

文章编号: 1674-8190(2024)02-049-07

# 基于语谱图的管制员疲劳状态检测研究

杨昌其, 冯筱晴, 张雨萱, 蔡子牛

(中国民用航空飞行学院 空中交通管理学院, 广汉 618307)

**摘要:** 现阶段利用陆空通话语音对管制员疲劳状态的研究中, 大多只考虑了语音在时域或频域的变化, 而忽视了疲劳会同时在时域与频域上产生影响。将三种疲劳状态下的陆空通话语音分别转化为可同时反映时域与频域特性的语音频谱图像, 利用灰度共生矩阵提取四维典型的特征参数, 对比管制员在不同状态下特征参数的变化情况, 构建管制员疲劳检测模型并对输入特征进行检测。结果表明: 利用语谱图特征结合传统特征作为输入特征的检测准确率最高, 达到 95.49%, 较单一使用传统特征的检测准确率高出 4%; 管制员疲劳状态的变化会直观地反映在语谱图上, 会对其特征值产生影响, 利用这种影响对管制员疲劳状态进行检测, 可以得到良好的检测结果。

**关键词:** 管制员; 疲劳检测; 语谱图; 灰度共生矩阵; 机器学习

**中图分类号:** V355

**文献标识码:** A

**DOI:** 10.16615/j.cnki.1674-8190.2024.02.06

## Research on controller fatigue state detection based on speech spectrogram

YANG Changqi, FENG Xiaoqing, ZHANG Yuxuan, CAI Ziniu

(College of Air Traffic Management, Civil Aviation Flight University of China, Guanghan 618307, China)

**Abstract:** In the current research on the fatigue state of controllers using radiotelephony communication, most of them only consider the changes of voice in the time domain or frequency domain, while ignore that fatigue will affect the time domain and frequency domain at the same time. The voice of radiotelephony communication in the three fatigue states is converted into speech spectrum images that can reflect the characteristics of both the time domain and frequency domain, and the grayscale co-existence matrix is used to extract the typical feature parameters in four dimensions. The changes of the characteristic parameters of the controllers in different states are compared, and the controller fatigue detection model is constructed to detect the input features. The results show that the detection accuracy of using the spectrogram features combined with the traditional features as the input features is the highest (95.49%), which is 4% higher than that of the traditional features alone. The change of controllers fatigue state can intuitively reflected on the spectrogram and has an impact on its eigenvalues, and the better detection results can be obtained by using this influence to detect the controllers fatigue state.

**Key words:** controller; fatigue detection; spectrogram; grayscale co-existence matrix; machine learning

收稿日期: 2023-03-07; 修回日期: 2023-07-13

通信作者: 冯筱晴(1998-), 女, 硕士研究生。E-mail: 2993885139@qq.com

引用格式: 杨昌其, 冯筱晴, 张雨萱, 等. 基于语谱图的管制员疲劳状态检测研究[J]. 航空工程进展, 2024, 15(2): 49-55.

YANG Changqi, FENG Xiaoqing, ZHANG Yuxuan, et al. Research on controller fatigue state detection based on speech spectrogram[J]. Advances in Aeronautical Science and Engineering, 2024, 15(2): 49-55. (in Chinese)

## 0 引言

在导致航空器事故或事故征候的原因中,人为因素是其中最重要的原因之一<sup>[1]</sup>,为了避免事故的发生,应对可能出现的由人为因素导致的差错进行识别。近年来,随着航班量的持续增加,空域态势变得更为复杂,管制员的工作负荷随着发布的指令数量,尤其是随着非固定管制指令数量的增加而增大<sup>[2]</sup>,同时由于存在同一终端不同扇区工作负荷分配不均的情况<sup>[3]</sup>,负责高负荷扇区指挥工作的管制员更易产生疲劳积累。据美国联邦航空管理局的统计,每年约有 1/5 的航空器安全事故与管制员疲劳有关<sup>[4]</sup>。因此,准确识别管制员的疲劳状态,并采取措施进行解决,可以有效降低运行风险。

陆空通话作为管制员指挥工作中不可或缺的一环,选择对陆空通话语音作为研究对象进行特征提取更易获得可用性高的数据,国内外研究者对此进行了大量的研究。

目前,最常用作语音疲劳状态检测的特征是由 Cowie 等<sup>[5]</sup>总结的频谱、梅尔频率倒谱系数特征(MFCC)系数、共振峰以及韵律特征;Wu Nan 等<sup>[6]</sup>提出一种基于自适应量子遗传算法异构学习的集成学习方法,将多种机器学习方法进行异构集成,对提取到的 MFCC 进行疲劳检测;Shen Zhiyuan 等<sup>[7]</sup>首先对语音信号进行小波分解和分型差分(FD)计算,提取语音频域内的特征并利用支持向量机进行检测,得到 92.82% 的检测准确率;赵强<sup>[8]</sup>通过神经网络选取了语音短时能量、短时平均过零率、梅尔频率倒谱系数、基音频率和第一共振峰五种语音特征对语音进行疲劳状态检测,并对比不同疲劳状态下特征的变化情况,得出五种语音特征随疲劳程度的加深呈明显下降的趋势,但其对语音疲劳程度的判定来源于说话人的主观感受,并未对其进行量化。

目前已有包含识别管制员的语速、语调、声纹等特征的“智能无感式空中交通管制员状态监测告警系统”在华北空管局进行验证,证明了通过陆空通话语音对管制员疲劳状态识别的现实可行性。

疲劳状态会在音频的时域与频域中同时产生影响,而上述研究仅单独考虑了时域与频域上的

变化。考虑到语音频谱图像(简称语谱图)能够在二维图像中同时反映音频与时频特性且在语音情感分析领域得到了较为成熟的应用及较好的成果,Darekar 等<sup>[9]</sup>使用卷积神经网络(CNN)从被分割成帧的信号中恢复语谱图特征,并将其与提取的声学特征形成混合特征向量,通过 E-Relief 混合选择最重要的特征并利用分类器集合建模进行语音情感识别,其检测结果的准确率较传统模型提高了 17.5%。

目前将语音频谱图像结合传统特征进行疲劳状态识别的应用较少。本文将管制员复训的模拟机音频作为研究对象,通过模拟管制的方式获取不同疲劳状态下的音频数据并生成对应的语谱图;利用灰度共生矩阵在四个方向进行特征提取,构建基于陆空通话数据的管制员疲劳状态检测模型,通过检测结果验证所选特征对管制员疲劳状态的区分能力。

## 1 语音数据采集及语谱图的生成

### 1.1 陆空通话语音数据采集及标记

为满足实验需求,采集管制员使用模拟机进行管制训练时的语音数据,以此为实验数据,由于模拟机很大程度上还原了管制员日常工作内容,因此将模训数据作为实验数据具有很高的可靠性。

在开始采集语音数据前,需保证录制环境无特别明显的噪音,录制音频格式设定为 32 bit, 48 000 Hz,录制内容为管制员日常标准陆空通话。

只简单将管制员的状态判定为疲劳或不疲劳并不能满足实际的检测需求,应当对管制员疲劳等级进行细化分类<sup>[10]</sup>,因此设计语音样本的类别标记标准如下:要求被测人员填写满分为百分制的疲劳自测量表,在其数值大于 60 分时,通过精神运动警觉性任务(PVT)测试获得人员反应时间,该测试通过获取人员对屏幕上出现色块颜色发生变化时的点击速度,测试人员的反应时间。该测试持续 5 min,以其结果作为人员平均反应时间,将其视为正常状态下的正常反应时间。音频录制开始后要求被测人员在模拟训练时每半小时填写一次疲劳自测量表并进行反应时间测试,当疲劳自测量表结果小于 50 分、进行时长为 5 min 的

PVT测试的平均反应时间大于正常状态的平均反应时间时视为一般疲劳状态,此时受测人员的状态是感到疲劳但基本不会对管制工作造成影响;当疲劳自测量表结果小于40分、进行时长为5 min的PVT测试的平均反应时间远大于正常状态的平均反应时间(超过80 ms)视为重度疲劳状态,此时受测人员反应速度下降,思维迟缓,会对管制工作产生影响。

## 1.2 语音频谱图像的生成

为了消除低频与高频能量分布不均匀的问题,首先要对信号进行预加重,起到对语音低频信号起到抑制、对高频信号产生增强的作用<sup>[11]</sup>;然后使用窗函数对时域内的语音数据进行截断,这一步被称为分帧与加窗;最后采用短时傅里叶变换(STFT)获得每帧信号的频谱图,对其每一帧的功率谱取对数并在时间维度上拼接,获得管制员陆空通话的语音频谱图像<sup>[12]</sup>,语谱图生成过程如图1所示。

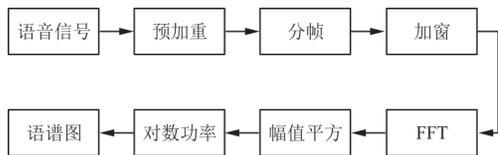


图1 语音频谱图像生成过程

Fig. 1 Voice spectrum image generation process

在语谱图中,幅值是由颜色的不同深度反映的,因此语谱图可以在二维的图像上同时反映时间、频率、幅值三个维度的信息。

## 1.3 疲劳状态下语谱图变化

当人处于疲劳状态时,会使身体在一段时间内产生无法控制的变化,对发生的影响从宏观的角度看包括语调升高、语速减慢。这些变化同样会反映在语音频谱图像中,由于控制发声部位的肌肉松弛以及声道壁与气流的摩擦,会导致共振峰降低,且作用范围在200~3 000 Hz处最为明显<sup>[13]</sup>,因此将输出图片的频率范围设置为0~3 500 Hz,可以有效放大易受疲劳影响作用的区域,方便后续提取图片中的纹理特征,增加疲劳检测的准确性。

从语谱图中可以直观地表现出音频的频域随着时间的变化并展示出部分特征变化,选取管制员PVT测试结果分别为383、421、504 ms,其标记分别为正常状态、一般疲劳以及重度疲劳三种状态下同一句“可以着陆跑道36R”(帧长取2 048,帧移取50%,加汉宁窗)的管制指令,分别生成正常频率范围及0~3 500 Hz的语谱图图像,如图2所示。

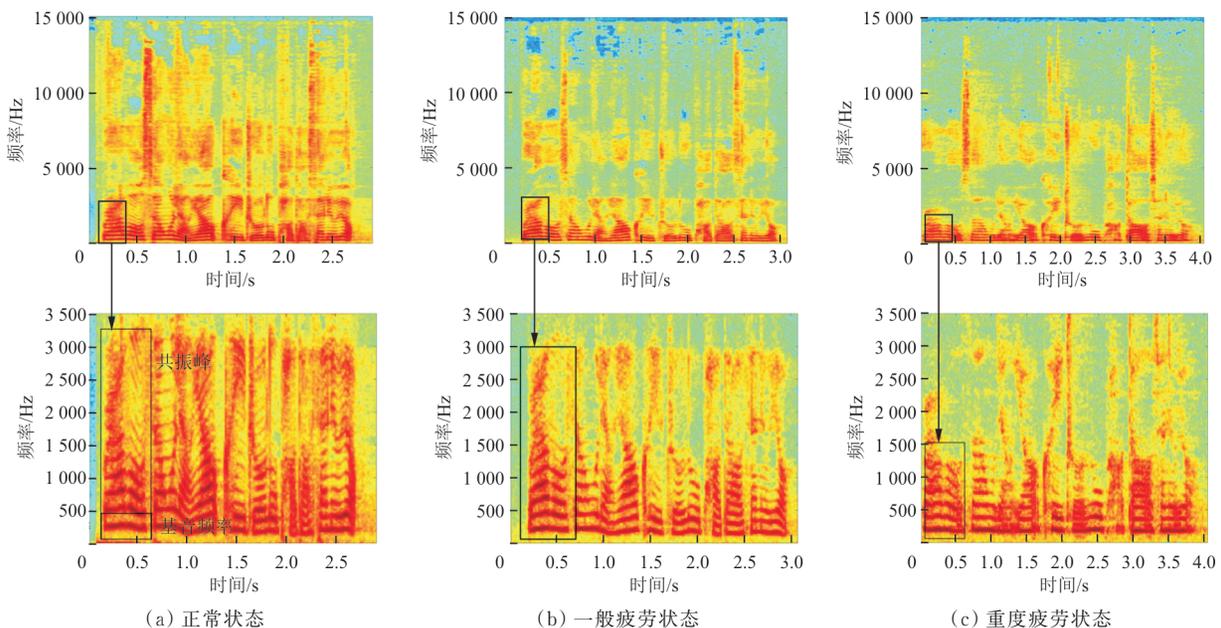


图2 不同状态下的语音频谱图像对比

Fig. 2 Comparison of speech spectrum images in different states

对比三种不同疲劳状态对应的语谱图,随着管制员疲劳程度的加深,发布相同指令所使用的时长增加;颜色深度降低,反映了音频幅值的降低。在频率为 0~3 500 Hz 的三张图中,最低的横纹与其谐波分别反映了基音频率与共振峰的变化,对比黑色框线部分可知,共振峰及基音频率随着疲劳程度的加深呈下降趋势。

## 2 灰度共生矩阵

灰度共生矩阵是一种典型的统计方法,描述空间上从灰度为  $i$  的像素点出发,达到固定位置  $d$ 、方向为  $\theta$  的点的灰度值为  $j$  的概率  $P^{[13]}$ 。通过灰度共生矩阵可以得到语谱图上纹理特征的规律及变化规则<sup>[14]</sup>。

能量可以表现语谱图的纹理粗细度、判断灰度分布是否均匀的特征。若灰度共生矩阵的元素值接近,则得出的能量值较小;若元素值差异较大,则能量值较大<sup>[15]</sup>。

$$A = \sum_i \sum_j \{P_d(i, j, d, \theta)\}^2 \quad (1)$$

熵可以反映陆空通话语音的频谱图像的纹理复杂性,其纹理复杂性与熵值成正比例关系<sup>[16]</sup>。

$$E = - \sum_i \sum_j P_d(i, j, d, \theta) \log P_d(i, j, d, \theta) \quad (2)$$

对比度可以反映矩阵中值的分布,展示陆空通话的语谱图中局部的变化,其值的大小反映了图片纹理沟纹的深浅与效果<sup>[15]</sup>。

$$C = \sum_i \sum_j (i - j)^2 P_d(i, j, d, \theta) \quad (3)$$

相关性可以在一定程度上反映灰度共生矩阵各元素在行和列方向上的相似性,相关性与矩阵元素之间的差值成正比关系。

$$C_{or} = \frac{\sum_i \sum_j ij P_d(i, j, d, \theta) - \mu_1 \mu_2}{\sigma_1 \sigma_2} \quad (4)$$

其中:

$$\mu_1 = \sum_{i=0} \sum_{j=0} P_d(i, j, d, \theta) \quad (5)$$

$$\mu_2 = \sum_{i=0} \sum_{j=0} P_d(i, j, d, \theta) \quad (6)$$

$$\sigma_1^2 = \sum_{i=0} (i - \mu_1)^2 \sum_{j=0} P_d(i, j, d, \theta) \quad (7)$$

$$\sigma_2^2 = \sum_{j=0} (j - \mu_2)^2 \sum_{i=0} P_d(i, j, d, \theta) \quad (8)$$

## 3 实验

### 3.1 模型构建

考虑到无法在管制员执勤时对其进行测试以确定当下时刻指令对应的疲劳标记,选取管制员使用模拟机训练时搜集到的数据进行分析,探究管制员处于不同疲劳程度时,语音频谱图像上产生的变化及所选取特征的区分度。以受测人疲劳自测量表结果及 PVT 测试结果作为标签,以灰度共生矩阵作为提取语谱图上特征的手段,将语谱图特征与传统特征进行结合,采用不同分类器对特征进行分类,以达到对受测人疲劳状态检测的目的,管制员疲劳检测模型如图 3 所示。

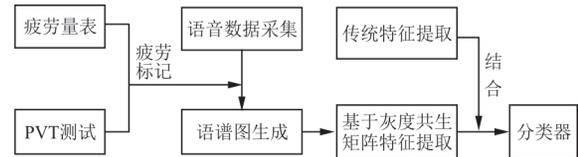


图3 管制员疲劳检测模型

Fig. 3 Controller fatigue detection model

具体步骤如下:

1) 语音数据采集与标记:在安静、无明显噪声的环境下架设外置录音设备进行语音数据的采集,按前述方式对音频进行标记,录制格式固定为 32 bit、48 000 Hz。共采集陆空通话语音 1 219 条,其中正常状态下的陆空通话语音 396 条,一般疲劳状态下的陆空通话语音 464 条,重度疲劳状态下的陆空通话语音 359 条。

2) 数据预处理及语音频谱图像的生成:将采集到的语音数据进行预处理,经降噪、分帧、加窗后,生成每条语音对应的语音频谱图像,其中帧长取 2 048,帧移取 50%,加汉宁窗,频率选取为 0~3 500 Hz。

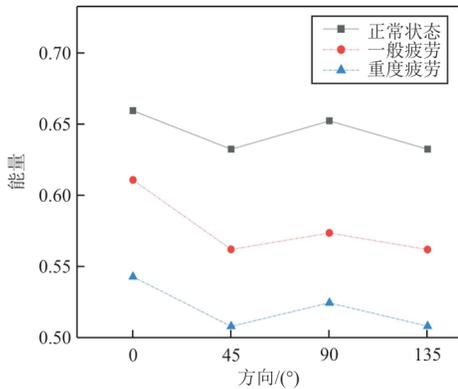
3) 特征提取:利用灰度共生矩阵,提取出对需要检测的三种不同状态有良好区分度的特征,以提高分类准确性。其中位置关系  $d$  取 1,灰度级为 128,选取  $0^\circ$ 、 $45^\circ$ 、 $90^\circ$ 、 $135^\circ$  四个方向的结果并取其均值作为特征值作为后续分类器的输入特征。

4) 检测分类:将提取到的语谱图特征作为输入特征,并与传统特征结合,选取不同分类器进行训练,获取可以对受测人处于何种疲劳状态的分类能力及准确性,并以此作为所选特征区分度是

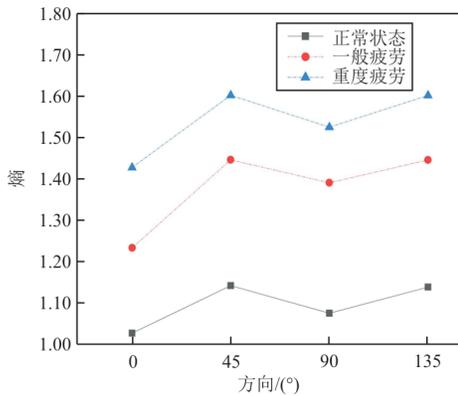
否明显的判断依据。

### 3.2 灰度共生矩阵特征值提取及分析

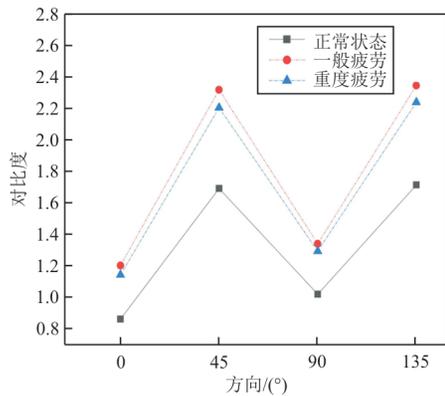
利用 Matlab 2017a 进行编程,对采集到的音频数据生成的语谱图取  $\theta$  为  $0^\circ$ 、 $45^\circ$ 、 $90^\circ$ 、 $135^\circ$  四个方向的灰度共生矩阵进行计算,取位置关系  $d=1$ ,灰度级为 128,并基于灰度矩阵对以上四维纹理特征进行计算,选取通过上述三种不同状态的图像得来的特征值,得到四个方向的特征变化结果如图 4 所示。



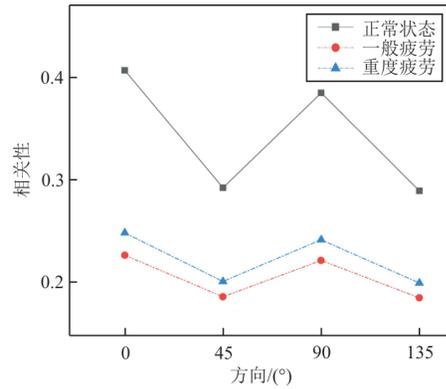
(a) 能量



(b) 熵



(c) 对比度



(d) 相关性

图 4 不同状态下 GLCM 四种特征提取结果

Fig. 4 Four features of GLCM extracted in different states

从图 4 可以看出:随着疲劳状态的加深,能量值明显降低,熵值则明显增加,说明随着疲劳程度的加深,语谱图的分布变得更加不均匀,且纹理复杂度增加,可以通过能量值与熵值在后续的研究中对三种状态进行区分。

正常状态在对比度和相关性上与疲劳状态区分度明显,说明正常状态的纹理沟纹相比疲劳状态较浅但是灰度相关性更大,可以通过对比度与相关性对正常状态与疲劳状态进行区分。虽然依据对比度与相关性不易区分一般疲劳状态与重度疲劳状态,但这两种特征可以提供四个方向的局部变化规律,因此将这两种特征结合其他特征进行一般疲劳状态与重度疲劳状态的分类识别。

通过上述分析,使用能量、熵值、对比度、相关性四个特征能够有效对三种状态的语音频谱图像进行分类识别。为了提高分类结果的准确率,采用计算四个方向参数值的均值与方差以抑制生成方向  $\theta$  对特征提取结果的影响,得到的结果如表 1~表 2 所示。

表 1 四维特征参数均值

Table 1 Four-dimensional feature parameter mean

疲劳状态	能量	熵	对比度	相关性
正常	0.643 912	1.094 999	1.318 752	0.343 609
一般疲劳	0.576 745	1.379 191	1.797 979	0.204 334
重度疲劳	0.520 440	1.539 960	1.716 782	0.222 257

表2 四维特征参数方差

疲劳状态	能量	熵	对比度	相关性
正常	0.0139 090	0.055 269	0.444 617	0.061 611
一般疲劳	0.023 186	0.101 048	0.615 323	0.022 477
重度疲劳	0.016 592	0.083 024	0.582 761	0.026 138

### 3.3 检测结果分析

除所提取的能量、熵、对比度、相关性四维语谱图特征外,本文另外选取了已经在文献[6]中证实具有区分度的短时平均能量、短时过零率、共振峰、平均基音频率四维传统特征,将传统特征与语谱图特征组成的八维特征同时作为分类器输入特征进行检测。

表3 不同分类器测试结果对比

Table 3 Comparison of different classifier test results

分类器	疲劳状态	语谱图特征		传统特征		两种特征结合	
		准确率/%	平均准确率/%	准确率/%	平均准确率/%	准确率/%	平均准确率/%
SVM	正常	79.00		75.66		78.44	
	一般疲劳	73.01	80.42	66.72	77.05	73.69	82.08
	重度疲劳	89.25		88.77		94.01	
KNN	正常	89.91		84.92		90.49	
	一般疲劳	87.53	90.65	81.64	86.01	88.11	91.36
	重度疲劳	94.5		91.48		95.49	
朴素贝叶斯	正常	75.88		72.46		72.21	
	一般疲劳	67.76	76.32	60.41	67.54	66.07	75.08
	重度疲劳	85.32		69.75		86.97	

从表3可以看出:传统特征可以对管制员的三种不同状态有较好的检测效果,使用KNN作为分类器时平均准确率可达86.01%;采用本文选取的四维语谱图特征作为输入特征进行检测时,准确率可以提高3%~4%;将本文选用的语谱图特征与传统特征结合,作为输入特征进行检测时,准确率较传统特征提高5%~6%;采用KNN作为分类器进行检测,其平均准确率可达91.36%,对重度疲劳的检测准确率达到95.49%,对正常状态及一般疲劳状态的检测准确率也分别达到了88.11%和90.49%。由此可以说明,利用并结合语音频谱图像上的特征对管制员的疲劳状态进行检测,相较于使用单一的传统特征进行检测,可得到更优

的检测结果;同时也说明将语谱图特征与传统特征相结合,在区分管制员的疲劳程度方面有着更好的分类能力。

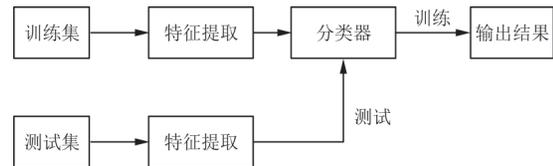


图5 分类器工作流程

Fig. 5 Classifier workflow

在分类器的选取方面,本文同时选取在机器学习领域较为成熟的支持向量机(SVM)、最近邻(KNN)以及朴素贝叶斯算法分别进行对比实验,实验结果如表3所示。

## 4 结论

1) 本文将管制员陆空通话语音转化为语音频谱图像,利用灰度共生矩阵提取了四个典型特征(能量,熵,对比度,相关性)进行分析研究,在确认所选特征对于管制员的三种疲劳状态具有一定区分性后,将提取到的语谱图特征与传统特征相结合,输入到管制员疲劳检测模型中。

2) 利用KNN作为分类器进行检测时,检测准确率可达91.3%,其中对重度疲劳的检测效果达

到95.49%,检测效果良好。

考虑到本文存在选取特征维度较少以及数据来源为模拟机而非真实数据的问题,在后续的研究中将增加特征维度并对参数进行调整,进一步提高所选特征对不同疲劳状态的区分能力,并选用真实的管制员陆空通话数据作为数据库,使研究更具有现实意义。

### 参考文献

- [1] 张建祥,甘旭升,吴亚荣,等.空管人因失效概率的THERP-CREAM预测方法[J].航空工程进展,2022,13(6):59-68.  
ZHANG Jianxiang, GAN Xusheng, WU Yarong, et al. THERP-CREAM prediction method for failure probability of human factor of air traffic control [J]. Advances in Aeronautical Science and Engineering, 2022, 13(6): 59-68. (in Chinese)
- [2] 何媛,赵晓伟,甘旭升,等.基于管制员工作负荷的扇区开合阀设置研究[J].航空工程进展,2022,13(3):127-136.  
HE Yuan, ZHAO Xiaowei, GAN Xusheng, et al. Research on setting of sector opening and closing valve based on controller workload [J]. Advances in Aeronautical Science and Engineering, 2022, 13(3): 127-136. (in Chinese)
- [3] 令璐璐,罗军.基于管制员负荷的西安终端区扇区优化[J].航空工程进展,2017,8(3):293-298.  
LING Lulu, LUO Jun. Research on Xi'an terminal sector optimization base on controller workload [J]. Advances in Aeronautical Science and Engineering, 2017, 8(3): 293-298. (in Chinese)
- [4] International Civil Aviation Organization. Fatigue risk management systems manual for regulators: DOC9966 [R]. US: International Civil Aviation Organization, 2016.
- [5] COWIE R, COX C, MARTIN J C, et al. Issues in data labelling[M]. Berlin: Emotion-Oriented Systems, 2011.
- [6] WU Nan, SUN Jingjuan. Fatigue detection of air traffic controllers based on radiotelephony communications and self-adaption quantum genetic algorithm optimization ensemble learning[J]. Applied Sciences, 2022, 12(20): 1-12.
- [7] SHEN Zhiyuan, PAN Guozhuang, YAN Yonggang. A high-precision fatigue detecting method for air traffic controllers based on revised fractal dimension feature [J]. Mathematical Problems in Engineering, 2020, 10: 1-8.
- [8] 赵强.基于神经网络的语音疲劳度检测[D].北京:北京交通大学,2019.  
ZHAO Qiang. Speech fatigue detection based on neural network [D]. Beijing: Beijing Jiaotong University, 2019. (in Chinese)
- [9] DAREKAR R V, CHAVAND M S, SHARANYAA S, et al. A hybrid meta-heuristic ensemble based classification technique speech emotion recognition[J]. Advances in Engineering Software, 2023, 180: 234-241.
- [10] 沈志远,魏义涛,闫永刚,等.空中交通管制员疲劳检测与管理综述[J].航空工程进展,2021,12(6):26-38.  
SHEN Zhiyuan, WEI Yitao, YAN Yonggang, et al. A review on fatigue detection and management for air traffic controllers[J]. Advances in Aeronautical Science and Engineering, 2021, 12(6): 26-38. (in Chinese)
- [11] 罗海涛.基于C语言的语谱图生成[J].现代计算机,2022,28(14):71-74.  
LUO Haitao. Speech spectrum image generation based on C language[J]. Modern Computer, 2022, 28(14): 71-74. (in Chinese)
- [12] 王鹏,周成成,吴文琪,等.基于语谱图的高频分量进行音频篡改研究[J].信息技术与信息化,2022(9):62-65.  
WANG Peng, ZHOU Chengcheng, WU Wenqi, et al. Research on audio tampering based on the high-frequency components of the spectrogram [J]. Information Technology and Informatization, 2022(9): 62-65. (in Chinese)
- [13] 李响,李国正,邓明君,等.基于语音频谱图像特征的人体疲劳检测方法[J].仪器仪表学报,2021,42(2):123-132.  
LI Xiang, LI Guozheng, DENG Mingjun, et al. A human fatigue detection method based on speech spectrogram features[J]. Chinese Journal of Scientific Instrument, 2021, 42(2): 123-132. (in Chinese)
- [14] DWAICH H A, ABDULBAQI H A. Signature texture features extraction using GLCM approach in android studio[J]. Journal of Physics: Conference Series, 2021(1): 1804-1810.
- [15] 刘涛,尹志强,雷经发,等.金属疲劳损伤过程中表面形貌的多分辨特征提取[J].激光与光电子学进展,2022,59(8):444-453.  
LIU Tao, YIN Zhiqiang, LEI Jingfa, et al. Multiresolution feature extraction of surface topography in metal fatigue damage process[J]. Laser & Optoelectronics Progress, 2022, 59(8): 444-453. (in Chinese)
- [16] 刘安琪,刘华勇,王焕宝.改进的SLIC超像素图像分割与合并算法[J].安徽建筑大学学报,2020,28(4):39-46.  
LIU Anqi, LIU Huayong, WANG Huanbao. Improved super-pixel image segmentation and merging algorithm [J]. Journal of Anhui Jianzhu University, 2020, 28(4): 39-46. (in Chinese)

(编辑:丛艳娟)