

## 结合 EHHT 和 CI 的精神分裂症语音自动检测算法

田维维<sup>1</sup> 周悦<sup>1</sup> 尹旺<sup>1</sup> 何凌<sup>1</sup> 邓丽华<sup>1</sup> 李元媛<sup>2</sup>

1 四川大学电气工程学院 成都 610065

2 四川大学华西医院心理卫生中心 成都 610041

(2285675739@qq.com)

**摘要** 通过对精神分裂症语音的临床特点进行研究,实验采集了 14 个精神分裂症患者的 686 个元音数据样本和 14 个与之性别、年龄、文化程度相匹配的健康对照组的 793 个元音数据样本,来建立病理语音数据库,利用结合集成希尔伯特黄变换(Ensemble Hilbert-Huang Transform, EHHT)和倒谱内插(Cepstrum Interpolation, CI)的改进共振峰提取算法,来获取反映精神分裂症语音音质情感变化的声学特征参数集,结合支持向量机(Support Vector Machine, SVM)分类器来进行分类,实现了精神分裂症患者语音和健康对照组语音的自动检测,并设计实验讨论了白噪声的次数和方差、IMF 分量个数、窗长 4 个因素对检测效果的影响,以及与经典的共振峰估算方法的比较。实验结果表明,文中提出的算法的检测正确率可以达到 98.8%,精神分裂症患者在体现音质特征的共振峰语音声学参数上与健康对照组存在显著差异,并有可能为精神分裂症的临床辅助诊断研究提供一个全新、客观、定量和高效的指标。

**关键词:** 精神分裂症语音;音质特征;共振峰;集成希尔伯特黄变换;倒谱内插

中图分类号 TP391.9

## Automatic Voice Detection Algorithm for Schizophrenic Combining EHHT and CI

TIAN Wei-wei<sup>1</sup>, ZHOU Yue<sup>1</sup>, YIN Wang<sup>1</sup>, HE Ling<sup>1</sup>, DENG Li-hua<sup>1</sup> and LI Yuan-yuan<sup>2</sup>

1 College of Electrical Engineering, Sichuan University, Chengdu 610065, China

2 Mental Health Center of West China Hospital, Sichuan University, Chengdu 610041, China

**Abstract** Through studying the clinical characteristics of schizophrenic speech, the experiment collected 686 vowel data samples from 14 schizophrenic patients and 793 vowel data samples from 14 healthy controls matched with gender, age and education level to establish a pathological voice database. Using the improved formant extraction algorithm combining Ensemble Hilbert-Huang Transform (EHHT) and Cepstrum Interpolation (CI) to obtain the acoustic feature parameter set reflecting emotion change of schizophrenic voice quality, then combined with the Support Vector Machine (SVM) classifier for classification, automatic voice detection of schizophrenic patients and the healthy controls is achieved. Besides, it designed experiments to discuss the influence of the four factors, that is, the number and variance of white noise, the number of IMF components and the window length, on the detection effect, and compared with the classical formant estimation methods. Experimental results show that the detection accuracy of the proposed algorithm can reach 98.8%. The patients with schizophrenia have a significant difference in the acoustical parameters of the formants represent the sound quality feature with the healthy controls, and it may provide a new objective, quantitative and efficient indicator for the clinical assistant diagnostic research of schizophrenia.

**Keywords** Schizophrenic voice, Sound quality feature, Formant, Ensemble hilbert-huang transform, Cepstral interpolation

## 1 引言

精神分裂症是一种病程迁延的慢性神经退化性障碍,具有复发率高、致残率高、依从性差等特点<sup>[1]</sup>,常常伴随患者终生,严重损害其生活质量和社会认知,给患者和家属造成了极大的生产力损失。研究表明,其主要临床表现包含幻听、妄想、

言语(思维)紊乱、显著的紧张性或异常行为、阴性症状(如语言贫乏、情感淡漠、动机缺乏等)五个维度中的一个或多个<sup>[2]</sup>。我国精神科医师的数量存在严重缺口,精神卫生资源分布不合理等问题,往往导致患者错失最佳治疗时机。若能在精神症状发作的早期采取合理有效的自动检测手段,及时加以干预治疗,将有助于改善患者的病况,减少其社会功能的损害。

本文已加入开放科学计划(OSID),请扫描上方二维码获取补充信息。

基金项目:成都市科技惠民技术研发项目(2015-HM01-00430-SF);国家自然科学基金青年科学基金(61503264);四川大学创新火花库项目(2082604401189);四川省科技厅项目(2019YFS0236)

This work was supported by the Chengdu Science and Technology Benefiting People Technology Research and Development Project, Sichuan Province, China (2015-HM01-00430-SF), National Natural Science Foundation of China (61503264), Sichuan University Innovation SparkBank Project, China (2082604401189) and Science and Technology Department Project of Sichuan Province, China (2019YFS0236).

通信作者:李元媛(guojipangxie@126.com)

目前临床上精神分裂症的诊断和监测评估主要依靠以下6种方式:1)临床量表,包含阴性症状量表(The Scale for the Assessment of Negative Symptoms, SANS)、阳性症状量表(The Scale for the Assessment of Positive Symptoms, SAPS)、阳性与阴性量表(Positive and Negative Syndrome Scale, PANSS)等<sup>[3]</sup>,其严重依赖于患者自述及家属供史,易受评分者的主观因素干扰,如问诊技巧、临床经验等,缺乏标准化的评估体系;2)脑影像学,功能核磁共振(Functional Nuclear Magnetic Resonance, fMRI)脑区图像主要集中于精神分裂症患者大脑的脑实质及脑功能的改变,仪器操作复杂,检测费用高昂,且缺乏客观的生物学标记<sup>[4-5]</sup>;3)脑电信号,EEG中包含丰富的生理状态信息,充分反映了大脑的活动,但 $\alpha$ 波、 $\beta$ 波均无特异性,对诊断的价值有限<sup>[6]</sup>;4)视频分析,相比正常人,精神分裂症患者具有更少的身体运动和更呆滞的面部表情<sup>[7]</sup>,可以通过分析提取受试者的表情动作特征来实现自动检测,但目前的研究瓶颈在于缺乏统一的实验范式,其尚处于起步阶段,未达到临床辅助诊断水平,还需要进一步的深入探讨;5)基因组学,全基因外显子区中存在着与精神分裂症紧密相关的变异体,但在基因测序捕捉时仍存在假阳性和假阴性的问题<sup>[8]</sup>;6)语音信号,作为一个全新的领域,研究表明<sup>[9]</sup>精神分裂症的阴性症状与语音情感表达密切相关。随着大数据、人工智能、语音信号处理算法的飞速发展,以及医学、计算机科学等多学科的交叉融合研究,利用计算机辅助分析和诊断技术提取的语音情感特征正被逐渐应用于精神分裂症的研究。

Compton等<sup>[10]</sup>通过研究发现精神分裂症患者存在慢性发音表达障碍,具有情感淡漠扁平化、旋律范围狭窄、音质异常等显著阴性症状。Gold等<sup>[11]</sup>研究了与情绪感知相关的刺激对92名患者和73名对照受试者的影响,结果显示精神分裂症患者的语音情感识别能力存在明显缺陷,且与潜在的声学特征的损伤有关,同时也对社会认知功能造成了显著影响。Xu<sup>[12]</sup>等利用自然语言处理技术自动提取50名精神分裂症患者和25名年龄匹配的健康对照组在访谈记录中的词汇特征和文档向量,并使用集成机器学习算法将患者和对照组分类,最高准确率为78.7%。Cohen等<sup>[13]</sup>提取了4个基本声学特征——基音频率(F0)、第一共振峰(F1)、第二共振峰(F2)和强度来量化精神分裂症患者的语音信号,发现激活和怀有敌意的患者的舌头的运动更为明显,而F1和F2的形成分别与舌头上和前后方向的运动有关。Zhang等<sup>[14]</sup>通过评估26名患者(入组时和入组一周后)和30名健康对照样本(入组时)的10项语音学参数发现,患者组线性预测系数评分明显高于对照组,但Mel频率倒谱系数评分明显低于对照组;且在与17项临床特征构成的170个相关系数中,有10项达到了显著性水平。上述研究表明,语音声学特征参数与精神病理学症状之间存在着潜在的关系,还需要对更加复杂多样的患者进行随机深入的研究。目前国内关于精神分裂症的语音学研究进展缓慢,具有广阔的理论研究价值和临床应用价值。

本文结合集成希尔伯特黄变换(EHHT)和倒谱内插(CI)算法,提取了反映语音情感音质特征变化的共振峰参数,并利用SVM分类器,自动检测精神分裂症语音和正常人语音,从而有利于早期识别并追踪精神分裂症患者的情感变化,可以为辅助指导精神科医师对患者病情的诊断和治疗效果的监测提供新的科学依据,能在一定程度上帮助患者恢复社会功能与提高生活质量。

## 2 精神分裂症语音自动检测系统

本文通过分析精神分裂症患者的临床语音特点,提出了结合集成希尔伯特黄变换和倒谱内插的精神分裂症语音自动检测系统。该系统的流程图如图1所示。

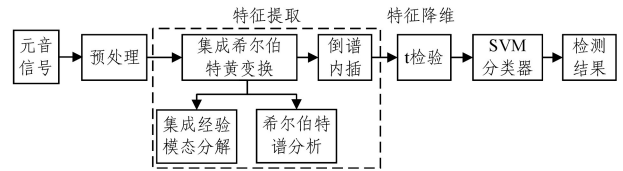


图1 精神分裂症语音自动检测系统的流程图

Fig.1 Flow chart of automatic voice detection system for schizophrenic

精神分裂症语音自动检测算法包含以下4个部分:1)语音信号预处理,对输入的元音信号进行去直流、归一化和预加重处理,提升语音信号的高频分量,避免口齿辐射对提取共振峰分量的干扰;2)特征提取,通过对语音信号多次加入随机高斯白噪声来进行经验模态分解(Empirical Mode Decomposition, EMD),对多次分解结果求均值得到最终的固有模式函数(Intrinsic Mode Function, IMF),计算各IMF分量对应的Hilbert边缘谱的归一化频带能熵比的比值,筛选出包含共振峰的IMF分量来重构语音信号,分帧加窗处理后,再利用倒谱内插法提取共振峰的81个特征参数,即前3个共振峰的频率、带宽、幅值所对应的峰值个数、均值、方差、四分位差、中位数、众数、极差、偏度和峭度;3)特征降维,利用t检验提取具有显著差异的参数,用于分类检测;4)模式识别,结合提取的共振峰参数向量集和精神科医师校验确认的标签类别,利用SVM分类器实现精神分裂症患者语音和健康对照组语音的自动分类检测。

### 2.1 共振峰

精神分裂症患者因脑区存在缺陷,导致其语言认知功能损伤和情感表达出现异常,常常表现出情感淡漠、言语贫乏等阴性症状,易在言语交流中削弱或抑制情感表达的范围和强度,研究表明<sup>[15-16]</sup>语音情感的表达和声音质量的变化密切相关,因此可以通过声音质量的异常来区分精神分裂症患者和正常人的语音。音质作为一种主观的语音评估指标,可用于衡量语音是否清晰、纯净、容易区分等。影响音质的声学行为包括哽咽、颤音、喘气等<sup>[17]</sup>,精神分裂症患者语音音质的不稳定可能成为监测和追踪精神病学临床症状的重要指标。本文基于精神分裂症患者的计算语音学分析,提取的共振峰参数代表着音质特征的变化。

语音是由肺部呼出的气流经从喉头到嘴唇的各个器官相互作用而产生的,喉头(包括声带的部分)叫作声门,从声门到嘴唇的通路称为声道,如图2所示。声道的形状主要由舌头、腭、嘴唇的位置来决定,随声道形状的改变而发出不同性质的语音<sup>[18]</sup>。将声道看作一个谐振腔,共振峰指语音频谱中能量的相对集中区域,本质表示声腔的共振频率,是反映声道谐振特性的重要物理特征,代表了发音信息最直接的来源,承载着语音信号中语气语调的理论。声门准周期脉冲激励经过声腔的调制,谐波振幅形成起伏变化的频谱包络曲线,其突出峰值位置的频率值与声腔共振频率一致,这些共振峰值与共振腔的材料、大小和形状密切相关<sup>[19]</sup>。将语音信号进行谱分解形

成的共振峰可以识别语音形成过程中的许多运动,尤其是元音,因此共振峰是语音信号处理中重要的声学特征参数之一,已被广泛应用于语音编码、语音合成与语音识别等领域。

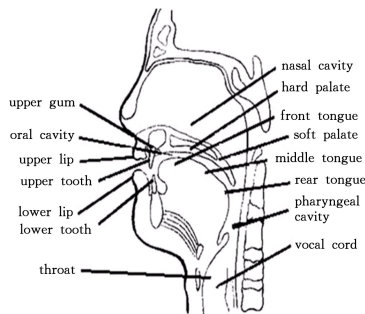
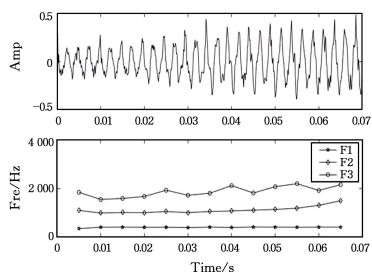


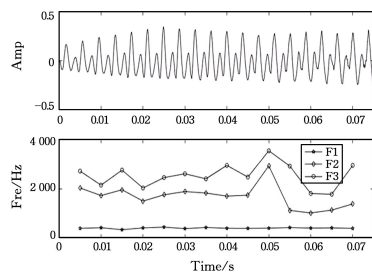
图2 声道结构示意图

Fig. 2 Structure diagram of vocal tract

研究学者发现,精神分裂症患者语音的共振峰特性与正常人有所不同。Compton 等<sup>[10]</sup>和 Delattre 等<sup>[20]</sup>通过设计实验范式,让被测者进行即兴演讲和自由朗读散文节选,发现第一共振峰 F1(表示口腔通道开口度)、第二共振峰 F2(表示舌头后缩和唇缘圆角)和第三共振峰 F3(表示软腭高度)与不同元音感知发音清晰度的差异有关,而具有失语症的患者相比控制组的 F2 标准差更小。Covington 等<sup>[21]</sup>使用 WaveSurfer 1.8.5 从 25 名首发精神分裂症患者语音样本中提取共振峰特征,并利用皮尔森相关系数检验特征参数与阴性症状量表评分之间的关系,结果表明反映舌头前后运动减少的测量参数 F2 的标准偏差与阴性症状的严重程度在统计学上具有显著相关性( $r = -0.446, p = 0.03$ )。Bernardini 等<sup>[22]</sup>研究发现,因语言的音韵学不同而导致意大利语的 F1 和 F2 的标准偏差与阴性症状严重程度之间的相关性比英语更为显著。图 3(a)、图 3(b)分别展示了精神分裂症患者发元音/u/和正常人发元音/u/时的时域波形及对应的语谱图,从图中可以发现,精神分裂症患者和正常人元音信号共振峰频率的分布和波动范围均有明显的差异和复杂的时变性质,这主要是由发音时声道结构的改变所造成的。



(a) 精神分裂症患者发出的元音/u/



(b) 健康对照组发出的元音/u/

图3 时域波形及语谱图

Fig. 3 Time-domain waveform and spectrogram

为了研究共振峰特性在精神分裂症患者语音与健康对照组语音之间的区别,本文提出了一种基于倒谱内插的共振峰提取算法,相比传统的倒谱法,其更加逼近于语音信号的频谱包络。但由于噪声和高频谐波等的干扰,实际获取的谱包络峰值仍可能偏离共振峰的位置。为了进一步提升在干扰环境下提取共振峰的精度,本文采用了集成希尔伯特黄变换和倒谱内插相结合的新算法。

## 2.2 集成希尔伯特黄变换

基于精神分裂症患者和正常人语音信号的非线性非平稳模型——声道的调频-调幅模型,本文提出了集成希尔伯特黄变换的重构语音信号算法,包含集成经验模态分解(Ensemble Empirical Mode Decomposition, EEMD)和希尔伯特谱分析(Hilbert Spectrum Analysis, HSA)两部分。通过 EEMD 提取语音信号固有的一簇 IMF 分量,对每个 IMF 分量进行希尔伯特变换导出时变的幅值和频率,得到语音信号关于幅值的时间-频率二维希尔伯特谱,根据归一化频带能熵比的比值筛选出反映共振峰结构特征的 IMF 分量重构语音信号。

### 2.2.1 集成经验模态分解

EMD 的本质是一组具有自适应性的带通滤波器,其滤波器的组数、中心频率和带宽都取决于信号自身的特点。设定语音信号为  $x(t)$ ,经过以数据为驱动的 EMD 算法筛选后被分解成  $N$  个具有不同频率尺度的固有模态分量  $c_k(t), k = 1, 2, \dots, N$  和一个剩余分量  $s_N(t)$ ,即:

$$x(t) = \sum_{k=1}^N c_k(t) + s_N(t) \quad (1)$$

IMF 分量必须满足以下两个条件:1)极值点(包括极大值点和极小值点)的个数与过零点的个数相等或最多相差一个;2)在任意时间点上,由局部极大值点和局部极小值点确定的上、下包络线的均值为零。 $s_N(t)$  则是一个非振荡的单调序列,反映了信号中的缓慢变化<sup>[23]</sup>。同时 EMD 具有多分辨率的特点,不同的 IMF 分量反映了信号在不同时间尺度上振动模式特征的演变过程,在频域上表现为从高频到低频的逐层过滤,有效突出了信号的细节变化和局部特性,有助于快速有效地掌握语音信号的动态特征<sup>[24]</sup>。

理论上,每个 IMF 分量只包含一个频率成分,但在实际分解过程中,信号局部极值点的分布情况常常因间歇性小幅度高频干扰的引入而在极短的时间内发生多次跳变,从而产生模态混叠现象,导致一个 IMF 分量中同时包含显著不同的特征时间尺度,或相近的特征时间尺度分布于不同的 IMF 中。针对这个问题,本文采用一种噪声协助的时频域分析方法——集成经验模态分解<sup>[25-26]</sup>,EEMD 的筛选过程如下:

(1)在原始信号中加入均值为 0、方差为  $std$  的具有相同分布属性的不同高斯白噪声,合成目标信号  $y(t)$ ,令  $k=1$ ;

(2)提取  $y(t)$  的所有极大值点和极小值点,分别采用三次样条插值得到上、下包络;

(3)求出上、下包络的均值  $m_k(t)$ ,则  $h_k(t) = y(t) - m_k(t)$ ,将  $h_k(t)$  视为  $y(t)$  返回步骤(2)直到  $h_k(t)$  满足 IMF 分量的限制条件,则第  $k$  个 IMF 分量  $c_k(t) = h_k(t)$ ;

(4)令  $s_k(t) = y(t) - c_k(t), k = k+1$ ,并将  $r_k(t)$  视为  $y(t)$ ,返回步骤(2)直到筛选次数达到  $\log_2(L) - 1$ ,其中  $L$  表示信号长度,固定筛选次数可以减弱局部扰动和新增数据的影响;

(5)循环上述步骤  $M$  次;

(6)对上述结果中对应的 IMF 分量求均值,得到最终的分解结果。

数,  $H_k$  表示信号第  $k$  个 IMF 分量的 Hilbert 边际谱熵。

对应的能熵比表示为:

$$EEF_k = \sqrt{1 + |E_k/H_k|} \quad (12)$$

近年来,随着语音信号分析和合成算法的发展,实验研究表明<sup>[29]</sup>,前 3 个共振峰频率的范围一般为 200~3000 Hz。设第  $k$  个 IMF 分量的总能熵比为  $EEF_k$ , 200~3000 Hz 频带内的能熵比为  $EEF_{bb}$ ,  $k=1, 2, \dots, N$ , 将 IMF 分量在 200~3000 Hz 频带内的能熵比与整个频域上能熵比的比值  $EEF_{bb}/$

$EEF_k$  按从大到小的顺序排序,选取前  $n$  个予以保留,用于重构信号。剩下的 IMF 分量则被认为不包含共振峰分量,只是高频谐波和噪声等的干扰。包含共振峰信息的信号  $r(t)$  由  $n$  个 IMF 分量重构而成,即:

$$r(t) = \sum_{k=1}^n \overline{IMF_k}(t) \quad (13)$$

其中,  $\overline{IMF_k}(t)$  表示按归一化频带能熵比比值降序排列的 IMF 分量,且各个 IMF 分量相互独立。

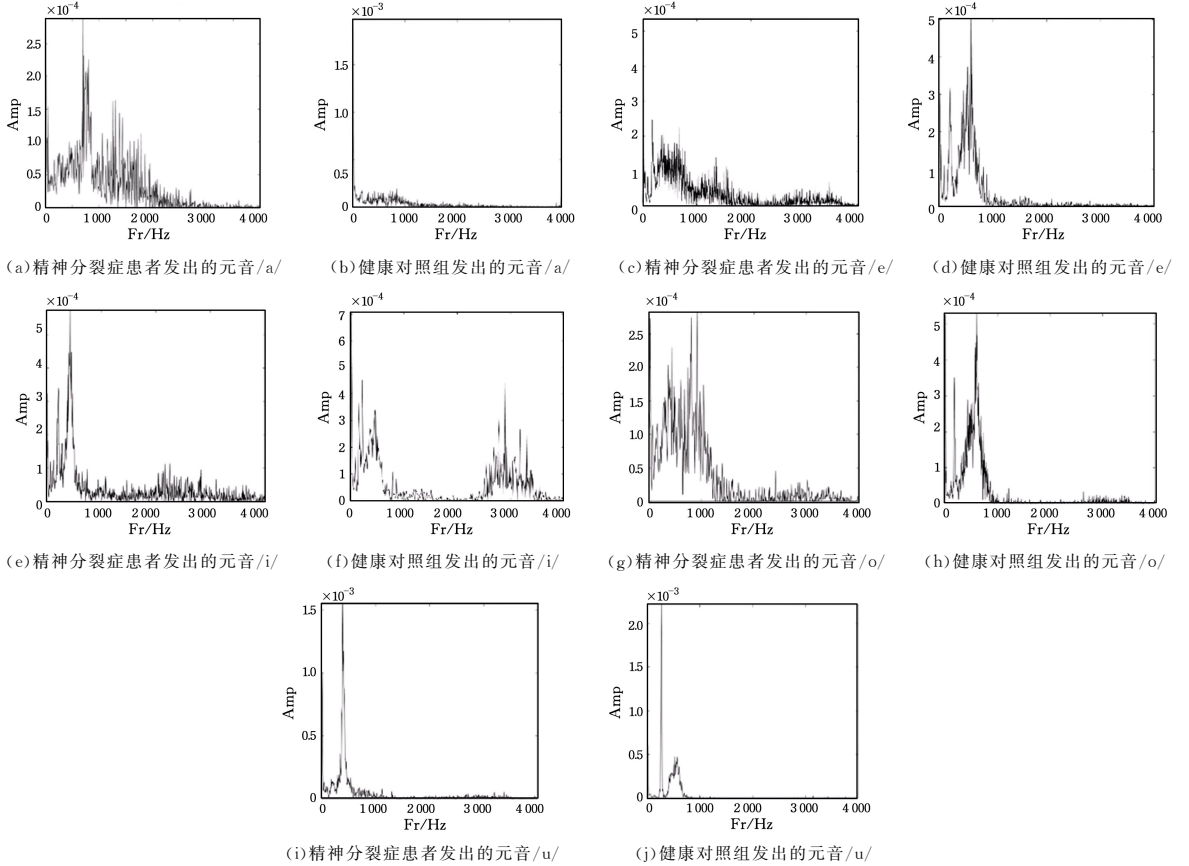


图 5 Hilbert 边际谱的分布情况

Fig. 5 Distribution of Hilbert's marginal spectrum

### 2.3 倒谱内插

重构的精神分裂症患者和正常人的语音信号不仅保留了原始语音信号中的共振峰成分,还有效避免了噪声、高频谐波等对提取谱包络的干扰,提高了非线性非平稳噪声环境下的共振峰提取精度。共振峰代表着频谱中的重要特性,与传统的倒谱法相比,对重构的语音信号采用倒谱内插算法,可以更加精确地提取共振峰相关的声学参数。

倒谱定义为分帧加窗后的语音信号  $r_t(n)$  的对数功率谱的逆傅里叶变换<sup>[18]</sup>,  $t$  表示第  $t$  帧,即:

$$\hat{c}_t(n) = IDFT(\log(|DFT(r_t(n))|^2)) \quad (14)$$

设计一个窗长为  $cepstL$  的低通窗函数  $window(n)$ , 通过截取其倒谱上的声道冲激响应序列得到:

$$\hat{h}_t(n) = \hat{c}_t(n) \times window(n) \quad (15)$$

将  $\hat{h}_t(n)$  经过傅里叶变换取指数,得到语音信号  $r_t(n)$  的功率谱包络线  $H_t(k)$ 。假设包络线上的峰值点对应于  $m\Delta f$  ( $\Delta f$  表示频率分辨率), 作为共振峰频率的初选值, 其邻近点分别表示为  $(m-1)\Delta f$  和  $(m+1)\Delta f$ , 对应于功率谱包络线上的幅值分别为  $H(m-1)$ ,  $H(m)$ ,  $H(m+1)$ 。运用抛物线内

插的方法可以得到精确的共振峰频率  $F_i$ 、幅值  $H_i$  和带宽  $B_i$ , 分别为:

$$F_i = (-b/2a + m) \times \Delta f \quad (16)$$

$$H_i = -b/4a + c \quad (17)$$

$$B_i = -\frac{\sqrt{b^2 - 4a(c - 0.5H_i)}}{a} \times \Delta f \quad (18)$$

其中,

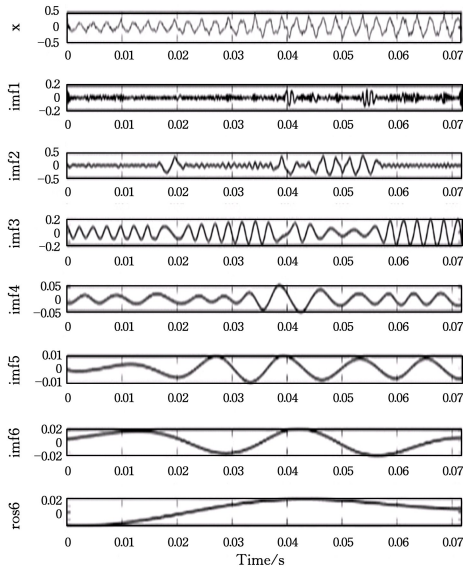
$$a = \frac{H(m+1) + H(m-1)}{2} - H(m) \quad (19)$$

$$b = \frac{H(m+1) - H(m-1)}{2} \quad (20)$$

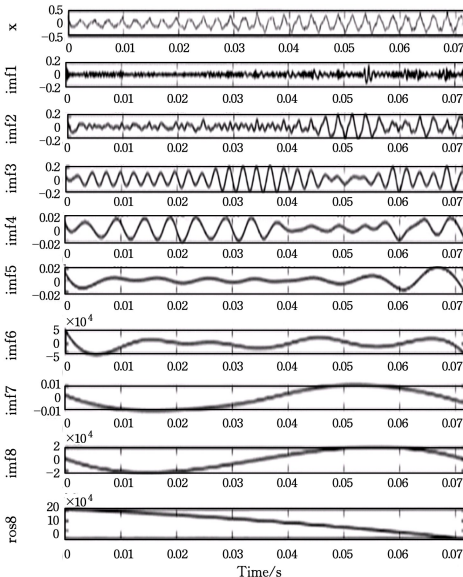
$$c = H(m) \quad (21)$$

语音实验表明<sup>[30]</sup>,对元音而言,主要信号包含于前 3 个共振峰中。图 6 为精神分裂症患者发出的元音/u/的一帧语音信号的功率谱包络线和共振峰值,本文提取了前 3 个共振峰( $F_1$ 、 $F_2$ 、 $F_3$ )频率(Fre)、带宽(Band)和幅值(Amp)的统计特征参数,包括峰数(max)、均值( $\mu$ )、方差( $\sigma^2$ )、四分位差(iqr)、中位数(med)、众数(mod)、极差(rang)、偏度( $\alpha$ )和峭度( $\beta$ )。

EEMD 在数据分析中有效地利用了有限个幅值非无穷小的高斯白噪声,使得在集成过程中遍历所有的解决方案,为 EMD 提供一个相对统一的参考尺寸,使得最终在适宜的 IMF 分量中分布不同尺度的信号,且分解对添加的噪声不敏感,在多次不基于先验主观评选标准的实验中消除引入的白噪声对真实 IMF 分量的影响,最终剩下具有实际物理含义的信号,为提取共振峰信息提供了一种独特和鲁棒性好的算法。以精神分裂症患者发出的元音/u/为例,图 4(a)给出了语音信号经过 EMD 分解得到的 6 个 IMF 分量和 1 个剩余分量,从 imf2 的波形可以明显看出,受干扰的影响,imf2 中存在两种不同频率信号的混叠,并会延续至剩余分量。图 4(b)给出了 EEMD 分解的结果,其中白噪声的方差为 0.1,次数为 100,对比图 4(a)发现,高频干扰和低频目标信号得以分离,有效避免了由 EMD 分解引入的模式混叠效应。



(a) 经验模式分解



(b) 集成经验模式分解

图 4 精神分裂症患者发出的元音/u/的分解结果

Fig. 4 Decomposition result of vowel /u/ spoken by a schi-zophrenic patient

### 2.2.2 希尔伯特谱分析

设  $d_k(t)$  是 IMF 分量  $c_k(t)$  的希尔伯特变换,得到:

$$b_k(t) = c_k(t) + jd_k(t) = a_k(t)e^{j\theta_k(t)} \quad (2)$$

再由式(3)一式(5)分别计算出具有局域性质的基本量—瞬时幅值  $a_k(t)$ 、瞬时相位  $\theta_k(t)$  和瞬时频率  $\omega_k(t)$ :

$$a_k(t) = \sqrt{c_k^2(t) + d_k^2(t)} \quad (3)$$

$$\theta_k(t) = \arctan\left(\frac{d_k(t)}{c_k(t)}\right) \quad (4)$$

$$\omega_k(t) = \frac{d\theta_k(t)}{dt} \quad (5)$$

根据 IMF 分量的瞬时幅值和瞬时频率,可将原始信号  $x(t)$  用 Hilbert 时频谱表示为:

$$x(t) = H(\omega, t) = \sum_{k=1}^N a_k(t) e^{j\omega_k(t)dt} \quad (6)$$

与式(7)中 Fourier 谱中固定的幅值和频率相比, Hilbert 时频谱具有时变的幅值和频率,可被看作是 Fourier 谱的推广,更具有普遍意义。

$$x(t) = \sum_{k=1}^{\infty} a_k e^{j\omega_k t} \quad (7)$$

谱是对传统时域及空间中难以直接发现的分布规律的扩充,通过谱能够定位到精神分裂症患者语音信号与正常人语音信号在不同谱空间的特征表现差异。Hilbert 时频谱表示信号的幅值随时间和频率变化的整体分布规律,这种时频联合分布谱在每个时间频率点上的加权值表示信号的局部幅值<sup>[27]</sup>。基于 HSA 的时频分析方法,将时变的非线性非平稳信号中频率和时间之间的关系直观地转换成二维数字图像来进行分析处理,特征描述更加完整、准确。进而,通过 Hilbert 时频谱对时间求积分得到 Hilbert 边际谱<sup>[28]</sup>,式(8)表示信号进行 EEMD 分解后的某一个 IMF 分量的局部 Hilbert 边际谱。

$$h_k(\omega) = \int_0^T H_k(\omega, t) dt \quad (8)$$

其中,  $T$  表示信号的采样时间,  $k=1, 2, \dots, N$ 。

在傅里叶分析中,信号在某一频率处的幅值表示一个同频率的正弦或余弦波在整个信号的时间长度上均存在,它延伸至整个时间轴。而边际谱中信号在某一频率处的幅值表示在整个时间轴上,可能有这个频率的振动波在局部出现过。图 5 给出了精神分裂症患者和健康对照组发出的/a/, /e/, /i/, /o/, /u/ 5 个元音的 Hilbert 边际谱分布情况,发现相比健康对照组,患者的频率分布范围更广,幅值略低一些,表示相应频率的局域波存在的概率更小,是由于精神分裂症患者具有不稳定的音质特征所导致的。

共振峰表示频谱中的能量集中区域,同时语音信号具有非线性和混沌特性,通过熵的变化能够反映语音信号的失真程度,即精神分裂症患者的发音相比正常人发音的时频域测度的差距。为了选取适合的含有共振峰信息的 IMF 分量进行语音信号重构,需要计算出各 IMF 分量的能熵比。Hilbert 边际谱能量表示为:

$$E_k = \sum_{\omega=0}^{W-1} h_k^2(\omega) \quad (9)$$

其中,  $h_k(\omega)$  代表语音信号第  $k$  个 IMF 分量的 Hilbert 边际谱,  $W$  代表信号频带的宽度,  $E_k$  表示信号第  $k$  个 IMF 分量的 Hilbert 边际谱能量。Hilbert 边际谱熵表示为:

$$p_k = \frac{E_k}{\sum_{k=1}^n E_k} \quad (10)$$

$$H_k = -p_k \log p_k \quad (11)$$

其中,  $p_k$  表示信号第  $k$  个 IMF 分量的归一化谱概率密度函

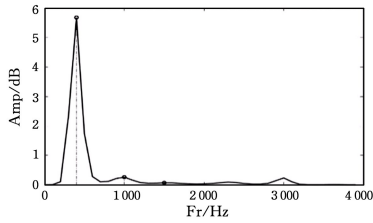


图6 功率谱包络线和共振峰值

Fig. 6 Envelop of power spectrum and peak of formant

峰数表示每个共振峰波动起伏的峰值个数,即局部最大值的个数。

均值是反映每个共振峰数据分布的平均水平,表示为:

$$\mu = \frac{1}{L} \sum_{l=0}^{L-1} x_l \quad (22)$$

方差是用于衡量每个共振峰分布均匀性的一种尺度标准,是各个共振峰离散分布情况的度量,表示为:

$$\sigma^2 = \frac{1}{N-1} \sum_{l=0}^{L-1} (x_l - \mu)^2 \quad (23)$$

四分位差表示分别使随机变量为  $X$  的分布函数  $F(x) = P\{X \leq x\} = 1/4$  和  $3/4$  的  $x$  之间的距离的一半,反映了各个共振峰中间部分数据的集中程度。

极差是各个共振峰序列  $x = (x_1, x_2, \dots, x_L)$  的最大值与最小值之差,代表各个共振峰的变化范围。

偏度描述的是各个共振峰总体取值分布的对称程度。偏度为 0 表示其数据分布形态与正态分布的偏斜程度相同;偏度大于 0 表示正偏,即其数据右端有较多极端值;偏度小于 0 表示负偏,即其数据左端有较多极端值。其绝对值越大表示各个共振峰分布形态的偏斜程度越大,表示为:

$$\alpha = \frac{1}{\sigma^3} \sum_{l=0}^{L-1} (x_l - \mu)^3 \quad (24)$$

峭度描述的是各个共振峰总体取值分布形态的尖锐程度,表示各个共振峰分布的峰值是否突兀或平坦。峭度同样也需要与正态分布相比较,峭度为 0 表示其总体数据分布与正态分布的尖锐程度相同;峭度越大代表峰值越尖锐;反之代表峰值越平坦,表示为:

$$\beta = \frac{1}{\sigma^4} \sum_{l=0}^{L-1} (x_l - \mu)^4 \quad (25)$$

## 2.4 特征降维

t 检验是一种利用 t 分布理论来计算差异发生的概率,进而判断两个或多个数据集之间是否存在差异以及差异是否显著的一种统计假设方法。在本实验中,理论上正常人和精神分裂症患者的语音信号所反映出来的共振峰特征参数是不同的,设定显著性水平  $\alpha$  为 0.05,计算相对应的共振峰特征参数统计量  $t$ ,若  $t$  值存在显著意义,则说明两组样本数据的特征之间的区分度明显。因此,从现有的 81 维声学特征参数中提取区分精神分裂症患者语音和健康对照组语音能力最优的特征子集,即把 t 检验中具有显著差异的特征集用于分类,可获得更好的分类检测率。同时,采用 t 检验的方法可以降低特征向量的维度,减小计算量,提高检验效率。

## 2.5 SVM 分类模型的构建

本文采用 SVM 的模式识别方法对具有显著意义的特征进行分类检测,它可以利用目前已知的算法来寻找目标函数的全局最小值,在解决小样本数据集、高维和非线性模式识别问题中独具优势<sup>[17]</sup>,因此已成为最广泛使用的分类算法之一。

SVM 算法是由 Vapnik 等于 1995 年提出的基于统计学模式识别理论的方法<sup>[31]</sup>,通过有监督的机器学习,自动找出具有很好分类效果的支持向量(即位于边界平面的数据集中点),由此构建一个如图 7 所示的最大化分割边缘的最佳超几何分布平面,最小化期望泛化误差,该方法仅由边界样本的类别来决定最终的分类结果,因此有较高的分辨力和适应性。

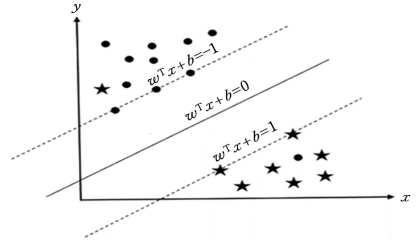


图7 SVM 划分算法示意图

Fig. 7 Schematic diagram of SVM partitioning algorithm

假设训练样本集表示为  $\{(x_l, y_l), l=1, 2, \dots, L, x \in R^L, y \in [-1, 1]\}$ ,若存在分类超平面  $w^T x + b = 0$  能够将样本集线性地准确分为两类,即满足:

$$\begin{cases} w^T x + b \geq 1, & y_l = 1 \\ w^T x + b \leq -1, & y_l = -1 \end{cases} \quad (26)$$

其中,  $w^T x + b = 1$  和  $w^T x + b = -1$  均代表边界平面。若输入空间数据非线性可分,则只需计算核函数  $k(x_1, x_2) = [\varphi(x_1)]^T \varphi(x_2)$  实现数据的非线性映射  $\varphi: R^L \rightarrow F$ ,然后在特征空间  $F$  中执行线性变换即可。在本实验中,我们采用对数据中的噪声有较强抗干扰能力的高斯径向基核函数,该函数将原始有限维空间映射为一个无限维空间,其表达式如下:

$$k(x_1, x_2) = \exp\left(-\frac{\|x_1 - x_2\|^2}{2\delta^2}\right) \quad (27)$$

其中,参数  $\delta$  取默认值 1。

## 3 实验结果与分析

### 3.1 实验数据

本实验根据元音的发声特点,分别选取了 14 个精神分裂症患者的 686 个元音数据样本和 14 个与之性别、年龄、文化程度相匹配的正常人的 793 个元音数据样本来建立病理语音数据库,如表 1 所列。其中,精神分裂症语音数据由四川大学华西医院心理卫生中心和四川省眉山市中医医院精神科提供。实验过程中,受试者被告知自由朗读包含普通话音节里面所有元音的标准化文本。数据以 8000 Hz 采样频率、16 位量化、单声道保存为 Wave 格式。数据采集后,由精神科医师进行数据校验筛选,以保证数据的有效性。

表1 实验语音数据库

Table 1 Experimental voice database

Vowel	Normal male	Normal female	Sick male	Sick female
/a/	98	95	85	64
/e/	108	107	101	92
/i/	113	119	108	95
/o/	42	42	40	34
/u/	34	35	35	32

为了验证生成模型的可靠性和实用性,本文采用十折交叉验证的方法,将数据分成十份,每次将 90% 的数据用于训练,而 10% 的数据留作测试,重复实验十次的均值作为算法检测率的估算。

### 3.2 特征降维

利用结合集成希尔伯特黄变换和倒谱内插的算法对精神分裂症患者和健康对照组的语音信号进行了共振峰特征参数的提取,并使用  $t$  检验筛选出具有显著差异的特征向量。设定显著差异水平为 0.05,从表 2—表 4 的  $t$  检验结果可以看出,粗体特征相比其余的特征  $p$  值更大,表明这些特征如果

用于区分患者和正常人将获得较差的分类准确率,因此在实际分类时应当舍弃粗体特征。同时注意到,峰值个数特征(max)在前 3 个共振峰的分类检测上均具有显著差异。如图 8 所示,可以明显发现两类语音信号中峰值个数特征分布差异明显,进一步表明了将该特征用于分类检测的有效性。

表 2 第一共振峰的各特征参数  $p$  值

Table 2  $p$ -values of characteristic parameters of first formant

Parameter	<i>max</i>	$\mu$	$\sigma^2$	<i>iqr</i>	<i>med</i>	<i>mod</i>	<i>rang</i>	$\alpha$	$\beta$
Fre	$7.3 \times 10^{-15}$	$2.64 \times 10^{-4}$	<b>0.3063</b>	0.0016	$6.1 \times 10^{-5}$	<b>0.2176</b>	<b>0.0727</b>	0.0017	<b>0.2038</b>
Band	$9.4 \times 10^{-10}$	$2.15 \times 10^{-4}$	<b>0.4639</b>	$2.9 \times 10^{-6}$	0.0011	0.0213	<b>0.1986</b>	<b>0.3886</b>	<b>0.3388</b>
Amp	0.0046	<b>0.2433</b>	<b>0.0670</b>	<b>0.0735</b>	<b>0.0520</b>	0.0083	0.0043	<b>0.6656</b>	<b>0.9609</b>

表 3 第二共振峰的各特征参数  $p$  值

Table 3  $p$ -values of characteristic parameters of second formant

Parameter	<i>max</i>	$\mu$	$\sigma^2$	<i>iqr</i>	<i>med</i>	<i>mod</i>	<i>rang</i>	$\alpha$	$\beta$
Fre	$1.6 \times 10^{-12}$	0.0194	$9.9 \times 10^{-10}$	0.0066	0.0301	0.0027	$8.1 \times 10^{-10}$	<b>0.0961</b>	<b>0.1225</b>
Band	$1.4 \times 10^{-12}$	<b>0.8978</b>	<b>0.3183</b>	<b>0.3738</b>	<b>0.5046</b>	0.0297	<b>0.9416</b>	<b>0.8604</b>	<b>0.7777</b>
Amp	$2.3 \times 10^{-12}$	<b>0.1480</b>	<b>0.1758</b>	<b>0.1105</b>	<b>0.8259</b>	0.0018	0.0032	0.0277	0.0399

表 4 第三共振峰的各特征参数  $p$  值

Table 4  $p$ -values of characteristic parameters of third formant

Parameter	<i>max</i>	$\mu$	$\sigma^2$	<i>iqr</i>	<i>med</i>	<i>mod</i>	<i>rang</i>	$\alpha$	$\beta$
Fre	$4.2 \times 10^{-7}$	$3.4 \times 10^{-7}$	$1.7 \times 10^{-5}$	$2.5 \times 10^{-6}$	$3.4 \times 10^{-4}$	$3.1 \times 10^{-8}$	$2.7 \times 10^{-12}$	$2.6 \times 10^{-4}$	0.0013
Band	$1.7 \times 10^{-14}$	0.0154	0.0338	<b>0.2673</b>	0.0431	$1.0 \times 10^{-6}$	0.0015	$8.8 \times 10^{-4}$	0.0052
Amp	$6.7 \times 10^{-14}$	<b>0.7142</b>	<b>0.3960</b>	<b>0.8555</b>	0.0370	<b>0.9625</b>	<b>0.2934</b>	<b>0.3028</b>	<b>0.4513</b>

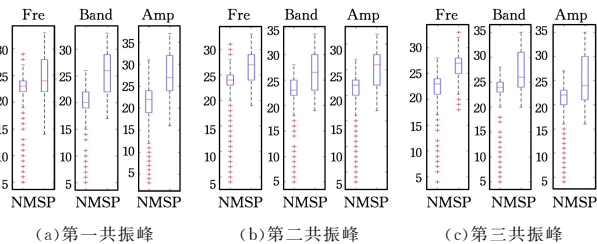


图 8 峰值个数特征箱体图

Fig. 8 Box map of peak number feature

### 3.3 精神分裂症语音自动分类检测结果的影响因素

利用 SVM 分类器对具有显著差异的共振峰特征参数进行分类,实验结果表明,考虑到数据运算量的限制,对精神分裂症患者语音的自动检测正确率最高可以达到 98.8%。本文分别测试了白噪声的集成次数  $M$  和方差  $std$ 、IMF 分量个数  $n$ 、窗长  $cepstL$  4 个因素对检测正确率的影响。

#### 3.3.1 白噪声的集成次数和方差对精神分裂症语音自动分类检测结果的影响

设定 IMF 的分量个数  $n=3$ ,窗长  $cepstL=8$ 。由图 9 可知,白噪声的集成次数  $M=100$ 、方差  $std=0.1$  时,对精神分裂症患者语音检测的效果最佳。根据 EHHT 的原理,随着白噪声集成次数的增加,高频间断信号逐渐被噪声淹没,与共振峰有关的低频信号被不断提取出来,但计算量也随之增加。同时考虑到白噪声的方差较小时,将不会改变 EMD 分解过程中所依赖的上下极值点的分布,因此不能较好地解决模式混叠的问题。而当白噪声方差较大时,必须增加白噪声的集成次数来减小引入的白噪声对分解结果的影响,导致计算效率降低。因此,在限定的计算机数据运算量范围内,白噪声的集成次数越多,削弱模式混叠响应的效果越明显,区分精神分裂症

患者语音和健康对照组语音的正确率就越高,在白噪声的集成次数之后,再选定白噪声的方差。综合考虑,本文选取的最佳白噪声的集成次数  $M=100$ ,方差  $std=0.1$ 。

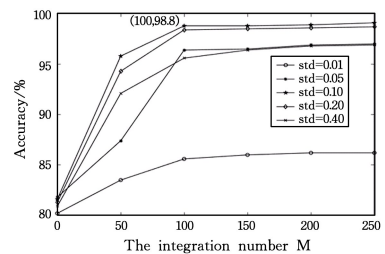


图 9 白噪声的集成次数和方差对检测结果的影响

Fig. 9 Effect of integration number and variances of white noise on detection results

#### 3.3.2 IMF 分量个数对精神分裂症语音自动分类检测结果的影响

设白噪声的集成次数  $M=100$ ,白噪声的方差  $std=0.1$ ,窗长  $cepstL=8$ 。IMF 分量个数对检测结果的影响如表 5 所列。

表 5 IMF 分量个数对检测结果的影响

Table 5 Effect of number of IMF components on detection results

IMF Number	Accuracy/%
1	96.7
2	97.8
3	98.8
4	96.6

由表 5 可知,IMF 分量个数  $n=3$  时,检测率最高。将各个 IMF 分量在 200~3000 Hz 频带内的能谱比占整个频带能谱比的比值降序排列,选取前  $n$  个 IMF 分量重构语音信号。

当  $n$  的取值过小时,会导致包含共振峰信息的 IMF 分量被舍弃;而当  $n$  的取值过大时,会引入无用的 IMF 分量,不利于信号的重构。实验结果表明,当  $n$  取值为 3 时,包含共振峰信息的信号提取效果最佳。EMD 分解算法中存在频带混叠效应,则一个共振峰可能分布在多个 IMF 分量中。本文采用的 EEMD 分解算法有效地将同一共振峰的多个 IMF 分量组成一个新的 IMF 分量,从而保证每个 IMF 分量中只包含一个共振峰。即归一化后能熵比的比值最大的前 3 个 IMF 分量,反映了语音信号前 3 个共振峰的结构特征。采用这些 IMF 分量来重构信号时,精神分裂症语音和健康对照组语音的分类检测效果最佳。

### 3.3.3 窗长对精神分裂症语音自动分类检测结果的影响

设白噪声的集成次数  $M=100$ ,白噪声的方差  $std=0.1$ , IMF 分量个数  $n=3$ 。

表 6 窗长对检测结果的影响

Table 6 Effect of window length on detection results

Window Length	Accuracy/%
6	94.5
7	97.6
8	98.8
9	96.8
10	95.7

由表 6 可知,窗长  $cepstL=8$  时,检测正确率最高。窗长和倒频率的分辨率有关,即和 FFT 的长度以及采样频率有关。在倒频谱域中包含共振峰信息的声道响应倒谱和包含基音频率信息的声门脉冲激励倒谱是相互分离的,通过调节窗长来确定包含共振峰信息的最佳倒谱分量。 $cepstL$  过长时,则将引入基频部分; $cepstL$  过短时,将导致共振峰信息丢失,在  $cepstL=8$  时语音分类检测效果最佳。

### 3.3.4 精神分裂症语音自动分类检测结果讨论

实验结果表明,分别讨论白噪声的集成次数  $M$  和方差  $std$ 、IMF 分量个数  $n$ 、窗长  $cepstL$  4 个因素对精神分裂症语音自动分类检测正确率的影响,精神分裂症语音的平均检测率约为 93.0%。通过对每个因素的分类讨论,逐步确定最优取值,代入下一个因素的讨论中,同时考虑到数据运算量的限制,最终确定当白噪声的集成次数  $M=100$ ,白噪声的方差  $std=0.1$ ,IMF 分量个数  $n=3$ , $cepstL=8$  时,精神分裂症语音检测正确率最佳,达到了 98.8%。表 7 和表 8 分别列出了各影响因素取最优值时各元音和各类参数的分类检测正确率,表明元音 /a/, /e/, /i/ 的检测正确率明显比元音 /o/, /u/ 的检测正确率高一些,且在男女之间存在差异,与各元音发音时声腔的形状和大小(如开口度、圆唇度和收紧程度等)不同而导致精神分裂症患者和健康对照组两类语音各共振峰分布差异有关。

表 7 各元音检测正确率

Table 7 Detection accuracy of vowels

Vowel	(单位:%)		
	Male	Female	All
/a/	95.6	96.6	95.9
/e/	94.7	99.5	99.0
/i/	98.2	87.9	96.1
/o/	86.6	82.9	91.1
/u/	98.6	88.1	91.2

表 8 各类参数检测正确率

Table 8 Detection accuracy of various parameters

(单位:%)

Parameter	F1	F2	F3	Fre	Band	Amp
Accuracy	87.7	97.5	97.0	97.0	92.1	79.3

同时发现单独使用第一共振峰和幅值比其余类型参数的检测正确率略低,进而表明精神分裂症患者和健康对照组发音时舌头前后运动、圆唇度、软腭高度和频谱能量的集中程度的区分度更明显,结合多个具有显著差异的参数能够提高整体的精神分裂症语音检测正确率,从而达到最佳的检测效果。

### 3.4 共振峰估算方法的对比分析

基于快速傅里叶变换(FFT)的传统的共振峰提取算法包括倒谱(Cepstrum, CEP)法和线性预测系数(Linear Prediction Coefficient, LPC)法。倒谱法<sup>[32]</sup>运用同态解卷技术将基频谐波和声道响应分离,从而获取信号倒谱包络线得到共振峰信息。线性预测系数法<sup>[33]</sup>采用 LPC 对语音信号进行解卷积,获取信号的全极点模型,进而通过 Newton-Raphson 方法得到线性预测系数,确定共振峰中心频率和 3dB 带宽。随着现代信号处理技术的发展,可以利用语音信号的动态特性从连续几帧语音信号中提取共振峰特征参数,如希尔伯特黄变换<sup>[34-35]</sup>(Hilbert-Huang Transform, HHT)法采用 FFT 来初步确定语音信号各阶共振峰对应的带通滤波器,经滤波后进行 EMD 分解得到一簇 IMF,计算瞬时频率和瞬时幅值得到共振峰频率参数。

为了定量评估用于实现精神分裂症患者语音与健康对照组语音自动分类检测的共振峰提取方法,本文设计实验,对比分析了结合集成希尔伯特黄变换(EHHT)和倒谱内插(CI)的共振峰提取算法与经典的倒谱法、线性预测系数法、希尔伯特黄变换法的性能。

经过测试,得到的实验结果如表 9 所列,从表中可以看出,本文提出的 EHHT+CI 算法的检测正确率普遍高于经典算法,其原因如下:

(1)使用传统的语音信号分析方法需要满足两个假设,即人的发声系统是线性的和语音信号是短时平稳的。而精神分裂症患者的语音信号具有明显的非线性特征以及混沌性质的不确定性。倒谱法中共振峰频率对应频域功率谱的极大值点,当精神分裂症患者语音的共振峰较接近或受噪声(谐波)干扰时,峰值检测法存在共振峰合并和虚假共振峰等问题,导致检测结果不精确。线性预测系数法的动态范围较大且可能存在滤波不稳定性问题,因此不适宜作为精神分裂症患者语音的共振峰检测方法。

(2)希尔伯特黄变换法虽然适用于非平稳、非线性的语音信号处理,但不足之处在于利用 FFT 谱初步估计共振峰存在人为的不确定误差,自适应性较差;同时出现的模态混叠效应导致不能有效提取包含精神分裂症患者语音共振峰的有效 IMF 分量。

(3)本文提出的结合集成希尔伯特黄变换+倒谱内插的共振峰提取算法,利用 EEMD 有效地避免了模态混叠问题,根据归一化 IMF 能熵比的比值筛选出包含精神分裂症语音共振峰的分量,用于重构语音信号,再结合 CI 算法提取共振峰参数,弥补了共振峰合并和虚假共振峰等缺陷。

表 9 共振峰估算方法的检测正确率比较

Table 9 Comparison of detection accuracy of formant estimation methods

Method	(单位:%)		
	Male	Female	All
CEP	91.3	88.2	89.7
LPC	65.0	76.4	67.4
HHT	86.1	77.1	83.7
EHHT+CI	99.9	99.4	98.8

**结束语** 精神分裂症病因复杂、病程迁长。本文根据精神分裂症患者语音的音质特点,提出基于集成希尔伯特黄变换(EHHT)和倒谱内插(IC)相结合的算法,提取共振峰的声学特征参数,并利用 SVM 分类器来实现精神分裂症患者语音的自动检测。利用精神分裂症患者共 686 个元音样本数据和健康对照组共 793 个元音样本数据对本文算法进行检测,并设计了两组实验。通过分类讨论 4 个因素对精神分裂症语音检测结果的影响,在考虑数据运算量的情况下,其检测正确率最高可以达到 98.8%。与经典的倒谱法、线性预测系数法和希尔伯特黄变换法相比,本文的共振峰估算方法具有更好的鲁棒性和稳健性。

本研究致力于实现精神分裂症疾病的“早发现、早确诊、早干预、早治疗”和长期监测追踪的目的。如何评估精神分裂症患者的语音参数和临床症状严重程度之间的映射关系有待进一步的深入探讨,目前存在的问题还包括样本量较少,缺乏一定的代表性,除了横向地扩大样本量以外,还需要纵向地长期监测追踪精神分裂症患者的语音信号情感变化。从共振峰特征入手,还可以结合其他声学特征,实现精神分裂症语音更加全面客观的检测。今后还欲将相关的研究成果应用到其他精神病学和神经系统疾病的研究中,从而促进精神卫生事业的长远发展,切实保障人民的身心健康和社会的和谐稳定。

## 参考文献

- [1] QIAO Y S,ZHANG S Q,CUI W D,et al. The Effect of Mental Health Education Diverse Family Therapy on Social and Family Functioning in Patients with Schizophrenia[J]. China Journal of Health Psychology,2019,27(2):11-15.
- [2] DELATTRE P. The Physiological Interpretation of Sound Spectrograms[J]. Publications of the Modern Language Association of America,1951,66(5):864-875.
- [3] JIN Y F,XIONG L N,CAI C L,et al. Research Progress on Psychotic Symptom Assessment Tools in Patients with Schizophrenia[J]. Journal of Nursing,2015,22(16):17-21.
- [4] WANG X H. Diagnostic Research Method of Schizophrenia Based on fMRI Technology[J]. Imaging Research and Medical Applications,2019,3(2):59-60.
- [5] WANG J J,WANG P F,QUAN W X,et al. The Characteristics of Language Cognition and Its Neural Basis in Schizophrenia[J]. Progress in Biochemistry and Biophysics,2015,42(1):49-55.
- [6] BREAKSPEAR M. The Nonlinear Theory of Schizophrenia[J]. Australian and New Zealand Journal of Psychiatry,2006,40(1):20-35.
- [7] CHAKRABORTY D,TAHIR Y,YANG Z,et al. Assessment and Prediction of Negative Symptoms of Schizophrenia from RGB+D Movement Signals[C]//2017 IEEE 19th International Workshop on Multimedia Signal Processing (MMSP). New York:IEEE Press,2017:1-6.
- [8] XUE C B,CHEN Q L,ZHU J,et al. Progress in Schizophrenic Genomics Research[J]. International Journal of Psychiatry,2014,41(1):17-20.
- [9] RAPCAN V,D'ARCY S,YEAP S,et al. Acoustic and Temporal Analysis of Speech:A Potential Biomarker for Schizophrenia[J]. Medical Engineering and Physics,2010,32(9):1074-1079.
- [10] COMPTON,MICHAEL T,ANYA L,et al. The Aprosody of Schizophrenia:Computationally Derived Acoustic Phonetic Underpinnings of Monotone Speech[J]. Schizophrenic Research,2018,197:392-399.
- [11] GOLD R,BUTLER P,REVHEIM N,et al. Auditory Emotion Recognition Impairments in Schizophrenia:Relationship to Acoustic Features and Cognition[J]. American Journal of Psychiatry,2012,169(4):424-432.
- [12] XU S,YANG Z,CHAKRABORTY D,et al. Automatic Verbal Analysis of Interviews with Schizophrenic Patients[C]//2018 IEEE 23rd International Conference on Digital Signal Processing (DSP). New York:IEEE Press,2018:1-5.
- [13] COHEN A S,MITCHELL K R,DOCHERTY N M,et al. Vocal Expression in Schizophrenia:Less Than Meets the Ear[J]. Journal of Abnormal Psychology,2016,125(2):299-309.
- [14] ZHANG J,PAN Z D,GUI C,et al. Clinical Investigation of Speech Signal Features Among Patients with Schizophrenia[J]. Shanghai Archives of Psychiatry,2016,28(2):95-102.
- [15] HAN W J,LI H F,RUAN H B,et al. Review on Speech Emotion Recognition[J]. Journal of Software,2014,25(1):37-50.
- [16] GOBL C,CHASAIDE A N. The Role of Voice Quality in Communicating Emotion, Mood and Attitude[J]. Speech Communication,2003,40(1/2):189-212.
- [17] JIANG H H,HU B. Speech Emotion Recognition in Mandarin Based on PCA and SVM[J]. Computer Science,2015,42(11):270-273.
- [18] SONG Z Y. Application of MATLAB in Speech Signal Analysis and Synthesis[M]. Beijing:Beijing University of Aeronautics and Astronautics Press,2013:1-344.
- [19] YANG J X,SHE Y M,FU M J,et al. Formant Estimation of Isolated Words in the Wa Language Based on Adaptive Variational Mode Decomposition[J]. Journal of Yunnan Minzu University (Natural Science Edition),2019,28(3):83-91.
- [20] FRANCES M A,PINCUS H A,FIRST M B. Diagnostic and Statistical Manual of Mental Disorders[J]. BMC Med,2013,17:133-137.
- [21] COVINGTON M A,LUNDEN S L A,CRISTOFARO S L,et al. Phonetic Measures of Reduced Tongue Movement Correlate with Negative Symptom Severity in Hospitalized Patients with First-Episode Schizophrenia-Spectrum Disorders[J]. Schizophrenia Research,2012,142(1/2/3):93-95.
- [22] BERNARDINI F,LUNDEN A,COVINGTON M,et al. Associations of Acoustically Measured Tongue/Jaw Movements and Portion of Time Speaking with Negative Symptom Severity in Patients with Schizophrenia in Italy and the United States[J]. Psychiatry Research,2016,239:253-258.

national Conference on Mechatronics and Automation (ICMA). Changsha:IEEE,2017:1000-1007.

- [18] MA S, YANG G, ZHANG H, et al. Target threat level assessment technology based on cloud model and Bayesian revision in air combat simulation[C]//2016 35th Chinese Control Conference (CCC). Chengdu:IEEE,2016:9753-9757.
- [19] GAO Y Y, YU M J, WANG Z B L. A new method of multi-target threat assessment for air combat[C]//2017 IEEE International Conference on Cybernetics and Intelligent Systems (CIS) and IEEE Conference on Robotics, Automation and Mechatronics (RAM). Ningbo:IEEE,2017:779-784.
- [20] ALI M, FALCONE P, SJÖBERG J. Model-based threat assessment in semi-autonomous vehicles with model parameter uncertainties[C]//2011 50th IEEE Conference on Decision and Control and European Control Conference. Orlando:IEEE,2011:6822-6827.
- [21] MATTEOLI S, DIANI M, CORSINI G. Automatic Target Recognition Within Anomalous Regions of Interest in Hyperspectral Images[J]. IEEE Journal of Selected Topics in Applied

Earth Observations and Remote Sensing, 2018, 11(4):1056-1069.

- [22] CHO H, CHO Y, SONG W. Image Matting for Automatic Target Recognition[J]. IEEE Transactions on Aerospace and Electronic Systems, 2017, 53(5):2233-2250.



**WU Hong-tao**, Ph.D, engineer. His main research interests include computer vision and analysis and modeling of intelligent transportation.



**LI Lu-kai**, postgraduate, lecturer. Her main research interests include 3D art design and so on.

(上接第 195 页)

- [23] HAN Z Y, WANG J, WANG D, et al. Dynamic Feature Extraction for Speech Signal Based on Formant Curve[J]. Computer Technology and Development, 2017, 27(6):72-80.
- [24] FENG J C, PAN S Y. Extraction Algorithm of Vital Signals Based on Empirical Mode Decomposition[J]. Journal of South China University of Technology (Natural Science Edition), 2010, 38(10):1-6.
- [25] WU Z, HUANG N E. Ensemble Empirical Mode Decomposition: A Noise-Assisted Data Analysis Method[J]. Advances in Adaptive Data Analysis, 2009, 1(1):1-41.
- [26] WANG Y H, YE H C H, YOUNG H W V, et al. On the Computational Complexity of the Empirical Mode Decomposition Algorithm[J]. Physica A: Statistical Mechanics and Its Applications, 2014, 400:159-167.
- [27] FU K. Research on Feature Extraction Methods and Its Application to Nonstationary Signal Based on HHT[D]. Chongqing: Chongqing University, 2015.
- [28] MA T L. Improvement of HHT Method and Its Application in Ground Tilting Tide Signal[D]. Kunming: Kunming University of Technology, 2013.
- [29] BAO H Q, LIN M C. Experimental Phonetics Summary (Updated Version) [M]. Beijing: Peking University Press, 2014:107-113.
- [30] ZHAO T T, YANG H W. Formant Extraction Algorithm of Speech Signal by Combining EMD and WMCEP[J]. Computer Engineering and Application, 2015, 51(9):207-212.
- [31] ZHUO J W, WANG H J. MATLAB Mathematical Modeling

Method and Practice (3rd Edition) [M]. Beijing: Beijing University of Aeronautics and Astronautics Press, 2018:66-69.

- [32] ZHAO Y, YIN X F, CHEN K A. A New Formant Detection Algorithm Based on Cepstrum [J]. Applied Acoustics, 2010, 29(6):416-424.
- [33] HAN F, ZHENG J J. Improved Resonance Peak Detection Algorithm Based on LPC[J]. Electronic Design Engineering, 2017, 25(17):85-89.
- [34] HUANG H, CHEN X X. Speech Formant Frequency Estimation Based on Hilbert Huang Transform [J]. Journal of Zhejiang University (Engineering Science), 2006, 40(11):1926-1930.
- [35] YU F Q, XIAO Z. Finding Speech Formant by Using the Character of Hilbert-Huang Transform as An Adaptive Band-Filter[J]. Technical Acoustics, 2008, 27(2):266-270.



**TIAN Wei-wei**, born in 1998, postgraduate. Her main research interests include speech signal processing and so on.



**LI Yuan-yuan**, born in 1984, Ph.D. attending doctor. Her main research interests include psychiatry and mental health.