第 39 卷 第 2 期
2020 年 3 月

应用声学
Journal of Applied Acoustics

Vol. 39, No. 2
March, 2020

◇ 研究报告 ◇

语音情感识别中的特征选择方法*

褚 钰[1†] 李田港[1] 叶 硕[1] 叶光明[2]

(1 武汉邮电科学研究院 武汉 430000)

(2 武汉烽火众智数字技术有限责任公司 武汉 430000)

摘要: 语音情感识别在许多领域具有重要研究价值,不同声学情感特征在使用不同分类器进行分类时,识别效果具有明显差异。与语音情感有关的声学特征包括谱特征、韵律学特征、音质特征。该文提出一种特征融合的方法,将 3 种声学特征中具有最好识别能力的特征进行融合:保留在实验中表现稳定且有较高识别率的谱特征的全部特征,提取韵律学、音质特征的相关统计量作为辅助特征融合于谱特征中。实验表明,该文所提出的融合特征在使用同一分类器进行分类时,识别率优于单一特征;当使用不同分类器时,融合特征依然具有较好的识别能力,且识别性能稳定,3 个数据集上均有较好的识别率,基本实现跨数据集识别。

关键词: 语音识别;情感识别;特征选择;特征融合

中图法分类号: TP183 **文献标识码:** A **文章编号:** 1000-310X(2020)02-0216-07

DOI: 10.11684/j.issn.1000-310X.2020.02.007

Research on feature selection method in speech emotion recognition

CHU Yu[1] LI Tiangang[1] YE Shuo[1] YE Guangming[2]

(1 *Wuhan Research Institute of Posts and Telecommunications, Wuhan* 430000, *China*)

(2 *Wuhan Fiberhome Wisdom Digital Technology Co. Ltd., Wuhan* 430000, *China*)

Abstract: Speech emotion recognition is of great value in many fields. The recognition effect of different emotion acoustic features is obviously different when different classifiers are used for classification. Acoustic features related to speech emotions include spectral features, rhythmic features and quality features. This paper proposes a method of feature fusion, which combines the features of the three acoustic features with the best recognition ability: all the features of the spectral features that are stable in the experiment and have a high recognition rate are retained, and the relevant statistics of the rhythmic features and quality features are extracted as auxiliary features and integrated into the spectral features. Experiments show that the fusion feature proposed in this paper is better than the single feature when using the same classifier for classification; when using different classifiers, the fusion feature still has better recognition ability and stable recognition performance. It has better recognition rate on three data sets and basically realizes cross-dataset recognition.

Keywords: Speech recognition; Emotion recognition; Feature selection; Feature fusion

2019-05-06 收稿; 2019-09-25 定稿

*湖北省科技厅 2018 年度湖北省技术创新专项重大项目 (2018AAA063)

作者简介: 褚钰 (1995–), 男, 河北张家口人, 硕士研究生, 研究方向: 机器学习, 语音识别。

†通信作者 E-mail: 18811309895@163.com

0 引言

语音情感识别是语音识别的重要组成部分，随着人工智能领域的发展与延伸，进一步了解语音，发掘语音下隐含的情绪信息，在安防、监控、医疗看护等领域具有重要的价值。目前与语音情感有关的声学特征主要分为3类，分别为基于谱的相关特征、韵律学特征、音质特征[1]。这些特征又分为常见低级描述和高级描述的水平统计函数[2−3]，低级描述主要包括：基音频率(Fundamental frequency)、能量(Energy)、过零率(Zero-crossing)、抖动(Jitter)、梅尔滤波特征(Mel-filterbank features)、共振峰位置/带宽(Formant locations/bandwidths)、谐波噪声比(Harmonics-to-noise ratio) 等；高级描述主要包括：均值(Mean)、方差(Variance)、最小值(min)、最大值(max)、范围(Range)、高阶矩(偏度、峰度)(Higher order moments(Skewness, Kurtosis))、线性回归系数(Linear regression coefficients)等。

近年来，Koolagudi 等[4]提出非个性化语音情感特征，不受说话人个人特征影响，主要包括无声部分时间与有声部分时间比率、基频平均变化率等。

不同特征对情感的分类结果有不同程度的影响，直接使用数量庞大的情感特征，往往导致运算速度降低、建模效果不理想等问题，如何在离散语音情感识别任务中找到有效的情感特征，并通过这些特征来表达情感信息，是研究者面临的一大问题；此外，同一情感特征在不同语音数据集中的表现也存在较大差异，适用于某一数据集的情感特征在其他数据集上表现并不一定理想。因此，寻找一种更为普遍、并能跨数据集实现情感识别的特征成为了当前语音情感识别的重点。

特征融合是一种优化参数的手段，在特征选择与特征融合问题上，Cao等[5]利用随机森林算法分析提取的声学特征，并去除包含多余情感信息的特征，以此进行特征选择；刘博等[6]提取语音谱特征，得到一个高斯混合模型，进一步拼接得到该语音的超向量；张文克[7]将两个不同的谱特征进行合并，求取融合后的特征参数序列，在此基础上，王忠民等[8]使用多核学习算法将谱特征与语音的语谱图特征进行融合，提高了分类精度与识别准确率；此外，基于深度学习，通过融合谱特征和基于音高的超韵律特征也可以显著提高识别准确率[9]。本文提出一种特征融合算法，保留了在实验中表现稳定且有较高识别率的谱特征的全部特征，提取韵律学特征基音频率、音质特征共振峰的相关统计量作为辅助特征融合于谱特征中。

1 情感特征提取

1.1 谱相关特征

谱特征被认为是声道形状变化和发声运动之间相关性的体现[10]。研究者发现，语音中的情感内容对频谱能量在各个频谱区间的分布有着明显的影响[11]。由于人听到的声音高低和频率大小不呈线性正比关系，而梅尔倒谱系数(Mel frequency cepstrum coefficient, MFCC)特征基于人耳听觉特性，因此在语音情感分类中具有良好的鲁棒性和准确度，其计算公式满足：

$$\mathrm{Mel}(f)=2595\times\lg\left(1+\frac{f}{700}\right),\tag{1}$$

式(1)中，f为声音频率，单位Hz。

为进一步反映语音的动态特性，本文提取语音MFCC特征的一阶、二阶差分，计算公式如下：

$$d_t=\begin{cases}C_{t+1}-C_t, & t<K,\\ \dfrac{\sum_{k=1}^{K}k(C_{t+k}-C_{t-k})}{\sqrt{2\sum_{k=1}^{K}k^2}}, & \text{其他},\\ C_t-C_{t-1}, & t\geqslant Q-K,\end{cases}\tag{2}$$

其中，d_t为第t个一阶差分，C_t为第t个倒谱系数，Q为倒谱系数的阶数，K为一阶导数的时间差，可取1或2。将式(2)中结果再代入就可以得到二阶差分的参数。

逆梅尔倒谱系数(Inverted MFCC, IMFCC)[12]是一种针对高频信息的语音特征，与Mel滤波器组在低频部分具有较高分辨率的特点相反，IMFCC特征在高频区域使用较窄的滤波器获得高频信息，强调不同频率带之间的差异。其表达式[13]为

$$\mathrm{IMel}(f)=2146.1-1127\ln\left(1+\frac{4000-f_{\mathrm{Hz}}}{700}\right).\tag{3}$$

感觉加权线性预测 (Perceptual linear predictive, PLP)参数是一种基于听觉模型的特征参数，具有更强的噪声鲁棒性[14]。它在临界频带分析处

理、等响度曲线预加重以及信号强度-听觉响度变换3个方面，模仿人耳的听觉感知机理[15]：首先通过傅里叶变换得到语音信号的频谱，取其实部和虚部的平方和得到语音信号的短时能量谱$P(f)$[16]，其表达式为

$$P(f)=\{\mathrm{Re}[X(f)]\}^2+\{\mathrm{Im}[X(f)]\}^2. \quad (4)$$

进而获得临界带宽听觉谱，接着进行等响度预加重，最后模拟强度与响度间的非线性关系，对预加重后的响度开立方根，通过逆傅里叶变换与线性预测得到PLP特征参数。RASTA滤波器[17]是一种用于抑制非语言学背景噪声的无限脉冲响应数字滤波器，经过RASTA滤波器处理后的PLP特征具有更好的语音识别效果，它的传输函数为

$$H(z)=0.1\times\frac{2+z^{-1}-z^{-3}-2z^{-4}}{z^{-4}\times(1-\rho z^{-1})}. \quad (5)$$

本文提取RASTA-PLP特征以及该参数的一阶、二阶差分。

1.2 韵律学特征

韵律相比于文本内容，其所包含的信息相对较少，但韵律注重音高、快慢以及轻重等方面的变化[18]，能够帮助听者更好地理解语音信息中的重点，使听者能够更快地获取有效信息，因而在情感识别上具有明显的优势。本文提取语音的基音频率、过零率、短时能量作为情感特征。

声带具有清音与浊音两种振动方式，其中浊音振动具有周期性，其振动频率便是基音频率。本文使用短时自相关函数法求取基音频率，对语音进行分帧，通过比较原始信号与时延后的信号间的类似性来求取该特征，自相关函数如下：

$$R_i(k)=\sum_{m=1}^{N-m}x_i(m)x_i(m+k), \quad (6)$$

其中，k是时延。使用式(7)和式(8)求取语音短时能量和短时过零率：

$$E(i)=\sum_{m=0}^{N-1}x_i^2(m), \quad (7)$$

$$Z(i)=\frac{1}{2}\sum_{m=0}^{N-1}|\mathrm{sgn}[x_i(m)]-\mathrm{sgn}[x_i(m-1)]|, \quad (8)$$

其中，$x_i(m)$为第i帧语音信号，N为帧长。

1.3 音质特征

声音质量是一种用于衡量语音是否具有纯净、清晰、容易辨识等特点的主观评价指标[19]。其特征一般有共振峰频率及其带宽、频率微扰和振幅微扰、声门参数等。语音信号作为一种非平稳信号，其生成与3个系统有关[20]，声带系统负责产生激励振动，声道系统负责气流通过，辐射系统则是指由嘴唇完成的语音辐射，形成“话”。

共振峰是指在频谱中能量相对集中的一些区域，体现的是声道的信息，其频率的分布特性决定语音的音色[21]，在语音识别方面具有重要的作用。常用的提取语音共振峰的方法为倒谱法和线性预测编码(Linear predictive coding, LPC)法，倒谱法对语音信号进行离散傅里叶变换，进一步求得共振峰参数；LPC法则是通过线性预测的方法推导出声道滤波器进而找出共振峰。本文采用LPC法提取共振峰参数。

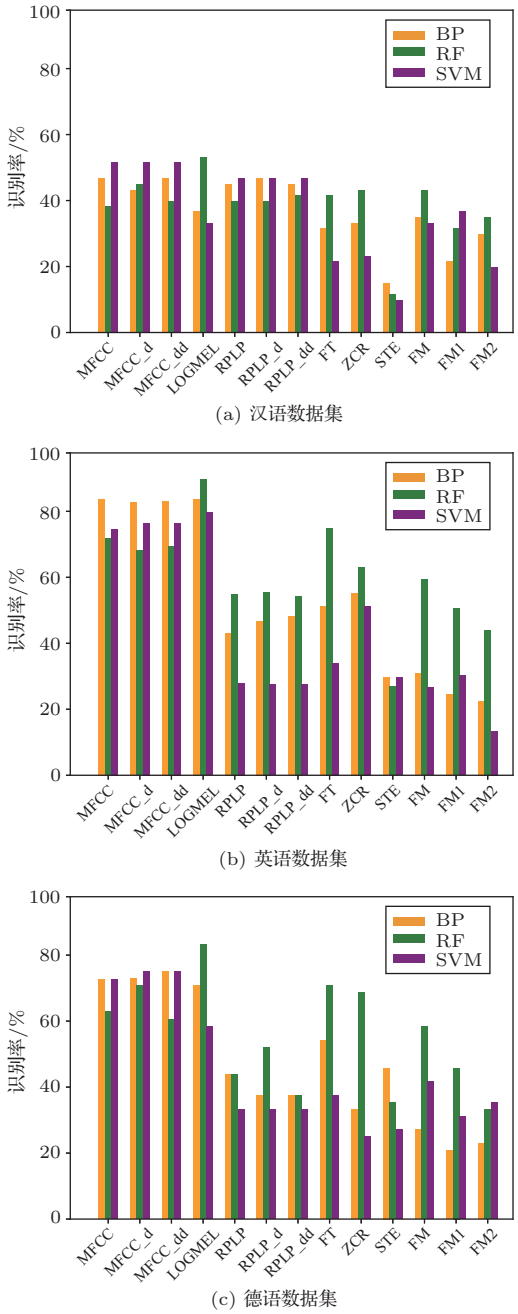
2 特征融合

图1是组图，分别由汉语数据集、英语数据集、德语数据集测试得来。每一幅图的纵坐标都表示识别率，横坐标为情感特征，从左到右依次是谱特征：MFCC，MFCC及其一阶、二阶差分，逆梅尔对数频谱系数，RASTA-PLP，RASTA-PLP及其一阶、二阶差分；韵律特征：基音频率，过零率，短时能量；音质特征：共振峰，共振峰一阶抖动，共振峰二阶抖动。

不同颜色的柱状图表示当前特征在不同分类器上的识别结果。本文采用BP神经网络(Back propagation neutral network, BPNN)、随机森林(Random forest, RF)、支持向量机(Support vector machine, SVM)3种方式进行分类。可以看出，情感特征对分类器敏感，同一特征在不同分类器上的表现具有明显差异：谱相关特征在神经网络中具有较好的分类效果，而韵律学特征则对随机森林敏感；此外，语种也对识别率具有一定影响，汉语的识别率稍低于英语、德语，这应该与日常的情感表达习惯有一定关系。

本文对3个数据集上具有最好表现的声学特征：MFCC、基音频率、共振峰进行融合。该算法首先将提取出的MFCC特征矩阵$\boldsymbol{M}$、基音频率特征向量$\boldsymbol{T}$、共振峰特征向量$\boldsymbol{F}$作为输入，之后将$\boldsymbol{M}$转

化为一维列向量 $\boldsymbol{M}'$，对 $\boldsymbol{T}$ 分别求取最大值、最小值、均值、标准差得到一维列向量 $\boldsymbol{T}'$，对 $\boldsymbol{F}$ 分别求取最大值、最小值、均值、标准差得到一维列向量 $\boldsymbol{F}'$，由于 MFCC 特征是情感识别中较为有效的频谱特征，在实验中表现稳定且在不同数据集上均具有较高的识别率，因此在之后的操作中保留 MFCC 的全部 13 维特征，将之前得到的 $\boldsymbol{T}'$ 和 $\boldsymbol{F}'$ 添加到 $\boldsymbol{M}'$ 之后，即得到融合特征向量。



(a) 汉语数据集

(b) 英语数据集

(c) 德语数据集

图 1　不同特征在不同分类器上的识别结果

Fig. 1　The recognition result of each feature on different classifiers

3　实验测试

3.1　数据集

本实验在 3 种语种的公开数据集上进行：中国科学院汉语数据集、EmoV-DB 英语情感数据集 [22]、德国柏林德语语料库 [23]。

汉语数据集共有语音 300 条，采样频率为 16 kHz，16 bit 量化，语音有 angry、fear、happiness、neutral、sad、surprise 共 6 种情感，每种情感各 50 条语音；EmoV-DB 英语情感数据集共有语音 1817 条，采样频率为 16 kHz，16 bit 量化，语音包含 amused、angry、disgust、neutral、sleepiness 共 5 种情感；德国柏林德语语料库中包含 7 种情感，共 535 句情感语音信号，本文从中选择了 angry、happy、neutral、sad 四种情感，每种情感随机选择 60 条语音，共 240 条用于识别，音频采样频率为 16 kHz，16 bit 量化。

本文共选择 2357 条语音用于构建实验数据集，总时长 2 h 50 min，其中训练集时长 2 h 16 min，包含语音 1886 条。

3.2　实验设计

为验证本文所提特征融合算法，实验分为两个部分：第一部分验证特征融合算法的有效性；第二部分验证本文所提融合特征较之于其他融合特征，具有更稳定的识别能力。

在第一部分的实验中，选取 3 个数据集上具有最好表现的声学特征：MFCC、基音频率、共振峰进行融合。将得到的融合特征分别使用 BP 神经网络、随机森林、支持向量机 3 种算法在 3 个数据集上进行情感识别，与 MFCC、基音频率、共振峰这 3 个单一特征的识别率进行比较。在第二部分的实验中设计多组对照试验，随机选取 3 个特征进行融合并在 3 个数据集上进行情感识别，将得到的识别率与本文提出的融合特征进行比较。

3.3　实验结果

本文在汉语、英语、德语 3 个数据集上测试所提融合特征的识别率，并使用在这 3 个数据集上具有最好表现的不同声学特征作为参照。实验结果如图 2 所示，其中蓝色为本文所提出的融合特征，橘色为频谱特征 MFCC，灰色为韵律特征基音频率，黄色为音质特征共振峰。可以看出，本文所提的融合

特征在3个数据集上均有较好的识别率，在使用相同分类器进行分类时，识别率几乎达到最优；当使用不同分类器时，融合特征依然具有较好的识别能力，且识别性能稳定。具体识别率如表1所示。

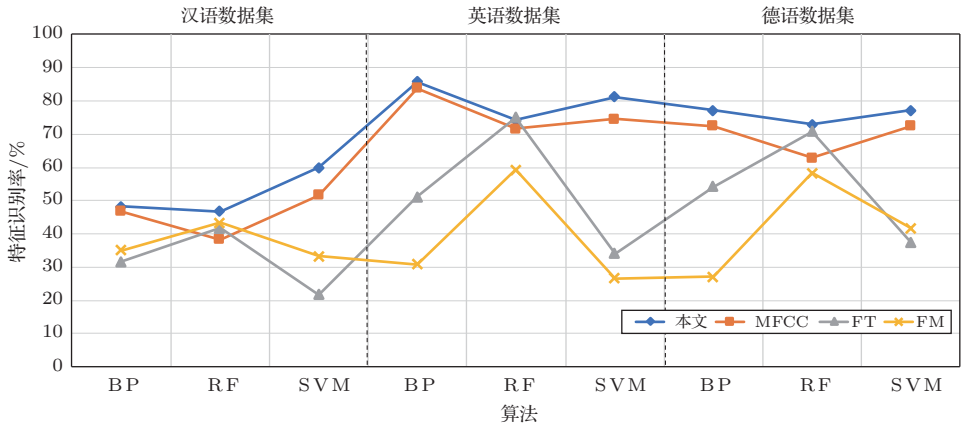


图2 不同数据集上融合特征的识别率

Fig. 2 The recognition rate of fusion features on different dataset

表1 不同数据集不同特征的识别率

Table 1 The recognition rate of fusion features on different dataset

	汉语数据集			英语数据集			德语数据集		
	BP	RF	SVM	BP	RF	SVM	BP	RF	SVM
OURS	**0.483**	**0.467**	**0.6**	**0.857**	0.742	**0.813**	**0.771**	**0.729**	**0.771**
MFCC	0.467	0.383	0.517	0.838	0.717	0.746	0.725	0.629	0.725
FT	0.317	0.417	0.217	0.511	**0.75**	0.341	0.542	0.708	0.375
FM	0.35	0.433	0.333	0.308	0.593	0.266	0.271	0.583	0.417

同时实验表明，本文所提出的融合特征在3个数据集上的识别率绝大多数优于单一特征，且识别性能稳定，能基本实现跨数据集的语音情感识别。

为进一步验证本文特征融合算法，另设计4组对照实验，随机选取3个单个特征进行融合，并与本文提出的融合特征的识别率进行比对，实验结果如图3所示。其中高亮显示的蓝色折线为本文提出的融合特征，橘色折线为以MFCC为主要特征、基音频率和共振峰二阶抖动为辅助特征得到的融合特征，灰色折线为以MFCC为主要特征、过零率和共振峰一阶抖动为辅助特征得到的融合特征，黄色折线为以基音频率为主要特征、过零率和RASTA-PLP为辅助特征得到的融合特征，绿色折线为以共振峰为主要特征、MFCC和RASTA-PLP为辅助特征得到的融合特征。可以看出，本文所提融合方法在英语和德语数据集上基本具有最好的识别率，在汉语数据集上与其他融合特征的识别率大致相当，总体上看，本文提出的融合特征识别效果最为稳定。具体识别率如表2所示。

观察图3中的蓝色折线、橘色折线和灰色折线可以发现，3条折线都是以MFCC特征为主要特征，选取其他两个特征作为辅助特征得到的融合特征，它们都保留了MFCC的全部13维特征；而黄色折线和绿色折线选用非MFCC特征作为主要特征，随机两个特征作为辅助特征进行融合。

可以看出，后者的识别率明显降低且表现不稳定，分析认为，MFCC谱特征与其他特征的相关统计量具有互补性，MFCC特征在情感识别中具有良好的鲁棒性，在其基础上通过添加辅助特征的识别效果来提高总体的识别率是一种行之有效的手段。而在选择辅助特征时，选择单个识别率稳定且表现最好的特征具有更好的效果。如果选用非谱特征作为主要特征进行融合，反而会失去主要特征的识别优势，辅助特征也无法起到互补的效果。

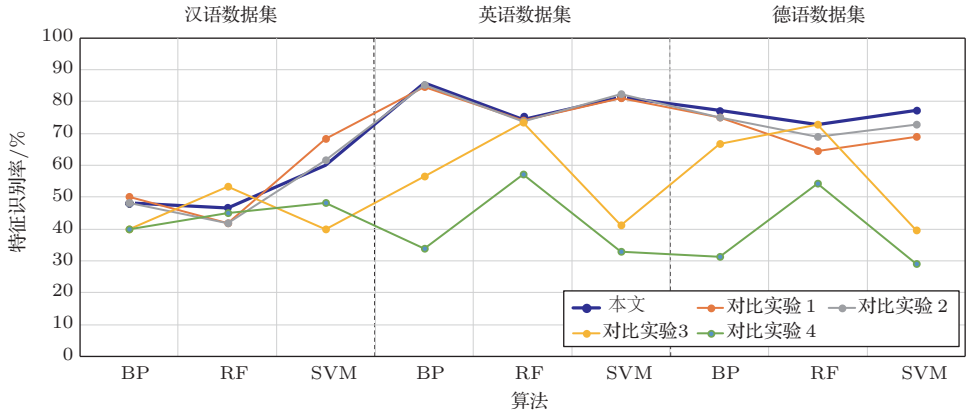


图 3　不同数据集上 5 组融合特征识别率

Fig. 3　The recognition rate of five sets of fusion features on different dataset

表 2　不同数据集上 5 组融合特征的识别率

Table 2　The recognition rate of five sets of fusion features on different dataset

	汉语数据集			英语数据集			德语数据集		
	BP	RF	SVM	BP	RF	SVM	BP	RF	SVM
OURS	0.483	0.467	0.6	**0.857**	**0.742**	0.813	**0.771**	**0.729**	**0.771**
对比实验 1	**0.5**	0.417	**0.683**	0.846	0.739	0.81	0.75	0.646	0.688
对比实验 2	0.483	0.417	0.617	0.852	0.736	**0.824**	0.75	0.688	0.729
对比实验 3	0.4	**0.533**	0.4	0.566	0.734	0.412	0.667	0.729	0.396
对比实验 4	0.4	0.45	0.483	0.338	0.571	0.33	0.312	0.542	0.292

辅助特征是以不同特征相关统计量的形式添加到主要特征中的，这使得主要特征的特征向量中出现了与原始数据差异较大的元素。在使用融合特征进行语音情感识别时，相比单个特征进行识别，实验时间有一定程度的增长。对于 BP 神经网络和随机森林算法，识别效率影响不大，但是对于 SVM 算法，实验时间明显增长，这是由于 SVM 算法会将数据输入到高维空间进行分类，差异较大的新元素的加入会使 SVM 运算量显著加大，从而导致实验时间的延长。

4　结论

研究发现，语音情感特征在不同分类器下具有不同的识别能力，本文提出的特征融合算法，保留了不同特征的优点，较好地实现了不同分类方式下的稳定识别，且在不同数据集上均能较好地完成识别。

目前语音情感识别依旧具有一定难度，不同语种数据集的识别率存在明显差异，这应该与文化、地域等诸多因素有关，如何在不同数据集上均能实现稳定的高识别率，这需要寻找一种更具有普遍性的声学特征。除此之外，人类情感具有模糊的时间边界，且一句话中很可能包含多种情感，如何实现长时语音的复杂情感识别，也是未来的研究方向。

参　考　文　献

[1] Nwe T L, Foo S W, de Silva L C. Speech emotion recognition using hidden Markov models[J]. Speech Communication, 2003, 41(4): 603–623.

[2] Mirsamadi S, Barsoum E, Zhang C. Automatic speech emotion recognition using recurrent neural networks with local attention[C]. 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), New Orleans, LA, 2017: 2227–2231.

[3] Hsiao P, Chen C. Effective attention mechanism in dynamic models for speech emotion recognition[C]. 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Calgary, AB, 2018: 2526–2530.

[4] Koolagudi S G, Rao K S. Emotion recognition from speech: a review[J]. International Journal of Speech Technology, 2012, 15(2): 99–117.

[5] Cao W, Xu J, Liu Z. Speaker-independent speech emotion recognition based on random forest feature selection algorithm[C]. 2017 36th Chinese Control Conference (CCC), Dalian, 2017: 10995–10998.

[6] 刘博, 范钰超, 徐明星. 基于特征级决策级双层融合的语音情感识别 [C]//中国中文信息学会语音信息专业委员会. 第十三届全国人机语音通讯学术会议 (NCMMSC2015) 论文集, 2015: 6.

[7] 张文克. 融合 LPCC 和 MFCC 特征参数的语音识别技术的研究 [D]. 长沙: 湘潭大学, 2016.

[8] 王忠民, 刘戈, 宋辉. 基于多核学习特征融合的语音情感识别方法 [J]. 计算机工程, 2019, 45(8): 248–254.
Wang Zhongmin, Liu Ge, Song Hui. Feature fusion based on multiple kernel learning for speech emotion recognition[J]. Computer Engineering, 2019, 45(8): 248–254.

[9] Liu G, He W, Jin B. Feature fusion of speech emotion recognition based on deep learning[C]. 2018 International Conference on Network Infrastructure and Digital Content (IC-NIDC), Guiyang, 2018: 193–197.

[10] 宋静, 张雪英, 孙颖, 等. 基于 PAD 情绪模型的情感语音识别 [J]. 微电子学与计算机, 2016, 33(9): 128–131.
Song Jing, Zhang Xueying, Sun Ying, et al. Emotional speech recognition based on PAD emotion model[J]. Microelectronics & Computer, 2016, 33(9): 128–131.

[11] Benesty J, Sondhi M M, Huang Y. Springer handbook of speech processing[M]. Berlin: Springer-Verlag, 2008.

[12] Sandipan C, Anindya R, Sourav M. Capturing complementaiy information via reversed filter bank and parallel implementation with MFCC for improved text-independent speaker identification[C]//Proceedings of the 2007 International Conference on Computing: Theory and Application. Piscataway: IEEE, 2007: 463–467.

[13] 鲜晓东, 樊宇星. 基于 Fisher 比的梅尔倒谱系数混合特征提取方法 [J]. 计算机应用, 2014, 34(2): 558–561, 579.
Xian Xiaodong, Fan Yuxing. Parameter extraction method for Mel frequency cepstral coefficients based on Fisher criterion[J]. Journal of Computer Applications, 2014, 34(2): 558–561, 579.

[14] 魏艳, 张雪英. 噪声条件下的语音特征 PLP 参数的提取 [J]. 太原理工大学学报, 2009, 40(3): 222–224.
Wei Yan, Zhang Xueying. A PLP speech feature extraction method in noisy environment[J]. Journal of Taiyuan University of Technology, 2009, 40(3): 222–224.

[15] Haque S, Togneri R, Zaknich A. Perceptual features for automatic speech recognition in noisy environments[J]. Speech Communication, 2008, 51(1): 15–25.

[16] 魏艳. 改进 RASTA-PLP 语音特征参数提取算法研究 [D]. 太原: 太原理工大学, 2009.

[17] Hermansky H, Morgan N, Bayya A, et al. RASTA-PLP speech analysis technique[C]// IEEE International Conference on Acoustics. IEEE, 1992.

[18] 韩文静, 李海峰, 阮华斌, 等. 语音情感识别研究进展综述 [J]. 软件学报, 2014, 25(1): 37–50.
Han Wenjing, Li Haifeng, Ruan Huabin, et al. Review on speech emotion recognition[J]. Journal of Software, 2014, 25(1): 37–50.

[19] 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.

[20] 高慧, 苏广川, 陈善广. 不同情绪状态下汉语语音的声学特征分析 [J]. 航天医学与医学工程, 2005(5): 350–354.
Gao Hui, Su Guangchuan, Chen Shanguang. Acoustic features analysis of mandarin speech under various emotional status[J]. Space Medicine & Medical Engineering, 2005(5): 350–354.

[21] 赵力. 语音信号处理 [M]. 北京: 机械工业出版社, 2016: 11–14.

[22] Adigwe A, Tits N, EI Haddad K, et al. The emotional voices database: towards controlling the emotion dimension in voice generation systems[J]. arXiv: 1806.09514, 2018.

[23] Burkhardt F, Paeschke A, Rolfes M, et al. A database of German emotional speech[C]. In: Proc. of the 2005 INTERSPEECH. Lisbon: ISCA, 2005: 1517–1520.