹鉴定工作站在实现对语音的采集、编辑、播放等基础功能的基础上,主要提供数字化声谱分析、真实性检验、降噪与增益三大功能,在本实验中,利用该工作站对欺骗性语音样本进行频谱分析。

3 实验结果与分析

3.1听觉差异性分析。首先进行听觉测验,以中国刑事警察学院的陈子谦在《合成语音的语音同一认定研究》中对合成语音进行听觉检验的方法为例。该实验基于转换语音与合成语音的语料一致性,从听觉特征中选取14个特征制作听觉比对检验表。这些特征为共鸣方式、嗓音纯度、平均音高、音域、语速、节奏、清晰度、流畅度、响亮度、舌位、变调轻重、音调、儿化音和音节完整性。每个特征根据正常语音与欺骗性语音的相符程度分为5个等级与相应5个分数,选取5位听音人对这些特征进行打分并分析结果,打分标准如表1所示:

表1 听觉差异分析评分标准表

分数	正常语音与变声语音的一致性
1分	完全不一致
2分	多数部分不一致
3分	不能确定
4分	多数部分一致
5分	完全一致

统计结果显示,欺骗性语音和正常语音在变调、流畅度、儿化音以及轻重音等听觉特征上表现出较高一致性,而在嗓音纯度、平均音高、音域以及清晰度等一致性较低。具体表现在:自然度不够,整体上语音机械度较强。例如“我肯定不会骗你”,原始语音的“肯定”两字更加坚定、富有情感,且语音整体有一定的节奏感,而欺骗性语音则较生硬;合成语音朗诵的文本不会出现“b”“p”“h”“f”等原始语音会出现的喷麦情况。

3.2转换语音类声纹特征分析。(1)能量曲线形态特征。声学特征可通过声音能量来表示,即声波振幅大小。本研究选取要求朗诵的20句话作为研究对象,发现转换语音与原始语音的能量曲线形态

存在较为相似的情况(见图1),而相同音节出现能量大小的差异可能是由于转换软件对声强进行调整导致转换语音的能量曲线形态特征较为微弱。(2)基频特征分析。基频指基音频率,基频走向能够反映一段时间内基频变化。以图2为例观察,同一人基频曲线会发生整体升高或降低,但曲线单调区间基本保持一致,因此转换语音基频曲线主要特征表现为转换后会发生整体升高或降低,此外,将基频提高到300HZ以上时基频曲线会出现纵向拉伸。(3)共振峰走向特征。选取“钱”“要”两个音节中复合元音[ian][iau]作为研究对象,如图2反映的宽带图中观察到的共振峰走向来看,低次共振峰走向表现出较高的一致性和稳定性,而“要”字的高位共振峰走向表现较为模糊,观察图2中“找”及其他音节亦可见此现象。因此,对男性语音使用软件进行转换时,高次共振峰走向特征易发生改变。

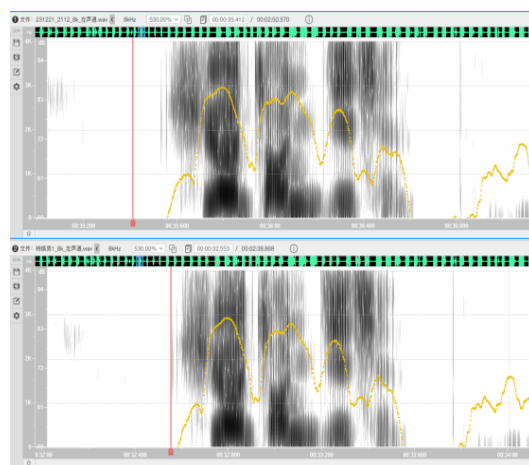


图1 转换语音能量曲线特征以“需要找你借钱”为例

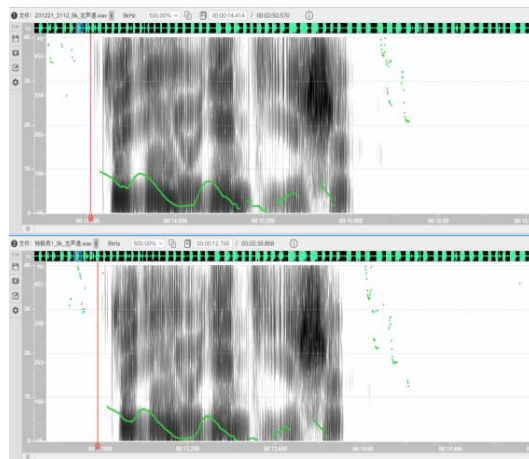


图2 基频曲线特征以“最近有件急事”为例

3.3合成语音类声纹特征分析。(1)共振峰特征。共振峰频率特征反映发音过程中或某一时段、瞬间声道的形状特点,共振峰走向特征则反映动态形态。从合成语音材料及原语音选出15组相同音节进行观察。就国音智能分析平台设置的4kHz谱范围而言,13组相同音存在显著差异且大部分均集中在2kHz以上,以图3某男性合成语音及原始语音“mei”音节4kHz的宽带图为例,可观察到共振峰频谱特征的差别主要表现在共振峰位置不同。

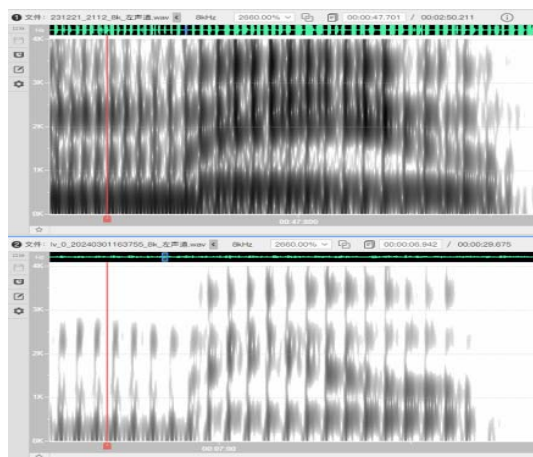


图3 合成语音及原始语音“mei”音节4kHz谱范围下宽带图

高位共振峰差异主要反映在共振峰频率、走向等特征上。图4中粉、绿两条曲线分别代表某男性合成语音及其原型语音的“bu”“pian”两音节共振峰瞬时功率图,可观察到,两者F1较契合但材料语音第三共振峰为2.7kHz、2.3kHz而原型语音第三共振峰为2.2kHz、1.8kHz,前者该共振峰明显偏高。经过大量观察,发现共振峰位置差异是合成语音主要且较为稳定的差异。

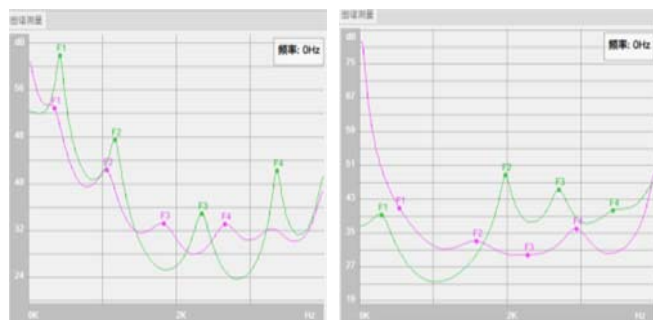


图4 合成语音及原始语音“bu”“pian”音节瞬时功率图

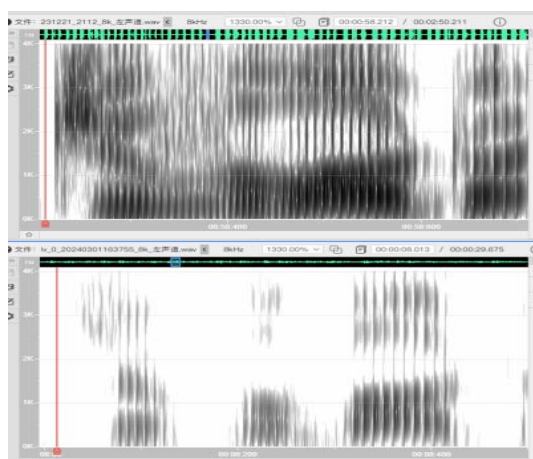


图5 合成语音及原始语音“支付宝”音节宽带图

此外,部分语音在音节过渡段形态上存在差异,以图5“支付宝”的音节宽带图为例,合成语音音节宽带图在元音、辅音连接处存在一定的空缺或割裂,而原始语音表现为过渡自然,当前语音合成技术尚不能高质量呈现语音过渡段。

(2)能量曲线形态特征。根据听觉检验结果,由于合成语音机械感较强、情感反映较弱,其能量曲线形态表现为变弱或变强且具有整体一致性。以合成及原始语音图6“我肯定没有骗你”的能量曲线为例,可明显观察到前者能量集中分布在72db以下区间,而后者各个音节间变换相对较大且能量分布区间明显高于合成语音。

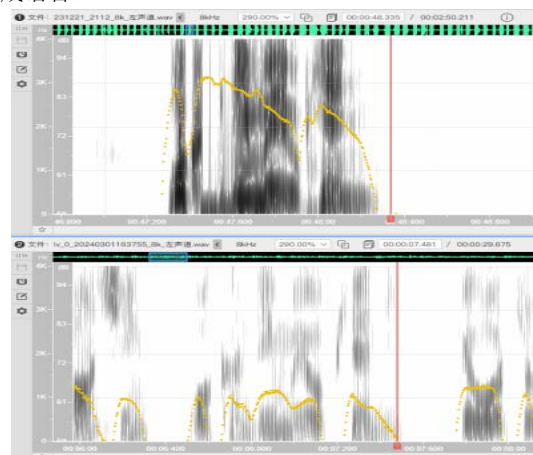


图6 合成语音能量曲线以“我肯定没有骗你”为例

4 总结与展望

据上述实验结果,可得阶段性转换、合成技术能改变发音人大部分声学特征,部分呈现出一定规律。在听觉特征中,变调、流畅度、儿化音、音节完整性、轻重音等方面呈现出较高稳定性,变化程度相对较小。在嗓音纯度、平均音高、音域、清晰度等方面则表现为一致性较低。合成语音明显表现出缺乏感情和缺少断句等问题。在频谱特征中,转换语音频谱表现为在共振峰频率、走向特征上呈现较高稳定性。转换后基频走向存在一定变化规律,即随频率的增加,基频走向曲线纵向逐渐延长,同时对转换语音发音速度产生影响。根据前四个共振峰测量结果,不同人或音素间存在发音上相似性,大部分情况下存在明显差异,并且缺乏特定规律。此外,低次共振峰的变化幅度较小,而高次变化较显著且容易,部分音节还存在音节内辅音和元音过渡段缺失的情况。

进一步的工作有望在语音转换速度变化方面展开,探讨该变化是否存在一般规律性。我们将继续研究共振峰变化规律性,着重探讨同一人在合成语音中共共振峰变化程度的差异。同时计划扩大研究样本,以期进一步探索转换语音普遍规律。

[基金项目]

2023年国家级大学生创新创业训练计划项目《基于部分欺骗性语音研究的相关应用》(S202311534026)。

[参考文献]

- [1]魏骏北.欺骗性语音的检测算法研究及其系统实现[D].广东技术师范大学,2023.
- [2]滕永盛.汉语信息处理中若干基础问题的研究[D].安徽:中国科学技术大学,1998.
- [3]周晔.基于深度学习的语音欺骗检测[D].杭州电子科技大学,2024.