

## 基于多延迟四阶累积量倍频程谱线的腭裂语音咽擦音自动检测算法

何 飞<sup>1</sup> 孟雨璇<sup>1</sup> 田维维<sup>1</sup> 王熙月<sup>1</sup> 何 凌<sup>1</sup> 尹 恒<sup>2</sup>

1 四川大学电气信息学院 成都 610065

2 口腔疾病研究国家重点实验室 成都 610041

(2804373578@qq.com)



**摘 要** 为了实现对腭裂语音咽擦音及正常音节的自动分类检测,通过对腭裂咽擦患者发音特点的研究,提出了基于多延迟四阶累积量倍频程谱线(Fourth-order Cumulant One-third Octave Spectra Line,FTSL)的腭裂语音咽擦音自动检测算法。目前,咽擦音的研究多基于咽擦音的辅音时长及其在频域的能量分布等特征,实现了咽擦音及正常擦音自动检测的其他研究较少。文中实验基于腭裂语音咽擦音的发音特性,通过研究语音信号的多延迟四阶累积量,利用1/3倍频程算法提取特征谱线,实现了腭裂语音咽擦音与正常擦音的自动分类检测。实验提取了200个正常擦音辅音和194个腭裂语音咽擦音辅音的FTSL特征谱线,使用SVM(Support Vector Machine)分类器进行分类,并设计了FTSL谱线与其他传统语音特征的对比实验,进行了充分的分析讨论。实验结果表明,FTSL谱线对咽擦音的自动分类检测正确率高达92.7%,具有较优的性能,能为临床腭咽功能评估提供有效、客观、无创的辅助依据。

**关键词** 咽擦音;FTSL谱线;四阶累积量;1/3倍频程谱线

中图法分类号 TP391.9

## Automatic Detection Algorithm of Pharyngeal Fricative in Cleft Palate Speech Based on Multi-delay Fourth-order Cumulant Octave Spectral Line

HE Fei<sup>1</sup>, MENG Yu-xuan<sup>1</sup>, TIAN Wei-wei<sup>1</sup>, WANG Xi-yue<sup>1</sup>, HE Ling<sup>1</sup> and YIN Heng<sup>2</sup>

1 School of Electrical Engineering and Information, Sichuan University, Chengdu 610065, China

2 State Key Laboratory of Oral Diseases, Chengdu 610041, China

**Abstract** In order to realize the automatic classification and detection of palate pharyngeal fricative and normal speech, an automatic pharyngeal fricative detection algorithm based on multi-delay fourth-order cumulant one-third octave spectral line (FTSL) was proposed by studying the pronunciation characteristics of cleft palate patients with pharyngeal fricative. Currently, most researches involved with the detection of pharyngeal fricatives are based on the length of consonants and the energy distribution of speech in frequency-domain. There exist few researches which have achieved automatic classification of pharyngeal fricatives and normal speech. This experiment is based on the pronunciation characteristics of pharyngeal fricative. Each frame's multi-delay fourth-order cumulant is computed, and then one-third octave is used to extract the FTSL. Automatic classification of pharyngeal fricative and normal speech is realized by FTSL. In this experiment, the FTSL of 200 normal consonants and 194 consonants of pharyngeal fricative are extracted, and the SVM classifier is used to classify. Besides, comparative experiments were conducted on FTSL feature and traditional acoustic features, and the results were fully analyzed and discussed in this paper. The experimental results show that the proposed FTSL has an accurate rate of 92.7% for the automatic classification of pharyngeal speeches, and it has excellent performance and can provide an effective, objective and non-invasive auxiliary basis for clinical pharyngeal state assessment.

**Keywords** Pharyngeal fricative, FTSL spectral line, Fourth-order cumulant, One-third octave spectral line

### 1 引言

腭裂语音咽擦音作为常见的代偿性构音错误之一,发生率仅次于声门塞音<sup>[1]</sup>,同声门塞音一样会影响音节的声学特性表达<sup>[2]</sup>。咽擦音的错误发音形式表现为构音位置的后移<sup>[3]</sup>。腭裂咽擦患者发辅音时舌尾后缩至后咽壁,导致咽部

变窄,声道形状改变,来自肺部的气流通过该狭窄通道,经摩擦发声,而正常人发擦音时气流从缩小的口腔通路中挤出。咽擦音常见于唇腭裂患者以及咽咽功能障碍患者的发音中,以替换有声或是无声的摩擦<sup>[4]</sup>。咽擦音会降低腭裂患者的发音清晰度<sup>[3]</sup>,严重影响患者的正常生活。

临床上对腭裂语音咽擦音的研究主要集中在两个方面:

收稿日期:2018-07-21 返修日期:2018-11-18 本文已加入开放科学计划(OSID),请扫描上方二维码获取补充信息。

基金项目:国家自然科学基金青年科学基金(61503264)

This work was supported by the Young Scientists Fund of the National Natural Science Foundation of China (61503264).

通信作者:尹恒(yinheng@scu.edu.cn)

1)辅音声学特性;2)患者发音时发音器官的运动。关于辅音声学特性的研究,邓四海等<sup>[5]</sup>研究了腭裂术后患者的发音特点,得到腭裂患者发擦音时存在的辅音异常,该异常表现为辅音弱化。蒋莉萍等<sup>[6]</sup>通过研究声门塞音及咽擦音发现,腭裂患者在咽成形术后,仍以腭咽闭合不全型的代偿性语音错误为主,主要表现为辅音的脱落与弱化。在发音器官的运动变化上,马莲<sup>[7]</sup>通过观察腭裂患者发咽擦音时的舌运动,认为咽擦音频域出现的噪声能量可能与咽擦音的严重程度及舌尖运动有关。临床实践中专业语音师根据腭裂患者的发音缺陷来进行咽擦音有无的判听,语音师判听时依赖于主观经验,结果具有较大的主观性。在咽擦的治疗过程中,常通过入侵式设备、CT等影像手段来进行进一步诊断,但这种方式会给病人带来一定的创伤。

近年来,国内外学者致力于研究无创且更为客观的数字化处理方法,将其用于咽擦音的评估与诊断,以减轻病人治疗时的痛苦。目前,咽擦音的研究主要分为两类:1)基于数字化检测仪器的咽擦音特征分析。Garcia等<sup>[8]</sup>使用鼻音计6400(Kay PENTAX)来检测患者发音时的鼻音分数,发现有咽擦音和高鼻音的患者在发辅音/f/,/s/时,具有较高的鼻音分数。王国民等<sup>[9]</sup>利用动态音频频谱分析仪(TSL)观察患者咽喉摩擦音的第一、第二、第三共振峰及辅音起始时间,发现正常音与咽擦音的共振峰出现位置不同且起始时间存在差异。张楚涵等<sup>[10]</sup>使用TSL-4150B计算机语音声学系统对咽擦音等代偿性构音错误进行分析后发现,术后患者治疗的关键在于发音时口腔内蓄足足够的压力。马莲<sup>[7]</sup>结合主观判听及USSA语音频谱分析系统观察咽擦音的声学特点,发现咽擦音在语图上50~1100 Hz区域内出现噪声能量,认为该能量分布可能与咽擦音的严重程度及舌尖运动有关。朱云山等<sup>[11]</sup>利用USSA语音频谱分析系统研究患者发辅音时咽喉部的除阻过程。这些研究利用数字化的医用检测仪器测量鼻音分数、观察语音频谱等参数变化,医生根据这些参数变化给出诊断结果,但各个参数的变化仍由医生主观评估,存在主观性强和不准确的问题。2)基于数字信号处理技术的咽擦音时频特征分析。Xiao等<sup>[1]</sup>研究了咽擦音与正常擦音,并研究了咽擦音辅音时长以及其频域能量分布,采用方差分析(ANOVA)的统计分析方法来计算咽擦音与正常擦音在谱能量上的差异,得出两者具有显著差异的结论。国内外关于咽擦音在数字信号处理方面的研究较少,还需要不断地进行深入探讨。

本文结合语音信号的高阶统计特性及其谱分析,提出了能够区分咽擦音及正常擦音的FTSL谱线特征,并利用SVM分类器,实现了咽擦音与正常音的自动分类检测。自动检测腭裂患者是否存在咽擦音能为语言病理专家提供更为科学的评估<sup>[12]</sup>,同时为医生提供临床辅助诊断。

## 2 腭裂语音咽擦音自动检测系统

本文通过研究咽擦音的特性,提出了基于多延迟四阶累积量倍频程谱线的腭裂语音咽擦音自动分类检测系统,该系统的实现流程如图1所示。

腭裂语音咽擦音自动检测算法由以下3个部分组成:1)语音信号预处理。对输入辅音信号进行分帧、预加重及幅值归一化。2)特征提取。计算每一帧辅音信号的多延迟四阶

累积量,并计算该多延迟四阶累积量的傅里叶空间谱,结合1/3倍频程谱线提取算法得到特征谱线FTSL。3)腭裂语音咽擦音自动分类检测。以临床语音师给出的实验语音类别标签为金标准,根据本实验计算得到的FTSL特征谱线,利用SVM分类器实现腭裂语音咽擦音的自动分类检测。

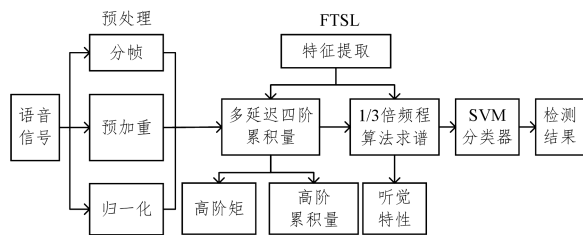


图1 腭裂语音咽擦音自动检测系统流程图

Fig. 1 Flow chart of automatic detection system for pharyngeal fricative of cleft palate speech

### 2.1 腭裂语音咽擦音的发音机理及特点

#### 2.1.1 发音机理

儿童普遍在5岁前就习得了93%的辅音<sup>[13]</sup>,辅音对人的发音清晰度有着及其关键的作用。辅音对口腔发音器官的结构和运动非常敏感,要求很高。腭裂咽擦患者发辅音时总是试图用一些代偿性动作在气流通过腭咽口进入鼻腔前阻塞气流。与正常辅音的产生相比,代偿性辅音的形成位置更靠近口腔后部和上声道的下部。大多数咽擦患者发辅音时口内气流压力低,导致发音不清晰,甚至出现音素遗漏<sup>[14]</sup>。本文正是通过研究这些差异在信号高阶的表现,来实现咽擦音及正常音的分类。

图2为腭裂患者发咽擦音时对应的生理现象的侧视图<sup>[15]</sup>。如图2中箭头所示,患者舌后缩,导致腭咽口变窄,此时来自肺部的气流经过该狭窄通道经摩擦发声,可以直观地看出此时患者的构音位置整体后移。

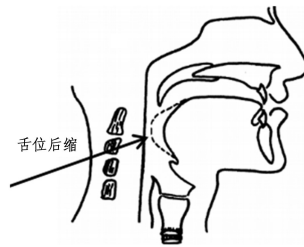


图2 咽擦代偿音发音生理现象的侧视示意图

Fig. 2 Lateral view schematic illustration of pharyngeal fricative compensatory articulation

#### 2.1.2 咽擦音与正常音的区别

咽擦患者发音时发音器官运动的改变使得语音信号的声学特性发生改变,该差异主要表现在对发音器官运动敏感的辅音上。大部分咽擦患者在发辅音时由于腭咽口闭合不全或腭裂导致的前牙缺失等问题,无法在口腔内的正确位置形成摩擦,从而产生足够的口腔压力,因此患者总是试图将舌位后置,与后咽壁产生摩擦,从而产生擦音,导致所发出的咽擦音与正常擦音相比出现辅音弱化的现象,甚至遗漏整个辅音音节。图3(a)、图3(b)分别展示了咽擦音/si/与正常擦音/si/的时域波形及对应的语谱图。从图中可以观察到,咽擦音辅音与正常擦音辅音相比存在弱化的现象;再观察对应的语谱

图可知,咽擦音辅音部分的能量在语谱图上表现为能量的弱化及强频集中区的改变,即能量集中分布的频带和强度伴有对应的变化。这些差异主要是由咽擦患者发音与正常人发音时的发音器官运动及共振位置的不同所造成的,本文通过更进一步的实验来研究这些变化在语音中的呈现。

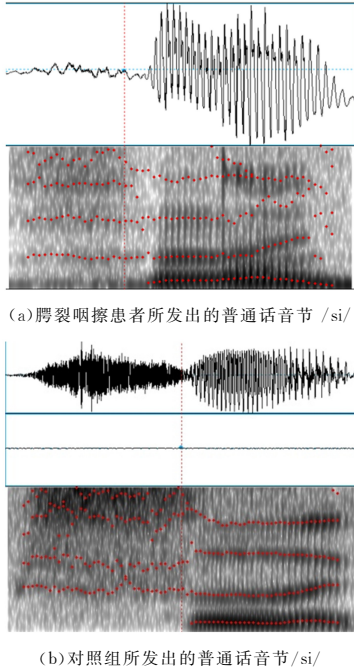


图3 咽擦患者及正常人发出音节/si/的语音时域波形及语谱图  
Fig. 3 Time domain waveform and spectrogram of syllable /si/ spoken by cleft palate speaker with pharyngeal fricative and normal speaker

## 2.2 腭裂语音咽擦音高阶特性-多延迟四阶累积量

在传统的语音信号处理中,研究者常使用二阶统计量作为数学分析工具,例如相关函数与功率谱,但相关函数与功率谱具有等价性,不能辨识最小非相位系统,且对加性白噪声敏感,一般只能处理加性白噪声的观测数据。并且基于二阶统计量的超分辨率算法通常把信号当成高斯分布,仅用到信号的二阶信息,且假设信号中的噪声为白噪声。实际上,非高斯信号是更普遍的信号,对于非高斯信号,二阶统计量仅表达了其中的一部分信息<sup>[16]</sup>。为了弥补这些缺点,必须使用三阶或更高阶的统计量,称其为高阶统计量。

基于高阶统计量的信号分析称为信号的高阶统计分析<sup>[17]</sup>,与二阶统计量信号处理算法不同,高阶累积量可以得到更多的信号源信息,并且高阶累积量对高斯过程不敏感,因此将高阶累积量与高分辨率算法相结合,可以在加性噪声为有色高斯分布时,使算法能很好地发挥效用<sup>[18]</sup>。高阶累积量包含丰富的高阶信息,能够提供更多的累积量项,从而具有阵列扩展的作用,与二阶矩相比,可以构造更大的阵列矩阵,处理更多的信号。由于充分对称分布的序列的奇数阶累积量为零,因此在信号处理中最常用的是四阶累积量。樊养余<sup>[19]</sup>通过研究谐波信号的高阶累积量发现,四阶累积量比二阶、三阶累积量所含的信息更丰富,可以分析出谐波信号各分量间频率及相位存在的关系。因此,本文关于咽擦音的研究也都是基于四阶累积量展开的。

四阶累积量作为高阶统计量的一种,能很好地表达信号在高阶的统计特性。四阶累积量主要有以下几个优点:1)包含了信号丰富的相位信息,与系统盲相位的一、二阶累积量相比,可以辨识系统的相位信息,还可以检验和表征信号中的非线性并辨识非线性系统,体现信号中的循环平稳性以及分析和处理循环平稳信号<sup>[20]</sup>;2)具有很强的噪声抵消能力,能从混合信号中提取出非高斯信号的累积量信息<sup>[18]</sup>,可以度量非高斯信号的高斯偏离度<sup>[21]</sup>。这些优点使得四阶累积量被广泛应用于各个需要进行数字信号处理的领域<sup>[22-24]</sup>。

在语音信号处理中,四阶累积量常用于信号的检测与分离。梁红等<sup>[25]</sup>利用四阶累积量实现高斯色噪声中对称分布的非高斯信号的检测。文献<sup>[26]</sup>基于四阶互累积量的特性,提出了一套适应于非平稳信号盲分离的算法。Elias等<sup>[27]</sup>通过四阶累积量建立新的参数空间,提出了一种利用时域最优滤波器和四阶累积量的语音增强新方法。Qian等<sup>[28]</sup>利用高阶累积量对高斯噪声的盲特性,提出了一种基于经验模态分解(Empirical Mode Decomposition, EMD)和高阶累积量统计特性的噪声环境下语音流检测算法。本文利用四阶累积量来研究咽擦音和正常音的高阶统计差异。

腭裂咽擦患者发音时由于口腔内压力不足,难以将发辅音时的成阻、持阻、除阻3个阶段全部完成<sup>[29]</sup>。这些因素导致语音信号的高阶矩及其累积量等反映信号变化趋势及总体分布的统计特性发生了改变。语音信号是典型的非高斯信号,其低阶统计量只能反映信号的部分信息,而四阶累积量能够更为全面地包含语音信号源的信息,衡量语音信号的高斯偏离度,从而突出咽擦音信号以及正常语音信号的分布形态的陡缓程度。本文在计算FTSL谱线时首先计算每帧语音信号的多延迟四阶累积量,它是每一延迟对应的四阶累积量的序列组合,计算时需要计算出每一延迟对应的四阶累积量。

设每一帧信号为 $x(n)$ ,因为语音信号具有短时平稳性,所以可将 $x(n)$ 做零均值 $k$ 阶平稳随机过程处理,则该过程的 $k$ 阶累积量定义为随机变量 $\{x(n), x(n+\tau_1), \dots, x(n+\tau_{k-1})\}$ 的 $k$ 阶联合累积量。

$$C_{k,x}(\tau_1, \tau_2, \dots, \tau_{k-1}) = cum(x(n), x(n+\tau_1), \dots, x(n+\tau_{k-1})) \quad (1)$$

其中, $C_{k,x}(\tau_1, \tau_2, \dots, \tau_{k-1})$ 为计算得到的 $k$ 阶累积量; $\tau_1, \tau_2, \dots, \tau_{k-1}$ 分别为各阶对应的延迟, $cum(\cdot)$ 为 $k$ 阶联合累积量。

该零均值 $k$ 阶平稳随机过程的 $k$ 阶矩如下:

$$m_{k,x}(\tau_1, \tau_2, \dots, \tau_{k-1}) = mom(x(n), x(n+\tau_1), \dots, x(n+\tau_{k-1})) \quad (2)$$

其中, $m_{k,x}(\tau_1, \tau_2, \dots, \tau_{k-1})$ 为 $k$ 阶矩, $mom(\cdot)$ 为 $k$ 阶联合矩。

由于 $x(n)$ 是 $k$ 阶平稳的,因此 $x(n)$ 的 $k$ 阶累积量和 $k$ 阶矩仅是时延 $\tau_1, \tau_2, \dots, \tau_{k-1}$ 的函数,而与时刻 $n$ 无关,其二阶、三阶、四阶累积量分别为:

$$C_{2,x}(\tau) = E[x(n)x(n+\tau)] \quad (3)$$

$$C_{3,x}(\tau_1, \tau_2) = E[x(n)x(n+\tau_1)x(n+\tau_2)] \quad (4)$$

$$C_{4,x}(\tau_1, \tau_2, \tau_3) = E[x(n)x(n+\tau_1)x(n+\tau_2)x(n+\tau_3)] - C_{2,x}(\tau_1)C_{2,x}(\tau_2 - \tau_1) \quad (5)$$

$k$ 阶累积量可由 $k$ 阶及 $k$ 阶以下的各阶矩表示,四阶累

积量的完整计算式如式(6)所示:

$$C_{4,x}(\tau_1, \tau_2, \tau_3) = m_{4,x}(\tau_1, \tau_2, \tau_3) - m_{2,x}(\tau_1)m_{2,x}(\tau_3 - \tau_2) - m_{2,x}(\tau_2)m_{2,x}(\tau_3 - \tau_1) - m_{2,x}(\tau_3)m_{2,x}(\tau_2 - \tau_1) \quad (6)$$

其中,  $m_{2,x}, m_{4,x}$  为信号对应的二阶矩及四阶矩。在一帧语音信号为零均值平稳随机过程的假设下, 四阶累积量的计算保留了信号的高阶特性, 同时降低了计算的复杂度。文献[30-31]采用递推的方式来逐级减少高阶累积量的运算量, 其中文献[30]还进一步研究了五阶累积量的优化与仿真估计。

利用上述公式, 通过设置合适的延迟  $\tau_1, \tau_2, \tau_3$ , 来计算每一帧信号的四阶累积量, 本研究中  $\tau_1 = 1, \tau_2 = 3, \tau_3 = m$ ,  $m$  的取值将在后文进行讨论。研究上述计算过程可知, 四阶累积量的计算与其前三阶矩的累积量密切相关, 综合了各阶统计特性, 可以很好地反映咽擦音及正常擦音的区别。

通过式(6)计算的一帧语音信号在最大延迟  $m = 100$  时组成的多延迟四阶累积量信号如图4所示。可以看出, 正常音及咽擦音在  $m = 100$  附近对应的累积量具有显著的波动差异, 咽擦音对应的曲线较正常音的曲线更加平滑。

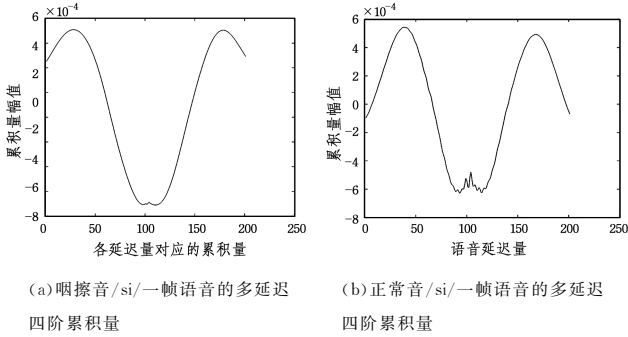


图4 语音信号的多延迟四阶累积量

Fig. 4 Multi-delay fourth-order cumulant of speech signal

图4中, 四阶累积量以  $m = 100$  为中轴, 将每帧信号以1个采样点分别向正负方向进行时延, 时延最大为  $m$  个单位, 依次计算各延迟对应的累积量, 从而组合得到每帧信号的多延迟累积量。得到的正常擦音与咽擦音的四阶累积量波形虽然相似, 但是咽擦音与正常擦音累积量的幅值及波形峰值的分布位置均有差异, 特别是中轴附近的波动差异, 这些差异在频域会更加明显, 因此在得到每帧语音的多延迟四阶累积量后, 本文对其频谱进行了更深入的分析。

### 2.3 腭裂语音咽擦音的 FTSL 谱线计算

综合上述分析, 咽擦音的辅音与正常擦音的辅音在多延迟四阶累积量分布上的差异体现在峰值分布及幅值波动上, 该特性在频域上体现为能量在频带上的幅值与位置分布差异。根据信号的以上特点, 本文将信号的多延迟四阶累积量转换到频域进行分析。

人耳听音的频率范围为 20 Hz 到 20 kHz, 对整个频域范围内的所有频率成分进行分析的运算量大且数据冗余, 因此将整个听觉频率范围划分为多个频带, 每一个频带为一个频程。对于频程划分, 通常采用的划分规律是恒定带宽比, 即保持频带的上限和下限的频率比率为常数。设定每一频带的上限频率都是下限频率的两倍, 即频率之比为 2, 据此划分得到

的每一个频带称为 1 倍频程, 简称倍频程。在一个倍频程的上限和下限频率之间再插入 2 个频率点, 得到包括上、下限频率在内的 4 个频率点, 并使这 4 个频率之间的比值相同(即相邻两频率比值=倍), 这样就将一个倍频程划分成了 3 个频程, 据此划分得到的频程称为 1/3 倍频程。1/3 倍频程谱是由一系列的频率点以及这些频率点附近频带内信号的平均幅值组成的, 构成 1/3 倍频程谱的这些频率点称为中心频率  $f_c$ , 每一频带具有上限频率  $f_2$  和下限频率  $f_1$ , 带宽为  $(f_2 - f_1)$ 。中心频率  $f_c$  附近的频带处于  $f_2$  和  $f_1$  之间, 中心频率如下:

$$f_c = \sqrt{f_1 \times f_2} \quad (7)$$

一般  $n$  倍频程带滤波器中,  $n$  的定义由式(8)确定:

$$\frac{f_2}{f_1} = 2^n \quad (8)$$

当  $n=1$  时,  $f_2/f_1 = 2$ , 这样的频率比值所确定的频程称为 1/1 倍频程, 简称为倍频程, 与此类似, 当  $n=1/3$  时,  $f_2/f_1 = 2^{1/3}$ , 该频率比值所确定的频程称为 1/3 倍频程, 在此频带  $(f_2 - f_1)$  的频谱称为 1/3 倍频程谱。

表1列出了 20 kHz 以内, 1/3 倍频程滤波带的中心频率以及带宽。

表1 中心频率和带宽

Table 1 Center frequency and bandwidth

(单位: Hz)

频带号	中心频率	倍频程带宽	频带号	中心频率	倍频程带宽
1	1.25	1.12~1.41	23	200	178~224
2	1.6	1.41~1.78	24	250	224~282
3	2	1.78~2.24	25	315	282~335
4	2.5	2.24~2.82	26	400	335~447
5	3.15	2.82~3.55	27	500	447~562
6	4	3.55~4.47	28	630	562~708
7	5	4.47~5.62	29	800	708~891
8	6.3	5.62~7.08	30	1000	891~1120
9	8	7.08~8.91	31	1250	1120~1410
10	10	8.91~11.2	32	1600	1410~1780
11	12.5	11.2~14.1	33	2000	1780~2240
12	16	14.1~17.8	34	2500	2240~2820
13	20	17.8~22.4	35	3150	2820~3550
14	25	22.4~28.2	36	4000	3550~4470
15	31.5	28.2~35.5	37	5000	4470~5620
16	40	35.5~44.7	38	6300	5620~7080
17	50	44.7~56.2	39	8000	7080~8910
18	63	56.2~70.8	40	10k	8910~11.2k
19	80	70.8~89.1	41	12.5k	11.2k~14.1k
20	100	89.1~112	42	16k	14.1k~17.8k
21	125	112~141	43	20k	17.8k~22.4k
22	160	141~178			

从表1中可以得到, 以第一个频带中心频率为起点, 每隔两个频带, 中心频率的值变为原来的两倍, 因此 1/3 倍频程谱中每个带滤波器的中心频率与上、下限频率之间的关系为:

$$f_2/f_1 = 2^{1/3}, f_2/f_c = 2^{1/6}, f_c/f_1 = 2^{1/6} \quad (9)$$

本文基于 1/3 倍频程谱提取 FTSL 谱线, 将每一帧辅音信号的多延迟四阶累积量进行快速傅里叶变换(FFT), 由此得到频率由低到高对应的频谱, 利用频谱计算多延迟四阶累积量对应的能量谱, 然后根据表1中心频率及其带宽的对应关系, 计算每个中心频率对应频带的能量谱均值, 该值即为该频带的 FTSL 值, 将每个频带的 FTSL 值按照表1中频带的

顺序进行组合,得到该帧语音信号的 FTSL 谱线。

综上,本文基于高阶累积量的统计特性及 1/3 倍频程的谱特性,首先计算了每帧语音信号的多延迟四阶累积量,多延迟四阶累积量反映了语音信号由于发声机理的改变而导致的高阶统计特性差异;然后将该差异转换到频域做进一步处理,结合 1/3 倍频程谱算法,计算得到 FTSL 谱线。

## 2.4 SVM 分类器

研究咽喉音的自动检测与分类主要考虑以下两个方面:

1) 咽喉音在临床上的样本量较少,无法做大样本的训练与学习;2) 咽喉音及正常音的自动检测与分类是将咽喉音及正常音检测出来并赋予新的检测标签,为典型的二分类问题。针对这两个问题,本文选用 SVM 分类器进行腭裂语音咽喉音以及正常音的检测分类,理由如下:

1) 咽喉音临床样本较少,传统的模式识别或人工神经网络方法都要求有较多的训练样本,因此不适用于咽喉音的检测分类。经典的模式识别方法、各种统计方法以及近年来出现的各种神经网络算法都被应用于估计这种数据中固有的相关性,试图建立一个能预测未来数据的模型。这些方法有其固有的算法缺陷:由于不知道数据的分布密度以及有限的训练数据,常常得到病态的训练结果,训练结果最好的模型不一定是预报能力最好的模型<sup>[32]</sup>。SVM 的重要特征之一是解的稀疏性,即多数最优值为 0,只有少量的最优值不为 0,也就是说只需少量样本(支持向量)就可构成最优分类器,从而解决样本较少的难题<sup>[33]</sup>。

2) SVM 算法没有以传统的经验风险最小化原则为基础,而是建立在结构风险最小化(Structural Risk Minimization, SRM)原理的基础之上,发展成为了一种新型的结构化学习方法。它能很好地解决有限数量样本的高维模型的构造问题<sup>[34]</sup>,还能有效地预测短期负荷,更好地验证算法精度<sup>[35]</sup>。

3) SVM 算法中的最优分类面(Optimal Hyperplane)是基于线性可分的情况提出的。最优分类面要求分类面不但能将两类样本点尽可能无错误地分开,而且能使两类的分类空隙最大,在二分类信号处理中具有突出的优势。

总而言之,SVM 是典型的适用于两类识别的分类器,在解决小样本、非线性及高维模式识别中表现出了许多特有的优势。SVM 也被广泛应用于数字信号的处理中<sup>[36-38]</sup>,其基本模型定义为使特征参数之间的间隔最大的线性分类器,其对特征参数的分类如图 5 所示。

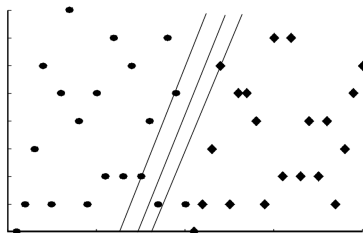


图 5 超平面的选取

Fig. 5 Selection of hyperplane

本文将信号提取的 FTSL 谱线作为特征值,通过 SVM 分类器对输入的参数进行学习,该线性分类模型由图

5 中两类特征间的线性分类函数决定,这些线性函数被统称为超平面,该超平面的定义为:

$$g(x) = \omega^T x + b \quad (10)$$

其中, $\omega$  为 SVM 的支持向量, $b$  代表该超平面的偏置。当输入任一测试样本  $x$  时,超平面与样本的关系可定义为:

$$result = \begin{cases} C_1, & g(x) > 0 \\ C_2, & g(x) < 0 \end{cases} \quad (11)$$

分类面为  $\omega^T x + b = 0$ , 即当  $g(x) = 0$  时,分类器拒绝对输入的测试样本做判断。由图 5 还可以得出对样本进行测试分类的超平面不只一个,当旋转该超平面时,仍然可以得到样本的分类结果。分类器在进行模型训练时,会不断根据输入的特征参数进行超平面的调整,因此 SVM 分类器在两类的识别分类中对样本具有很好的鲁棒性。

## 3 实验结果与分析

### 3.1 实验数据

本实验采用的语音数据来自于四川大学华西口腔医院唇腭裂外科,数据按照“四川大学华西口腔医院语音矫治室普通话语音测量表”进行录制。该表包括普通话音节中的所有辅音,根据咽喉音的辅音发声特点,选取含辅音/h/, /sh/, /c/, /x/的词语共 394 个,例如“鲨鱼”“西瓜”等词。语音数据前期由专业的语音师进行判听,将其人工判定咽喉的有无作为“金标准”。语音数据具体包含的词语类别及对应的个数如表 2 所列。

表 2 样本及数量

Table 2 Samples and quantities

辅音	患有咽喉	正常	总计
/sh/	48	50	98
/x/	52	50	102
/c/	60	50	110
/h/	34	50	84

### 3.2 腭裂语音咽喉音自动分类检测结果的影响因子

本文提取咽喉音辅音和正常擦音辅音的 FTSL 谱线作为特征值,利用 SVM 分类器进行咽喉音辅音与正常擦音辅音的自动分类检测。实验采用 10 折交叉验证<sup>[39]</sup>计算腭裂语音咽喉音的自动分类的正确率,得到准确率最高为 92.7%。实验测试了多延迟四阶累积量的最大延迟量  $m$  及 FTSL 谱线的组合对腭裂语音咽喉音自动分类检测结果的影响。

#### 3.2.1 多延迟四阶累积量最大延迟范围对分类结果的影响

本文在计算四阶累积量时, $\tau_1, \tau_2$  为固定值( $\tau_1 = 1, \tau_2 = 3$ ),因此每帧信号的多延迟四阶累积量由延迟  $\tau_3 = m$  决定。通过设置最大延迟  $m$ ,就可以计算得到语音信号与各延迟相对应的累积量矩阵,即本文研究的由辅音信号转换而来的新信号序列,该序列长度由设置的最大延迟  $m$  决定,即  $(2 * m + 1)$ 。因此  $m$  越大,每帧语音信号对应计算的多延迟四阶累积量信号具有越高的维度。

在讨论最大延迟  $m$  对识别准确率的影响时,参照表 1,取 FTSL 谱线的第 20 到 43 根谱线。实验分析了  $m$  分别为 50, 70, 100, 150, 200 时的咽喉音自动识别检测结果。在改变延迟量时,选取的频带谱线保持不变。

表3 延迟量  $m$  对应的分类正确率Table 3 Classification accuracy rate corresponding to delay  $m$ 

延迟量 $m$	识别正确率/%
50	84.3
70	90.1
100	92.7
150	85.2
200	84.3

通过对表3的分析可得,当延迟量  $m=100$  时,样本的分类正确率最高为 92.7%。在计算语音的累积量时不同的延迟对应着不同的累积量输出,因此通过设置最大延迟量  $m$ ,计算  $m$  内的所有延迟量对应的累积量,能够反映语音信号的高阶统计特性的差异。当  $m$  较小时,延迟范围小,对应图4,只能取到中部的波动部分,语音信息表达不全面,因此做频谱后,只利用了一部分有效信息。当  $m$  较大时,波形向两边扩展,幅度变缓,这与数据过拟合原理类似,反而弱化了语音之间的差别,且在计算多延迟四阶累积量时,最大延迟取值还需满足  $(2 * m + 1)$  小于语音分帧长度  $L$  的条件,否则延迟会超出语音长度从而导致错误的累积量结果。因此,只有选择一个合适的  $m$  值,才能很好地体现咽擦音与正常音之间的差别,在本次实验中  $m=100$  时的分类准确率最高。

图6为相同谱线、不同延迟下的分类器分类结果的ROC曲线(Receiver Operating Characteristic Curve),它们分别对应的AUC(Area Under the Curve)值如表4所列。AUC是对不同的ROC曲线进行比较的一个指标,即ROC的曲线下面积,它体现了分类器的平均性能。根据表4,  $m=100$  时,分类器具有更优的平均性能,同前面讨论的结果一致。

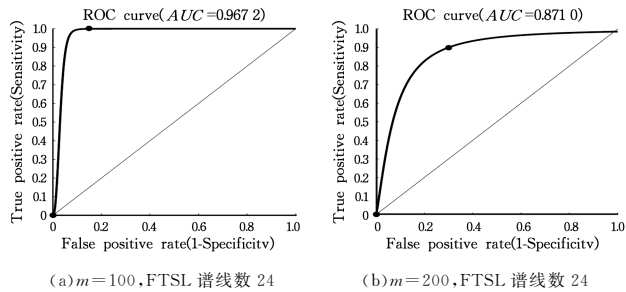


图6 SVM分类器的ROC曲线

Fig. 6 ROC curve of SVM classifier

表4 不同延迟  $m$  对应的ROC曲线下的AUC值Table 4 AUC value corresponding to ROC curve in different  $m$ 

特征设置	AUC值
FTSL=后24条, $m=100$	0.9672
FTSL=后24条, $m=200$	0.8710

### 3.2.2 FTSL谱线组合对腭裂语音咽擦音自动检测识别的影响

表1列出了采样率为44.1 kHz的信号的分1/3倍频程频带及其中心频率。由表1可得,用1/3倍频程谱算法计算FTSL谱线总共可得43条谱线,可根据语音特性提取相应的谱线。本文在控制变量的情况下,分别选取了前1到25根谱线、20到43根谱线、1到43根全谱线等3种谱线组合作为特征输入分类器,再对每一组下的分类情况进行了分析。

根据上述累积量的最大延迟对样本分类正确率的影响,

在改变谱线数量时,  $m$  取值100。

图7展示了咽擦音及正常音/si/的辅音部分任意一帧信号提取的24条FTSL谱线。由图7可以发现,在这两帧信号中,咽擦音与正常擦音相比,其多延迟四阶累积量的能量在低频部分更集中,在高频部分能量变化趋于平缓。观察上文给出的语音信号对应的多延迟四阶累积量曲线可知,其细节波动较少,因此能量仍然大部分集中于低频。咽擦音对应的多延迟四阶累积量曲线变化平缓,不存在明显波动,观察图7(a)可知,咽擦音的能量谱中第一根谱线高度明显高于其他的谱线,即表现为低频能量的高度集中。表5列出了不同的谱线序列组合对腭裂语音咽擦音自动识别检测的准确率。

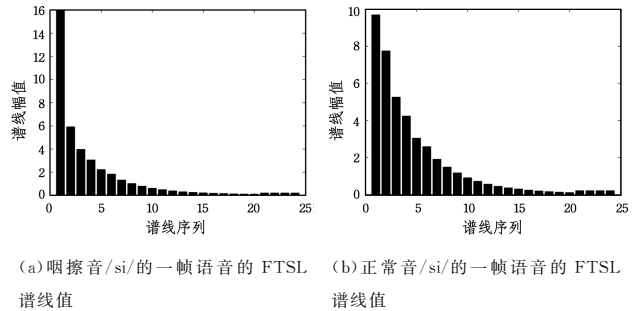


图7 FTSL特征谱线

Fig. 7 FTSL characteristic spectral lines

表5 谱线数目对应的识别正确率

Table 5 Classification accuracy corresponding to number of spectral lines

FTSL谱线数目	识别正确率/%
24(后24条)	92.7
25(前25条)	80.7
43(整合频谱谱线)	90.4

由表5可知,当  $m=100$  时,选择第20到43条谱线,腭裂语音咽擦音自动分类检测的准确率最高,为92.7%。通过分析表1的各个频带,得到其对应的谱线在低频部分较密集,第20条以后的频带都较宽,因此低频部分的谱线在频域分辨率较低的情况下存在重合的情况,即数据冗余;通过观察表1中的频带信息,可以发现这20条谱线只包含了0~112 Hz频率范围的信息,高频部分被忽略,未能体现信号之间的细节差异,因此识别准确率低。选择全谱线时,因为特征矩阵为一个辅音信号各帧FTSL谱线的拼接组合,具有较大的维度,分类器进行分类时存在过拟合(overfitting)问题,使得测试集对训练集的模型不适应,所以导致了准确率的下降。而从第20条谱线开始选取时,每一谱线对应较宽的频带,避免了同一谱线代表多个频带的现象,且包括了大部分低频谱线及所有高频谱线,更能够突出辅音信号多延迟四阶累积量能量分布差异的特点,使得咽擦音辅音与正常擦音辅音的自动识别检测分类效果更好。

图8给出了相同延迟量、不同谱线组合下的分类器分类结果的ROC曲线,它们分别对应的AUC值如表6所列。由表6可得,当  $m=100$ ,FTSL谱线选择全谱线和后24条谱线时,相比只表达了低频部分信息的前25条谱线的组合,分类器均具有较优的平均性能。

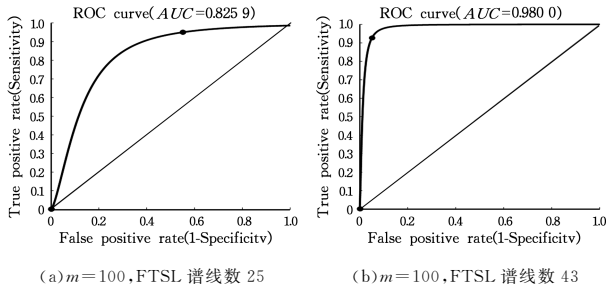


图 8 SVM 分类器的 ROC 曲线  
Fig. 8 ROC curve of SVM classifier

表 6 不同谱线组合 ROC 曲线下的 AUC 值

Table 6 AUC values under ROC curves of different groups of spectral lines

特征设置	AUC 值
FTSL=前 25 条, $m=100$	0.8259
FTSL=全谱线, $m=100$	0.9800
FTSL=后 24 条, $m=100$	0.9672

### 3.3 实验对比分析

#### 3.3.1 低阶累积量对应的 FTSL 测试结果

本文在四阶累积量的基础上计算 FTSL 谱线,将其作为特征值输入分类器进行自动识别检测。除了四阶累积量以外,二、三阶累积量也常用于信号处理领域。本节讨论了二、三阶累积量下计算的 FTSL 谱线对咽擦音的识别结果。图 9 给出了一幅咽擦语音在最大延迟  $m=100$  时对应的多延迟二阶累积量曲线及多延迟三阶累积量曲线。从图中可以看出,三阶累积量对应的曲线与二阶累积量对应的曲线相比有更多的波动。

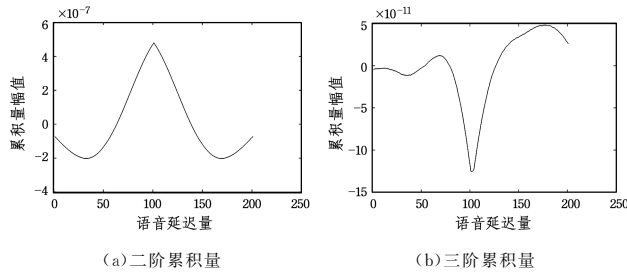


图 9 一幅咽擦语音的二阶、三阶累积量

Fig. 9 Two and three order cumulants of a frame of Pharyngeal fricative

表 7 列出了后 24 条 FTSL 谱线在不同延迟量下对应的咽擦音的自动识别正确率。

表 7 低阶累积量的识别正确率

Table 7 Recognition accuracy of low-order cumulants

低阶累积量 FTSL 谱线	正确率/%
$m=70$ , 二阶累积量	68.3
$m=70$ , 三阶累积量	70.4
$m=100$ , 二阶累积量	68.3
$m=100$ , 三阶累积量	72.1
$m=150$ , 二阶累积量	68.3
$m=150$ , 三阶累积量	39.8

从表 7 中可以看出,二、三阶累积量下的 FTSL 谱线对应的识别正确率总体偏低,且在二阶累积量基础上改变最大延

迟量,识别准确率没有发生变化。用三阶累积量计算谱线时改变了延迟量,正确率在低延迟量下相比二阶累积量的结果有所提升,当最大延迟  $m$  为 150 时,正确率锐减为 39.8%。结合前文对各阶累积量的探讨以及对四阶累积量下 FTSL 测试结果的讨论,二、三阶累积量所包含的信息较四阶累积量少,特别是其中的二阶累积量不包含信号本身的相位信息,而语音信号是典型的非高斯信号,其只能表达语音信号很少的一部分信息,因此在识别咽擦音及正常音时准确率偏低。当最大延迟设置过大时,表达信息冗余,低阶累积量包含的有效信息较少,当存在信息冗余时,有效信息便被覆盖,因此当最大延迟为 150 时,基于三阶累积量的 FTSL 谱线的识别正确率突降。

#### 3.3.2 FTSL 谱线与 TSL 谱线的对比分析

Xiao 等<sup>[1]</sup>提出了一种将语音信号分帧后直接利用傅里叶变换计算频谱,再利用倍频程算法提取每一帧语音信号的谱线作为咽擦音以及正常擦音分类特征的算法。该文仅以辅音/sh/作为研究对象,对咽擦音的辅音时长及频域能量分布进行特征计算。本文实验基于多延迟四阶累积量,研究了咽擦音和正常擦音在高阶统计量上的特征差异,且训练样本为多个辅音音节,与许多传统特征做了对比分析,实验设计更为全面。为了更好地呈现两种特征的对比结果,将文献[1]中研究的特征命名为 TSL(Traditional-Spectral-Line)。实验对比分析了两种特征取不同谱线数目时对腭裂语音咽擦音的自动检测结果,对比结果如表 8 所列。

表 8 TSL 与 FTSL 谱线对应的识别正确率

Table 8 Recognition accuracy of TSL and FTSL spectral lines

特征类型	参数设置	正确率/%
FTSL 谱线	24(后 24 条)	95.16
	25(前 25 条)	80.65
	43(全谱)	90.32
TSL 谱线	24(后 24 条)	73.83
	25(前 25 条)	83.19
	43(全谱)	82.54

在进行 FTSL 及 TSL 谱线的比较时可以发现,FTSL 谱线在提取后 24 条及整个频谱谱线时正确率显著高于 TSL 谱线,在取前 25 条谱线时略低于 TSL 谱线。因为 FTSL 谱线体现的是语音信号统计特性的差异,这些差异主要存在于语音信号细节的波动上;TSL 谱线代表的是咽擦音信号在频域的能量分布,语音信号的大部分能量集中于低频部分,也就是序号位于前端的谱线所代表的部分。因此,当取两种谱线最前端 25 条谱线作为特征时,代表能量的 TSL 谱线包含更多的有用信息,而 FTSL 谱线主要的差异信息存在于后面的高频谱线部分,所以此时 TSL 对咽擦音的识别正确率较高;当取后 24 条谱线或者全谱时,FTSL 谱线中关于咽擦音以及正常音的细节波动差异得以表达,对咽擦音以及正常音的识别效果更好。

#### 3.3.3 传统语音特征检测结果分析

在时域与频域中有许多可以反映语音信号的经典特征,如反映信号非线性程度的信息熵、香农熵和赫斯特指数,反映信号能量分布以及集中程度的短时能量、自相关函数主副峰比值、MFCC 倒谱距离等。本节实验对传统特征对咽擦音的

自动分类识别准确率进行了测试。各个语音传统特征的测试结果如表 9 所列。

表 9 传统特征对咽擦音的识别正确率

Table 9 Recognition accuracy of pharyngeal fricative by traditional speech features

传统语音特征	正确率/%
香农熵	80.2
信息熵	77.4
自相关(主副峰比值)	81.8
MFCC 倒谱距离	72.1
短时能量	84.4
赫斯特指数	76.0

观察表 8 中相应的语音特征及测试准确率,本文提出的 FTSL 特征谱线的分类识别正确率均优于表中的传统声学特征,即短时能量、自相关函数求取的共振峰主副峰比值、非线性特征中的熵以及赫斯特指数等,这些传统特征识别效果较差的主要原因如下。

1)短时能量:短时能量能反映语音信号能量的变化,对咽擦音的区分度较好,这是因为咽擦音存在辅音弱化的现象,在能量分布上与正常音的集中区域不同。但是因为短时能量是一个存在个体差异的参数,它与人的性别、年龄、音调等因素有关,这些不可控因素会影响咽擦音的分类结果,因此识别准确率依然偏低。

2)MFCC 倒谱距离:提取 MFCC 特征参数采用的是三角滤波器组,造成相邻频带之间频谱能量的泄露。在腭裂语音咽擦音中,腭咽闭合功能的缺陷也会造成频谱能量的泄露,在计算倒谱距离时弱化了参数本身,从而降低了识别正确率。并且咽擦音与正常音的差异主要是由发音器官运动以及气流变化引起的,MFCC 无法很好地表达构音变化的差异。

3)自相关(主副峰比值):利用自相关函数求取语音信号的主副峰比值的算法主要用于语音信号中的声韵母切分,因为辅音强频集中区和元音的共振峰峰值及频带具有较大差异。咽擦音与正常音的差异主要体现在辅音上,且研究的辅音并不全是浊辅音,强频集中区不一定存在。因此,使用该特征来区分咽擦音以及正常擦音时,效果不佳。

4)非线性特征参数:语音信号的产生是一个复杂的非线性过程,用声学 and 空气动力学理论分析,语音不仅有声门的非线性振动过程,还受舌、声道形状的变化影响。因此,由于发声时气流及气流运动腔道的改变,使得咽擦音与正常音在非线形特征上的呈现具有差异,观察表 9 中的测试结果可知,两种熵以及赫斯特指数等非线性参数对咽擦音的识别准确率均为 80%左右,能在一定程度上识别出咽擦音以及正常音。但是这些非线性表达的信息不够全面,因此识别准确率较低。

通过与传统参数的比较发现,传统语音特征虽然能在一定程度上区分咽擦音以及正常音,但是准确率不高。FTSL 谱线通过计算每一帧信号的多延迟四阶累积量,突出体现了信号的波动差异,与传统语音特征相比,能够更好地突出咽擦音与正常音的区别。

**结束语** 本文通过计算语音信号的多延迟四阶累积量频谱,使用 1/3 倍频程算法提取 FTSL 特征谱线,该谱线突出了咽擦音辅音与正常擦音辅音在统计特征分布上的差异。与传

统的方法相比,实验从信号发音机理出发,分析了咽擦音与正常擦音在高阶统计量上的分布差异,并使用 SVM 分类器实现了咽擦音及正常擦音的自动分类检测。实验还进一步讨论了四阶累积量的延迟范围设置及谱线组合对咽擦音分类结果的影响,并对比讨论了低阶累积量对应的谱线特征及传统语音特征对咽擦音的检测结果。实验分别对 394 个语音样本进行了特征参数提取及模式识别,得到腭裂语音咽擦音的自动分类检测准确率。经研究发现,在基于多延迟四阶累积量的 FTSL 谱线特征下检测正确率最高,为 92.7%,这充分说明了本文提出的 FTSL 谱线对咽擦音与正常擦音的自动分类检测具有很好的鲁棒性,对辅助临床诊疗具有一定的应用价值。

**致谢** 腭裂咽擦音患者的语音数据由华西医院唇腭裂治疗中心提供,由于咽擦样本量较少,医生们时刻关注新增患者并及时提供新的样本,并对这些样本进行了人工评估,使得本文的研究得以顺利进行,因此衷心感谢华西医院唇腭裂治疗中心的医生们为本文做出的贡献。

## 参 考 文 献

- [1] XIAO Y, LIANG M G. Automatic Detection of Pharyngeal Fricatives in Cleft Palate Speech[C]// Proceedings of the 4th International Conference on Computer Engineering and Networks. Springer International Publishing, 2015: 861-868.
- [2] REN Z, ZHOU X, MA L, et al. Comparison Study of Vocal Attack Time in Patients With Cleft Palate With and Without Glottal Stop in Mandarin[J]. Journal of Voice: Official Journal of the Voice Foundation, 2018, 33(5): 803. e15-803. e21.
- [3] MA S W, REN Z P, WEN Y X, et al. Compensatory articulation in patients with repaired cleft palate and the speech therapy approach[J]. Journal of Practical Stomatology, 2012, 28(5): 619-622.
- [4] GUERRA T A, MARINO V C C, ROCHA D C, et al. Nasalance at presence and absence of pharyngeal fricative[J]. Speech, Language, Hearing Sciences and Education Journal, 2016, 18(2): 449-458.
- [5] DENG S H, WANG T S, HUANG R C, et al. Study on the characteristics of the speech of persons with disorder and sound after Postoperatively in cleft palate[J]. China Prac Med, 2017, 12(2): 194-195.
- [6] JIANG L P, WANG G M, YANG Y S, et al. The study on articulation characteristics of the patients after pharyngoplasty[J]. China Journal of Oral Maxillofacial Surgery, 2005(1): 56-58.
- [7] MA L. The acoustic features and the articulation character of tongue movement of pharyngeal fricatives [C]// Abstracts of the 2nd Chinese International Congress on Oral and Maxillofacial Surgery in Conjunction with 5th National Congress on Oral and Maxillofacial Surgery. 1998: 267-268.
- [8] GARCIA A F, MARINO V C, PEGORARO-KROOK M I, et al. Nasalance during use of pharyngeal and glottal place of production[J]. Cotas, 2014, 26(5): 395-401.
- [9] WANG G M, CHEN Y, QIU W L, et al. Clinical application and evaluation in analysis of articulation disorders WITH TSL[J]. J. Oral Maxil. Surg., 2000(3): 189-197.

- [10] ZHANG C H, ZHOU H Y, JIAO X H. Phonetic features of before and after posterior pharyngeal flap surgery in older parents with velopharyngeal insufficiency[J]. Journal of Harbin Medical University, 2016, 50(2): 162-165.
- [11] ZHU Y S, SHI J J. A acoustic technology analysis of misarticulation in patients with cleft palate [J]. Journal of Practical Stomatology, 2004(3): 364-366.
- [12] ALAM M K, ZULKIPLI A S, HAQUE S, et al. A perceptual evaluation of speech disorders in children with repaired unilateral cleft lip and palate in Hospital Universiti Sains Malaysia[J]. Angladesh Journal of Medical Science, 2018, 17(3): 470-478.
- [13] MCLEOD S, CROWE K. Children's Consonant Acquisition in 27 Languages: A Cross-Linguistic Review[J]. Am. J. Speech Lang Pathol., 2018, 27: 1546-1571.
- [14] 雷丽. 腭裂语音治疗学[M]. 武汉: 湖北科学技术出版社, 2004: 24-37.
- [15] TROST J E. Articulatory additions to the classical description of the speech of persons with cleft palate[J]. Cleft Palate Journal, 1981, 18(3): 193-203.
- [16] 张贤达. 时间序列分析—高阶统计量方法[M]. 北京: 清华大学出版社, 1999.
- [17] 张贤达. 现代信号处理(2版)[M]. 北京: 清华大学出版社, 2002.
- [18] DONG X H. Application of MUSIC algorithm based on fourth-order cumulants in high frequency ground wave radar[D]. Wuhan: Wuhan University, 2004.
- [19] FAN Y Y. High order statistics feature extraction of ship noise and its response[D]. Xi'an: Northwestern Polytechnical University, 1999.
- [20] VOSOUGHI E, JAVAHERIAN A. Parameters effective on estimating a nonstationary mixed-phase wavelet using cumulant matching approach[J]. Journal of Applied Geophysics, 2018, 148: 83-97.
- [21] LV J Y. High Order Statistics Analysis and its Applications [D]. Beijing: Beijing University of Posts and Telecommunications, 2014.
- [22] ANANTHRAM S, GEORGIOS B, et al. Bibliography on higher-order statistics[J]. Signal Processing, 1997(60): 65-66.
- [23] MENDEL J M. Tutorial on higher order statistics (spectra) in signal processing and system theory: Theoretical results and some applications[J]. Proc. IEEE, 1991, 79(3): 278-305.
- [24] ALBATAINEH Z. Robust blind channel estimation algorithm for linear STBC systems using fourth order cumulant matrices [J]. Telecommunication Systems, 2018, 68(3): 573-582.
- [25] LIANG H, YANG C S. A Signal Detection Algorithm Based on Fourth-order Cumulant[J]. Torpedo Technology, 2007(5): 48-50.
- [26] MEI T M. Blind signal separation algorithm based on symmetric fourth-order mutual cumulant[C]// Signal Processing Branch of China Electronics Society and Signal Processing Branch of China Institute of Instruments and Instruments, 2003: 4.
- [27] ELIAS N, RAFIK G, SAMY M. Speech enhancement using fourth-order cumulants and optimum filters in the subband domain[J]. Speech Communication, 2002, 36(3): 219-246.
- [28] QIAN Z, LI X Y, ZHANG R B, et al. Speech-stream detection in short-wave channel based on empirical mode decomposition and higher-order statistics[J]. Journal of Harbin Institute of Technology, 2009, 16(5): 713-716.
- [29] BAO H Q. A brief introduction to phonological and acoustic analysis of Putonghua (continued 1) [J]. Journal of Audiology and Speech Pathology, 2004(4): 285-286.
- [30] CHENG J, LI G H, ZHOU G L. Simplified Calculating Simulation of Fourth-Order Cumulants [J]. Computer Simulation, 2009, 26(8): 80-83.
- [31] ZHANG A Q, ZHANG X H. Recursive estimation of fourth-order cumulants and application[J]. Signal Processing, 2002(1): 88-90.
- [32] LU W C, CHEN N Y, YE C Z, et al. Introduction to Support Vector Machine Algorithms and Software ChemSVM[J]. Computer and Applied Chemistry, 2002(6): 697-702.
- [33] FAN X W. Research and application of support vector machine algorithm [D]. Hangzhou: Zhejiang University, 2003.
- [34] QIN Y Q, ZHANG X Y. Speech signal emotion recognition based on SVM[J]. Journal of Circuits and Systems, 2012, 17(5): 55-59.
- [35] NAZEER O, JAVAID N, et al. Short Term Load Forecasting Using Heuristic Algorithm and Support Vector Machine[C]// 12th International Conference on Complex, Intelligent, and Software Intensive Systems (CISIS), 2019: 791-799.
- [36] LUO R L. Research on text independent speaker recognition algorithm based on SVM [D]. Lanzhou: Lanzhou University of Technology, 2009.
- [37] LUO R L. Text independent speaker recognition algorithm based on SVM [D]. Lanzhou: Lanzhou University of Technology, 2009.
- [38] TANG J T, HU D, GONG Z M. Research on image texture classification based on SVM[J]. Computer Engineering and Science, 2008(8): 44-45, 48.
- [39] GANDEK B, WARE J E, AARONSON N K, et al. Cross-validation of item selection and scoring for the SF-12 Health Survey in nine countries; results from the IQOLA Project[J]. Journal of clinical epidemiology, 1998, 51(11): 1171-1178.



**HE Fei**, born in 1998, postgraduate. Her main research interests include speech signal processing and image processing.



**YIN Heng**, born in 1971, master. Her main research interests include evaluation of cleft palate speech.