Specific Two Words Chinese Lexical Recognition Based on Broadband and Narrowband Spectrogram Feature Fusion with Zoning Projection

WEI Ying1 WANG Shuang-wei1 PAN Di3 ZHANG Tingfa3 LIANG Shili3

(Department of Physics, Northeast Normal University, Changchun 130024, China)1

(College of Science, Jilin University of Science and Technology, Changchun 130022, China)3

(Key Laboratory of Photo Electronic Imaging Technology and Systems, Ministry of Education of China,
Beijing Institute of Technology, Beijing 100081, China)3

Abstract A method based on broadband and narrowband spectrogram feature fusion with zoning projection of specific two words Chinese lexical recognition was presented. In the process of image feature extraction, the image processing technique is applied to the speech recognition field. Firstly, equal width zoning line projection and binary width zoning line projection are carried out in the narrowband spectrogram, and they are set respectively as the narrowband spectrogram of the first characteristic set and the second characteristic set. Meanwhile, equal width zoning line projection is carried out again to the broadband spectrogram after Fourier transform, treating it as the third feature set. Then, equal width column projection is carried out to the broadband spectrogram, regarding it as the fourth feature set. The above three feature sets are used as feature vectors to support vector machine (SVM) as a classifier for the overall recognition of specific two words Chinese vocabulary. 1000 voice samples are used in simulation experiment. The results show that the correct recognition rate of the two words Chinese word recognition by the first three feature sets is 92.4%. The correct recognition rate of two words vocabulary recognition using fourth feature sets is 80%. The correct recognition rate of the two words Chinese word recognition by using the feature value fusion of the above four features can reach 95.4%. This method of feature fusion provides a new way of thinking of Chinese vocabulary overall recognition.

Keywords Speech recognition, Spectrogram, Feature fusion, Line projection, Column projection, Support vector machine (SVM)
图像处理技术的发展使得语音信号的语谱图域参数的提取成为识别语音信号的关键。语音信号的语谱图特征在语音识别中的应用已经得到了广泛的研究。

对于语音信号的语谱图特征，研究者提出将语音信号变换成语谱图图像并对语谱图行平均向量进行说话人识别。这种方法为语音识别提供了一个全新的视角。

2.1 快速傅里叶变换语谱图

快速傅里叶变换（FFT）是将原始信息中频率信息变换成语音语谱图的常见方法。通过对语音语谱图的音素分割和模板匹配方式，可以提高语音识别的鲁棒性。

2.2 纠正

近年来，针对语音信号的语谱图特征的识别方法，研究人员提出将语音信号的语谱图特征融合得到的特征值对特定人二字汉语词汇进行识别。这种方法可以实现特定人二字汉语词汇的识别。

2.3 视觉语谱图

视觉语谱图是一种新的思路，将视觉语谱图应用到自动语音识别系统中，使得语谱图与傅里叶变换更好地结合，提高了语音识别的鲁棒性。

2.4 纠正

由于窄带语谱图有较高的频域分辨率而宽带语谱图有较高的时域分辨率，因此本文选用窄带语谱图行投影值和宽带语谱图分带列投影得到的特征值对特定人二字汉语词汇的识别率可达95.4%。
图中显示每个语音帧长分别为14.4032,s 16.4753,s ... 数字信号的时域波形图 (Hamming)。数据文件部分数据检验结果表为相应的宽带语谱图，图中显示其频率分辨率为4 k Hz。数据之间的识别差异为样本所对应的语谱图矩阵，以下部分数据检验发现，每一幅语谱图的矩

3.2

3.2.1

\[ a = \frac{S_{14}}{N_{1}}, \quad b = \frac{S_{15}}{N_{1}}. \]
3.2.2 多带投影分析

将语谱图矩阵分带投影到第 100 带，每带宽 256, 257-258, 259-260, 261-262, 263-264 带。每带的带宽可以是可选的，分为不同大小的带，如窄带（< 1000Hz）和宽带（> 1000Hz）。每带含有特定频率范围的信息，通过窄带和宽带的结合，可以更好地反映出语音信号的整体性。例如，语音信号的低频部分可以被分到图像的中间部分，而高频部分则可以被分到图像的上部。

3.2.3 支持向量机

本文对语音识别系统的容错能力进行了研究，通过窄带和宽带的融合，可以有效地提高语音识别的准确性和可靠性。实验表明，窄带和宽带的融合可以显著提高语音识别的效果，如图所示。

4.1 实验结果

实验表明，宽窄带语谱图融合的特征数据集合可以被很好地应用于窄带和宽带的语音识别。窄带和宽带的特征数据融合可以提高语音识别的准确性，如图所示。实验结果表明，窄带和宽带的特征数据融合可以提高语音识别的准确性。
4.2
4.2.1

测样本分带行投影和二进宽度分带行投影频域分辨率而宽带语谱图有较高的时间分辨率。对宽带语谱图矩阵进行等宽度分带列投影组数据将投影值作为特征集合。

特征量得出对语词汇识别的仿真结果。一个人的相同词汇前，对宽带语谱图矩阵进行等宽度分带列投影。一组数据作为训练样本数据特征值，分别相应地将后，一组训练样本数据特征值。因此将窄带语谱图矩阵进行等宽度行投影和二进宽度行投影，后，一组训练样本数据特征值。由于窄带语谱图有较高的，一组训练样本数据特征值。这已经这些没有自己中国可以问题工作生活这样。高频带等宽度行投影，一组训练样本数据特征值。数据和检测样本数据给予显示。

4.2.2

对窄带语谱图矩阵等宽度行投影和二进宽度行投影，一组训练样本数据特征值。组数据放入训练好的系统中进行检测。实验结果表明，对于多字词汇量语音的识别还有。实验结果表明，对于多字词汇量语音的识别还有。
1 0 1 3


[27] 2010, 328-332

[28] 2008

[29] Chang C C, Lin C J. A Library for Support Vector Machines [M], National Taiwan University, 2001

• 232 •