病态嗓音特征的小波变换提取及识别研究

于燕平[1,2],胡维平[1]
YU Yan-ping[1,2],HU Wei-ping[1]

1.广西师范大学 物理与电子工程学院,广西 桂林 541004
2.柳州铁道职业技术学院 电子工程系,广西 桂林 545007
1.College of Physics and Electronic Engineering,Guangxi Normal University,Guilin,Guangxi 541004,China
2.Department of Electronic Engineering,Liuzhou Railway Vocational Technical College,Liuzhou,Guangxi 545007,China
E-mail:yuyp156@126.com

YU Yan-ping,HU Wei-ping.Research of extracting of pathological voice's characteristics and recognition based on wavelet transformation and Gaussian mixture model.Computer Engineering and Applications,2009,45(22):194-196.

Abstract: Considering the voice pronunciation mechanism,the different performances of the abnormal voice and the normal voice in the field of frequency,the paper proposes a new method for extracting characteristics that is Entropy Coefficient based on De-noise,Decomposition of Multi-scale Analysis (ECDDMA) using the wavelet decomposition to find the pathological voice's characteristics,and comparative analysis of the effective speech characteristics MFCC.242 normal voices samples and 234 abnormal samples are recognized with MFCC and the new extracted characteristics ECDDMA based on Gaussian Mixture Model (GMM).The result indicates that,the parameters of ECDDMA are more advantageous to the normal and abnormal voice recognition than the traditional MFCC and the dynamic characteristic which mimic the human ears non-linear characteristic with frequency,and improves the abnormal and normal voice's recognition result.

Key words: Gaussian Mixture Model(GMM);pathological voice;Mel Frequency Cepstrum Coefficient(MFCC);wavelet transformation

摘 要:通过分析嗓音的发音机理、病态嗓音与正常嗓音在频域的表现差异,利用小波变换对信号进行分解,突出病态嗓音的特点,提出了基于多尺度分析的小波降噪、分解的熵系数(Entropy Coefficient based on De-noise,Decomposition of Multi-scale Analysis,ECDDMA)作为识别的特征矢量集。并对比分析了语音识别中经典特征参数 Mel 倒谱系数(MFCC),分别运用这两种特征参数对 242 例正常嗓音和 234 例病态嗓音运用高斯混合模型(GMM)进行了识别。结果显示:ECDDMA 系数较传统的模拟人耳听觉非线性特性的 MFCC 及其动态特征能更准确地表征正常与病态嗓音之间的差异,有利于同时提高病态和正常嗓音的识别率。

关键词:高斯混合模型(GMM);病态嗓音;Mel 倒谱系数(MFCC);小波变换
DOI:10.3778/j.issn.1002-8331.2009.22.062 **文章编号**:1002-8331(2009)22-0194-03 **文献标识码**:A **中图分类号**:TN912

1 引言

声带的各种病理性改变导致其振动和闭合异常,使得喉声源声学性质发生改变,出现不同程度的声音嘶哑[1]。目前国内较为常用的喉功能检查方法是利用计算机技术, 采用 Dr.speech 软件进行正常嗓音及病态嗓音的多种声学参数分析,同时和电声图结合能进一步地对基频(F0)、频率微扰(Jitter)、振幅微扰(Shimmer)、规范化噪声能量(NNE)等各种声学参数进行检测,但它们在有效检测病态嗓音方面都存在一定局限[2-3]。而计算机病态嗓音识别方法对临床实现无痛无损伤化嗓音检查具有重要的意义[4]。

语音信号是一种非线性、非平稳的信号,以往求取特征的方法都是通过加窗傅里叶变换,这种变换最大的缺点就是不能同时提高时间与频率的分辨率,而小波变换则可以很好的克服这个缺点,它可以灵活地调整时-频窗,同时对时频分辨率作出贡献。在传统计算机语音识别方法中,因 MFCC 能比较充分利用人耳特殊感知特性而获得了广泛的应用[5],但有分析认为人耳在最初识别声音时使用的是小波变换[6],结合病态嗓音在不同频率范围表现的差异[7],提出了基于多尺度分析的小波降噪、分解的熵系数(Entropy Coefficient based on De-noise,Decomposition of Multi-scale Analysis,ECDDMA), 并用实验证明了 ECDDMA 在病态嗓音识别中与传统特征相比的优越性。近年来,HMM 广泛地用于语音识别,神经网络也被用于病态嗓音评估[8],作为状态数为 1 的连续型 HMM 的高斯混合模型(GMM)也同样得到了广泛的应用。由于在一个状态中可以包含多个高斯密度函数,不存在状态转移概率,因此在计算量上 GMM 就比 HMM 要小得多。使用 GMM 对病态嗓音和正常嗓音

基金项目:广西自然科学基金(the Natural Science Foundation of Guangxi of China under Grant No.0448035)。
作者简介:于燕平(1982-),女,硕士研究生,主要研究方向为语音信号处理;胡维平(1963-),男,教授,博士,主要研究方向为生物医学信号处理和图像处理等。
收稿日期:2008-04-23 **修回日期**:2008-07-24

尝试识别率。特征使用目前广泛应用于语音识别的 MFCC 参数及提出的 ECDDMA 参数,并比较其识别效果。

2　数据来源

实验数据来源于临床病例,采集数据时的环境要求在安静的室内进行;采样频率为 16 kHz,时间 1.5 s 至 3 s;受试声样为汉语元音'a',分别对正常人和患有各类喉科疾病的对象进行语音采样。正常对照组 242 例,年龄 18~40 周岁,平均年龄 25 周岁,经询问近期无喉部疾病者;病态嗓音组 234 例,年龄 15~50 周岁,平均年龄 27 周岁,为前来医院就诊临床病例。采集后用 cooledit 软件进行语音分割,得到实验用语音库。

3　高斯混合模型

高斯混合模型作为高斯概率密度函数的一个线性组合,只要有足够数目的混合分量,就可以逼近任意一种密度函数。一个 M 阶混合高斯模型的概率密度函数是由 M 个高斯概率密度函数加权求和得到的,所示如下:

$$P(\boldsymbol{X}/\lambda)=\sum_{i=1}^{M}\omega_i b_i(\boldsymbol{X}) \tag{1}$$

其中,$\boldsymbol{X}$ 是一个 D 维随机向量,$b_i(X_t)$,$i=1,\cdots,M$ 是子分布,ω_i,$i=1,\cdots,M$ 是混合权重。每个子分布是 D 维的联合高斯概率分布,可表示为:

$$N(x_t,\boldsymbol{\mu}_i,\boldsymbol{\Sigma}_i)=\frac{1}{\sqrt{(2\pi)^D|\boldsymbol{\Sigma}_i|}}\exp\left[-\frac{1}{2}(\boldsymbol{X}-\boldsymbol{\mu}_i)^t\boldsymbol{\Sigma}_i^{-1}(\boldsymbol{X}-\boldsymbol{\mu}_i)\right] \tag{2}$$

其中 $\boldsymbol{\mu}_i$ 代表此密度函数的均值向量,$\boldsymbol{\Sigma}_i$ 则代表此密度函数的协方差矩阵,混合权重需满足:

$$\sum_{i}^{M}\omega_i=1 \tag{3}$$

完整的混合高斯模型有参数表示为:

$$\lambda=\{\omega_i,\boldsymbol{\mu}_i,\boldsymbol{\Sigma}_i\},i=1,\cdots,M \tag{4}$$

GMM 模型参数估计最常用的参数估计方法是最大似然(Maximum Likelihood,ML)估计。对于一组长度为 T 的训练矢量序列 $\boldsymbol{X}=\{X_t\}$,$t=1,2,\cdots,T$,GMM 的似然度可以表示为:

$$P(\boldsymbol{X}/\lambda)=\prod_{t=1}^{T}P(X_t/\lambda) \tag{5}$$

由于上式是参数 λ 的非线性函数,很难直接求出上式的最大值。因此,常常采用 EM(Expection Maximization,EM)算法估计参数 λ。对于多观察序列迭代的 ML 重估公式为:

$$\bar{\omega}_k=\frac{\sum_{c=1}^{c}\sum_{t=1}^{T_c}\gamma_t^c(k)}{\sum_{k=1}^{K}\sum_{c=1}^{c}\sum_{t=1}^{T_c}\gamma_t^c(k)} \tag{6}$$

$$\bar{\mu}_k=\frac{\sum_{c=1}^{c}\sum_{t=1}^{T_c}\gamma_t^c(k)x_t^c}{\sum_{c=1}^{c}\sum_{t=1}^{T_c}\gamma_t^c(k)} \tag{7}$$

$$\bar{\Sigma}_k=\frac{\sum_{c=1}^{c}\sum_{t=1}^{T_c}\gamma_t^c(k)(x_t^c-\mu_k)^2}{\sum_{c=1}^{c}\sum_{t=1}^{T_c}\gamma_t^c(k)} \tag{8}$$

其中 c 为观察序列的数目,T_c 是模型的第 c 个观察序列的长度,$\gamma_t^c(k)$是第 c 个观察序列的第 k 个混合成分在时间 t 的概率:

$$\gamma_t^c(k)=\frac{N(x_t^c,\mu_k,\Sigma_k)}{\sum_{k=1}^{K}N(x_t^c,\mu_k,\Sigma_k)} \tag{9}$$

4　基于小波变换的特征提取

4.1　二进离散小波变换原理及其 Mallat 算法

为了便于计算机的处理,在对嗓音信号进行小波分析时,不仅对尺度参数 a 和时移参数 b 进行离散化处理,而且信号在时间上也是离散的,并表示为 $f(n)(n\in Z)$,这种情况下的母小波和相应的小波都应该是离散时间的,分别用 $\psi(n)$和 $\psi_{j,k}(n)$表示:

$$\psi_{j,k}(n)=2^{-j/2}\psi(2^{-j}n-k),j,k\in Z \tag{10}$$

则 $f(n)$关于 $\psi_{j,k}(n)$的离散二进小波变换(DWT)表示为:

$$DWT_\psi f(2^j,k)=\sum_{n=-\infty}^{\infty}f(n)\overline{\psi}_{j,k}(n)=2^{-j/2}\sum_{n=-\infty}^{\infty}f(n)\overline{\psi}(2^{-j}n-k),j,k\in Z \tag{11}$$

其中$\overline{\psi}_{j,k}(n)$是 $\psi_{j,k}(n)$的共轭。

Mallat 分解算法数学式为:

$$\begin{cases}s^{j+1}(n)=\sum_k s^j(k)h(k-2n),n=0,1,\cdots,2^{N-j}-1\\ d^{j+1}(n)=\sum_k s^j(k)g(k-2n),j=0,1,\cdots,M-1\end{cases} \tag{12}$$

式中,$h(k)$、$g(k)$为分解低通 H、高通滤波器 G 的冲激响应;M 为分解层数;s^j、d^j 分别为第 j 尺度下的近似系数和细节系数,如图 1。

Mallat 算法重构示意见图 2,重构式为:

$$s^j=\sum_k\{h^+[n-2k]s^{j+1}(k)+g^+[n-2k]d^{j+1}(k)\},\quad j=M-1,M-2,\cdots,0 \tag{13}$$

式中,$h^+(k)$、$g^+(k)$为重构低通、高通滤波器的冲激响应。

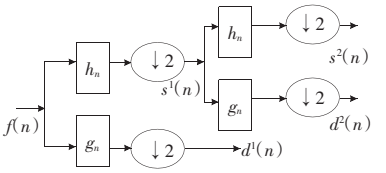

图 1　离散小波的分解算法

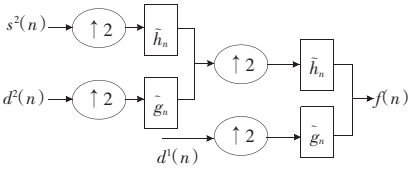

图 2　离散小波的重构算法

4.2　特征提取

由于声带病理性的改变,其振动与闭合异常将导致产生的

嗓音中存在的噪声随着症状的加重而不断的增多。文献[9]中提到:轻度喉病对声带振动的影响的是噪声首先在大于 1 700 Hz 的频谱高端出现,随着病情的加重,噪声才逐步出现在低于 1 700 Hz 的频段中。在病情发展过程中(在发展到非常严重之前),一般低频段谐波都能保持正常,基本上极少有噪声出现。因此在病情发展初期噪声增加不是很明显的情况下,时域上病态与正常嗓音差异甚微,要直接判断出病态与正常嗓音就显得非常的困难。通过以上对病态嗓音特性的分析,完全可以有理由认为如果对病态嗓音与正常嗓音进行滤波的预处理,可以突出其频域的区别。考虑到信号和噪声在小波域中有不同的形态表现,它们的小波系数幅值随尺度变化的趋势不同,随着尺度的增加,噪声系数的幅值很快衰减为零,而真实信号系数的幅值基本不变,采用小波阈值降噪的方法进行预滤波处理。

由于正常嗓音与病态嗓音之间的这些差别及噪声随着小波分解层数增加而迅速衰减的特性,首先对语音进行小波降噪预处理,提取出降噪后的语音,根据正常与病态语音能量集中的频段不同,其在各频段表现出的不稳定性也不同,这里再对降噪后语音作小波分解,分解出信号的各频率细节,为了更加准确的描述正常与病态嗓音的各频段幅度变化的差别,将不使用求方差的作法,而采用一种描述复杂度的量–近似熵[10]。

由于 ECDDMA 参数的求取是基于小波变换,而小波变换本身是一个灵活的时–频窗,它不需像以往的方法做加窗处理来满足语音的短时平稳性,因此在这种方法中不需要对语音信号分帧。有实验表明去噪的效果与选取的小波和阈值有关,硬阈值在逼近程度方面要优于软阈值,而在光滑程度方面却劣于软阈值[11],由于硬阈值难以确定,去噪时选取软阈值。ECDDMA 参数提取过程如下:

(1)对输入语音 X 使用 sym8 小波对语音进行 5 层分解,返回小波分解向量 $\boldsymbol{C}$ 及相应的记录向量 $\boldsymbol{L}$;

(2)对小波分解后的信号的细节系数作软阈值自动降噪,重构降噪后的有用信号 Xd;

(3)对降噪后的信号用 db4 小波做小波分解(对小波做 10 层分解),求取低频系数向量(cA)和高频系数向量(cD);

(4)重构低频信号 $A10$ 和各层的高频信号 $D10$、$D9$、$D8$、$D7$、$D6$、$D5$、$D4$、$D3$、$D2$、$D1$;

(5)对低频信号和各高频信号按文献[10]中求近似熵(共 11 维)。

5 基于 GMM 的病态嗓音识别实验结果及分析

实验采用 GMM 模型作为识别系统,在数据库的 242 例正常嗓音和 234 例病态嗓音中,随机选取正常嗓音和病态嗓音各 80 例作为模型训练集;其余作为测试集,各为 162 例和 154 例。

(1)为了表明预滤波的有效性,将语音提取 ECDDMA 特征参数时去掉去噪的部分提取出没有去噪的参数,与 ECDDMA 参数(都为 11 维)的识别结果对比如表 1。

表 1 去噪前参数与 ECDDMA 参数识别结果对比(错误/正确)

特征	训练组正常样本识别率/(%)	训练组病态样本识别率/(%)	测试组正常样本识别率/(%)	测试组病态样本识别率/(%)
去噪前参数	97.50 (2/78)	97.50 (2/78)	91.97 (13/149)	87.66 (19/135)
ECDDMA 参数	100 (0/80)	100 (0/80)	95.06 (8/154)	92.20 (12/142)

(2)表 2 给出了使用传统特征 MFCC(12 维)及其差分参数(24 维)的识别结果、ECDDMA 参数(11 维)以及 A–F 标准差参数[12](6 维)的识别结果数对照表,其中模型混合数选取各种参数最佳的混合数。

表 2 两种特征识别结果对照表(错误/正确)

特征	训练组正常样本识别率/(%)	训练组病态样本识别率/(%)	测试组正常样本识别率/(%)	测试组病态样本识别率/(%)
MFCC 参数	100 (0/80)	98.75 (1/79)	69.75 (49/113)	92.85 (11/143)
MFCC+ △MFCC	100 (0/80)	100 (0/80)	70.98 (47/115)	92.85 (11/143)
ECDDMA 参数	100 (0/80)	100 (0/80)	95.06 (8/154)	92.20 (12/142)
A–f 标准差参数	97.5 (2/78)	98.75 (1/79)	87.65 (20/142)	92.85 (11/143)

由于声带的病理性变化或功能性问题引起声门非完全关闭,特别是声带麻痹时声带固定,声门难于闭合,发声时由于声门空气泄露造成的骚动噪声增加[13]。从病态嗓音中不同频段表现及其噪声成分特性出发,使用先降噪再用小波分解的方法。表 1 中 ECDDMA 参数的识别效果要优于未去噪而直接作小波分解提取的参数,测试组正常样本和病态样本识别率分别高出 3.09%,4.54%,结果表明:(1)语音信号经小波变换,分解到不同频率范围后求取近似熵得到的参数能有效地鉴别正常嗓音与病态嗓音;(2)在提取参数之前对语音作预滤波是能提高识别率的。表 2 识别结果显示:ECDDMA 参数能够克服传统特征无法同时提高正常和病态样本识别结果的缺点,因此,ECDDMA 能较准确地表征正常与病态嗓音之间的差别,测试集有效识别率正常组为 95.06%,病态组为 92.20%,正常组较 MFCC 高出 24.08%;较文献[12]中提出的 A–f 标准差参数其正常识别率也高 7.41%。由以上的分析看出:不论是模拟人耳特殊感知特性的 MFCC 及其动态参数,还是基于 HHT 提取出的 A–f 参数,提出的参数在病态嗓音识别中都显现出了优越性。

6 总结

这种特征提取的方法最大的优势在于:由于小波变换不依赖语音的短时平稳性,测试语音就不需要分帧;虽然 A–f 标准差参数特征维数为 6 维,但由于其提取过程中需对语音进行分帧处理,因此实际特征维数为 6* 帧数,文中最终特征维数为 11 维,也就是说,一段复杂的语音只需几个简单的特征就足以描述,其计算量要比传统方法要快得多,作为一种无损伤的检测方法,这种能有效识别正常与病态嗓音的方法将对喉病临床诊断具有一定的实用价值,给临床医生提供了有力的参考。这种特征提取的方法也将为其他信号处理领域提供新的思路。但由于小波变换需要解决如何选取合适的小波基等问题,因此在取得好的结果之前总是要经过多次的实验来确定,这也是一直以来小波变换研究待解决的问题。

参考文献:

[1] 徐洁洁,乔宗海.病态嗓音的计算机定量评估[J].南京医科大学学报,2000,20(2):121–124.

(下转 205 页)