

基于小波变换和倒谱分析的腭裂高鼻音等级自动识别

赵利博 刘 奇 付方玲 何 凌

(四川大学电气信息学院 成都 610065)

摘 要 为实现对腭裂高鼻音等级的自动识别,通过对语音信号小波处理和特征提取方法的综合研究,提出基于小波分解系数倒谱特征的腭裂高鼻音等级自动识别算法。目前,研究人员对腭裂语音的研究多基于 MFCC、Teager 能量、香农能量等特征,识别正确率偏低,且计算量过大。文中对 4 种等级腭裂高鼻音的 1789 个元音\ a\ 语音数据提取小波分解系数倒谱特征参数,使用 KNN 分类器对 4 种不同等级的高鼻音进行自动识别,将识别结果与 MFCC、LPCC、基音周期、共振峰和短时能量共 5 种经典声学特征的识别结果作比较,同时使用 SVM 分类器对不同等级的腭裂高鼻音进行自动识别,并与 KNN 分类器进行对比。实验结果表明,基于小波分解系数倒谱特征的识别结果优于经典声学特征,且 KNN 分类器的识别结果优于 SVM 分类器。小波分解系数倒谱特征在 KNN 中的识别率最高达到 91.67%,在 SVM 中达到 87.60%,经典声学特征在 KNN 分类器中的识别率为 21.69%~84.54%,在 SVM 中的识别率为 30.61%~78.24%。

关键词 腭裂,高鼻音,分类系统,小波分解系数倒谱

中图分类号 TP391.9

文献标识码 A

DOI 10.11896/j.issn.1002-137X.2018.04.047

Automatic Detection of Hypernasality Grades Based on Discrete Wavelet Transformation and Cepstrum Analysis

ZHAO Li-bo LIU Qi FU Fang-ling HE Ling

(College of Electronics and Information Engineering, Sichuan University, Chengdu 610065, China)

Abstract This paper proposed an automatic hypernasality grades classification algorithm in cleft palate speech based on discrete wavelet decomposition coefficients and cepstrum analysis. Currently, the widely used features to classify hypernasality grades include MFCC, Teager energy, Shannon energy and so on. However, the classification accuracy is low, and the computation amount is large. The speech data tested in this work include 1789 Mandarin syllables with the final \ a \, which are spoken by cleft palate patients with four grades of hypernasality. The wavelet decomposition coefficient-cepstrum was extracted as the acoustic feature, and then KNN classifier was applied to identify four grades of hypernasality. The classification performance was compared with five acoustic features: MFCC, LPCC, pitch period, formant and short-time energy. Meanwhile, the performance of KNN was compared with SVM classifier. The experimental results indicate that the recognition accuracy obtained by using wavelet decomposition coefficient cepstrum feature is higher than that obtained by using five classical acoustics features. The classification accuracy is higher when using KNN than SVM classifier. Recognition accuracy obtained by using wavelet decomposition coefficient cepstrum feature combined with KNN is 91.67%, and 87.60% combined with SVM. Recognition accuracy using classical acoustics features combined with KNN is only 21.69%~84.54%, and 30.61%~78.24% combined with SVM.

Keywords Cleft palate, Hypernasality, Recognition system, Wavelet decomposition coefficient cepstrum

1 引言

腭裂是口腔颌面部常见的先天畸形。腭裂的明显临床表现为软腭与咽壁无法紧密接触,造成患者口鼻腔不能完全分离,发音时控音和共振装置出现问题,语音时会伴随浓厚的鼻音,发生语言障碍,产生高鼻音症状。腭裂患者无法像正常人一样清楚地发音和交流,这对患者生理和心理造成了严重的创伤。在中国,腭裂的发病率高达 0.182%^[1]。目前,临床上主要通过语音师的临床经验和 X 线影像、鼻咽纤维镜评估、鼻音计估计等技术实现对腭裂高鼻音的评估。这些检测、评

估手段存在易受个人主观因素和环境影响、易对被测者身体造成辐射损伤、增加心理压力等缺点。利用计算机信号处理和模式识别方法实现对腭裂语音等级的自动评估,具有无创、准确率高、检测方法简单的优点。

目前,对腭裂语音的研究主要包括共振异常、高鼻音检测、鼻漏气检测和代偿性构音。其中,高鼻音是腭裂语音的一种突出的临床表现。不同程度的高鼻音等级对应不同程度的腭咽闭合状态。并且,对于高鼻音的正确检测及高鼻音等级的评估,能有效评价腭咽闭合的完整性。同时,高鼻音的辨别与分类能有效地辅助医生诊断腭裂病情,因此其在临床上得

到稿日期:2017-03-06 返修日期:2017-06-17 本文受国家自然科学基金青年科学基金项目(61503264)资助。

赵利博(1993—),男,硕士生,主要研究方向为语音信号处理,E-mail:ling.he@scu.edu.cn;刘 奇(1967—),男,博士,教授,主要研究方向为医学信号和图像处理;付方玲(1996—),女,硕士生,主要研究方向为语音信号处理;何 凌(1981—),女,博士,讲师,主要研究方向为语音信号处理,E-mail:ling.he@scu.edu.cn(通信作者)。

到了更为广泛的应用^[2-4]。近年来,国内外许多研究人员利用了高鼻音的声学特征,并利用不同的分类器和特征提取方法对腭裂高鼻音及其等级进行识别。Maier 等人^[5]结合线性判别分类器、决策树模型分类器、SVM 分类器,使用 Teager 能量和 MFCC 特征对腭裂儿童是否存在高鼻音进行识别,正确率达到 75.80%,但数据量较少,且模型复杂度较高;Murillo 等人^[6]针对儿童的腭裂高鼻音进行研究,通过主成分分析方法对语音特征进行降维,使用贝叶斯分类器识别高鼻音等级,识别正确率达到 88.82%,但该方法识别过程复杂,特征降维过程繁琐,同时主成分分析方法会丢弃一些对识别有效的特征,降低识别结果的准确率;He 等^[7]使用短时香农能量、Teager 能量、MFCC 和前四阶能量频带,基于 GMM 模型分类器对 4 种腭裂高鼻音等级(正常、轻度、中度、重度)进行识别,正确率为 80.40%,但 GMM 模型的复杂度较高。

本文提出一种基于小波变换和倒谱分析的腭裂高鼻音声学特征提取算法,并应用 KNN 和 SVM 两种分类器实现高鼻音等级的自动识别,将提取的小波分解系数倒谱和倒谱低阶参数作为腭裂高鼻音的识别声学特征参数,同时与经典声学特征进行了比较。小波分解是一种多分辨率分析方法,尤其适合非平稳特性的语音信号,对语音信号进行小波分解能提取出反映语音信号不同频段特性的系数。同时,语音信号的倒谱参数基于同态分析理论,解卷积语音信号声门激励信息和声道响应信息,提取倒谱低阶参数可分离出声道响应信息,其能很好地反映腭裂患者腭咽闭合不全的特性。实验结果表明,使用小波分解系数倒谱及其低阶参数声学特征参数,结合 KNN 分类器对 4 类(正常、轻度、中度和重度)高鼻音自动识别的正确率为 91.67%。

2 腭裂高鼻音等级自动识别系统

本文提取腭裂高鼻音语音信号小波分解系数倒谱特征和经典声学特征作为区分高鼻音等级的特征,使用 KNN 和 SVM 两种分类器对腭裂高鼻音等级进行识别,识别系统主要包括语音信号输入、声学特征提取、模型训练、识别分类 4 个部分。算法流程图如图 1 所示。

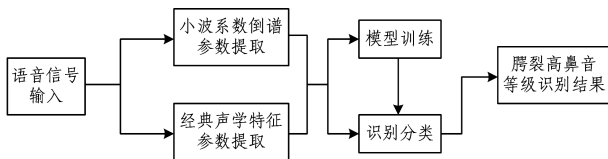


图 1 腭裂高鼻音等级自动识别算法流程图

Fig. 1 Flowchart of automatic cleft palate hypernasality grades recognition method

将提取的腭裂高鼻音信号的小波分解系数倒谱特征和经典声学特征分别分为训练数据和测试数据。训练数据对分类器进行模型训练,测试数据测试训练好的分类器,以得到高鼻音等级的判定结果。

3 腭裂高鼻音特征提取

特征提取是自动识别系统最重要的一个步骤,提取的特征参数的优劣直接影响着系统识别正确率的高低,甚至对识别结果有决定性作用。腭裂高鼻音的特征提取是将能够区分高鼻音等级的显著特征提取出来,并作为高鼻音等级自动识别系统的输入特征,通过对特征的训练,得到能够自动识别

高鼻音等级的系统。

通过对腭裂高鼻音的研究发现,腭裂患者与正常人发元音时,有着不同的语音特点:鼻化元音在频谱特征上的低频区明显存在着额外的共振峰,在高频区的能量相比于正常元音,出现了明显衰减,其依据高鼻音的严重程度而呈现不同的表现。因此,本实验采用单元音\`a\`的语音数据进行腭裂高鼻音等级识别。图 2 为 4 种等级的\`a\`语音的波形图。

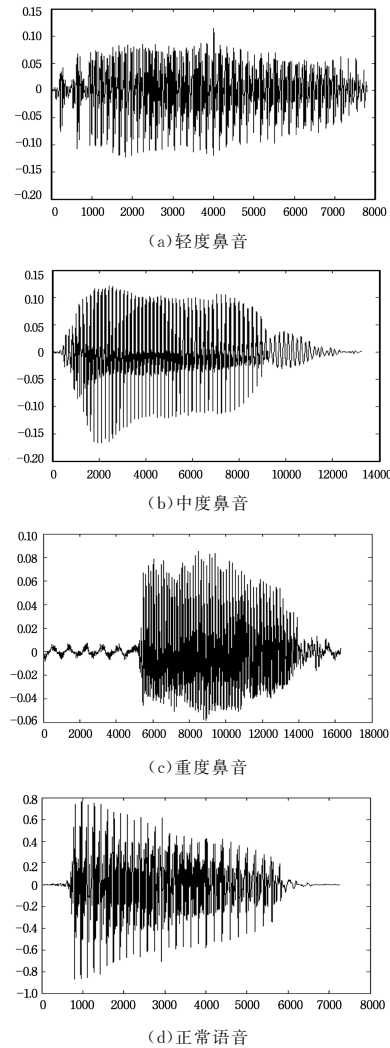


图 2 不同等级的高鼻音语音时域波形图

Fig. 2 Time domain waveforms of nasal speech in different hypernasality grades

3.1 基于小波变换和倒谱分析的腭裂高鼻音特征提取

3.1.1 小波变换原理

小波变换^[8]是一种十分有效的调和工具,分为连续小波变换和离散小波变换。小波变换采用面积不变、形状改变的分析窗口对信号进行变换,可对信号进行多分辨率分析,尤其适合非平稳信号^[9]。

1) 连续小波变换

小波分析主要针对函数空间 $L^2(R)$ 内平方可积的函数,即函数应满足条件:

$$\int |f(t)|^2 dt < +\infty \quad (1)$$

其中, $f(t)$ 表示能量有限信号。

小波变换中,母小波 $\psi(t) \in L^2(R)$,对母小波进行伸缩和平移,得到小波序列:

$$\psi_{a,\tau}(t) = |a|^{-\frac{1}{2}} \psi\left(\frac{t-\tau}{a}\right) \quad (2)$$

其中, $a, \tau \in R, a$ 表示伸缩因子且 $a \neq 0, \tau$ 表示平移因子, 则母小波 $\psi(t)$ 的连续小波变换定义为:

$$W_\psi(a, \tau) = |a|^{-\frac{1}{2}} \int_{-\infty}^{+\infty} f(t) \overline{\psi\left(\frac{t-\tau}{a}\right)} dt \quad (3)$$

其中, $\overline{\psi\left(\frac{t-\tau}{a}\right)}$ 表示对 $\psi\left(\frac{t-\tau}{a}\right)$ 的共轭运算。

2) 离散小波变换

离散小波变换是在连续小波变换的基础上, 对伸缩因子和平移因子离散化处理得到, 即取:

$$a = a_0^m, \tau = n\tau_0 a_0^m, m, n \in Z \quad (4)$$

从而得到相应的离散小波:

$$\psi_{m,n}(t) = |a_0|^{-\frac{m}{2}} \psi(a_0^{-m} t - n\tau_0), m, n \in Z \quad (5)$$

离散小波变换为:

$$W_\psi(a, \tau) = |a_0|^{-\frac{m}{2}} \int_{-\infty}^{+\infty} f(t) \overline{\psi(a_0^{-m} t - n\tau_0)} dt \quad (6)$$

图 3 为信号的离散小波变换结果。

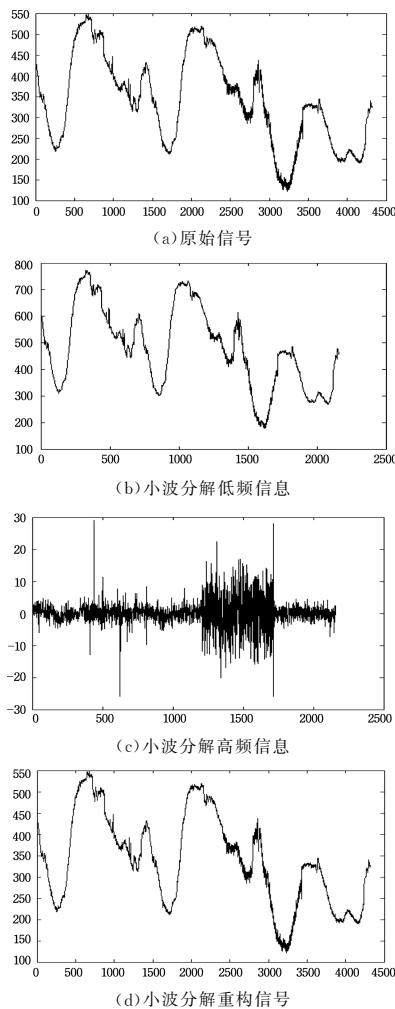


图 3 信号小波分解波形图

Fig. 3 Wavelet decomposition waveforms of speech signal

基于 Mallat 算法^[10-11]的离散小波变换对非平稳语音信号进行处理, 变换后的信号可分为两部分: 低频信号和高频信号。选择不同尺度的小波函数, 可以在任意尺度下观察语音信号。

3.1.2 腭裂高鼻音小波变换参数提取

对腭裂高鼻音语音信号进行小波变换, 提取高频和低频

参数。设高鼻音信号为 $x(n)$, 对其进行小波变换:

$$W_\psi(x) = |a_1|^{-\frac{m}{2}} \int_{-\infty}^{+\infty} x(n) \overline{\psi(a_1^{-m} n - n_1 b_1)} dn \quad (7)$$

腭裂高鼻音信号经过小波变换后得到小波系数, 对小波系数做进一步的滤波处理, 提取出能表示基音和语义的低频系数和表示噪声和突变信号的高频系数, 这些参数特征能够很好地反映出不同等级的高鼻音的特点。

Daubechies 小波函数^[12]经常作为滤波函数, 其具有良好的正则性, 可进行连续小波变换和离散小波变换, 具有随着阶数增加, 频域局部化能力增强的特性, 但计算量会明显增大。选用 2 阶 Daubechies 小波函数作为滤波函数, 既能减少计算量, 也能完成高鼻音信号的高通和低通滤波:

$$WH(x) = W_\psi(x) * H(x) \quad (8)$$

其中, $WH(x)$ 表示小波变换系数高频分量, $H(x)$ 表示高通小波滤波器。

$$WL(x) = W_\psi(x) * L(x) \quad (9)$$

其中, $WL(x)$ 表示小波变换系数低频分量, $L(x)$ 表示低通小波滤波器。

图 4 和图 5 分别表示正常人和重度高鼻音腭裂患者的 /ca/ 语音信号提取一帧的离散小波变换结果。可以看出, 正常人与重度高鼻音腭裂患者的语音信号波形图存在明显的不同。

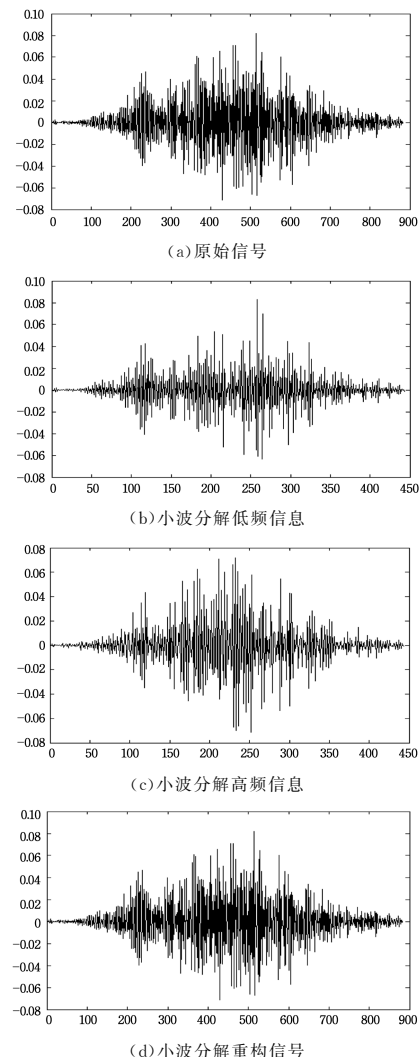


图 4 正常语音信号波形图

Fig. 4 Time domain waveform of normal speech signal

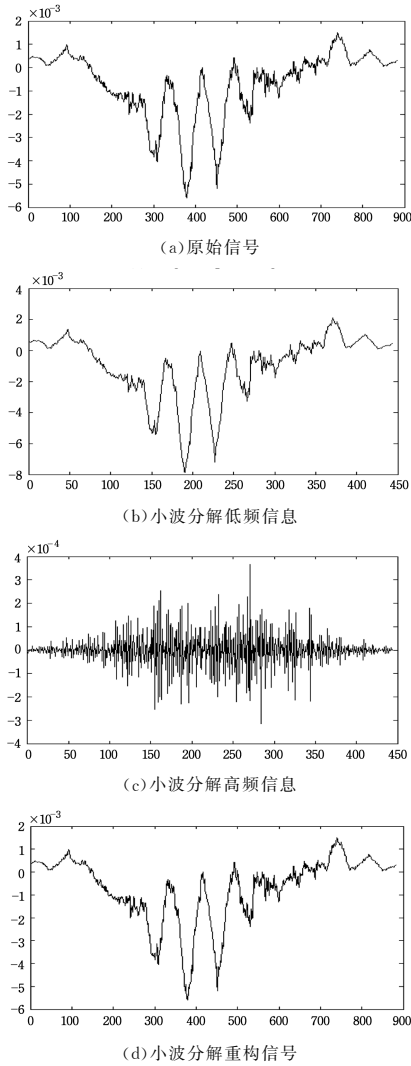


图 5 重度高鼻音信号波形图

Fig. 5 Time domain waveform of severe nasal speech signal

3.1.3 腭裂高鼻音小波变换系数倒谱

倒谱^[13]反映语音信号的声门激励和声道响应信息,对语音信号倒谱进行提取和研究,有助于进一步了解语音的产生和声道内的传播,还可对发声器官的正常性、声道结构的完整性等做出评估。腭裂患者因为腭咽结构的不健全,声道内骨组织和软组织缺损会造成口鼻腔相通,发音时引起严重的口鼻腔共振,产生高鼻音,并且声道的不同受损程度会产生不同等级的高鼻音,因此,倒谱特征可用来作为区分高鼻音等级的一种特征。对上一节中求出的腭裂高鼻音小波分解系数求取倒谱,先对小波变换系数分量进行傅里叶变换:

$$F_H[WH(x)] = \sum_{n=0}^{N-1} WH(x) e^{-j\frac{2\pi}{N}kn}, 0 \leq n \leq N-1 \quad (10)$$

$$F_L[WL(x)] = \sum_{n=0}^{N-1} WL(x) e^{-j\frac{2\pi}{N}kn}, 0 \leq n \leq N-1 \quad (11)$$

其中, F 表示离散傅里叶变换,对其取对数得到:

$$\hat{F}_H = \ln|F_H[WH(x)]| + j\arg\{F_H[WH(x)]\} \quad (12)$$

$$\hat{F}_L = \ln|F_L[WL(x)]| + j\arg\{F_L[WL(x)]\} \quad (13)$$

对上式取实部:

$$Re_H(\hat{F}_H) = \ln|F_H[WH(x)]| \quad (14)$$

$$Re_L(\hat{F}_L) = \ln|F_L[WL(x)]| \quad (15)$$

Re 表示取实部。然后进行离散傅里叶逆变换:

$$\hat{C}_H(n) = F^{-1}[\ln|F_H[WH(x)]|] \quad (16)$$

$$\hat{C}_L(n) = F^{-1}[\ln|F_L[WL(x)]|] \quad (17)$$

其中, $\hat{C}(n)$ 即为倒谱, F^{-1} 表示离散傅里叶逆变换。

3.1.4 腭裂高鼻音小波变换系数倒谱低阶参数

腭裂高鼻音小波变换系数倒谱反映了语音信号的声门激励信息和声道响应信息,提取倒谱低阶参数可以分离声道响应信息。腭裂患者发音不清楚是由于声道结构不完全,腭咽不能完全闭合造成的,通过对声道响应信息的分析,可以得到患者的腭裂高鼻音的严重程度的信息。对腭裂高鼻音小波变换系数倒谱提取低阶参数的过程如下:

$$C_{H\text{-para}}(n) = \hat{C}_H(n_{low}) \quad (18)$$

$$C_{L\text{-para}}(n) = \hat{C}_L(n_{low}) \quad (19)$$

其中, $C_{H\text{-para}}(n)$ 表示小波变换高频系数倒谱低阶参数, $C_{L\text{-para}}(n)$ 表示小波变换低频系数倒谱低阶参数。

3.2 经典声学特征

本文提取的腭裂高鼻音经典声学特征^[13-14]包括 MFCC、线性预测倒谱系数(LPCC)、基音周期、共振峰和短时能量。腭裂患者因为口腔结构的不完整性产生的高鼻音与正常人语音在以上 5 种特征方面存在不同,通过提取经典声学特征,能够实现对其高鼻音等级的分类识别。

4 KNN 和 SVM 分类器

4.1 KNN 分类器

KNN^[15]通过计算待分类样本点与其他所有样本点的距离,取出与待分类样本点最近的 k 个样本点,统计 k 个样本点中所属分类比例最大的种类,将待分类样本点划归此类,即可得到分类结果。

本文中的腭裂高鼻音数据共有 4 类:轻度、中度、重度和正常。将其每一类记为 $N_i, i=1,2,3,4$,依次代表 4 类。所有样本的属性维度为 n 。

1) 计算测试样本点与训练样本集中所有样本点的欧氏距离:

$$dis(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (20)$$

其中, x 表示测试样本点, y 表示训练样本点。

2) 将距离按升序排列,选出 k 个最小值。

3) 根据式(21)求出 k 个最小距离测试样本点与训练样本点之间的相似度:

$$Simulate(x, y) = \frac{1}{1 + dis(x, y)} \quad (21)$$

4) 统计最近邻训练样本点中属于不同类的个数:

$$Num(x, N_i) = \sum_{j=1}^k Simulate(x, y_j) \quad (22)$$

5) 将测试样本点分在最近邻所属类别个数最多的类中:

$$i = \max_i (Num(x, N_i)) \quad (23)$$

4.2 SVM 分类器

支持向量机(SVM)^[16]是一种具有良好鲁棒性的分类器,常用于二分类。SVM 多分类系统中,在任意两类样本之间设计一个 SVM,即可通过投票的方式判定测试样本点所属类别。本文中有 4 个高鼻音类别,需要设计 6 个 SVM,每个 SVM 的分类超平面为:

$$\sum_{j=1}^m w_j x + b_j = 0 \quad (24)$$

其中, w 表示 SVM 支持向量, m 表示支持向量的个数, b 表示超平面偏置。

当输入一个测试样本点时:

1) 计算测试样本点与超平面之间的关系:

$$f(X) = \sum_{j=1}^m w_j X + b_j \quad (25)$$

2) 根据判别公式判定测试样本点所属类别:

$$\delta(x) = \begin{cases} 1, & f(X) > 0 \\ 0, & f(X) \leq 0 \end{cases} \quad (26)$$

3) 统计 6 个 SVM 分类器的投票结果:

$$Num(i) = \sum_{j=1}^6 f(X_j) \quad (27)$$

4) 得票最多的即为测试样本点的类别:

$$i = \max(Num(i)) \quad (28)$$

5 腭裂高鼻音等级识别结果及分析

5.1 实验数据

目前, 腭裂语音信号研究的瓶颈是腭裂语音信号的采集问题。腭裂语音的采集受到多方面的限制: 腭裂病人的数量稀少、腭裂病人的口音参差不齐以及反映腭裂语音特点的词汇列表设计难度大等。此外, 腭裂患者大多为儿童, 儿童不容易配合医生进行数据采集。本文的数据采集工作是由四川大学华西口腔医院唇腭裂外科“腭裂语音治疗中心”的专业语音师团队完成。该中心是国内最大的唇腭裂治疗中心, 有近十年的腭裂治疗经验, 存储了大量的腭裂语音数据, 并建立有语音数据库, 形成了对腭裂语音完善的治疗和评估机制, 数据具有很高的可靠性。该数据库包含儿童、少年和成年腭裂患者的语音样本。数据库中的所有语音都由语音师团队通过人工听来评估和标注。由于数据量的庞大和腭裂病人特有的共鸣和构音的语言障碍, 人工评估工作量巨大, 并且要求保证极高的准确率。语音师评估和标注的结果被全国各级医院认定为最权威的“金标准”, 因此将其作为本文自动高鼻音等级识别的参考标准。

本次实验数据来自 80 个腭裂病人。其中, 男性 40 人分, 为 4 组, 每组 10 人, 分别对应正常、轻度、中度和重度 4 个高鼻音等级; 同样地, 女性 40 人, 分为 4 组, 每组 10 人, 分别对应正常、轻度、中度和重度 4 个高鼻音等级。

5.2 KNN 识别结果及分析

对 KNN 分类系统进行训练, 将腭裂高鼻音的小波分解系数倒谱及其低阶参数和 5 种经典声学特征分别分成训练数据和测试数据, 并将其作为分类器的输入。训练后的 KNN 分类器对高鼻音的识别正确率如表 1—表 4 所列。

表 1 女性腭裂高鼻音等级识别结果(KNN)

Table 1 Automatic hypernasality grades classification accuracy in cleft palate speech spoken by female patients(KNN)

(单位: %)	
语音特征	正确率
小波分解高频系数倒谱	81.95
小波分解低频系数倒谱	80.92
MFCC	82.39
LPCC	74.40
基音周期	21.69
共振峰	39.36
短时能量	66.00

表 2 男性腭裂高鼻音等级识别结果(KNN)

Table 2 Automatic hypernasality grades classification accuracy in cleft palate speech spoken by male patients(KNN)

(单位: %)	
语音特征	正确率
小波分解高频系数倒谱	83.49
小波分解低频系数倒谱	84.41
MFCC	84.54
LPCC	67.82
基音周期	27.24
共振峰	35.79
短时能量	67.23

表 3 使用低阶参数的女性腭裂高鼻音等级识别结果(KNN)

Table 3 Automatic hypernasality grades classification accuracy in cleft palate speech spoken by female patients using parameters with lower order(KNN)

(单位: %)	
语音特征	正确率
高频系数倒谱低阶参数	86.11
低频系数倒谱低阶参数	91.67

表 4 使用低阶参数的男性腭裂高鼻音等级识别结果(KNN)

Table 4 Automatic hypernasality grades classification accuracy in cleft palate speech spoken by male patients using parameters with lower order(KNN)

(单位: %)	
语音特征	正确率
高频系数倒谱低阶参数	83.33
低频系数倒谱低阶参数	88.89

从表 1 和表 2 可以看出, KNN 分类器在女性腭裂高鼻音等级识别中, 使用小波分解高频系数倒谱的正确率为 81.95%, 而低频系数倒谱的正确率为 80.92%, 对于男性, 其正确率则分别为 83.49% 和 84.41%, 均优于其他经典声学特征的识别结果。

表 3 和表 4 给出了提取小波分解系数倒谱的低阶参数作为分类器特征得到的识别结果。分别使用小波分解高频和低频系数倒谱低阶参数作为特征进行训练和识别, 对女性腭裂高鼻音的分类正确率达到 86.11% 和 91.67%, 对男性腭裂高鼻音的分类正确率为 83.33% 和 88.89%。因此使用这两个特征优于使用其他特征得到的识别正确率。

5.3 SVM 识别结果及分析

对 SVM 分类系统进行训练, 将腭裂高鼻音的小波分解系数倒谱及其低阶参数和 5 种经典声学特征分别分成训练数据和测试数据, 并将其作为分类器的输入。训练后的 SVM 分类器对高鼻音的识别正确率如表 5—表 8 所列。

表 5 女性腭裂高鼻音等级识别结果(SVM)

Table 5 Automatic hypernasality grades classification accuracy in cleft palate speech spoken by female patient(SVM)

(单位: %)	
语音特征	正确率
小波分解高频系数倒谱	81.29
小波分解低频系数倒谱	82.08
MFCC	77.46
LPCC	67.02
基音周期	31.68
共振峰	30.72
短时能量	42.66

表 6 男性腭裂高鼻音等级识别结果(SVM)

Table 6 Automatic hypernasality grades classification accuracy in cleft palate speech spoken by male patients(SVM)

(单位:%)	
语音特征	正确率
小波分解高频系数倒谱	77.74
小波分解低频系数倒谱	78.82
MFCC	78.24
LPCC	58.36
基音周期	40.16
共振峰	30.61
短时能量	43.86

表 7 使用低阶参数的女性腭裂高鼻音等级识别结果(SVM)

Table 7 Automatic hypernasality grades classification accuracy in cleft palate speech spoken by female patients using parameters with lower order(SVM)

(单位:%)	
语音特征	正确率
高频系数倒谱低阶参数	85.95
低频系数倒谱低阶参数	87.60

表 8 使用低阶参数的男性腭裂高鼻音等级识别结果(SVM)

Table 8 Automatic hypernasality grades classification accuracy in cleft palate speech spoken by male patients using parameters with lower order(SVM)

(单位:%)	
语音特征	正确率/%
高频系数倒谱低阶参数	83.47
低频系数倒谱低阶参数	86.78

从表 5 和表 6 可以看出,在 SVM 分类器对女性腭裂高鼻音等级识别中,使用小波分解高频系数倒谱的正确率为 81.29%,而低频系数倒谱的正确率为 82.08%,男性的正确率则分别为 77.74%和 78.82%,优于其他经典的声学特征的识别结果。

表 7 和表 8 给出了提取小波分解系数倒谱的低阶参数作为分类器特征得到的识别结果。分别使用小波分解高频和低频系数倒谱低阶参数特征,对女性腭裂高鼻音的分类正确率达到 85.95%和 87.60%,对男性腭裂高鼻音的分类正确率为 83.47%和 86.78%,相比于使用其他特征所得到的识别结果,其得到了更高的识别正确率。

5.3.1 小波变换系数倒谱特征与经典声学特征的对比分析

从最终的识别正确率可以看出,在 KNN 和 SVM 两种分类器中,本文提出的腭裂高鼻音小波分解系数倒谱特征的分类识别正确率优于 5 种经典声学特征,即 MFCC、LPCC、基音周期、共振峰和短时能量,主要原因在于:

1) 本文提出的小波变换系数倒谱低阶参数特征充分利用了腭裂患者的声道信息,可以较好地反映出不同程度的腭裂患者的腭咽闭合情况。腭咽口闭合程度的不同会产生不同等级的高鼻音,闭合程度越小,高鼻音越严重。因此,使用小波变换系数倒谱低阶参数作为特征的识别正确率更高。

2) 提取 MFCC 特征参数采用的是三角滤波器组,会造成相邻频带之间频谱能量的泄露。在腭裂高鼻音中,腭咽闭合不全也会造成频谱能量泄露,从而降低识别正确率;并且本文

对腭裂高鼻音等级识别的研究,是利用正常人与不同高鼻音等级腭裂患者声道特征信息存在的不同来实现的,而 MFCC 不能详细描述声道信息。因此在腭裂高鼻音等级识别中,利用 MFCC 特征的识别效果有限。

3) LPCC 特征参数是根据人的声管模型得到的。语音信息主要集中在低频部分,而 LPCC 不能详细描述语音低频部分的局部特征。腭裂高鼻音等级识别需要提取声道信息的详细特征,而 LPCC 特征参数会丢失一部分语音的低频特征,从而降低识别正确率。

4) 基音周期特征是声带振动频率的倒数,正常人与不同高鼻音等级腭裂患者在发音时,声带振动频率差异性较小,而高鼻音等级的不同是由腭咽闭合程度的不同引起,主要是在声道模型上的差异。因此将基音周期作为腭裂高鼻音等级识别的特征,得到的识别正确率较低。

5) 共振峰特征是声道谐振的重要特性,共振峰信息包含在语音信号谱包络中,谱包络中的最大值一般被认为是共振峰。在腭裂高鼻音的等级识别中,腭咽闭合不全会造成腭裂高鼻音共振峰出现虚假峰,从而对识别正确率造成不良影响。

6) 短时能量特征反映了语音信号能量的变化,它与人的性别、年龄、音调等因素有关,这些因素会干扰对腭裂高鼻音等级使用短时能量特征的识别效果。因此,使用短时能量识别腭裂高鼻音的等级的正确率较低。

5.3.2 KNN 分类器与 SVM 分类器的对比分析

从实验结果可以看出,在腭裂高鼻音的等级识别中,KNN 分类器的识别正确率达到 91.67%(女性)和 88.89%(男性),均优于 SVM 分类器的识别正确率 87.60%(女性)和 86.78%(男性),主要原因在于:

1) 对于多分类问题,SVM 参数个数较多且设置复杂,且参数一般选用经验值,没有普适性,从而对腭裂高鼻音等级分类识别的准确率造成影响;同时,在识别过程中,SVM 分类器使用少数有价值的特征数据建立超平面,而不依赖所有特征数据,导致对分类面附近的特征数据的识别缺乏准确性。在腭裂高鼻音等级分类识别中,提取的特征参数数据量较大,在分类面附近的特征数据较多,造成分类超平面准确性低,从而降低了识别正确率。

2) KNN 分类器方法简单,它是依靠邻近有限的样本来确定样本所属类别,而不是通过判别类域的方法来确定,因此对于类域的重叠或交叉较多的待分类样本集而言,KNN 分类器更适合。在腭裂高鼻音等级识别中,高鼻音等级被分为 4 类,而在提取出的每一种特征参数中,不同等级的腭裂高鼻音之间存在交叉重叠现象,KNN 分类器能够降低特征参数域的交叉或重叠的影响,具有较好的分类能力,从而得到较高的腭裂高鼻音等级识别正确率。

结束语 在腭裂高鼻音的研究中,一直存在识别正确率不理想、数据量小、计算量大、模型复杂等问题,解决这些问题可以提高腭裂高鼻音等级分类识别的效果,并且能够有效地辅助医生对腭裂患者的诊断和治疗。

本文提出了基于小波分解和 KNN 分类器对腭裂语音高鼻音等级自动识别的方法,并使用该方法对 80 名测试者的语音数据进行特征提取和识别处理,得到的数据量大且具有统

计意义。本文使用 KNN 和 SVM 两种分类器对腭裂高鼻音等级进行识别分类,从而形成对比。所提方法模型简单易懂,计算量小,识别率可以达到 91.67%。结果表明,使用小波分解系数倒谱特征的识别正确率优于其他经典特征,KNN 分类器的识别率优于 SVM 分类器,这充分说明了本文提出的小波分解系数倒谱特征和 KNN 分类器对腭裂高鼻音等级进行识别的方法的有效性和可行性。今后,随着研究的不断深入,可针对腭裂患者的其他病症特征,如共振异常、鼻漏气等,建立腭裂高鼻音等级识别系统,实现高鼻音等级识别方法的多样性。

致谢 感谢华西口腔医学院专家尹恒语音师提供了腭裂语音数据,并对腭裂语音数据进行 4 种等级(正常、轻度、中度、重度)的人工评估所做出的贡献。

参 考 文 献

- [1] CHEN R J. The state and consider about speech therapy of cleft palate in China[J]. *International Journal of Stomatology*, 2012, 39(1):1-5.
- [2] ARIAS-LONDOÑO J D, GODINO-LLORENTE J I, SAENZ-LECHON N, et al. Automatic Detection of Pathological Voices Using Complexity Measures, Noise Parameters, and Mel-Cepstral Coefficients[J]. *IEEE Transactions on Bio-medical Engineering*, 2011, 58(2):370-379.
- [3] SMYTH A. Clinical grading system for submucous cleft palate [J]. *British Journal of Oral & Maxillofacial Surgery*, 2014, 52(3):275-276.
- [4] VILLAFUERTE GONZALEZ R, et al. Acoustic analysis of voice in children with cleft palate and velopharyngeal insufficiency[J]. *International Journal of Pediatric Otorhinolaryngology*, 2015, 79(7):1073-1076.
- [5] MAIER A, HONIG F, HACKER C, et al. Automatic evaluation of characteristic speech disorders in children with cleft lip and palate[C]//Conference of the International Speech Communication Association. Brisbane, Australia, 2008:270-278.
- [6] ARROYAVE J R O, BONILLA J F V, et al. Automatic detection of hypernasality in children[C]//International Conference on Interplay Between Natural and Artificial Computation: New Challenges on Bioinspired Applications. Spain, 2011:167-174.
- [7] HE L, ZHANG J, LIU Q, et al. Automatic Evaluation of Hypernasality and Consonant Misarticulation in Cleft Palate Speech [J]. *IEEE Signal Processing Letters*, 2014, 21(10):1298-1301.
- [8] KUMARI V S R, DEVARAKONDA D K. A Wavelet Based Denoising of Speech Signal[J]. *International Journal of Engineering Trends & Technology*, 2013, 5(2):107-115.
- [9] CHEN Z, WANG S, YIN F. A Time Delay Estimation Method Based on Wavelet Transform and Speech Envelope for Distributed Microphone Arrays[J]. *Advances in Electrical & Computer Engineering*, 2013, 13(3):39-44.
- [10] MALLAT S G. A Theory for Multiresolution Signal Decomposition: The Wavelet Representation [J]. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 1989, 11(7):674-693.
- [11] 成礼智, 王红霞. 小波的理论及应用[M]. 北京: 科学出版社, 2004.
- [12] 刘明才. 小波分析及其应用(第 2 版)[M]. 北京: 清华大学出版社, 2013.
- [13] ZHAO L. *Speech Signal Processing*[M]. Beijing: China Machine Press, 2012.
- [14] DAVE N. Feature Extraction Methods LPC, PLP and MFCC In Speech Recognition[J]. *Ijaret Org*, 2013, 1(6):1-5.
- [15] ALI Z, ABBAS A W, THASLEEMA A W, et al. Database development and automatic speech recognition of isolated Pashto spoken digits using MFCC and K-NN[J]. *International Journal of Speech Technology*, 2015, 18(2):271-275.
- [16] AARON M, GANESH B, RATNADEEP R. Automatic Speech Recognition and Verification using LPC, MFCC and SVM[J]. *International Journal of Computer Applications*, 2015, 127(8):47-52.
- [17] REN S, CAO X, WEI Y, et al. Face Alignment at 3000 FPS via Regressing Local Binary Features [C] // IEEE Conference on Computer Vision and Pattern Recognition. New York: IEEE Press, 2014.
- [18] ZHANG J, SHAN S, KAN M, et al. Coarse-to-Fine Auto-Encoder Networks(CFAN) for Real-Time Face Alignment[C]//European Conference on Computer Vision (ECCV). Zurich: Springer-Verlag, 2014.
- [19] KAZEMI V, SULLIVAN J. One millisecond face alignment with an ensemble of regression trees[C]//European Conference on Computer Vision(ECCV). Zurich: Springer-Verlag, 2014.
- [20] ZHU S Z, LI C, CHEN C L, et al. Face alignment by coarse-to-fine shape searching[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. New York: IEEE Press, 2015.
- [21] LE V, BRANDT J, LIN Z, et al. Interactive facial feature localization[C]//European Conference on Computer Vision(ECCV). Italy: Springer-Verlag, 2012.
- [22] RAMANAN D, ZHU X. Face detection, pose estimation, and landmark localization in the wild [C] // IEEE Conference on Computer Vision and Pattern Recognition. New York: IEEE Press, 2012.
- [23] MESSER K, MATAS J, KITTLER J, et al. Xm2vtsdb: the extended m2vts database[C]//Second International Conference on Audio and Video-based Biometric Person Authentication. Zurich: Springer-Verlag, 1999.
- [24] CHEN T. MXNet: A Flexible and Efficient Machine Learning Library for Heterogeneous Distributed Systems[J]. *Statistics*, 2015, 42(12):87-129.

(上接第 277 页)