基于机器学习的语音情感声学特征筛选

董文琪[1]，**王涵**[1]，**张璟玮**[12*]
[1]澳门大学人文学院中国语言文学系
[2]澳门大学认知与脑科学研究中心
yc37717@um.edu.mo；yc37729@um.edu.mo；jwzhang@um.edu.mo

摘要

筛选有效表达情感的声学特征对语音情感研究至关重要。对具有相同或相似声学特征的情感，声学研究中仅使用基频和时长无法有效区分。本研究扩大声学参数的种类和数量，使用三种机器学习方法，筛选出区分情感类型的多组有效声学参数，补充和完善语音情感声学研究的声学特征集。研究发现，区分不同情感所依赖的声学参数、参数数量、参数贡献都不相同，其中频谱和信噪参数发挥重要作用。本研究为语音情感声学分析的参数选择提供参考。

关键词： 声学特征；语音情感分类；机器学习；频谱；噪音水平

Acoustic Feature Selection for Speech Emotion Based on Machine Learning

DONG Wenqi[1], **WANG Han**[1], **ZHANG Jingwei**[12*]
[1]Department of Chinese Language and Literature, Faculty of Arts and Humanities, University of Macau
[2]Centre for Cognitive and Brian Sciences, University of Macau
yc37717@um.edu.mo;yc37729@um.edu.mo;jwzhang@um.edu.mo

Abstract

The selection of acoustic features that effectively convey emotions is crucial for research on speech emotion. For emotions with similar or identical acoustic characteristics, the differentiation solely through fundamental frequency and duration proves inadequate within acoustic studies. This study broadens the types and quantities of acoustic parameters, employing three machine learning methods to identify multiple sets of effective acoustic parameters for distinguishing emotion types. These augment and refine the acoustic feature set for speech emotion research. Findings reveal that the acoustic parameters, their quantities, and contributions vary in discriminating different emotions, with spectral and signal-to-noise parameters playing pivotal roles. This study offers insights into parameter selection for acoustic analysis of speech emotion.

Keywords: Acoustic features , Speech emotion classification , Machine learning , Spectrum , Noise level

通讯作者
基金项目：澳门特别行政区科学技术发展基金0036/2022/ITP

1 引言

语音能够传达说话者的非语言信息，包括生理特征（例如身高、体型、性别、年龄）、社会特征（例如社会阶层、受教育程度）和情感特征（例如愤怒、高兴、悲伤、恐惧）（Laver and Trudgill，1979）。不同领域对语音情感研究的侧重不同，例如计算机科学主要关注语音情感识别的算法和情感识别的准确性，声学语音学聚焦于不同情感状态下语音的声学特征。不论是研究语音情感识别还是语音情感的声学特征，都需要基于一组能有效区分不同情感类型的声学参数。

声学参数表示语音信号在频率、振幅、响度等方面的物理特性（Bhangale and Kothandaraman，2023），是声学特征的量化指标。早期的语音学研究发现，基频是区分不同情感最常用的声学参数之一（Cosmides，1983）。例如，愤怒和高兴的平均基频更高，基频变化范围更大（Pereira and Watson，1998；Yildirim et al.，2004）。与时长、能量相关的参数也是常用的声学参数。愤怒和高兴通常持续时长更长（Yildirim et al.，2004），恐惧则表现为持续时长缩短、语速变快（Kienast and Sendlmeier，2000）。Rong等（2009）的研究从84个特征中筛选出了16个最有效的声学特征，其中大多数与基频和能量相关。有的情感类型在基频等声学特征上有相同或相似的表现，例如高兴和愤怒（Pereira and Watson，1998；Zvarevashe and Olugbara，2020），因此仅通过基本的声学参数不能实现语音情感的良好区分。在语音情感识别中最常用的声学特征包括时域的韵律特征，例如基频、时长、能量，和频域的谱特征，例如线性预测倒谱系数（LPCC）、梅尔频率倒谱系数（MFCC）（Zvarevashe and Olugbara，2020；刘振焘等，2018；陶建华等，2023）。近年来，各领域的研究者也越来越注意到声门特征对语音情感识别的重要性（Sun and Zhang，2018）。表示声门特征的参数，例如基频抖动（Jitter）、振幅微扰（Shimmer）应用到了语音情感识别中（陶建华等，2023）。在语音学研究中，Birkholz等（2015）通过感知实验发现，嗓音发声类型对离散情感的分类起到重要作用，其中中性情感和快乐表现为模态发声（modal voice），恐惧表现出明显的气嗓音（breathy voice），愤怒表现为较紧的嗓音（pressed voice）。张锐锋（2016）从声门阻抗信号（EGG）中提取了Jitter、Shimmer和谐波噪音比（Harmonic Noise Ratio，HNR）参数，有效区分了高兴和愤怒两种情感。

近年来，许多语音情感识别的研究都聚焦于分类器选择和算法优化上，而对于许多机器学习算法来说，语音声学参数的选择和精简更为重要（徐欣等，2014；Sun，2020；Jha et al.，2022）。目前的研究已经得出从语音信号中提取的声学特征是语音情感识别的重要依据，但还不清楚哪些特征对语音情感识别更有效（Sun，2020；Zvarevashe and Olugbara，2020）。对于情感声学特征的研究，大多是比较不同情感在某一个或几个常用声学参数上的异同（董理等，2021），对情感的声学特征分析还不够系统全面。本研究将基于实验控制下录制的语音样本形成小型数据集，通过监督学习的方法，筛选出能够区分不同情感并且具有语言学意义的声学参数，补充完善语音情感声学研究的特征集，进而促进情感语音合成的研究。

2 实验方法与设计

2.1 录音材料

词是语音情感识别最小的有意义的单位(Schuller et al.，2011)。同时，不同音素的特征对情感识别也有影响(Sethu et al.，2008)。考虑到情感表达的完整性和元音对情感识别的影响，本研究选择“dada”这一不预设任何字形、字义和声调的无意义音节作为录音材料。录音时要求发音人分别以中性、高兴、愤怒和恐惧四种情感状态读出这个词。

2.2 发音人

共有21人进行了录音，11名男性10名女性，平均年龄为25.76±4.52岁。

2.3 录音设备及条件

录音地点是一间专业的隔音室，使用便携式录音笔(奥林巴斯LS-100)和电容式麦克风(AKG C-420)录制语音信号，采样频率为44,100 Hz，精度值为16 bit，保存为.wav格式，最后将录制好的语音导入联想Thinkpad T14型笔记本电脑进行预处理。录音时每个例词读两遍，两遍中间间隔时长为1秒。

2.4 数据预处理

首先，使用Praat对录制的语音样本进行切音和标注处理。切音时以双音节词为单位切分，标注时对每个音节都选主要元音段进行标注。为了使数据标注具有内部一致性，语音波形起点选择在声带规则振动的第二个周期的起始点，终点选择在基频曲线、振幅曲线、第一共振峰和第二共振峰结构相对稳定的点。经过预处理后，共得到336个样本(4种情感*2遍*2个音节*21人)，每种情感类型各有84个样本。

其次，通过基于Matlab的语音声学分析软件VoiceSauce提取声学参数。VoiceSauce是针对语音学研究而开发的声学参数测量、提取与分析软件(Shue et al.，2011；凌锋等，2019)。该软件包含33个与语音学相关的常用声学参数，可以分为以下四类：

(1)基本参数：基频(F0)、第一到第四共振峰(F1/F2/F3/F4)、能量(Energy)，从时域信号中提取的参数，是语音情感研究最基本、最常用的声学参数。其中，基频参数的测量常用的有Stright(strF0)和Praat(pF0)两种算法，共振峰参数的测量常用的有Snack(sF1/sF2/sF3/sF4)和Praat(pF1/pF2/ pF3/ pF4)两种算法。

(2)频谱参数：第一谐波(H1)、第二谐波(H2)、第四谐波(H4)、靠近2000 Hz的谐波(H2K)、靠近第一共振峰的谐波(A1)、靠近第二共振峰的谐波(A2)、靠近第三共振峰的谐波(A3)、第一、二谐波差(H1-H2)、第二、四谐波差(H2-H4)、第四与靠近2000 Hz的谐波差(H4-H2K)、靠近2000 Hz的谐波与靠近5000 Hz的谐波差(H2K- H5K)、第一谐波与第一共振峰谐波振幅差(H1-A1)、第一谐波与第二共振峰谐波差(H1-A2)、第一谐波与第三共振峰谐波差(H1-A3)、次谐波与谐波比(SHR)。从频域信号中提取的参数，反映的是声带闭合程度和声带闭合速率，进而反映语音信号在不同频段上的能量变化速度及嗓音发声类型(Stevens, 1977)。一般来说，谐波振幅差值越大，频谱斜率越大，声带闭合程度越低，闭合速率低，能量下降速度越快，表现为气嗓音。

(3)信噪参数：倒频谱突显峰值(CPP)、0-500 Hz频段的谐波噪音比(HNR05)、0-1500 Hz频段的谐波噪音比(HNR15)、0-2500 Hz频段的谐波噪音比(HNR25)、0-3500 Hz频段的谐波噪音比(HNR35)，反映语音中噪音成分的多少和信号周期性的好坏。CPP和HNR的值越大，语音中的噪音成分越少，信号的周期性越好。

(4)声源激励参数：激励期(epoch)、激励强度(SoE)，反映声门状态。激励期即声门闭合瞬间，激励强度受到声门闭合时长和闭合速率的影响。

2.5 算法选择与评估

每一种算法都有其优缺点。考虑到数据量的大小、分类的多少以及算法的难易程度，本研究选择K最近邻(K-Nearest Neighbors，以下简称“KNN”)(Petrushin，2000)、支持向量机(Support Vector Machine，以下简称“SVM”)(Schuller et al.，2004)和随机森林(Random Forest，以下简称“RF”)(Breiman，2001)三种经典机器学习算法对语音情感的声学特征进行筛选。KNN是最基本、最简单的分类算法，对语音情感数据的拟合性能较高(Basu et al.，2017；Jha et al.，2022)，SVM的二分类效果最佳，并且适合小样本训练集(Jha et al.，2022)，RF是基于决策树的集成学习算法，能够避免由较小数据集引起的过度拟合问题(Rong et al.，2009；Jha et al.，2022)。为了便于比较三种算法的分类效果，本研究将中性、高兴、愤怒、恐惧四种情感两两组合，均进行二分类。每种分类模型均按照70%的比例设置训练集，30%的样本作为测试集。采用准确率(Accuracy)这一最直观的衡量标准来比较各算法在不同情感组合中的分类准确性，它是预测正确的样本数与样本总数的比率(Zvarevashe and Olugbara，2020)。通过受试者工作特征曲线(Receiver Operating Characteristic Curve，以下简称“ROC曲线”)和ROC曲线下面积(Area Under Curve，以下简称“AUC”)来评估各算法的分类性能，ROC曲线和AUC是评估二分类模型性能的重要指标(Jha et al.，2022)。

3 结果

3.1 分类效果

在设置模型参数时，均通过定义一个超参数网格并通过网格搜索来寻找最佳的参数组合，以提高模型的性能。准确率也是在最佳参数组合下得出的。

从表1中可以看出，不同的算法在区分不同情感时的准确率不同。平均来看，三种算法区分愤怒和恐惧、高兴和恐惧的效果最好，平均准确率均超过85%；区分中性与恐惧的效果最

差，平均准确率只有64%。具体的，KNN和SVM区分愤怒和恐惧效果最好，KNN的准确率达到了96%，SVM的准确率达到92%；RF区分高兴和恐惧的准确率最高，为84%。其次，在三种算法中，相比较而言，高兴情感都能与其他三种情感进行良好区分，而且高兴与恐惧之间区分得最好，在KNN中准确率达到88%，SVM的准确率为84%，RF的准确率也是84%。不论哪一种算法，恐惧与中性情感的区分效果都比较差，在KNN和RF中，区分恐惧与中性情感的准确率都是最低的，只有60%；在SVM中，区分恐惧与中性情感的准确率为72%，只高于愤怒与中性情感的区分准确率。此外，在KNN和SVM中，愤怒与中性情感的区分效果也比较差，在KNN中的准确率最低，仅有64%。总之，在三种算法中，高兴、愤怒和恐惧三种情感之间能够进行较为准确的区分，恐惧和中性情感最难准确区分开。

情感类型	KNN	SVM	RF	平均准确率
中性/高兴	88%	76%	72%	78.67%
中性/愤怒	64%	68%	80%	70.67%
中性/恐惧	60%	72%	60%	64%
高兴/愤怒	84%	76%	80%	80%
高兴/恐惧	88%	84%	84%	85.33%
愤怒/恐惧	96%	92%	72%	86.67%

Table 1: 基于三种算法二分类的准确率

对中性/高兴、中性/愤怒、中性/恐惧、高兴/愤怒、高兴/恐惧、愤怒/恐惧六组情感都使用KNN、SVM和RF三种算法进行二分类，图1—图6是对基于三种算法的分类模型性能的评估结果，用ROC曲线表示。ROC曲线横轴为假正例率(False Positive Rate，以下简称“FPR”)，纵轴为真正例率，也称为召回率(True Positive Rate，以下简称“TPR”)。FPR表示将负例误判为正例的比例，TPR表示将正例正确判断为正例的比例。FPR越低，TPR越高，ROC曲线越靠近左上角，说明模型的性能越好。

通过图1可以看出，三种分类模型在区分中性-高兴时的FPR都为0，说明对两种情感分类时没有误判，KNN的TPR要高于SVM和RF的TPR，说明在区分中性-高兴时，KNN算法更灵敏，召回率更高。从图2中可以看出，相比较而言，KNN在区分中性-愤怒时的FPR较高，TPR较低，KNN算法区别中性-愤怒两类情感时效果较差。图3显示，三种分类模型在区分中性-恐惧时的FPR都较低，即对两种情感的误判较低，但TPR也很低，说明对中性和恐惧这两类情感的分类不灵敏。对比图1—图3可以发现，KNN、SVM和RF三种算法区分对中性与愤怒、中性与恐惧的区分效果不佳。

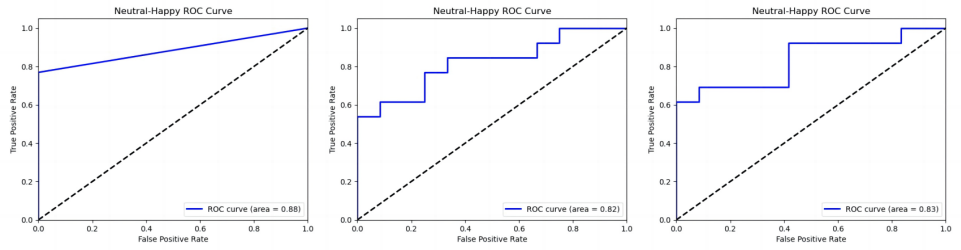

Figure 1: KNN、SVM、RF区分“中性/高兴”的ROC曲线

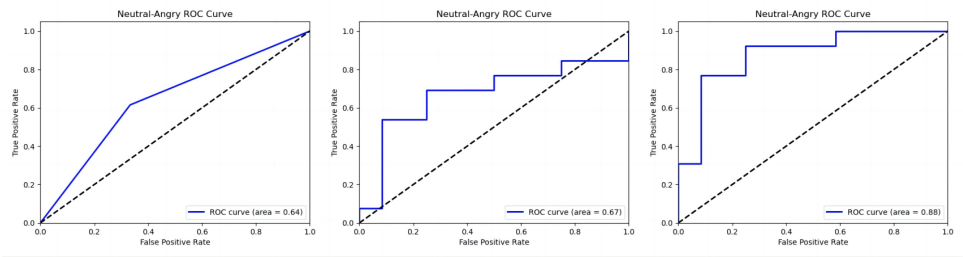

Figure 2: KNN、SVM、RF区分“中性/愤怒”的ROC曲线

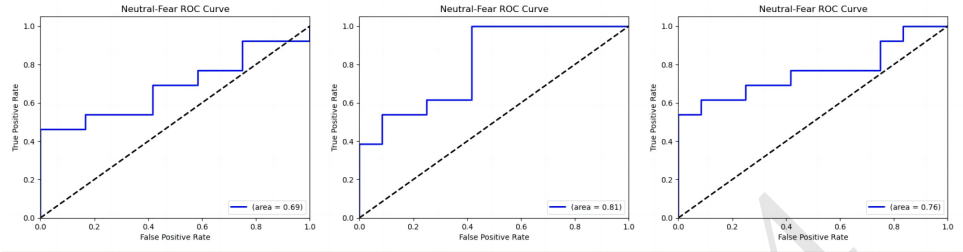

Figure 3: KNN、SVM、RF区分“中性/恐惧”的ROC曲线

通过图4—图6可以看出，相比较而言，不论是区分高兴-愤怒、高兴-恐惧还是愤怒-恐惧，基于KNN算法的分类模型都存在将两种情感误判情况。对高兴-愤怒的区分中，基于RF算法的模型分类效果最好，对高兴-恐惧的区分也是如此。区分愤怒-恐惧时，基于三种算法的分类模型效果都很好。

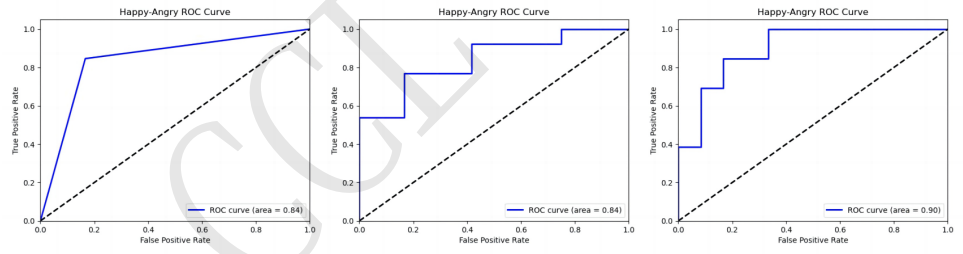

Figure 4: KNN、SVM、RF区分“高兴/愤怒”的ROC曲线

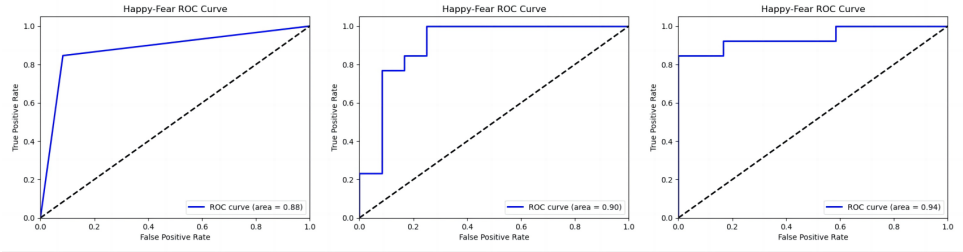

Figure 5: KNN、SVM、RF区分“高兴/恐惧”的ROC曲线

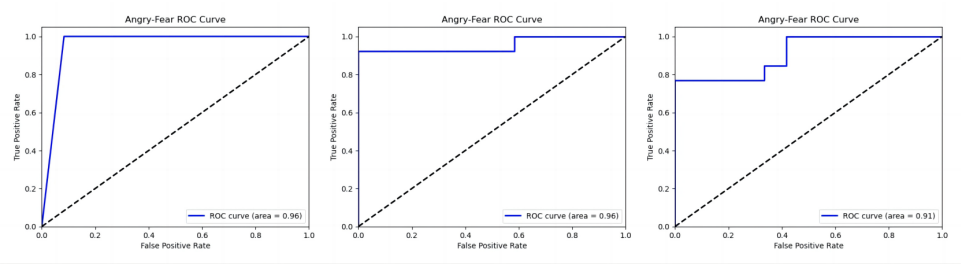

Figure 6: KNN、SVM、RF区分“愤怒/恐惧”的ROC曲线

3.2 特征筛选

基于不同算法的不同情感之间的分类都使用了相同的一组参数。在进行模型训练之前，首先通过独立样本T检验初步筛选出与情感分类不存在显著性差异(ps <0.05)的参数并剔除，这部分参数如附录A所示。

基于剩下的声学特征进行模型训练与测试。排列重要性(Permutation Importance)是衡量各个特征对模型性能影响程度的指标。通过这一方法可以得出不同声学参数对模型的贡献，进而筛选出帮助分类的最有效的一组声学参数。排列重要性是基于特定模型的，本研究中基于KNN、SVM和RF三种算法得出的各组声学参数的排序均不相同。我们已经得出评估不同模型的AUC(表2)，因此本研究提出了一种综合三种算法的排序的计算方法，对各声学参数在不同情感分类中的贡献进行综合计算和排序。计算公式如下：

$$\text{Weight} = k_{\text{KNN}} \cdot s_{\text{KNN}} + k_{\text{SVM}} \cdot s_{\text{SVM}} + k_{\text{RF}} \cdot s_{\text{RF}}$$

其中，“Weight”代表声学参数在排序中所占的权重，即贡献。“k”表示各算法占比系数，k值根据AUC计算，即对每一组情感，以三种算法中最小的AUC值(即模型效果最差)为基线，对三个AUC值进行归一化，归一化后的值即为k。“s”表示不同算法得出的各声学参数排序重要性的得分。声学参数的最终贡献为三个算法的占比系数与该算法下声学参数的重要性得分的乘积之和。

情感类型	KNN	SVM	RF
中性/高兴	88%	82%	83%
中性/愤怒	64%	67%	88%
中性/恐惧	69%	81%	76%
高兴/愤怒	84%	84%	89%
高兴/恐惧	88%	90%	94%
愤怒/恐惧	96%	96%	91%

Table 2: 评估三种算法二分类的AUC

基于上述计算方法，我们得出区分各组情感的声学参数重要性排序及其贡献，如表3所示。对于部分参数贡献出现负值的情况，根据Scikit-learn的解释，这可能是由于使用的数据集较小或者参数之间存在较强相关性导致的，未来可以通过聚类分析进行更细致的参数筛选改善(Pedregosa et al.，2011)。

从表3中可以看出，已有语音情感声学研究常用的基本声学参数基频(strF0和pF0)和时长(Time)对不同情感的区分没有贡献或贡献较低。基频在中性/高兴和高兴/恐惧两组情感分类中有贡献，时长对中性/恐惧的区分有贡献。在中性/高兴组中，贡献最大的H1-A2得分是基频的11倍，在高兴/恐惧组中，相比较而言，基于Stright算法得出的基频(strF0)贡献更大，但贡献最大的CPP得分约是它的1.6倍。在中性/恐惧组中，贡献最大的H4的得分约是时长的5.4倍。

其次，区分不同情感所依赖的声学参数、参数数量、参数的贡献都不相同。在区别中性/高兴、中性/愤怒时H1-A2的贡献最大，区别中性/恐惧时H4的贡献最大，区别高兴/愤怒时SHR的贡献最大，区别高兴/恐惧、愤怒/恐惧时CPP的贡献最大。H1-A2，即第一谐波与第二共振峰谐波振幅差，反映声带闭合的速率(Stevens，1977)，进而反映声带漏气情况和能量下

降的快慢。H4指频谱上的第四谐波，其频率是基频的四倍。在语音学研究中，谐波成分很少单独使用，常通过进一步计算谐波差来反映语音信号的频谱斜率(Gordon and Ladefoged，2001)，它反映声带的闭合情况，进而反映频谱能量的变化。SHR是次谐波与谐波比，影响语音信号非周期性振动段的音高，用来判断信号的周期性(凌锋等，2019)。CPP即倒频谱突显峰值，测量的是倒谱峰值到倒谱回归线的距离，反映语音信号中噪音成分的多少，进而反映语音信号周期性的好坏(Shue et al.，2011；凌锋等，2019)。这四个参数及其相关参数可以分为频谱参数和信噪参数两类，都是判断嗓音发声类型的常用参数，说明嗓音发声类型对情感识别发挥了重要作用。在语音情感的声学分析中，除了常用的基频、时长、能量等参数，还需要引入和分析反映嗓音发声类型的频谱参数和信噪参数。

对语音学研究来说，表3为语音情感的声学分析提供了更重要、更全面的声学特征参考。对中性、高兴、愤怒、恐惧四种情感进行声学分析时，可根据研究者的需要，选择贡献更大、排序更靠前的参数进行分析和声学特征比较。

排序	中性/高兴		中性/愤怒		中性/恐惧		高兴/愤怒		高兴/恐惧		愤怒/恐惧	
	特征	贡献	特征	贡献	特征	贡献	特征	贡献	特征	贡献	特征	贡献
1	H1-A2	0.22	H1-A2	0.30	H4	0.49	SHR	0.27	CPP	0.21	CPP	0.17
2	SHR	0.18	H2K	0.26	Energy	0.22	pF1	0.15	HNR05	0.18	A1	0.08
3	SoE	0.14	pF1	0.24	A2	0.18	H2K	0.11	strF0	0.13	Energy	0.08
4	pF3	0.11	Energy	0.22	H1-A1	0.13	A1	0.03	HNR15	0.09	H2	0.08
5	H2	0.09	H4	0.22	HNR15	0.12	Energy	-0.02	A1	0.09	HNR05	0.07
6	H1-A3	0.08	A2	0.19	A1	0.10	HNR35	-0.02	A3	0.06	H1-H2	0.06
7	H1-A1	0.06	HNR15	0.18	Time	0.09	HNR15	-0.02	pF0	0.06	pF1	0.06
8	pF0	0.02	HNR25	0.15	H2K	0.06	H4	-0.03	pF2	0.06	H1-A2	0.03
9	strF0	0.02	A1	0.13	A3	0.06	H2	-0.03	epoch	0.04	A2	0.02
10	H2-H4	0.02	A3	0.08	H1-A2	0.05	HNR25	-0.04	H2K	0.04	pF3	0.02
11	H1	0.00	H1-A1	0.08	H1-A3	0.02	HNR05	-0.04	HNR25	0.03	H1-A3	0.02
12	pF4	-0.02	HNR35	0.08			SoE	-0.06	pF3	0.03	SHR	0.02
13	HNR05	-0.07	H1-A3	0.01			H1	-0.06	sF2	0.03	sF2	0.02
14	epoch	-0.07					A3	-0.06	H1-A2	0.02	H4-H2K	0.02
15							A2	-0.10	HNR35	0.02	H4	0.00
16							epoch	-0.11	A2	0.02	A3	0.00
17							strF0	-0.15	H1-A3	0.01	pF2	0.00
18							pF0	-0.18	H4-H2K	0.0	H2K	-0.02
19									H1-A1	0.00	H1-A1	-0.04
20									SoE	0.00		
21									H2-H4	-0.02		
22									Energy	-0.04		

Table 3: 声学参数重要性排序及贡献(保留到小数点后两位)

4 结论

在语音情感研究中，不同领域的研究侧重点不同，语音学更关注语音中不同情感的声学表现。现有的语音情感的声学研究大多是探讨两到三种离散的情感在基频、时长等一个或少数几个常用参数上的不同表现，而有些情感在这些常用参数上存在许多共同特征，例如愤怒和高兴的平均基频相较于中性情感都变大，因此缺乏区别性的声学特征，筛选有效表达情感的声学特征至关重要。本研究通过扩大声学参数的种类和数量，使用机器学习的方法，筛选出区分情感类型的一组有效声学参数，以补充和完善语音情感声学研究的声学特征集。

研究采用机器学习的方法，将语音学研究常用的四类共33个声学参数放入基于KNN、SVM和RF三种算法的分类模型中，筛选出不同情感分类所依赖的多组声学参数。结果发现，语音情感声学分析常用的基频和时长参数对各组情感的分类贡献不大。区分不同情感所依赖的声学参数、参数数量、参数的贡献都不相同。H1-A2在区分中性/高兴、中性/愤怒时的贡献最大，H4在区分中性/恐惧时的贡献最大，SHR在区分高兴/愤怒时的贡献最大，CPP在区分高兴/恐惧、愤怒/恐惧时的贡献最大。这组参数表明频谱参数和信噪参数对区分不同情感类型具有重要作用。本研究得出的声学参数排序重要性为语音情感声学分析的参数选择提供参考。

第二十三届中国计算语言学大会论文集，第568页-第576页，太原，中国，2024年7月25日至28日。
卷1：主会论文

由于本研究所使用的数据集较小，在算法选择等方面受到限制，后续研究应在此基础上扩大样本量，以优化算法，从而获得更精简和精确的声学参数。此外，本研究通过数据归一化来消除性别等因素上的差异，已有研究发现性别影响不同情感的区分(Arias et al.，2021)，未来可以在获得大量数据的基础上，探讨性别对各组情感分类的影响。在语音学研究方面，本研究为语音情感声学分析选择哪些参数提供了参考，后续可以在此基础上进行声学分析，探讨不同情感在这些参数上的具体表现。

参考文献

Arias P. Rachman L. Liuni M. and Aucouturier J.-J. 2021. Beyond Correlation: Acoustic Transformation Methods for the Experimental Study of Emotional Voice and Speech. *Emotion Review*, 13(1), 12–24.

Bhangale K. and Kothandaraman M. 2023. Speech Emotion Recognition Based on Multiple Acoustic Features and Deep Convolutional Neural Network. *Electronics*, 12(4), Article 4.

Birkholz P. Martin L. Willmes K. Kröger B. J. and Neuschaefer-Rube C. 2015. The contribution of phonation type to the perception of vocal emotions in German: An articulatory synthesis study. *The Journal of the Acoustical Society of America*, 137(3), 1503–1512.

Breiman L. 2001. Random Forests. *Machine Learning*, 45(1), 5–32.

Cosmides L. 1983. Invariances in the acoustic expression of emotion during speech. *Journal of Experimental Psychology: Human Perception and Performance*, 9(6), 864–881.

Arias P. Rachman L. Liuni M. and Aucouturier J.-J. 2021. Beyond Correlation: Acoustic Transformation Methods for the Experimental Study of Emotional Voice and Speech. *Emotion Review*, 13(1), 12–24.

Gordon M. and Gordon P. 2001. Phonation types: a cross-linguistic overview. *Journal of phonetics*, 29(4), 383-406.

Jha T. Kavya R. Christopher J. and Arunachalam V. 2022. Machine learning techniques for speech emotion recognition using paralinguistic acoustic features. *International Journal of Speech Technology*, 25(3), 707–725.

Kienast M. and Sendlmeier W. F. 2000. Acoustical Analysis of Spectral and Temporal Changes in Emotional Speech. In *ISCA Tutorial and Research Workshop (ITRW) on Speech and Emotion.*

Laver J. and Trudgill P. 1979. Phonetic and linguistic markers in speech. In: Scherer, Klaus R.; Giles, Howard (Ed.). *Social markers in speech.* Cambridge: Cambridge University Press.p.1-32.

Pedregosa et al. 2011. Scikit-learn: Machine Learning in Python. *JMLR 12*, pp. 2825-2830.

Pereira C. and Watson C. 1998. Some acoustic characteristics of emotion. *5th International Conference on Spoken Language Processing (ICSLP 1998)*, paper 0684-0.

Petrushin V. A. 2000. Emotion recognition in speech signal: Experimental study, development, and application. *6th International Conference on Spoken Language Processing (ICSLP 2000)*, vol.2, 222-225–0.

Rong J. Li G. and Chen Y.-P. P. 2009. Acoustic feature selection for automatic emotion recognition from speech. *Information Processing Management*,45(3), 315–328.

Schuller B. Rigoll G. and Lang M. 2004. Speech emotion recognition combining acoustic features and linguistic information in a hybrid support vector machine-belief network architecture. *2004 IEEE International Conference on Acoustics, Speech, and Signal Processing*,1, I–577.

Schuller B. Batliner A. Steidl S. Seppi D. 2011. Recognising realistic emotions and affect in speech: State of the art and lessons learnt from the first challenge. *Speech Communication*, 53(9), 1062–1087.

Sethu V. Ambikairajah E. and Epps J. 2008. Phonetic and speaker variations in automatic emotion classification. *Interspeech 2008* , (pp.617-620).

Shue Y. L. Keating P. Vicenik C. and Yu K. 2011. VoiceSauce: A program for voice analysis. *ICPhS XVII* , (pp.1846-1849).

Stevens K. N. 1977. Physics of laryngeal behavior and larynx modes. *Phonetica*, 34(4), 264-279.

Sun T.-W. 2020. End-to-End Speech Emotion Recognition With Gender Information. *IEEE Access*, 8, 152423–152438.

Sun Y. and Zhang X-Y. 2018. Characteristics of human auditory model based on compensation of glottal features in speech emotion recognition. *Future Generation Computer Systems*, 81, 291–296.

Yildirim S. Bulut M. Lee C. M. Kazemzadeh A. Deng Z. Lee S. Narayanan S. and Busso C. 2004. An acoustic study of emotions expressed in speech. *Interspeech 2004*,2193–2196.

Zvarevashe K. and Olugbara O. 2020. Ensemble Learning of Hybrid Acoustic Features for Speech Emotion Recognition. *Algorithms*, 13(3), Article 3.

董理, 梁晓静, 黄慧怡. 2021. 昆曲女性行当情感念白时长特征. 语言学论丛(02),272-290.

凌锋, 史辉, 袁丹, 沈瑞清. 2019. 发声态研究的相关问题与VoiceSauce的使用. 方言(04),385-397.

刘振焘, 徐建平, 吴敏, 曹卫华, 陈略峰, 丁学文, 郝曼, 谢桥. 2018. 语音情感特征提取及其降维方法综述. 计算机学报(12),2833-2851.

陶建华, 陈俊杰, 李永伟. 2023. 语音情感识别综述. 信号处理(04),571-587.

徐欣, 李雅, 许小颖, 陶建华. 2014. 情感语音识别的区别性声学特征选择.第十一届中国语音学学术会议(PCC2014)论文集(pp.146).

张锐锋. 2016. 普通话情感语句中的发声——一项预试性研究. 语言学论丛(02),305-322.

附录A. 初步剔除的参数

情感类型	剔除的参数
中性/高兴	Time、H4、H2K、A1、A2、A3、H1-H2、H4-H2K、H2K-H5K、CPP、Energy、HNR15、HNR25、HNR35、sF1、sF2、sF3、sF4、pF1、pF2
中性/愤怒	Time、H1、H2、H1-H2、H2-H4、H4-H2K、H2K-H5K、CPP、HNR05、SHR、strF0、pF0、sF1、sF2、sF3、sF4、pF2、pF3、pF4、epoch、SoE
中性/恐惧	H1、H2、H4、H2K、A1、A2、A3、H1-H2、H2-H4、H4-H2K、H2K-H5K、H1-A1、H1-A2、H1-A3、strF0、pF0、sF1、sF3、sF4、pF1、pF4、epoch
高兴/愤怒	Time、H1-H2、H2-H4、H4-H2K、H2K-H5K、H1-A1、H1-A2、H1-A3、CPP、sF1、sF2、sF3、sF4、pF2、pF3、pF4
高兴/恐惧	Time、H1、H2、H4、H1-H2、H2K-H5K、SHR、sF1、sF3、sF4、pF1、pF4
愤怒/恐惧	Time、H1、H2-H4、H2K-H5K、HNR15、HNR25、HNR35、strF0、pF0、sF1、sF3、sF4、pF4、epoch、SoE