科学通报 2018 年 第 63 卷 第 20 期: 2081 ~ 2092
专辑: 抑郁症研究 论 文
《中国科学》杂志社
SCIENCE CHINA PRESS

# 基于语音的抑郁症识别

潘玮[1,2], 汪静莹[1,2], 刘天俐[3], 刘晓倩[1], 刘明明[1,2], 胡斌[4], 朱廷劭[1*]

1. 中国科学院心理研究所, 北京 100101;
2. 中国科学院大学心理学系, 北京 100049;
3. 北京大学人口研究所, 北京 100871;
4. 兰州大学信息科学与工程学院, 兰州 730000
* 联系人, E-mail: tszhu@psych.ac.cn

2018-04-02 收稿, 2018-05-16 修回, 2018-05-18 接受, 2018-07-11 网络版发表

**摘要** 抑郁症是世界范围内常见的精神疾病之一, 抑郁症患者往往长期伴随情绪低落, 如悲伤内疚、低自尊、兴趣丧失、功能减退等, 对个人、家庭及社会造成了巨大损失. 抑郁症的发病原因复杂, 临床诊断存在一定的困难, 有必要寻找一种更加便捷、客观、高效的方式来辅助抑郁症的快速识别. 语音作为一个相对客观且容易获得的变量, 具有其潜在的价值. 本研究旨在构建基于语音的抑郁症识别模型, 探究语音与抑郁症之间的关系. 收集了103名被试(45名抑郁症患者, 58名健康人)的语音数据, 实验组为临床确诊的抑郁症患者, 年龄在23.8~44.6岁之间, 控制组为健康人, 年龄为20.1~41.7岁. 我们采用了3(情绪状态: 正性、中性、负性)×3(任务类型: 语言问答、文本朗读、图片描述)的实验设计, 运用机器学习的分类算法——逻辑回归(LR)来构建抑郁识别模型. 实验结果表明, 语音的抑郁识别精度可以达到82.9%. 本文采用机器学习方法, 基于语音变量建立有效的抑郁症自动识别模型, 为抑郁症的辅助识别提供客观的指标和依据.

**关键词** 抑郁症, 语音特征, 分类算法, 逻辑回归

世界卫生组织(World Health Organization)2017年的数据显示, 全球有超过3亿人患有抑郁症, 且这一数字在逐年增加. 调查显示, 中国有5400多万抑郁症患者. 抑郁症不仅影响患者的身心健康, 更会导致高致残致死率. 有报告指出, 抑郁症患者的自杀率是普通人群自杀率的20倍以上, 且在已自杀者中, 抑郁症患者占1/2以上[1]. 抑郁症不仅给个人和家庭带来巨大的创伤, 其后果对社会也是一种巨大的损失.

抑郁症虽然危害巨大, 但它是一种可以被有效治疗和改善的疾病. 临床实践中, 药物治疗能够促使抑郁症患者康复, 此外还有心理治疗(如认知行为疗法、人际心理疗法、行为治疗、精神动力学治疗)以及物理治疗(如重复经颅磁刺激)等. 然而有相当一部分患者群体没有得到及时的诊断. 国内外受抑郁症影响的人群中有不到1/2(在许多国家, 只有不到10%)接受治疗[1]. 这一现象的原因大致有以下3点.

(ⅰ) 长期以来, 抑郁症被公众所误解, 常常被贴上“小心眼”“想不开”“矫情”的标签而缺少足够的重视. 事实上, 抑郁症是一种严重的心境障碍. 精神疾病, 往往同其他心理疾病一样, 公众缺乏足够的认识. 躯体疾病者求医会被认为是正常的, 而心理疾病, 尤其是精神疾病者, 会被大众贴上“精神不正常”的标签. 社会的高速发展、高强度的工作压力、社会身份的变更、人际关系甚至躯体病患的影响, 都会增加抑郁症患病的可能. 20~60岁是抑郁症的高发期, 青春期、职业人群、更年期以及老年阶段都是抑郁症

**引用格式**: 潘玮, 汪静莹, 刘天俐, 等. 基于语音的抑郁症识别. 科学通报, 2018, 63: 2081–2092
Pan W, Wang J Y, Liu T L, et al. Depression recognition based on speech analysis (in Chinese). Chin Sci Bull, 2018, 63: 2081–2092, doi: 10.1360/N972017-01250

的高发阶段. 抑郁症涉及人群广泛, 可能引发严重后果, 应该引起足够的重视.

(ⅱ) 抑郁症患者求助意愿不够强烈. 一方面, 抑郁症患者各项功能受损, 情绪体验较为消极, 兴趣减退, 缺乏积极改善现状的愿望. 另一方面, 抑郁症患者对抑郁症的认识不够, 更倾向于隐藏自己患病的事实, 讳疾忌医. 《2017年中国网民抑郁症调研报告》显示, 当个体出现抑郁情绪时, 仅5%的人表示会寻求专业机构或个人的帮助, 而其他95%则选择忍耐或者跟亲友倾诉.

(ⅲ) 相关资源的缺乏及医疗条件的限制. 一方面, 由于对精神疾病的重视不够, 相应的资金投入和专业医疗人士、心理服务工作者相对缺乏; 另一方面, 对抑郁症的有效筛查存在一定难度, 目前对抑郁症的诊断往往由专门的精神科医生借助精神疾病诊断与统计手册(the diagnostic and statistical manual of mental disorders, DSM)进行诊断, 该手册的使用需要经过严格训练, 测试时间往往较长, 并且应用场景也受限, 与此同时, 还存在量化不够精确等问题[2,3]. 此外, 对抑郁症的诊断存在误诊情况. 例如, 抑郁症患者往往伴随躯体不适, 因此他们会更多以躯体不适为主诉到综合医院就诊, 抑郁症患者到内科的就诊率远远高于到精神科的就诊率. 而综合医院的医生往往相对缺乏精神疾病诊断的临床经验. 即便是精神科医生, 在诊断过程中也难免会带有一定的主观倾向, 诊断的效力相对较低. 再者, 抑郁症与双相障碍是两种比较类似的心境障碍类疾病, 抑郁症和双相障碍的抑郁期往往较难区分[4]. 《全球疾病负担》(Global Burden of Disease, GBD)在2016年的多项针对中学生的流行病学的调查发现[5], 中学生抑郁症状的检出率为23.7%~54.4%, 其中重度抑郁症状检出率为3.3%~9.68%.

抑郁症治疗的关键在于前期的诊断筛查, 如果能够在相对安全、不过多涉及隐私的情况下快速有效地诊断出个体是否患有抑郁症, 将极大地降低抑郁症临床筛查的困难, 促使患者尽早接受治疗. 因此, 有必要引入更加客观便捷的测量方式, 打破传统医学诊断的限制, 以更加灵活高效的方式辅助临床抑郁症的早期筛查, 帮助患者得到及时有效的治疗.

语音是一种非侵入式、临床极易获取的信息. 目前已有关于语音与抑郁症的大量研究, 为探索语音作为临床抑郁症自动化诊断的工具提供了可能. 根据已有研究, 抑郁症病人的语音具有以下特点: 语速较慢、停顿次数多、停顿时间长[6,7]、语音特征的变化减少[8]、声音缺少抑扬顿挫、呆板[9]. 抑郁组个体与正常个体相比, 气息声更显著[10]. 从韵律特征的角度来看, 抑郁症患者的基频(fundamental frequency, F0)变化减少, 如带宽、振幅、能量等[6,7,11], 反映了抑郁症患者声音频率的变化变少. 频谱特征也与患者的抑郁程度有关, 有研究发现, 500 Hz以下和500~1000 Hz的声谱能量的变化程度和抑郁的严重程度增加相关[12]. 可见, 对语音进行特征抽取, 对相对客观且不容易被个体刻意掩盖的声学特征的捕捉, 将有助于更好地理解抑郁症. 针对抑郁症患者语音特点的研究大致有以下两大类.

(ⅰ) 追踪研究. 早期阶段, 研究者已经发现生理或者心理上的轻微变化会导致声音特征的显著改变[13]. 心理疾病, 或者说精神疾病会伴随病人声音特性(voice)和语音特征(speech, 包括词汇语法)等的改变[14,15]. 针对精神疾病患病群体的声学分析发现, 言语相关变量与症状测评结果存在着中等以上的相关[16,17]. 在这一阶段, 研究者主要进行追踪研究, 借助于治疗前后患者语音特点的变化情况进行探索. 例如, Darby和Hollien[18]最早提出借助实验方法考察精神疾病对病人言语的影响, 采用7种语音维度组成的语音声音量表, 结合前后测的方式, 发现抑郁症病人在治疗前后的声音有显著不同. 但是, 在这一类研究中, 存在追踪时间较短, 被试数量比较少或缺乏对照等问题. 例如, Szabadi等人[19]对8名被试(4名健康人和4名中度抑郁症患者)进行了两个月的追踪研究, 结果发现, 停顿时间的增长可以作为精神运动阻滞的客观指标, 而精神运动阻滞是抑郁症的重要特征之一. Mundt等人[6]对35名抑郁症患者进行了持续6周的电话访谈收集被试的语音. 抑郁症患者的治疗过程往往较长, 并且患者在这一过程中可能存在病情不稳的情况, 此外, 较少的被试量以及特殊的数据收集方式, 如电话访谈, 都可能会影响数据结果的说服力及推广性.

(ⅱ) 横向研究. 比较抑郁症患者群体与其他不同群体之间的语音特点. 例如, Flint等人[20]对比了重度抑郁症、帕金森患者与正常人语音特点的异同, 结果发现, 两类病人与正常人相比, 发音起始时间(voice onset time)更短, 第二共振峰迁移(second formant transition)减少. France等人[9]对比了正常人、抑

郁患者和有自杀倾向的人的语音特点, 结果发现共振峰(formant)和功率谱密度(power spectral density)特征在分类问题中是有效特征. 另有研究发现, 抑郁患者比正常人在语音上缺少韵律(prosody)变化[21]. 此外还有一些跨文化的有关研究. 例如, Alghowinem等人[22]对美国、德国、澳大利亚三个国家的两个语种(英语和德语)进行了实验, 结果发现语言种类和文化并不能显著影响结果的好坏. 鉴于这些研究的实验设计都相对简单. 一些研究开始设计更加复杂的任务来验证语音与抑郁症之间的关联是否具有跨任务的一致性. 例如, Alghowinem等人[23]指出自发语音的分类准确率比文本朗读的高. Mitra等人[24]的研究显示, 自然语音要比朗读获取的语音对是否抑郁的预测效果更好. 可见, 言语方式也会影响到实验的结果. 不同的言语方式表明说话人在信息收集、认知加工、言语组织等过程中存在差异, 这一点可能会反映在语音特征的变化上, 进而影响了实验结果.

此外, 还有从特征选取和算法改进方面来研究如何提高语音对抑郁症的预测效果[25~27].

更加复杂的实验设计使得实验结果更加具有说服力. 但是, 这些研究往往偏向于探索通过算法改进来提高语音对抑郁症的预测效果, 这些针对特定数据集的算法改进较少得到有效的重复性验证. 且这些研究较少对不同任务取得的不同结果进行更进一步的探讨. 有研究发现, 与健康人相比, 重度抑郁症患者对正性情绪(如喜悦)的反应减弱, 对负性情绪(如悲伤、恐惧、愤怒)的反应增强, 且这一偏差与抑郁症患者的病情严重程度与功能失调症状程度相关[28]. 另有研究发现, 抑郁症患者对积极的人脸图片反应减弱[29], 而对消极的面部表情反应增强, 并且对消极表情的记忆更好[30]. 这些研究表明, 抑郁症患者对不同效价的情绪刺激的反应出现钝化现象. 情绪低落是抑郁症患者的主要特点之一. Stasak等人[31]的研究发现, 增加情绪效价这一信息能够将语音特征对是否抑郁的预测准确率提高5%. 可见, 发音与情绪之间也存在着一定的关联. 如果能够同时考虑情绪效价和任务类型两种因素, 将能够更加深入地了解个体的发音特点及其与情绪的关联, 并对结果进行更好的解释. 与此同时, 也能够进一步验证不同任务类型与不同情绪状态下语音对抑郁症的预测效果是否一致.

为了考察研究结果是否具备跨任务的稳定性, 我们创设了3种实验情景: 语言问答、文本朗读和图片描述. 为了考察不同情绪状态下, 语音特征对抑郁症的预测效果是否一致, 本研究准备了3种效价的情绪启动材料, 旨在启动个体的不同情绪状态, 获得不同情绪状态下的语音. 同时, 本文旨在通过提取语音特征, 建立针对抑郁症辅助识别的预测模型, 提高对抑郁症的有效诊断. 目前国内有关研究正处于起步阶段, 本文将探究在不同任务和不同情绪状态下, 语音变量对抑郁症的预测效果, 探索语音作为临床抑郁症诊断辅助工具的潜在价值.

## 1 方法

### 1.1 数据采集

本研究共包含103名被试. 其中, 抑郁症患者来自北京安定医院与回龙观医院, 健康人由广告招募而来. 所有被试均经过经验丰富的心理专家和精神科医生依照《简明国际神经精神访谈》(the MINI-International Neuropsychiatric Interview, MINI)[32]和《心理障碍诊断与统计手册》(Diagnostic and Statistical Manual of Mental Disorders, DSM-IV)[33]进行诊断筛查. 实验招募了45名抑郁症患者与58名健康人, 前者年龄在23.8~44.6岁 ($M$=34.2, $SD$=10.4); 后者年龄为20.1~41.7岁 ($M$=30.9, $SD$=10.8). 从性别角度来看, 抑郁症患者中, 男性为22人, 女性为23人; 健康人中, 男性共27人, 女性共31人. 所有被试不存在物质滥用、物质依赖、人格障碍等其他精神疾病, 无严重的躯体疾病或自杀行为. 被试为小学以上文化水平.

本研究为3(情绪状态: 正、中、负)×3(任务类型: 语言问答、文本朗读、图片描述)的实验设计. 其中, 语言问答任务的3种情绪条件下各有3道相同情绪效价的问题, 文本朗读任务在各种情绪下各有一段文本材料, 图片描述任务中, 每种情绪下分别有两张图片, 一张为人脸图片, 一张为情景图片.

实验在光线充足、安静的房间内进行. 所有任务与指导语都在同一电脑程序中, 被试坐在距离一台21寸电脑屏幕正前方约1 m远的地方, 屏幕中会显示任务要求. 每一任务材料呈现后, 被试需要根据提示进行回答, 主试在一旁对程序进行操作, 并使用录音设备记录被试的语音. 在语言问答任务中, 被试需要根据播放的问题进行回答. 例如, 在正性情绪启动条

件下, 问题为: “请跟我们分享一段您认为美好的回忆, 大致描述下当时的情景.” 文本朗读任务中, 屏幕中将显示一段文本, 被试需要仔细浏览一遍文本, 之后进行朗读. 图片描述任务中, 屏幕中将显示一张图片和一句提示语: “该图片让你联想到什么?” 被试需根据图片和提示进行联想并进行语言描述. 具体说明见补充材料.

所有任务随机呈现且仅呈现一次. 所有材料均来自现有文献并且具有显著启动效果, 语音问答任务为被试根据不同的任务要求进行回答[34,35]. 在此过程中对被试的声音进行录音, 从而获得语音数据. 每个被试在每种条件下分别生成一份语音数据. 本研究经北京安定医院和回龙观医院伦理委员会审批并获得许可. 实验前被试均签署知情同意书.

### 1.2 数据处理

数据以.wav形式进行保存, 在此基础上进行特征提取. 结合前人研究, 我们选取了26个在抑郁症研究中应用较为广泛的语音特征作为研究对象, 包括强度(intensity)、响度(loudness)、过零率(zero-crossing rate)、清浊比率(voicing probability)、基频、基频包络(F0 envelope)、8个线性谱对(line spectral pairs, LSP)以及12个梅尔倒谱系数(mel-frequency cespstral coefficients, MFCC). 其中, 强度为声波在单位时间内作用在与其传递方向垂直的单位面积上的能量. 响度即音量, 由振幅决定. 过零率指信号的符号变化(如从正变到负)的比率. 清浊比率用于评估每个调波中清音和浊音能量的百分比. 基频指一段周期性声波的最低频率. 基频包络是与音色有关的基频特征. 线性频谱对是指声音通过频道传输时的线性预测系数. 而梅尔倒谱系数, 是基于声音频率其非线性梅尔刻度(melscale)的对数能量频谱的线性变换的系数[36]. 在这些静态特征的基础上, 计算了能够反映语音动态变化的这26个特征的一阶导数(delta value), 并进一步获得了能够反映语音整体变化的19个长时特征, 包括最大值、最小值、全距、均值、标准差、峰度和偏度等统计指标. 采用专门的特征提取软件openSMILE[36], 一共获得了988个语音特征.

其中, 短时特征的提取过程为按照帧移(frame step)10 ms, 帧长(frame size)25 ms将音频文件分帧, 可分为498帧. 对于不同的韵律特征, 有两种不同的方式来进一步提取. 例如, 对于基频、基频包络、清浊比率三个特征, 需要进一步使用汉明窗(Hamming)函数处理信号, 减少帧间信号的不连续性, 随后进行快速傅里叶变换(fast Fourier transform, FFT)以便从频域处理信号, 应用自相关函数(autocorrelation function, ACF)求得浊音的基音周期, 从每一帧信号中得到对应这三个短时特征(基频、基频包络、浊音概率), 如图1所示. 对于其他23个特征, 在音频文件被分为498帧之后, 对语音的高频部分进行预加重(pre-emphasis), 去除口唇辐射的影响, 增加语音的高频分辨率, 接着通过汉明窗运算、快速傅里叶变换和梅尔倒谱分析, 可以得到12个梅尔倒谱系数. 基于自回归模型, 即AR(auto-regressive)模型, 进行线性预测编码(linear predictive coding, LPC). 之后从每一帧信号中得到其余11个短时特征(如强度、响度、过零率、线性频谱对等). 流程如图2所示.

对于长时特征(统计特征), 对每一帧获得的26个短时特征进行平滑处理和一阶导数运算, 得到26个短时Delta特征, 至此, 每一帧共有52个短时特征. 对每一个短时特征, 在498帧上纵向进行统计运算, 如对基频(F0)这个短时特征, 在所有帧上取最大值(max)和

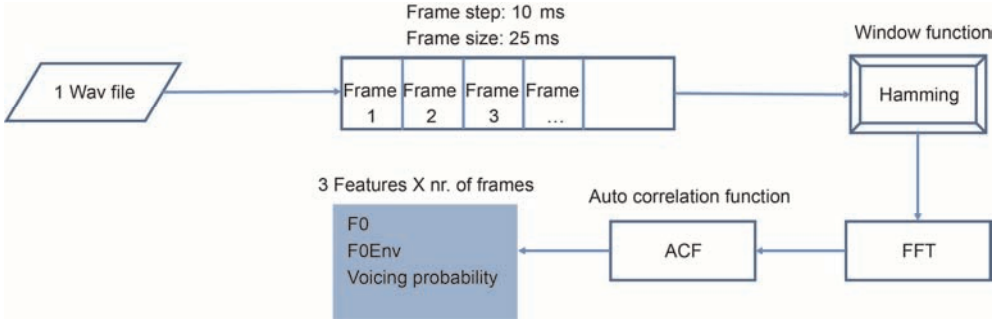


**图 1** (网络版彩色)3 个短时语音特征提取流程

**Figure 1** (Color online) Extracting procedure for 3 short-term voice features

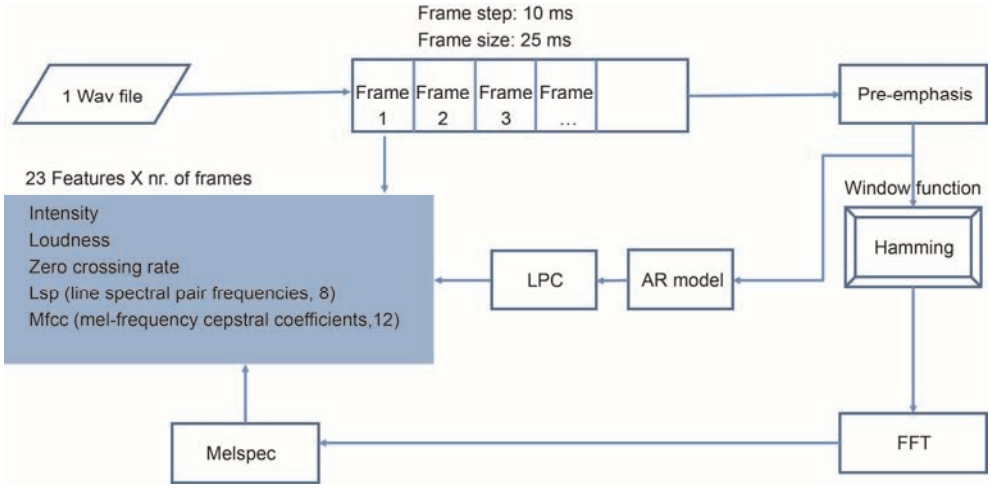


**图 2** (网络版彩色)23 个短时语音特征提取流程

**Figure 2** (Color online) Extracting procedure for 23 short-term voice features

最小值(min), 得到两个统计特征(长时特征). 按照上一步的思路, 对52个特征, 分别在所有帧上进行19种统计运算, 最后一个音频文件共得到988个长时特征. 具体处理流程如图3所示.

### 1.3 模型构建

本研究利用WEKA[37]软件构建抑郁症分类模型, 借助于逻辑回归(logistic regression)分析方法建立语音的预测模型.

逻辑回归[38]是线性回归的另一种形式, 线性回归用于对连续变量进行预测. 当需要对二分变量以及较复杂现象进行预测时, 就可以运用逻辑回归方法. 其核心是通过将数据拟合到logit函数中, 从而对事件发生的概率进行预测. 它是一种类似于朴素贝叶斯的算法, 逻辑回归的限制较前者宽松. 且该方法更加适用于复杂数据的分类问题.

### 1.4 评估指标

分类结果的好坏需要有评价指标与标准. 机器学习的分类算法中有以下几个分类结果好坏的评估指标: 准确率(precision)也称真阳性率, 也就是信号检测论中的击中/(击中+虚报)[39]; 召回率(recall)相当于击中率[39]; F值(F score)是准确率与召回率的加权调和平均数[39].

曲线下面积(area under curve, AUC): 受试者工作特征曲线(receiver operating characteristic curve, ROC曲线)下面积, AUC是二分类模型常用的评价指标[40]. 指给定一个正样本和一个负样本, 分类器输

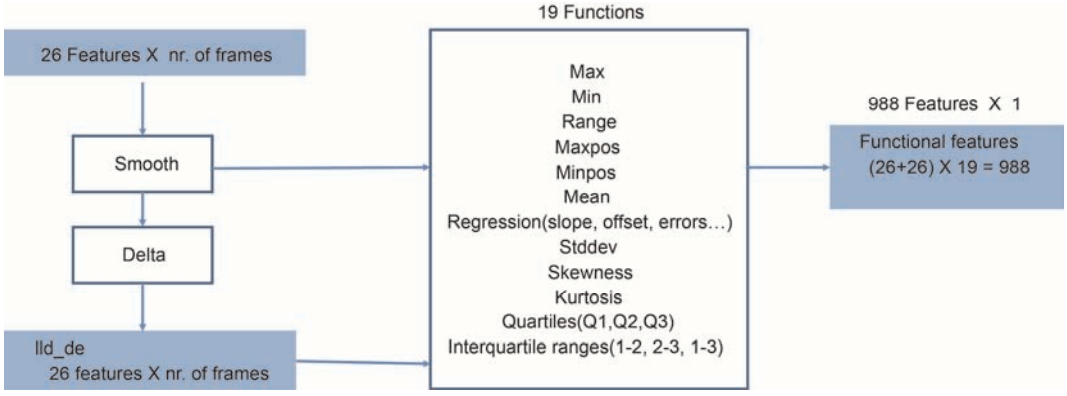


**图 3** (网络版彩色)26 个语音对应的 988 个长时特征提取流程

**Figure 3** (Color online) Extracting procedure for 988 long-term voice features

出正样本为正的概率比分类器输出负样本为正的概率要大的可能性. AUC越大, 表明分类模型的预测准确率越高, 分类器效果更好, 诊断工具的价值越大.

## 2 结果

借助逻辑回归算法, 分类预测结果显示, 在正性情绪启动下, 语音问答任务的预测效果最好, 其预测准确率为77.3%, 召回率为73.9%; 在中性情绪启动下, 图片描述任务的效果最好, 其预测准确率为76.1%, 召回率为74.5%; 负性材料中, 语音问答的效果最好, 其预测准确率为82.9%, 召回率为73.9%. 以中性情绪启动下图片描述任务的结果为例进行说明, 在该条件下的预测模型中, 所有被预测为抑郁的人中, 临床确诊为抑郁症的比例为76.1%; 在所有临床确诊为抑郁症的人中, 被预测为抑郁的比例为74.5%. *F*值是准确率与召回率的权衡. 在正性情绪启动下, 语音问答任务下的分类预测模型要优于文本朗读任务下的预测模型, 而后者又优于图片描述任务中的预测模型(*F*值: 75.6>66.7>62.9). 在中性情绪启动下, 图片描述任务中的预测模型要优于文本朗读任务结果, 而文本朗读任务下模型结果与语音问答任务下结果相近(*F*值: 75.3>62.4>60.2); 在负性情绪启动下, 语音问答任务的预测效果最理想, 文本朗读任务下的预测模型效果次之, 图片描述任务下结果最差(*F*值: 78.2>68.9>60.7). 从图4可以看出, 在语音问答任务中的负性情绪材料启动下, *F*值最大, 为78.2%. 在所有条件下, *F*值在60.2%~78.2%.

在语音问答任务中, 负性材料的预测效果最好, 准确率为82.9%, 召回率为73.9%, 其次是正性材料, 其预测准确率为77.3%, 召回率为73.9%, 然后是中性材料, 准确率为67.6%, 召回率为54.3%. 在文本朗读任务中, 类似的, 负性和正性材料的预测效果较好, 预测准确率分别为67.4%和66%, 召回率分别为70.5%和67.4%. 中性材料的预测效果最差, 预测准确率为61.7%, 召回率为63%. 但是在图片描述任务中, 中性材料的预测效果最好, 准确率为76.1%, 召回率为74.5%. 而正性和负性材料的预测效果相对较差, 二者预测准确率分别为60.9%和58.7%, 召回率分别为65.1%和62.8%. 同样地, *F*值与AUC得分表现出了同样的分布特点, 例如, 在语音问答任务中, 负性情绪启动下, 语音特征的抑郁症预测模型效果较好, AUC=0.81, 曲线下面积这一指标是模型好坏的重要评价标准, 一般AUC在0.5~1之间, 表明模型分类效果优于随机猜测, 具有一定的预测价值. 整体来看, 在不同实验条件下, AUC值在0.66~0.81之间, 均在0.6以上. 具体结果见表1及图4.

整体来看, 在不同的任务及启动条件下, 语音特征对个体是否抑郁的预测效果基本在60%以上.

## 3 讨论

本研究探讨了语音对抑郁症的预测作用. 结果显示, 语音数据的分类准确率基本在60%以上, 即针对语音数据, 以机器学习的方法建立分类模型, 模型对个体是否抑郁的预测结果具备跨情景的一致性, 且模型效果较为理想. 这与国外有关语音与抑郁症的研究结果相一致. Cummins等人[41]采用调制的频谱特征来预测抑郁症, 分类准确率可以达到67%. Cohn等人[42]通过提取抑郁患者的语音特征, 如基频和转换速度(switch duration), 采用留一交叉验证(leave-one-out validation)方式进行逻辑回归分析, 结果表明, 语音特征预测是否抑郁的正确率为79%, 且抑郁患者的F0在治疗后显著下降. 抑郁症患者语速的提高、响度增加、停顿间隔减少已经被认为是比较典型的抑郁症治疗效果的参数[43,44]. 语音能够预测抑郁症这

**表 1 不同条件下分类模型预测结果**(%)[a)]

**Table 1** The results of classification models under different conditions (%)[a)]

| | 语音问答 | | | 文本朗读 | | | 图片描述 | | |
|---|---|---|---|---|---|---|---|---|---|
| | 正性 | 中性 | 负性 | 正性 | 中性 | 负性 | 正性 | 中性 | 负性 |
| Precision | 77.3 | 67.6 | 82.9 | 66 | 61.7 | 67.4 | 60.9 | 76.1 | 58.7 |
| Recall | 73.9 | 54.3 | 73.9 | 67.4 | 63 | 70.5 | 65.1 | 74.5 | 62.8 |
| *F* score | 75.6 | 60.2 | 78.2 | 66.7 | 62.4 | 68.9 | 62.9 | 75.3 | 60.7 |
| AUC | 78.3 | 66.8 | 80.9 | 69.9 | 65.7 | 72.5 | 67.5 | 77.7 | 65.8 |

a) 从左至右各列数据分别为实验材料编号第1, 4, 7, 10, 11, 12, 13, 14, 18号材料获取的语音数据建立的逻辑回归预测模型. 详见网络版补充材料

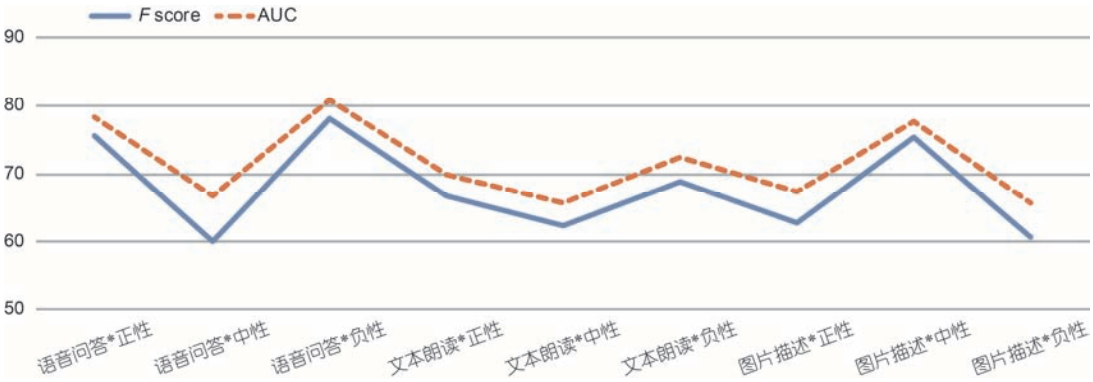


**图 4** (网络版彩色)不同实验条件下分类模型评价指标(%)

**Figure 4** (Color online) Evaluation indexs of classification results under different experimental conditions (%)

一结论表明, 抑郁症患者确实存在语音方面的异常.

从情绪的角度来看, 语音特征能够较好地区分在正中负三种情绪下个体是否患有抑郁症(基本均在60%以上). 这表明抑郁症患者与正常人相比, 确实存在情绪方面的异常, 而这一异常体现在了个体的声音特性上. 脑成像研究发现, 抑郁症患者中, 情感相关的神经系统在结构和功能上均存在异常[45]. 例如, 在杏仁核[46~48]、纹状体[49~51]和背外侧前额叶[52]、前扣带皮层[53]、顶叶上回、额眶皮层[54]、脑岛[55]出现对负性情绪刺激激活减弱. 这些结论共同说明, 抑郁症患者在情绪加工方面出现问题. 抑郁症患者在语音方面与健康人所存在的差异可能是由于抑郁症患者情绪加工方面出现异常, 而情绪对人体学习、生活、各项机能等都会有很大的影响, 这一影响就体现在了个体的声音特性上.

从特征的角度来看, 本文中所提取的语音特征均为情绪识别中常用的指标. 这进一步说明, 抑郁症患者语音与正常人语音的差异可能更多地反映了抑郁症患者的情绪状态以及认知方式的不同. 一项针对自杀行为的研究发现, 抑郁且已经自杀的个体, 与抑郁但未自杀的个体相比, 其生前情绪状态具有更多的负性情绪, 如绝望、无助、愤怒、焦虑、被抛弃、孤独感、羞愧感以及自我厌恶等情感[56]. 根据Beck的负性自动思维理论(negative independent thought theory)[57], 抑郁症患者更倾向于将正性、中性事件判断为负性事件, 存在着各种认知偏差, 如选择性注意、负性过度引申、灾难化、绝对化等认知特点. 这种偏向性情绪处理障碍可能是抑郁患者处于持续情绪低落并显现出语音异常的认知方面原因.

从任务的角度看, 正性与负性情绪效价下, 语音问答任务下模型较理想, 这可能是因为本研究的语音问答任务是需要被试根据自己的实际情况进行回答, 个体的以往经历更容易唤起类似的情绪体验, 由此致使健康人的正性或者负性情绪体验更强烈, 与抑郁患者的反差更大, 故该任务下的模型预测效果较好. 文本朗读也创设了一个情绪环境, 但是可能没有个体曾经的经历更能激发自己的情绪体验, 两组被试的情绪体验差别不大, 因此结果没有前者好. 图片描述由于提供的情绪环境更加受限制, 由此诱发的抑郁患者和健康人在正性和负性情绪体验上的差别就更小了. 而在中性情绪启动方面, 可能由于语言问答和文本朗读给个体创设了过于相似的情景, 而没有任何情绪体验, 故个体的语音特征差异不大, 不能很好地区分是否抑郁. 但是由于图片描述对个体的限定比较小, 因此能真实地反映个体在中性情绪下的语音特征, 这可能暗示了抑郁症患者和健康人存在认知加工方式的不同, 从而使得在面对相同的图片时产生了不同的情绪体验, 进而反映在了语音上. 虽然较为理想的结果主要是从实验材料编号第1, 4, 7, 10, 11, 12, 13, 14, 18这些材料获取的语音数据而建立的预测模型, 这表明并不是每种任务下每个任务类型都能获得比较好的启动效果, 已有研究发现, 不同的任务可能会导致不同的结果. 例如, Alghowinem等人[58]发现, 自发的语音(如访谈、图片描述)比自动语音(如朗读、数数)能够获得更好的结果, 且悲伤类的问题对应的语音特征分类效果比其他问题更好. 这与本研究结果基本一致. 从以往研究来看, 有研究布置专门的实验场景来收集语音数据[42], 另有研究则利用访谈或日常交流过程的录音进行研究[59]. 本研究中, 语音问答任务类似于自然语言访谈任务,

结合了自然情景下的访谈与实验室语音收集方式, 多角度地考察了不同任务下语音特征对是否抑郁的预测效果, 能够取得较好的效果, 该研究所建立的的语音预测模型具有一定的鲁棒性.

从算法模型的角度来看, 本研究所使用的逻辑回归方法是一种应用非常广泛的机器学习分类算法之一. 它能够利用较少的资源处理大规模数据, 属于比较高效的一种算法, 具有较好的鲁棒性. 同时, 它又是比较严谨的分类算法, 通过概率来预测分类结果[60].

鉴于临床对抑郁症的诊断存在的困难和主观偏差等问题. 语音作为一个相对客观且容易获得的指标, 具有一定的优越性. 例如, Alghowinem等人[61]采集了抑郁症患者和健康人的多项指标, 如语音(狭义的副语言特征(paralinguistic features))、眼睛注视、头部姿态等, 结果发现语音的预测效果要好于其他特征. 本研究证实了语音特征可以区分抑郁症患者和健康人, 说明语音这一客观且容易获得的变量在临床抑郁症诊断的快速识别方面具有的潜在价值, 为之后的进一步研究奠定了基础.

本研究存在以下不足. 首先, 未来研究需充分考虑和控制混淆变量, 找出对抑郁症具有高度关联的协变量, 并做到在实验组和控制组的严格匹配. 这是因为, 抑郁症是一种病因比较复杂、同时受生物学因素和社会因素影响的一种复杂的心境障碍疾病. 从社会的角度来说, 个体的年龄、受教育程度、社会等级、婚姻状况等因素均可能对个体是否患抑郁具有一定的影响. 他们之间可能存在着某种关联. 有研究发现, 首发年龄越早, 自杀倾向、自杀尝试越多, 神经质越高、睡眠越少、食欲越差、体重变化越大、抑郁症爆发次数越多、结婚率越低[62,63]. “中国精神障碍疾病负担及卫生服务利用的研究项目调查”显示, 抑郁症患者分居以及离婚率(9.47%)和丧偶率(5.93%)均高于已婚率(3.15%). 小学以下受教育程度者比率为3.13%, 大专及以上为1.39%. 另有研究发现, 受教育程度越高, 个体患抑郁症的概率更大, 而受教育程度低的患者自杀意念、自杀计划更多[64]. 抑郁症发作受多种复杂因素影响. 此外, 所在省份信息也应被纳入考察范围内, 语音往往具有显著的地域特色. 例如, 南方方言相比北方方言通常有更多的声调变化[65]. 我们未能在语音数据收集过程中对协变量进行很好的匹配. 排除潜在混淆变量的影响将有助于进一步理解抑郁症, 对结果做出更加全面的解释.

虽然在该样本中同时涵盖了男性与女性群体, 但是由于样本数量有限, 未能对两种性别进行单独考察. 男性、女性在语音上本就存在这较大的差别, 未来研究应该单独考虑男性与女性各自的语音特性.

而且本研究只考虑了抑郁症患者和健康人之间的区别. 在临床诊断中, 抑郁症往往会与双相障碍等其他精神障碍相混淆, 从而导致诊断的困难. 未来研究有必要划分更加细致的层次, 全面考察语音预测抑郁症在不同人群(如躯体疾病患者、其他精神障碍患者)的区分效果. 从而获得更加严谨的结论.

鉴于本研究验证了语音对抑郁症具有跨情景的预测作用, 未来研究可以尝试采用更加有效的处理高维数据的方法, 寻找对语音具有预测作用的最佳特征, 提高模型的泛化能力. 例如, 对特征进行降维. 有研究尝试采用瓶颈特征选择的方法来筛选特征, 借助于多重感知器、引入瓶颈层来降低特征的维度. 该方法对口音、环境噪声、设备差异等具有较好的鲁棒性[66]. 另有更加复杂的人工神经网络等算法, 如时间递归神经网络(long short-term memory)算法[67], 该方法能够智能化地、全面而非孤立地考虑前后信息之间的关联, 做出更加合理的预测.

## 4 结论

本研究采用机器学习的方法, 在不同实验条件下, 通过对高维语音数据建立是否抑郁的二分类的预测模型, 说明了语音作为临床抑郁症快速识别与诊断的工具的意义与价值. 鉴于本研究的结果和存在的不足, 未来研究可以一方面从与抑郁症存在显著关联的语音特征入手, 进行更加精确的定位和探讨; 另一方面, 可以借助更加精巧的计算机分类算法, 并进行有效的重复性验证, 提高模型的精度和泛化能力, 增强结果的可推广性.

## 参考文献

1 Lépine J P, Briley M. The increasing burden of depression. Neuropsychiatr Dis Treat, 2011, 7: 3

2 Mitchell A J, Vaze A, Rao S. Clinical diagnosis of depression in primary care: A meta-analysis. Lancet, 2009, 374: 609–619

3 Schumann I, Schneider A, Kantert C, et al. Physicians' attitudes, diagnostic process and barriers regarding depression diagnosis in primary care: A systematic review of qualitative studies. Fam Pract, 2011, 29: 255–263
4 Inoue T, Inagaki Y, Kimura T, et al. Prevalence and predictors of bipolar disorders in patients with a major depressive episode: The Japanese epidemiological trial with latest measure of bipolar disorder (JET-LMBP). J Affect Disord, 2015, 174: 535–541
5 Vos T, Allen C, Arora M, et al. Global, regional, and national incidence, prevalence, and years lived with disability for 310 diseases and injuries, 1990–2015: A systematic analysis for the Global Burden of Disease Study 2015. Lancet, 2016, 388: 1545–1602
6 Mundt J C, Snyder P J, Cannizzaro M S, et al. Voice acoustic measures of depression severity and treatment response collected via interactive voice response (IVR) technology. J Neurolinguist, 2007, 20: 50–64
7 Mundt J C, Vogel A P, Feltner D E, et al. Vocal acoustic biomarkers of depression severity and treatment response. Biol Psychiatry, 2012, 72: 580–587
8 Cannizzaro M, Harel B, Reilly N, et al. Voice acoustical measurement of the severity of major depression. Brain Cogn, 2004, 56: 30–35
9 France D J, Shiavi R G, Silverman S, et al. Acoustical properties of speech as indicators of depression and suicidal risk. IEEE T Bio-Med Eng, 2000, 47: 829–837
10 Scherer S, Stratou G, Lucas G, et al. Automatic audiovisual behavior descriptors for psychological disorder analysis. Image Vision Comput, 2014, 32: 648–658
11 Stassen H H. Speaking behavior and voice sound characteristics in depressive patients during recovery. J Psychiat Res, 1993, 27: 289–307
12 Tolkmitt F, Helfrich H, Standke R, et al. Vocal indicators of psychiatric treatment effects in depressives and schizophrenics. J Commun Disord, 1982, 15: 209–222
13 Scherer K R, Sundberg J, Fantini B, et al. The expression of emotion in the singing voice: Acoustic patterns in vocal performance. J Acoust Soc Am, 2017, 142: 1805–1815
14 Moses P J. The Voice of Neurosis. Oxford: Grune & Stratton, 1954. 5
15 Ostwald P F. Soundmaking. The Acoustic Communication of Emotion. Oxford: Charles C Thomas, 1963. 12
16 Cohen A S, Najolia G M, Kim Y, et al. On the boundaries of blunt affect/alogia across severe mental illness: Implications for research domain criteria. Schizophr Res, 2012, 140: 41–45
17 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. Schizophr Res, 2012, 142: 93–95
18 Darby J K, Hollien H. Vocal and speech patterns of depressive patients. Folia Phoniatr Logo, 1977, 29: 279–291
19 Szabadi E, Bradshaw C M, Besson J A. Elongation of pause-time in speech: A simple, objective measure of motor retardation in depression. Br J Psychiatry, 1976, 129: 592–597
20 Flint A J, Black S E, Campbell-Taylor I, et al. Abnormal speech articulation, psychomotor retardation, and subcortical dysfunction in major depression. J Psychiatry Res, 1993, 27: 309–319
21 Alpert M, Pouget E R, Silva R R. Reflections of depression in acoustic measures of the patient's speech. J Affect Disord, 2001, 66: 59–69
22 Alghowinem S, Goecke R, Epps J, et al. Cross-cultural depression recognition from vocal biomarkers. In: the 17th Annual Conference of the International Speech Communication Association. France: International Speech Communication Association, 2016. 1943–1947
23 Alghowinem S, Goecke R, Wagner M, et al. Detecting depression: A comparison between spontaneous and read speech. In: 2013 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). New York: IEEE, 2013. 7547–7551
24 Mitra V, Shriberg E. Effects of feature type, learning algorithm and speaking style for depression detection from speech. In: 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). New York: IEEE, 2015. 4774–4778
25 Kaya H, Eyben F, Salah A A, et al. CCA based feature selection with application to continuous depression recognition from acoustic speech features. In: 2014 IEEE International Conference on Acoustics, Speech and Signal Processing. New York: IEEE, 2014. 3729–3733
26 Torres J, Saad A, Moore E. Evaluation of objective features for classification of clinical depression in speech by genetic programming. In: Soft Computing in Industrial Applications. Berlin: Springer, 2007. 132–143
27 Torres J, Saad A, Moore E. Application of a GA/Bayesian filter-wrapper feature selection method to classification of clinical depression from speech data. In: Soft Computing in Industrial Applications. Berlin: Springer, 2007. 115–121
28 Péron J, El Tamer S, Grandjean D, et al. Major depressive disorder skews the recognition of emotional prosody. Prog Neuro Psych, 2011, 35: 987–996
29 Sloan D M, Bradley M M, Dimoulas E, et al. Looking at facial expressions: Dysphoria and facial EMG. Biol Psychol, 2002, 60: 79–90
30 Gilboa-Schechtman E, Erhard-Weiss D, Jeczemien P. Interpersonal deficits meet cognitive biases: Memory for facial expressions in depressed and anxious men and women. Psychiatry Res, 2002, 113: 279–293

31 Stasak B, Epps J, Cummins N, et al. An investigation of emotional speech in depression classification. In: the 17th Annual Conference of the International Speech Communication Association. France : International Speech Communication Association, 2016. 485–489

32 Hergueta T, Baker R, Dunbar G C. The mini-international neuropsychiatric interview (MINI): The development and validation of a structured diagnostic psychiatric interview for DSM-IVand ICD-10. J Clin Psychiatry, 1998, 59(Suppl 20): 2233

33 American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders (DSM-5). Valencia: American Psychiatric Publishing, 2013

34 Liu Z, Hu B, Li X, et al. Detecting depression in speech under different speaking styles and emotional valences. In: International Conference on Brain Informatics. Cham: Springer, 2017. 261–271

35 Wang J Y. An Exploratory study on auxiliary diagnosis of depression based on speech (in Chinese). Doctor Dissertation. Beijing: Graduate School of Chinese Academy of Sciences, 2017 [汪静莹. 抑郁症的辅助诊断研究——基于语音特征的探索. 博士学位论文. 北京: 中国科学院研究生院, 2017]

36 Eyben F, Weninger F, Gross F, et al. Recent developments in opensmile, the munich open-source multimedia feature extractor. In: Proceedings of the 21st ACM international conference on Multimedia. New York: ACM, 2013. 835–838

37 Hall M, Frank E, Holmes G, et al. The WEKA data mining software: An update. ACM SIGKDD Explor Newslett, 2009, 11: 10–18

38 Dreiseitl S, Ohno-Machado L. Logistic regression and artificial neural network classification models: A methodology review. J Biomed Inform, 2002, 35: 352–359

39 Powers D M. Evaluation: From precision, recall and F-measure to ROC, informedness, markedness and correlation. J Machine Learning Tech, 2011, 2: 37–83

40 Ling C X, Huang J, Zhang H. AUC: A statistically consistent and more discriminating measure than accuracy. In: Proceedings of the 18th International Joint Conference on Artificial Intelligence. San Francisco: Morgan Kaufmann Publishers Inc., 2003. 519–524

41 Cummins N, Epps J, Sethu V, et al. Modeling spectral variability for the classification of depressed speech. Genes Immun, 2013, 1: 857–861

42 Cohn J F, Kruez T S, Matthews I, et al. Detecting depression from facial actions and vocal prosody. Affective Computing and Intelligent Interaction and Workshops, 2009. ACII 2009. In: 3rd International Conference on. New York: IEEE, 2009. 1–7

43 Ellgring H, Scherer K R. Vocal indicators of mood change in depression. J Nonverbal Behav, 1996, 20: 83–110

44 Nilsonne A. Speech characteristics as indicators of depressive illness. Acta Psychiatr Scan, 1988, 77: 253–263

45 Savitz J, Drevets W C. Bipolar and major depressive disorder: Neuroimaging the developmental-degenerative divide. Neurosci Biobehav Rev, 2009, 33: 699–771

46 Peluso M A, Glahn D C, Matsuo K, et al. Amygdala hyperactivation in untreated depressed individuals. Psychiatry Res, 2009, 173: 158–161

47 Siegle G J, Steinhauer S R, Thase M E, et al. Can't shake that feeling: Event-related fMRI assessment of sustained amygdala activity in response to emotional information in depressed individuals. Biol Psychiatry, 2002, 51: 693–707

48 Thomas K M, Drevets W C, Dahl R E, et al. Amygdala response to fearful faces in anxious and depressed children. Arch Gen Psychiatry, 2001, 58: 1057–1063

49 Epstein J, Pan H, Kocsis J H, et al. Lack of ventral striatal response to positive stimuli in depressed versus normal subjects. Am J Psychiatry, 2006, 163: 1784–1790

50 Fu C H, Williams S C, Cleare A J, et al. Attenuation of the neural response to sad faces in major depression by antidepressant treatment: A prospective, event-related functional magnetic resonance imaging study. Arch Gen Psychiatry, 2004, 61: 877–889

51 Surguladze S, Brammer M J, Keedwell P, et al. A differential pattern of neural response toward sad versus happy facial expressions in major depressive disorder. Biol Psychiatry, 2005, 57: 201–209

52 Grimm S, Beck J, Schuepbach D, et al. Imbalance between left and right dorsolateral prefrontal cortex in major depression is linked to negative emotional judgment: An fMRI study in severe major depressive disorder. Biol Psychiatry, 2008, 63: 369–376

53 Drevets W C. Neuroimaging and neuropathological studies of depression: implications for the cognitive-emotional features of mood disorders. Curr Opin Neurobiol, 2001, 11: 240–249

54 Chan S W, Harmer C J, Goodwin G M, et al. Risk for depression is associated with neural biases in emotional categorisation. Neuropsychologia, 2008, 46: 2896–2903

55 Anand A, Li Y, Wang Y, et al. Activity and connectivity of brain mood regulating circuit in depression: A functional magnetic resonance study. Biol Psychiatry, 2005, 57: 1079–1088

56 Hendin H, Maltsberger J T, Szanto K. The role of intense affective states in signaling a suicide crisis. J Nerv Ment Dis, 2007, 195: 363–368

57 Beck A T. Depression: Clinical, Experimental, and Theoretical Aspects. Philadelphia: University of Pennsylvania Press, 1967

58 Alghowinem S, Goecke R, Wagner M, et al. A comparative study of different classifiers for detecting depression from spontaneous

speech. In: 2013 IEEE International Conference on Acoustics, Speech and Signal Processing. New York: IEEE, 2013. 8022–8026
59 Porritt L L, Zinser M C, Bachorowski J A, et al. Depression diagnoses and fundamental frequency-based acoustic cues in maternal infant-directed speech. Lang Learn Dev, 2014, 10: 51–67
60 Lever J, Krzywinski M, Altman N. Points of significance: Logistic regression. Nat Methods, 2016, 13: 541–542
61 Alghowinem S, Goecke R, Wagner M, et al. Multimodal depression detection: Fusion analysis of paralinguistic, head pose and eye gaze behaviors. In: IEEE Transactions on Affective Computing. New York: IEEE, 2016
62 Yang F, Li Y, Xie D, et al. Age at onset of major depressive disorder in Han Chinese women: Relationship with clinical features and family history. J Affect Disord, 2011, 135: 89–94
63 Yang F, Zhao H, Wang Z, et al. Age at onset of recurrent major depression in Han Chinese women–A replication study. J Affect Disord, 2014, 157: 72–79
64 Gan Z, Li Y, Xie D, et al. The impact of educational status on the clinical features of major depressive disorder among Chinese women. J Affect Disord, 2012, 136: 988–992
65 Wang D C. Introduction to Linguistics (in Chinese). Shanghai: Shanghai Foreign Language Education Press, 1997. 60–95 [王德春. 语言学概论. 上海: 上海外语教育出版社, 1997. 60–95]
66 Wang Y, Yang J A, Liu H, et al. Bottleneck feature extraction method based on hierarchical deep sparse belief network (in Chinese). Pattern Recogn Artif Intell, 2015, 28: 173–180 [王一, 杨俊安, 刘辉, 等. 基于层次稀疏 DBN 的瓶颈特征提取方法. 模式识别与人工智能, 2015, 28: 173–180]
67 Huang K Y, Wu C H, Kuo Y T, et al. Unipolar depression vs. bipolar disorder: An elicitation-based approach to short-term detection of mood disorder. Depression, 2016, 10: 12

## 补充材料

补充材料 1　实验使用材料
补充材料 2　结果图片

本文以上补充材料见网络版 csb.scichina.com. 补充材料为作者提供的原始数据, 作者对其学术质量和内容负责.

Summary for “基于语音的抑郁症识别”

# Depression recognition based on speech analysis

Wei Pan[1,2], Jingying Wang[1,2], Tianli Liu[3], Xiaoqian Liu[1], Mingming Liu[1,2], Bin Hu[4] & Tingshao Zhu[1*]

[1] *Institute of Psychology, Chinese Academy of Sciences, Beijing 100101, China;*
[2] *Department of Psychology, University of Chinese Academy of Sciences, Beijing 100049, China;*
[3] *Institute of Population Research, Peking University, Beijing 100871, China;*
[4] *School of Information Science and Engineering, Lanzhou University, Lanzhou 730000, China*
* Corresponding author, E-mail: tszhu@psych.ac.cn

Depression is one of the common mental diseases. Patients with depression often have depressed moods such as sadness, guilty, low self-esteem, loss of interest, hypofunction and so on. They suffer from serious emotional problems, unexplained suffering, which has caused enormous losses to individuals, families and society. According to the World Health Organization, there are aproximately 322 million people suffering from depression in the whole world in 2017. While there are about 54 million depressive patients in China.

Depression can be cured effciently. However, due to the complexity of the pathogenesis of depression, clinical diagnosis is accompanied with many difficulties. Firstly, the mental disease, especially depression, are not getting enough attention and even being misinterpreted by other people. Secondly, the depression patients are less willing to ask for help. Thirdly, it is hard to select and dignose the potential depression patients precisely, as well as there are limited medical resource for depression diagnosis.

It is necessary to find a more convenient, objective and efficient way to assist the fast identification of depression. As a relatively objective and easily accessible variable, speech has its potential value. The speech of patient is easy to acquire, and also, it has been proved that the sound of depressed patients have special charcteristics such as slow speech rate, lack of cadence and so on. The purpose of this paper is to explore the relationship between speech and depression by establishing classification models of voice feature and depression prediction. In this research, 3(emotion mood: positive, neutral, negative)×3(task type: question answering, text reading, picture description) experimental design was employed, and the voice data was collected from the speech of individuals recorded during different tasks. 103 participants were inculded in this study, including 45 depression patients (age: 23.8–44.6, $M$=34.2, $SD$=10.4, males=22, females=23) and 58 healthy ones (age: 20.1–41.7, $M$=30.9, $SD$=10.8, males=27, females=31). The former were recruited in the hospital in Beijing Anding Hospital and Huilongguan Hospital, while the latter were recruited by advertisement. All of them were diagnosed by specialist with DSM-IV and MINI interview. All participants did not have substance abuse, substance dependence, personality disorders and other mental diseases, no serious physical illness or suicidal behavior. The education level of subjects are all above the elementary school. 988 Voice features were extracted from the speech data using open SMILE software. Logistic regression, a machine learning method, was used to train the predicting models. Results showed that the precision rate of predicting can reach to 82.9%. Based on machine learning methods, this paper employed voice features to establish predicting models of depression. Results show the speech of depression patients has certain predicting effect, which paves the way for the further identification of depression in a more thorough way.

**depression, voice feature, classification algorithm, logistic regression**

doi: 10.1360/N972017-01250