

基于常见伪装语音的声纹鉴定研究

刘 莹 怡, 田 丰

甘肃政法大学司法警察学院, 甘肃 兰州

收稿日期: 2025年7月15日; 录用日期: 2025年8月15日; 发布日期: 2025年8月22日

摘 要

近年来电信网络诈骗案件呈现高发态势, 声纹鉴定对此类案件的侦破具有重要作用。由于在近期电信网络诈骗案件当中较高频次出现的人为伪装语音, 以及考虑到日常生活中人们佩戴口罩的频率较高情况。本文实验以一些日常生活中的实际案件里出现率较高的捏鼻伪装语音和戴口罩伪装语音两种人为语音伪装的方式为具体研究对象, 采用听觉分析和语谱图分析两种不同分析相结合的方法, 同时结合定量的分析结果, 运用成对比较T检验方法对两种变异语音中体现发音人本身稳定性的一些典型的声学特征进行研究分析。本文的研究可以为语音识别技术的实际应用提供一定的依据和参考, 为公安实战中基于变异语音案件的解决提供了一种新的思路。

关键词

声纹鉴定, 伪装语音, 声学特征, 语谱图分析, 成对T检验方法

Research on Voiceprint Authentication Based on Common Disguised Voices

Yuyi Liu, Feng Tian

School of Judicial Police, Gansu University of Political Science and Law, Lanzhou Gansu

Received: Jul. 15th, 2025; accepted: Aug. 15th, 2025; published: Aug. 22nd, 2025

Abstract

In recent years, telecom and internet fraud cases have shown a high incidence trend, where voiceprint identification plays a crucial role in solving such cases. Given the frequent occurrence of artificial voice impersonation in modern telecom fraud and considering the requirement for daily mask-wearing during COVID-19 prevention measures, this study focuses on two common voice imper-

sonation methods—nose-pinch impersonation and mask-wearing impersonation—found in real-life cases. By combining auditory analysis with spectral graph analysis, and integrating quantitative results, we employ pairwise t-test to investigate typical acoustic features reflecting speaker stability in these variants. This research provides practical references for voice recognition technology applications and offers a novel approach for solving voice variation cases in police investigations.

Keywords

Voice Print Identification, Disguised Speech, Acoustic Features, Speech Spectrum Analysis, Paired T-Test Method

Copyright © 2025 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

声纹鉴定技术在刑事侦查案件当中的广泛应用是传统侦查手段与现代信息技术的紧密联系与结合,是科技创新推动刑事侦查工作展开的具体表现之一,也是信息导侦的有效途径之一。在这个科学技术的不断进步与发展的时代,现代化通讯技术和录音录像技术也随之不断地普及与提高,这就导致了越来越多的需要声纹识别的录音或录像制品的出现,而随着科学技术的不断发展,声纹识别技术在刑事案件审查和侦查过程中所发挥的作用也将日益凸显。可见,声纹识别在法治社会的建设中,对于打击犯罪、维护公民权利等方面,起到了举足轻重的作用。总而言之,在有关于声响和语音等作为声音证据的司法活动过程当中,只有对声纹鉴定技术的作用与价值加强重视,才能够源源不断地发挥出刑事科学的力量,推动我国的声纹鉴定技术向着更高一层的水平发展和研究。

20 世纪 40 年代初期,语图仪发明现世,现代声纹鉴定技术也随着它的出现开始发展起来。当时在贝尔实验室工作的物理学家波特(Potter)等人开始研究利用声谱仪来分析语音[1]。国外对伪装语音的研究最早开始于二战前夕,一开始只是侧重于言语中所表现出来的不同情感状态的识别。20 世纪 70 年代,人们开始对伪装语音说话人识别进行研究。国内声纹鉴定技术的研究起步较晚,但是在这一领域中,各种技术的开发和应用大部分都已在国内几个主要的司法语音实验室中进行。在我国,许多学者对声纹鉴定技术进行了多方面的讨论,有些已经达到或接近了国际水准,有些还有待进一步探索和发展。早在 1995 年,广东省公安厅刑事技术中心的王英利、李卫平二人对利用鼻韵母音节过渡形态进行了声纹鉴定的研究[2];在 2010 年,中国人民公安大学的庄琳使用 VS-99 语音工作站对捏鼻语音的声学特点进行了分析归纳[3];在 2011 年,王英利、潘自勤、蓝常山三人对声纹鉴定中的音强特征进行的更深层次的研究[4];在 2016 年,郑方、李蓝天、张慧三人对声致识别技术及其应用现状进行了一定的阐述说明[5];在 2018 年,清华大学的李蓝天对说话人识别中的特征学习方法进行了一个系统的分析[6];在 2018 年,中国人民公安大学的陈维娜、曾庆发二人对不同录音环境对语音特征的影响进行了研究[7];在 2019 年,中国刑事警察学院的刘世杰、王虹二人对云录音声纹检验实验进行了研究归纳[8];在 2020 年,中国人民公安大学的王康宁、杨兵、马腾三人对声纹技术在网络犯罪侦查取证过程当中中的应用初探进行了归总[9];在 2021 年,山西大学的杨伟与山西警察学院的杨俊杰合作对于语言学音系例字的口音自动识别进行了研究[10];在 2022 年,湖南警察学院的张明健、张悦对基于语谱图以及深度置信网络的方言自动辨识与说话人识别进行了阐述说明[11]。

尽管上述学者对声纹鉴定技术的发展开展了多项研究,但是很多研究都是零散进行的,尚未形成较大的规模。对于不同的说话状态下,人类个体的发音会产生什么样的具体变化与不同,是否会对声纹鉴定的结果产生什么影响,目前为止仍然缺少相应的系统研究与证实。为积极应对当前电信网络诈骗类案件的高发趋势,合理利用语音证据来打击违法犯罪行为,亟需确定不同状态下个体的声纹特征的变化情况,探索并评估出这些变化对声纹鉴定产生的不同影响。本文实验以一些日常生活中的实际案件里出现率较高的捏鼻伪装语音和戴口罩伪装语音两种人为语音伪装的方式为具体研究对象,采用听觉分析和语谱图分析两种不同分析相结合的方法,提出基于语谱图的声纹鉴定方法和流程,同时,结合定量的分析结果,运用成对比较 T 检验方法对两种变异语音中体现发音人本身稳定性的一些典型的声学特征进行研究分析。本次研究为语音识别技术的实现与应用提供了一定的实验依据和参照,并且为公安实战当中对一些涉及伪装语音鉴别类案件的侦查与处理提供了一条全新的破案路径。

2. 相关理论

2.1. 声纹鉴定的概念

声纹指的是对语音纹理的描摹,在语言学界当中又被叫作语图。而在刑事技术和司法鉴定的领域里,人们参照了指纹的名称,从而生动地称其为声纹。语图是一种用声谱仪表示的、承载着声音信息的图像,叫作语声的频谱图。

声纹鉴定指的是利用与声学及语音学有关的基本知识,通过电声学的仪器设备,将有声话语通过语图仪或者是声纹仪用磁介质记录下来,并且对其频谱进行分析,把已有的语音检材与嫌疑人或当事人的语音样本用听觉分析与语图分析两种方法,将存在于声纹里的各种与语音有关的特殊信息结合进行鉴别与判断,从而得出语音检材与语音样本二者是否同一的结论。

2.2. 声纹鉴定的作用

在刑事侦查与司法领域当中,声纹鉴定技术起着提供证据与线索,以及审查证据真实性的作用。近年来,随着中国通讯技术现代化的进步发展以及录音设备的普及,通过电话实施的各种各样的犯罪案件数量只增不减,再加上人们的法律与证据意识有所提高,所以在法庭上逐渐出现了越来越多的录音证据。由此可以看来,随着科技的不断发展与进步,声纹鉴定技术在打击犯罪,维护社会治安以及保障公民合法权益等方面逐渐发挥着越来越重要的作用。

2.3. 语音的听觉特征

语音的听觉特征是个人发音器官特性、发音习惯特性以及讲话习惯特性在口语中的具体呈现征象。它反映了个体的独特性,就如同每个人都拥有独一无二的指纹一样。听觉特征主要体现在三个方面:嗓音音质,涵盖了声音的清脆或浑浊、尖细或低沉等特质;口头言语,包含语速的疾缓、语调的升降变化以及停顿的节奏把控;口语缺陷,诸如口吃、吐字不清或者某些特定的发音错误等。

2.4. 语音的语谱特征

首先原始信号进行分帧加窗,就可得到很多帧,然后对每一帧做 FFT (快速傅里叶变换),而傅里叶变换的作用是把时域信号转为频域信号,再将每一帧傅里叶变换以后的频域信号(频谱图)在时间上堆叠起来则可得到声谱图。原始信号对应的,无论是频谱还是功率谱,反应的都是信号在各个频率分量上的信息。但是这两者都丢失了信号在时间维度上的信息。从而为了将一个信号同时表达出信号在频率分量

上的信息，又要显示出信号随着时间变化的信息，这便是语谱图的本意。

2.5. 语音的声学特征

2.5.1. 时长比例特征

如果我们将一段语流视为一段音素序列，那么便会惊奇地发现，即便不同的人讲述的是同一段音素序列，各对应音素的时长通常也是存在差异的。这意味着每个人在发声时，对于相同的音素序列，其持续时间的把握是各不相同的。为了更有效地进行研究和分析，这里我们仅仅探讨稳定性相对较强的几个方面的特征，包括音节中声母韵母时长比例、鼻韵母中主要元音与鼻韵尾时长比例以及音节间时长比例这三个重要的方面。通过对这些相对稳定特征的研究，我们能够在一定程度上更深入地理解和把握语流中的音素时长规律。

2.5.2. 基频特征

基频乃是语音当中极为重要的声学参数之一。能够致使基频产生变化，且归属于话者个体的因素众多。其中，生理因素包含性别、年龄以及健康状态等。比如，男性和女性的基频通常就存在明显差异，而随着年龄的增长，基频也会有所改变，健康状况不佳时同样可能影响基频。再者，心理因素如情绪、心态等也会发挥作用。当一个人处于兴奋、紧张或平静等不同的心理状态时，基频会有所不同。此外，还有谈话环境、人际关系等广义语境因素。

2.5.3. 共振峰频率特征

共振峰是指在语音图谱当中能量相对集中的若干区域，而共振峰频率所指的是共振峰于带宽频谱图上所处的位置。为了便于计算和分析，通常将共振峰的宽度简化处理，取其中心线来当作共振峰的频率。通过 praat 这一工具，能够精确地计算出共振峰频率 F1、F2、F3 和 F4 的数值。共振峰频率特征具备较强的稳定性和差异性，这使得它既能充当认定同一的指标，也能成为否定同一的指标。

3. 实验

3.1. 实验环境

语音采集设备为 LSA-AN00 录音器，采样频率 48,000 Hz，录制格式.mp3。数据处理设备为基于 Window10 专业版 64 位的计算机。多功能语音学专业软件 Praat，版本号 v6.2.0.4。数据计算软件为 WPS office 的表格工具，版本号 2023 春季更新(14,036)。招募五名，身体状况良好且普通话发音标准的男性发音人，年龄均为 21~24 周岁，基本无发言口音和特殊发音习惯，具体情况详见表 1。

Table 1. Basic information of five speakers
表 1. 五位发音人的基本情况

编号	发音人	性别	年龄	出生地
1	甲	男	22	甘肃武威
2	乙	男	21	河南信阳
3	丙	男	24	浙江台州
4	丁	男	22	陕西榆林
5	戊	男	23	辽宁盘锦

录音的具体方法如下：发音人在正常发音、捏鼻子以及戴口罩这三种状态下发音，在静室(保持安静无噪声，且不会产生回声的密闭房间)平稳地念读实验语音材料各五遍，每天采样一次，共采样三次。并且要求发音人的嘴与麦克风采集口的距离控制在 5~10 厘米左右，切换不同的发音状态时候需间隔一分钟，念读的实验语音材料为中性语句，选取自作家三毛的语录：“飞蛾扑火时，一定是极快乐幸福的。”采集录制语音时确保外界环境基本稳定，无干扰因素，并对每一份样本进行备注编号。

3.2. 实验及分析

3.2.1. 听觉分析

在语音检验中，听觉分析是一种重要的检验方法。其主要目的是通过听音人的听觉感知，对案件里出现的语音检材以及样本语音材料，二者之间进行语音定性分析。人类的听觉功能有一个显著的特征，就是可以接受非常宽的动态范围的声音，并且，人耳对声音具有较强的解析和分辨能力。因为人耳对不同频率的声音都有一个相应的响度范围，并对各种不同响度的声音都有较强的分辨能力。所以，在变异语音检验中，听音人的听觉分析对于变异语音进行听觉鉴别显得尤为重要。

在正常状态下发音，五位发音人的语音样本均声音洪亮、吐字清晰平稳、普通话发音标准、语速正常且流畅。与正常发音的语音样本对比而言，发音人在捏鼻子的状态下发音，会出现明显且较长的鼻音，声音也变成较为沉闷，没有正常发音那么洪亮，语速变慢不流畅，吐字含糊不清。与正常发音的语音样本对比而言，戴口罩发音听着会比正常发音沉闷一点，吐字基本清晰流畅，声音的强度会比正常发音小一点，但是比捏鼻子发音的声音强度大，语速也比正常发音更加快一点。

3.2.2. 语谱图分析

在语谱分析图中，可以将语音特征转化为视觉图片特征，并且可以在一定程度地表现出来。甲某和乙某的正常发音、捏鼻子发音、戴口罩发音在波形图和宽带语谱图上有较大的差异。在图中，音强的高低程度通过黄色线条表现出来，基频的高低程度则通过蓝色线条表现出来。如图 1~3 所示，是发音人甲某在三种不同状态下发音的语音样本。如图 4~6 所示，为发音人乙某在三种不同状态下发音的语音样本。

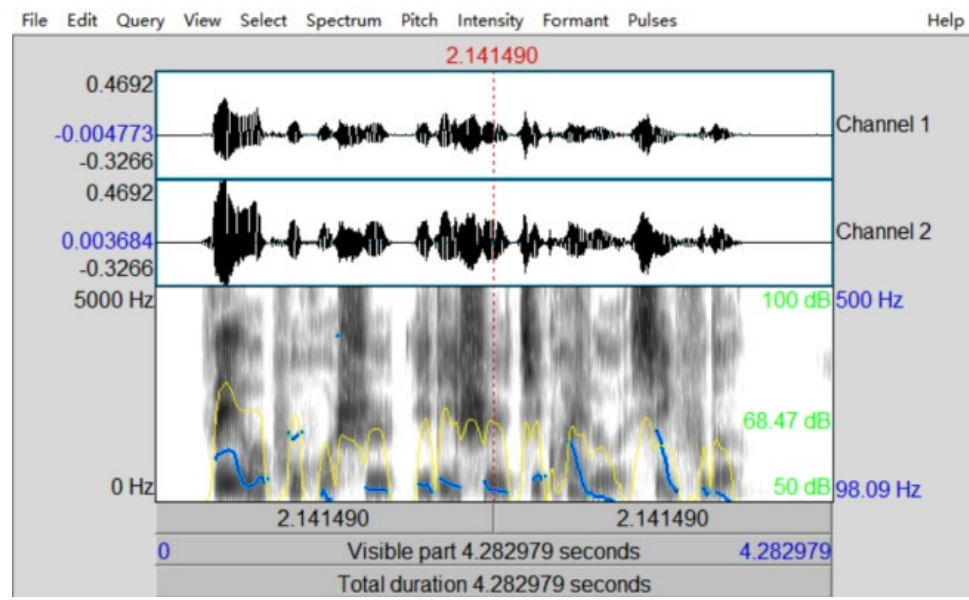


Figure 1. Pronunciation by speaker A pinching his nose the waveform diagram and the broadband speech spectrum diagram
图 1. 发音人甲某捏鼻子发音的波形图与宽带语谱图

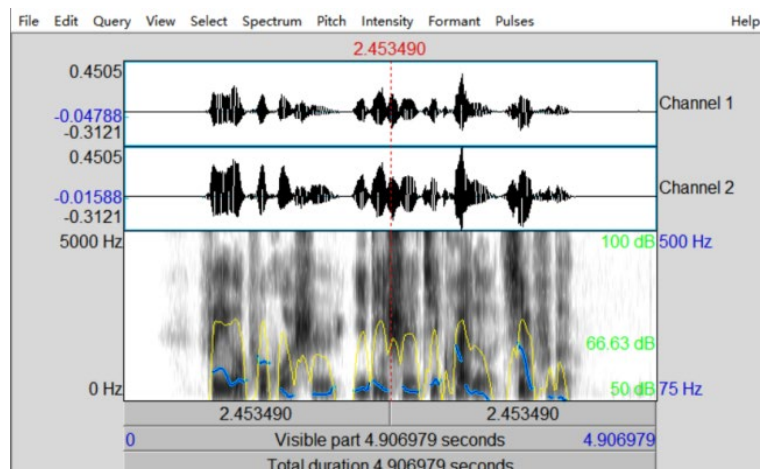


Figure 2. Pronunciation of speaker A wearing a mask the waveform diagram and the broadband speech spectrum diagram
图 2. 发音人甲某戴口罩发音的波形图与宽带语谱图

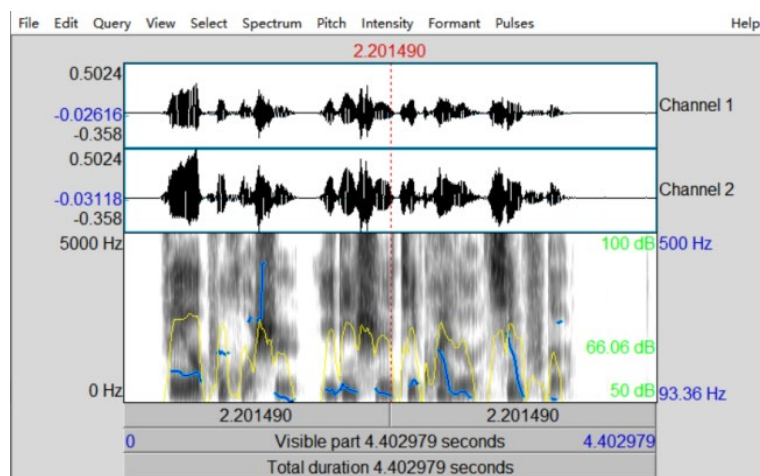


Figure 3. Waveform and broadband speech spectrum of normal pronunciation of speaker A
图 3. 发音人甲某正常发音的波形图与宽带语谱图

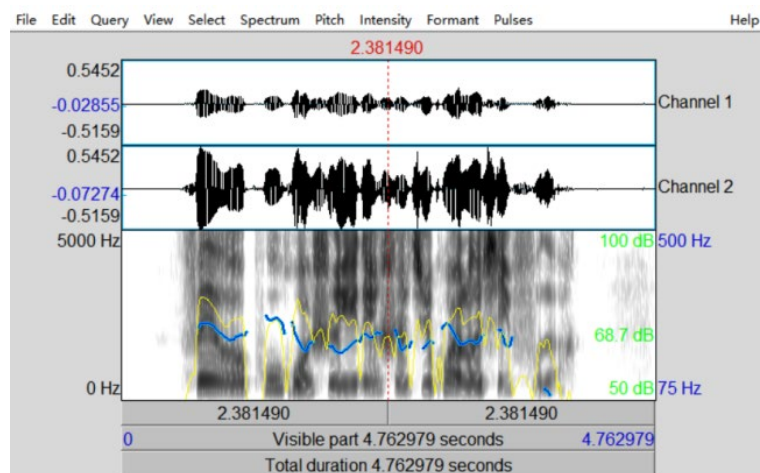


Figure 4. Pronunciation by speaker B pinching his nose the waveform diagram and the broadband speech spectrum diagram
图 4. 发音人乙某捏鼻子发音的波形图与宽带语谱图

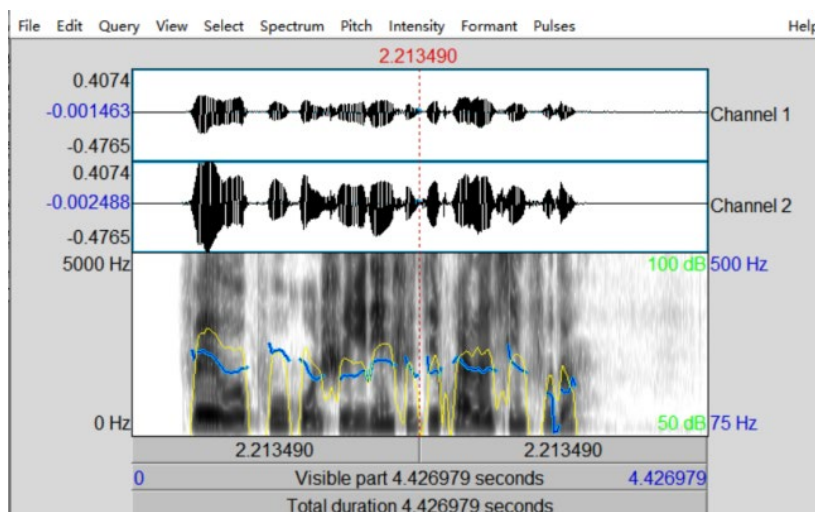


Figure 5. Pronunciation of speaker B wearing a mask the waveform diagram and the broadband speech spectrum diagram
图 5. 发音人乙某戴口罩发音的波形图与宽带语谱图

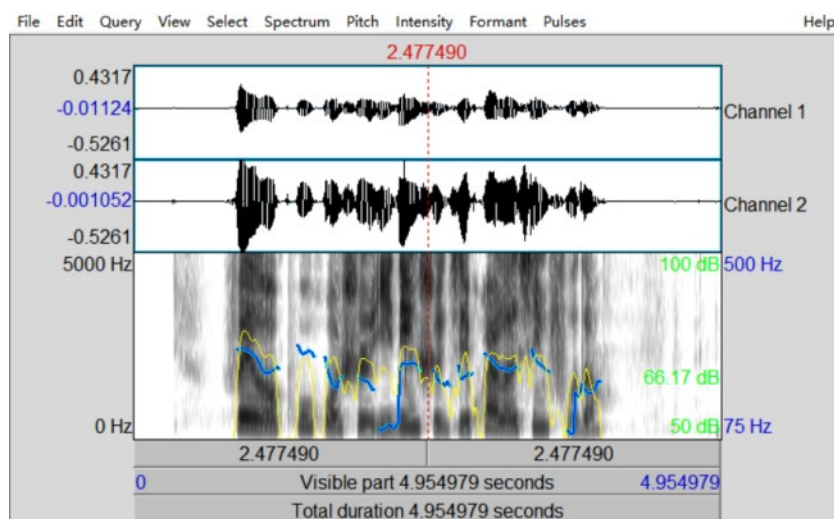


Figure 6. Waveform and broadband speech spectrum of normal pronunciation of speaker B
图 6. 发音人乙某正常发音的波形图与宽带语谱图

如图 1~3 的波形图所示, 正常发音、捏鼻子发音和戴口罩发音这三种不同状态下发音的语音样本, 它们的波形图之间存在着较为明显的差异点。在捏鼻子状态下发音的变异语音波形图竖线的密集程度较高, 在正常状态下发音的语音波形图竖线的密集程度较低, 而在戴口罩的状态下发音的变异语音波形图竖线密集程度则处在捏鼻发音与正常发音这二者之间。

如图 4~6 的宽带语谱图中音强线所示, 正常发音、捏鼻子发音和戴口罩发音这三种不同状态下发音的语音样本, 它们的音强走势以及大小之间存在着较为明显的差异点。

如图 1~3 的宽带语谱图中基频线所示, 正常发音、捏鼻子发音和戴口罩发音这三种不同状态下发音的语音样本, 它们的基频走势以及大小之间的差异则比较小, 不存在明显的差异点。

如图 4~6 的宽带语谱图中的共振峰所示, 正常发音与在戴口罩状态下发音的语音样本, 它们的共振峰都表现出向上的单向趋势, 此二者均与在捏鼻子状态下发音的语音样本的共振峰走势有着显著的差异, 捏鼻子发音的语音样本, 它的共振峰表现出向下的单向趋势。

3.2.3. 定量分析

选择语料文本中的“是极快乐幸福的”作为代表句音节，然后对发音人甲某与乙某的正常发音以及其他两种状态下发音的变异语音进行定量分析和比对。运用专业的语音分析软件 Praat，对整个代表句的七个音节进行切分并加以标注，各个音节都用其汉语拼音表示，从而提取出各个不同音节的时长、音强、基频和共振峰等不同参数，再使用均值、标准差、成对 T 检验方法进行统计与分析。

1) 音节时长

音节时长指的是音节发音持续时间的长短，它是声学差异分析其中的一个非常重要的参数。为了对在正常状态下发音以及在两种特殊状态下发音的变异语音的音节时长是否存在明显的差异进行评估与判断，所以对发音人甲某与乙某的正常发音以及两种特殊状态下发音的变异语音的音节时长数据进行了测量与统计，并且通过计算得到了其均值与标准差数据结果，具体数据情况如表 2、表 3 所示。

Table 2. Comparison of syllable duration of three pronunciation modes of speaker A (ms)

表 2. 发音人甲某三种发音方式语音的音节时长比对表(ms)

音节	正常发音	捏鼻子发音	戴口罩发音	均值	标准差
shi	159.5	143.6	145.6	149.6	7.1
ji	130.2	170.5	164.4	155.0	17.7
kuai	129.6	215.8	198.2	181.2	37.2
le	88.3	235.4	149.5	157.7	60.3
xing	213.2	206.3	173.3	197.6	17.4
fu	129.0	167.9	179.7	158.9	21.7
de	172.1	243.0	244.9	220.0	33.9
均值	146.0	197.5	179.4		21.3
标准差	36.9	34.7	31.5	24.4	

Table 3. Comparison of syllable duration of three pronunciation modes of speaker B (ms)

表 3. 发音人乙某三种发音方式语音的音节时长比对表(ms)

音节	正常发音	捏鼻子发音	戴口罩发音	均值	标准差
shi	137.2	182.4	175.2	164.9	19.8
ji	141.1	202.5	163.6	169.1	25.4
kuai	163.5	298.9	271.4	244.6	58.4
le	254.1	338.7	283.7	292.2	35.1
xing	141.8	253.3	163.3	186.1	48.3
fu	267.6	204.2	267.0	246.3	29.7
de	210.5	231.5	257.3	233.1	19.1
均值	188.0	244.5	225.9		23.5
标准差	51.7	52.5	51.3	43.8	

发音人在不同状态下的发音也会对音节的时长有影响,从表 2、表 3 中的数据里可以看来:在上述三种不同状态下的发音当中,同一个发音人在三种不同状态下的发音方式的同一音节发音时长并不相同,区别也大小不一。

从表 2、表 3 中所标明的每个音节时长的标准差里可以分析得出,标准差越大,各个发音方式之间的音节时长的差值越大,其中受发音方式影响最大的是表 2 中的“le”,它的时长标准差是 60.3 ms;受发音方式影响最小的是表 2 中的“shi”,它的时长标准差是 7.1 ms。从表 2、表 3 中的数据里可以分析出捏鼻子发音和戴口罩发音这两种特殊状态下的变异语音相对于正常发音而言,在音节时长上的表现情况存在着一定程度上的差异。

为了对在捏鼻子发音、戴口罩发音的状态下单个音节的时长与在正常发音的状态下单个音节的时长是否存在显著的差异进行分析判断,考虑到样本量较小,下面将运用成对比较 T 检验的方法对代表句“是极快乐幸福的”的这七个音节的时长进行统计与分析,得出具体的 P 值如下表 4、表 5 所示。

Table 4. Analysis results of t-test on syllable duration pairing of speaker A
表 4. 发音人甲某的音节时长配对 t 检验分析结果

配对编号	项	平均值	标准差	平均值差值	P
1	正常发音	146.0	36.9	51.5	0.026
	捏鼻子发音	197.5	34.7		
2	正常发音	146.0	36.9	33.4	0.045
	戴口罩发音	179.4	31.5		

置信空间 95%, $\alpha = 0.05$, $P < 0.05$ 。

Table 5. Analysis results of t-test on syllable duration pairing of speaker B
表 5. 发音人乙某的音节时长配对 t 检验分析结果

配对编号	项	平均值	标准差	平均值差值	P
1	正常发音	188.0	51.7	56.5	0.031
	捏鼻子发音	244.5	52.5		
2	正常发音	188.0	51.7	38.0	0.013
	戴口罩发音	225.9	51.3		

置信空间 95%, $\alpha = 0.05$, $P < 0.05$ 。

观察成对比较 T 检验的分析结果可以得到,若置信空间为 95%,则 $\alpha = 0.05$ 的时候,由于发音人甲某与乙某发音的单个音节的时长成对比较 T 检验显著概率 P 值均小于 0.05,这便可以说明,满足置信空间 95%,即正常发音和两种特殊状态下发音的音节时长存在着显著的差异。

2) 平均音强

音强就是语音的能量,指的是在单位时间内通过垂直于声波传播方向的单位面积的平均声能,又被称为声强。为了对在正常状态下发音以及在两种特殊状态下发音的变异语音的平均音强水平是否存在明显的差异进行评估与判断,所以对发音人甲某与乙某的正常发音以及两种特殊状态下发音的变异语音的各个音节的平均音强数据进行了测量与统计,并且通过计算得到了其均值与标准差数据结果,具体数据如表 6、表 7 所示。

Table 6. Average sound intensity comparison of three pronunciation modes of speaker A (dB)
表 6. 发音人甲某三种发音方式语音的平均音强对比表(dB)

音节	正常发音	捏鼻子发音	戴口罩发音	均值	标准差
shi	82.8	82.3	82.6	82.6	0.2
ji	79.5	78.6	79.6	79.2	0.4
kuai	82.3	76.9	77.3	78.8	2.5
le	83.5	82.0	83.9	83.1	0.8
xing	79.2	78.1	79.0	78.8	0.5
fu	80.6	74.5	76.4	77.2	2.5
de	81.9	73.3	76.3	77.2	3.6
均值	81.4	78.0	79.3		1.4
标准差	1.5	3.2	2.8	2.2	

Table 7. Average sound intensity comparison of three pronunciation modes of speaker B (dB)
表 7. 发音人乙某三种发音方式语音的平均音强对比表(dB)

音节	正常发音	捏鼻子发音	戴口罩发音	均值	标准差
shi	80.5	79.6	81.7	80.6	0.9
ji	81.2	78.6	80.3	80.0	1.1
kuai	78.5	77.4	81.7	79.2	1.8
le	81.2	82.2	81.8	81.7	0.4
xing	78.9	77.6	80.3	78.9	1.1
fu	79.7	75.4	80.0	78.4	2.1
de	81.4	72.0	82.1	78.5	4.6
均值	80.2	77.5	81.1		1.5
标准差	1.1	3.0	0.8	1.1	

发音人在不同状态下的发音也会对音节的平均音强有影响，从表 6、表 7 中的数据里可以看出，在整个代表句中各个音节的平均音强水平都有所不同，差异程度在各个音节的音强标准差大小上有所反应。

发音人甲某在正常状态下发音的语音样本，其单个音节的平均音强的标准差为 1.5 dB；在捏鼻子的状态下发音的语音样本，其单个音节的平均音强的标准差为 3.2 dB；在戴口罩的状态下发音的语音样本，其单个音节的平均音强的标准差为 2.8 dB。发音人乙某在正常状态下发音的语音样本，其单个音节的平均音强的标准差为 1.1 dB；在捏鼻子的状态下发音的语音样本，其单个音节的平均音强的标准差为 3.0 dB；在戴口罩的状态下发音的语音样本，其单个音节的平均音强的标准差为 0.8 dB。各个音节的音强都存在着一定程度上的差异。

为了对在捏鼻子发音、戴口罩发音的状态下单个音节的平均音强与在正常发音的状态下单个音节的平均音强是否存在显著的差异进行分析判断，考虑到样本量较小，下面将运用成对比较 T 检验的方法对代表句“是极快乐幸福的”的这七个音节的平均音强进行统计与分析，得出具体的 P 值如下表 8、表 9 所示。

Table 8. Analysis results of t-test on average sound intensity pairing of speaker A
表 8. 发音人甲某的平均音强配对 t 检验分析结果

配对编号	项	平均值	标准差	平均值差值	P
1	正常发音	81.4	1.5	3.4	0.015
	捏鼻子发音	78.0	3.2		
2	正常发音	81.4	1.5	2.1	0.042
	戴口罩发音	79.3	2.8		

置信空间 95%, $\alpha = 0.05$, $P < 0.05$ 。

Table 9. Analysis results of t-test on average sound intensity pairing of speaker B
表 9. 发音人乙某的平均音强配对 t 检验分析结果

配对编号	项	平均值	标准差	平均值差值	P
1	正常发音	80.2	1.1	2.7	0.042
	捏鼻子发音	77.5	3.0		
2	正常发音	80.2	1.1	0.9	0.048
	戴口罩发音	81.1	0.8		

置信空间 95%, $\alpha = 0.05$, $P < 0.05$ 。

观察成对比较 T 检验的分析结果可以得到, 若置信空间为 95%, 则 $\alpha = 0.05$ 的时候, 由于发音人甲某与乙某发音的单个音节的平均音强成对比较 T 检验显著概率 P 值均小于 0.05, 这便可以说明, 满足置信空间 95%, 即正常发音和两种特殊状态下发音的单个音节的平均音强存在着显著的差异。

3) 平均基频

基频指的是人在发音时声带振动的基本频率, 它也是声学差异分析中的一个至关重要的参数。在语音学中, 基频又被称为音高, 是决定汉语音节语义的重要声学参数。

为了对在正常状态下发音以及在两种特殊状态下发音的变异语音的平均基频是否存在明显的差异进行评估与判断, 所以对发音人甲某与乙某的正常发音以及两种特殊状态下发音的变异语音的各个音节的平均基频数据进行了测量与统计, 并且通过计算得到了其均值与标准差数据结果, 具体数据如表 10、表 11 所示。

Table 10. Average fundamental frequency comparison table of three pronunciation modes of speaker A (Hz)
表 10. 发音人甲某三种发音方式语音的平均基频对比表(Hz)

音节	正常发音	捏鼻子发音	戴口罩发音	均值	标准差
shi	102.5	94.5	105.6	100.9	4.7
ji	103.4	110.5	110.4	108.1	3.3
kuai	111.7	99.6	119.7	110.3	8.3
le	119.6	116.7	99.2	111.8	9.0
xing	105.0	101.4	89.5	98.6	6.6
fu	93.9	91.1	166.0	117.0	34.7
de	169.1	177.0	113.8	153.3	28.1
均值	115.0	113.0	114.9		0.9
标准差	23.3	27.4	22.8	17.0	

Table 11. Average fundamental frequency comparison table of three pronunciation modes of speaker B (Hz)
表 11. 发音人乙某三种发音方式语音的平均基频对比表(Hz)

音节	正常发音	捏鼻子发音	戴口罩发音	均值	标准差
shi	96.7	97.9	112.5	102.4	7.2
ji	105.3	103.7	111.6	106.9	3.4
kuai	176.4	184.8	103.4	154.9	36.6
le	172.5	105.5	107.0	128.3	31.2
xing	112.0	103.8	116.7	110.8	5.3
fu	92.1	95.1	98.6	95.3	2.7
de	134.8	128.7	143.8	135.8	6.2
均值	127.1	117.1	113.4		5.8
标准差	32.5	29.4	13.6	19.7	

对表 10、表 11 中的数据结果进行分析, 不同音节的平均基频有所不同。这与音节的声调有关, 因为汉语是声调语言, 而声调的声学参数就是基频。基频的高低与声调相对应, 音节的声调不同, 其基频的高低也就有所不同。但是各个音节基频的均值与整个句子的基频水平是基本一致的。所以整个句子的基频水平可以代表各个音节基频的平均水平, 并且三种发音方式所得出的全句基频差别不大。

为了对在捏鼻子发音、戴口罩发音的状态下单个音节的平均基频与在正常发音的状态下单个音节的平均基频是否存在显著的差异进行分析判断, 考虑到样本量较小, 下面将运用成对比较 T 检验的方法对代表句“是极快乐幸福的”的这七个音节的平均基频进行统计与分析, 得出具体的 P 值如下表 12、表 13 所示。

Table 12. Analysis results of t-test on average fundamental frequency pairing of speaker A
表 12. 发音人甲某的平均基频配对 t 检验分析结果

配对编号	项	平均值	标准差	平均值差值	P
1	正常发音	115.0	23.3	2.1	0.243
	捏鼻子发音	113.0	27.4		
2	正常发音	115.0	23.3	0.1	0.496
	戴口罩发音	114.9	22.8		

置信空间 95%, $\alpha = 0.05$, $P < 0.05$ 。

Table 13. Analysis results of t-test on average fundamental frequency pairing of speaker B
表 13. 发音人乙某的平均基频配对 t 检验分析结果

配对编号	项	平均值	标准差	平均值差值	P
1	正常发音	127.1	32.5	10.0	0.171
	捏鼻子发音	117.1	29.4		
2	正常发音	127.1	32.5	13.7	0.189
	戴口罩发音	113.4	13.6		

置信空间 95%, $\alpha = 0.05$, $P < 0.05$ 。

观察成对比较 T 检验的分析结果可以得到, 若置信空间为 95%, 则 $\alpha = 0.05$ 的时候, 由于发音人甲某与乙某发音的单个音节的平均基频成对比较 T 检验显著概率 P 值均大于 0.05, 这便可以说明, 无法满足置信空间 95%, 即正常发音和两种特殊状态下发音的单个音节的平均基频不存在着显著的差异。

4) 共振峰参数

共振峰是指在声音频谱中能量相对集中的一些区域, 共振峰不仅决定音质, 而且反映着声道的物理特性。共振峰的频率、带宽和强度是传统识别方法里最重要的识别参数。元音共振峰按照其功能, 可以分为两个档次: 初级共振峰 F1、F2, 是元音语言学信息的主要携带者; F3 及 F3 以上的高级共振峰, 是话者个体特征的主要携带者。

分析表 14、表 15 中数据可知, 同一发音人在三种不同状态下的发音对 F1、F2、F3、F4、F5 这五个共振峰参数的影响各不相同。在正常状态下发音与在捏鼻子的状态下发音, 二者产生的语音之间的 F1、F2、F3、F4、F5 这五个共振峰参数的差值明显要比在正常状态下发音与在戴口罩的状态下发音, 二者产生的语音之间的 F1、F2、F3、F4、F5 这五个共振峰参数的差值更大。以此可以判断, 在捏鼻子的状态下发音对共振峰的影响会比在戴口罩的状态下发音对共振峰的影响更加严重。

Table 14. Resonance peak parameters (Hz) of three pronunciation modes of speaker A

表 14. 发音人甲某三种发音方式语音的共振峰参数(Hz)

共振峰	正常发音	捏鼻子发音	戴口罩发音	均值	标准差
第一共振峰(F1)	547.0	664.3	638.5	616.6	50.3
第二共振峰(F2)	1739.4	1761.8	1676.8	1726.0	36.0
第三共振峰(F3)	2711.8	2855.5	2781.9	2783.1	58.7
第四共振峰(F4)	4041.4	4119.2	4053.5	4071.4	34.2
第五共振峰(F5)	4502.6	4621.8	4573.4	4565.9	48.9
均值	2708.5	2804.5	2744.8		39.6
标准差	1456.5	1463.0	1458.4	1459.1	

Table 15. Resonance peak parameters (Hz) of three pronunciation modes of speaker B

表 15. 发音人乙某三种发音方式语音的共振峰参数(Hz)

共振峰	正常发音	捏鼻子发音	戴口罩发音	均值	标准差
第一共振峰(F1)	689.5	845.6	737.1	757.4	65.4
第二共振峰(F2)	1697.9	1791.5	1692.2	1727.2	45.5
第三共振峰(F3)	2925.2	2977.5	2926.5	2943.1	24.4
第四共振峰(F4)	4017.9	4156.7	3985.4	4053.4	74.3
第五共振峰(F5)	4617.4	4703.3	4593.7	4638.1	47.1
均值	2789.6	2894.9	2787.0		50.3
标准差	1447.1	1434.3	1422.7	1434.6	

由表 16、表 17 的成对比较 T 检验结果表明: 若置信空间为 95%, 则 $\alpha = 0.05$ 的时候, 发音人甲某正常发音 - 捏鼻子发音 $P = 0.005$ 与发音人乙某正常发音 - 捏鼻子发音 $P = 0.002$ 均小于 0.05, 而发音人甲某正常发音 - 戴口罩发音 $P = 0.132$ 与发音人乙某正常发音 - 戴口罩发音 $P = 0.431$ 均大于 0.05。

Table 16. T-test analysis results of resonance peak parameter pairing of speaker A
表 16. 发音人甲某的共振峰参数配对 t 检验分析结果

配对编号	项	平均值	标准差	平均值差值	P
1	正常发音	2708.5	1456.5	96.1	0.005
	捏鼻子发音	2804.5	1463.0		
2	正常发音	2708.5	1456.5	36.4	0.132
	戴口罩发音	2744.8	1458.4		

置信空间 95%， $\alpha = 0.05$ ， $P < 0.05$ 。

Table 17. Results of t-test analysis on the pairing of resonance peak parameters of speaker B
表 17. 发音人乙某的共振峰参数配对 t 检验分析结果

配对编号	项	平均值	标准差	平均值差值	P
1	正常发音	2789.6	1447.1	105.4	0.002
	捏鼻子发音	2894.9	1434.3		
2	正常发音	2789.6	1447.1	2.6	0.431
	戴口罩发音	2787.0	1422.7		

置信空间 95%， $\alpha = 0.05$ ， $P < 0.05$ 。

说明总体上来说在捏鼻子的状态下发音相对于正常发音在共振峰数据上存在着显著的差异；而总体上来说在戴口罩的状态下发音相对于正常发音在共振峰数据上不存在显著的差异。

4. 结语

本文实验通过对五位男性发音人在正常状态下发音，以及在捏鼻子发音、戴口罩发音这两种变异语音的声学分析与讨论，可以发现同一发音人的正常发音与捏鼻子发音、戴口罩发音这两种变异语音在音节时长、平均音强、平均基频、共振峰参数上都存在这一定程度上的声学差异，有着不同的声学特征。在这三种不同状态下发音的语音中，相互比较而言，捏鼻子发音较为容易被辨识出来的，辨识度特点较高；而正常发音与戴口罩发音这二者的相似度较高，二者语音混杂其中，辨识难度较高，需要更精密、更细致的鉴别技术。

捏鼻子发音的变异语音与正常发音相比，在音节时长、平均音强、共振峰参数上的声学特征存在着显著的差异，但在平均基频方面的声学特征并不存在显著的差异。这则说明了发音人在捏鼻子状态下的发音会显著影响音节时长、平均音强和共振峰参数，而平均基频不会受到显著影响，仍然具有稳定性。因此，平均基频可以作为捏鼻子发音的变异语音同一认定中的一项重要指标来使用，音节时长、平均音强、共振峰参数并不能作为捏鼻子发音的变异语音同一认定的根据。

戴口罩发音的变异语音与正常发音相比，在音节时长、平均音强上的声学特征存在着显著的差异，但在平均基频、共振峰参数上的声学特征不存在显著的差异。这则说明了发音人在戴口罩状态下的发音会显著影响音节时长和平均音强，而平均基频和共振峰参数不会受到显著影响，仍然具有稳定性。因此，平均基频和共振峰参数可以作为戴口罩发音的变异语音同一认定中的一项重要指标来使用，音节时长和平均音强并不能作为戴口罩发音的变异语音同一认定的根据。

通过以上的论述，我们可以得知，只要将语音样本在一定的专业语音分析软件内进行声学分析与讨

论, 就可以利用其中有效的一些数据参数对正常发音与捏鼻子、戴口罩发音的两种变异语音进行同一认定, 并且其可以作为法庭证据来使用。这对于刑事侦查与司法鉴定的工作可以起到极其重要的作用。对于声纹鉴定工作来说, 我们还处于不断地学习及探索当中, 我们能够深刻地感受到声纹鉴定技术在司法领域所当中能够起到的作用是极其重要的。总而言之, 在有关于语音作为声音证据的司法活动过程中, 对于声纹鉴定技术的作用与价值加强重视, 能够为刑事科学的发展提供源源不断的动力。在此基础上, 也能够推动我国的声纹鉴定技术向更高水平发展研究。

参考文献

- [1] 王英利, 李敬阳, 曹洪林. 声纹鉴定技术综述[J]. 警察技术, 2012(4): 54-56.
- [2] 王英利, 李卫平. 利用鼻韵母音节前后音渡形态进行声纹鉴定的研究[J]. 中国刑警学院学报, 1995(2): 28-30.
- [3] 庄琳. 捏鼻语音的声纹鉴定研究[J]. 中国人民公安大学学报(自然科学版), 2010, 16(4): 1-7.
- [4] 王英利, 潘自勤, 蓝常山. 声纹鉴定中的音强特征研究[J]. 中国司法鉴定, 2011(5): 19-22.
- [5] 郑方, 李蓝天, 张慧, 艾斯卡尔·肉孜. 声纹识别技术及其应用现状[J]. 信息安全研究, 2016, 2(1): 44-57.
- [6] 李蓝天. 说话人识别中的特征学习方法研究[D]: [博士学位论文]. 北京: 清华大学, 2018.
- [7] 陈维娜, 曾庆发. 不同录音环境对语音特征的影响研究[J]. 中国人民公安大学学报: 自然科学版, 2018, 24(4): 18-23.
- [8] 刘世杰, 王虹. 云录音声纹检验实验研究[J]. 广东公安科技, 2019, 27(3): 15-18, 57.
- [9] 王康宁, 杨兵, 马腾. 声纹技术在网络犯罪侦查取证中的应用初探[J]. 湖南警察学院学报, 2020, 32(3): 45-51.
- [10] 杨伟, 杨俊杰. 基于语言学音系例字的口音自动识别探究[J]. 中国司法鉴定, 2021(2): 38-42.
- [11] 张明键, 张悦. 基于语谱图和深度置信网络的方言自动辨识与说话人识别[J]. 电子技术与软件工程, 2020(14): 151-154.