

小波分析和 MFCC 融合的声音信号端点检测算法

张芝漪 姚恩涛 石玉

(南京航空航天大学 自动化学院 南京 211106)

摘要: 声振法检测风机叶片的脱层损伤是一种可行也比较容易实现的方法,声信号端点检测是声振法检测中比较重要的环节。为了提高端点检测的适应性和鲁棒性,研究了小波分析及 MFCC 参数的提取,结合二者特点,利用小波变换改进 MFCC 提取过程,提出新参数 DWT-MFCC,并利用 SVM 进行声音信号端点检测。实验结果表明,相比于传统的小波参数和 MFCC 参数,在相同噪声环境下,DWT-MFCC 具有更高的端点检测率。

关键词: 小波变换;MFCC 参数;DWT-MFCC;SVM;端点检测

中图分类号: TN912.3 **文献标识码:** A **国家标准学科分类代码:** 080402

Audio endpoints detection algorithm based on wavelet analysis and MFCC

Zhang Zhiyi Yao Entao Shi Yu

(College of Automation Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China)

Abstract: Acoustic method to detect wind vane delaminating damage is a kind of feasible and relatively easy to implement, the method of acoustic signal endpoint detection is important link in acoustic resonance method to detect. In order to improve the adaptability and robustness of endpoint detection, wavelet transform and the MFCC parameters extraction are studied. Based on these research, a new parameter, DWT-MFCC, is extracted. And using the SVM to audio endpoint detection. The experimental results show that, compared with the traditional wavelet and MFCC parameters, in the same noise environment, DWT-MFCC has a higher rate of endpoint detection.

Keywords: wavelet transform; MFCC parameter; DWT-MFCC; SVM; endpoint detection

1 引言

风机叶片是由复合材料制成的一种薄壳结构。叶片在加工、使用和运输过程中容易产生脱层、纤维分层等损伤。复合材料脱层通常发生在材料内部,几乎不能通过目视法进行检测。目前复合材料损伤的无损检测技术很多,有射线法、涡流法、激光全息干涉法、热像技术、声全息、超声扫描、声发射、纤维光学法、应力波技术图等。这些方法所用的设备价格昂贵,操作复杂,且大多数方法对检测的环境要求比较严格,因此,在现场环境下很难完成检测。声音随着材料振动形式的变化而变化,物体有损伤时,敲击材料得到的振动形式发生变化,敲击发出的声音也相应发生变化^[1]。声音信号可以很容易的通过声音传感器获取。声音信号的频率范围为 20 Hz~20 kHz,但很容易受到外界噪声的影响,有效抑制声音信号中的噪声信号,然后对声音信号进行

处理和分析,运用信号处理的方法提取特征参数,以检测材料结构的健康状况。因此,声振法检测风力叶片的脱层损伤是一种可行也比较容易实现的方法。

声振法检测中比较重要的环节是声信号端点检测,把有效声信号从背景噪声信号中提取出来,便于后面对声音信号进行分析。传统的端点检测算法常采用短时能量、短时过零率和自相关最大值等方法^[2-4],这些方法在低噪声环境下均能取得比较好的结果,但在高噪声环境下性能则下降明显,甚至不能判断。还有一些方法是通过对声音信号的频域进行分析,如频带方差倒谱特征法和谱熵法等^[5-7],信噪比较高时,此类端点检测方法可以取得较好的检测结果,但在信噪比较低的情况下,这类端点检测方法的准确性依然很差。不同的端点检测方法有着各自的优点和不足,可以互补,近些年多种方法融合的端点检测算法不断出现,有效地提高了在信噪比较低、噪声环境复杂的情况下的端

点检测性能^[8]。

文献[9]提出了一种基于小波子带能量特征与支持向量机结合的语音端点检测方法,与传统的端点检测方法相比在信噪比较低的情况下也能得到较好的检测率,但这是相对于白噪声的,在其他突发性噪声或者有色噪声较大的情况下并不能得到较好的检测率。由于 MFCC(Mel 频率倒谱系数)较好地反映了人耳的听觉特性,本文提出了一种将小波分析和 MFCC 结合的方法(DWT-MFCC)并应用在声振法检测风机叶片脱层损伤中声音信号端点检测上。

2 基于小波子带能量的声信号特征提取

2.1 信号的小波分解

小波变换的基本思想是通过伸缩和平移一组小波基函数 $\varphi(t)$,使得信号 $f(t)$ 的时频窗口随频率的变化而变化,以实现低频分量和高频分量分别采用大视窗和小视窗分析。小波变换的定义为:

$$WT_f(a, \tau) = \frac{1}{\sqrt{a}} \int f(t) \varphi\left(\frac{t-\tau}{a}\right) dt \quad (1)$$

式中: a 表示伸缩尺度, τ 表示平移时间。

首先将传感器采集到的声音信号进行分帧处理,实验中采取 240 点为 1 帧,得到一系列的帧数据。对窗函数采用低频拉伸高频压缩的方法进行小波变换,然后对每帧信号 $f(n)$ 进行 5 层小波分解,分解公式如下所示:

$$f(n) \rightarrow a_1(n) \rightarrow a_2(n) \rightarrow \dots \rightarrow a_5(n) \\ d_2(n) \quad d_3(n) \quad \dots \quad d_5(n) \quad (2)$$

首先将信号分解到高频和低频空间,再将低频空间 $a_1(n)$ 分解到高频和低频空间,然后继续分解低频空间 $a_2(n)$,一直分解到 5 层后得到 $d_1(n) \sim d_5(n)$ 、 $a_1(n) \sim a_5(n)$ 共 10 个不同的频率信号,其中 $d_1(n) \sim d_5(n)$ 、 $a_5(n)$ 可表征原信号所有的频率信号。

2.2 特征量的提取

分析声音信号频域特征可知,声音信号在各个小波子带内的平均能量分布不是均匀分布的,在低频子带中信号能量较大,而高频子带相反。然而噪声(此处考虑噪声为白噪声)在各个子带内的平均能量分布比较均匀。

以下介绍声音信号的小波子带系数能量的提取过程:

1) 不同层小波子带系数的平均能量计算公式如下:

$$E_i = \frac{1}{N} \sum_{n=1}^N |s_i(n)| \quad (3)$$

式中: $s_i(n)$ 代表 $d_1(n) \sim d_5(n)$ 、 $a_5(n)$ 这些小波子带的小波系数, N 表示对应的小波子带中所含有的小波系数的个数。

2) 计算 6 个子带的平均能量的均值 E_m 和方差 σ^2 。

3) 对于每帧信号,由 6 个小波子带的平均能量 $E_1 \sim E_6$, 以及一个 E_m 和 σ^2 构成该帧语音信号的特征矢量:

$$\mathbf{X}(n) = [E_1, E_2, E_3, E_4, E_5, E_6, E_m, \sigma^2]^T \quad (4)$$

以这 8 个量为特征量,可以同时表现声音信号在时域

和频域上的能量分布情况,可以表征该帧信号的各种特点,通过这些特征参数可以判别该帧信号为噪声或有用信号。

2.3 小波能量特征提取的不足

基于小波子带能量特征的端点检测方法的主要思路就是提取信号小波子带能量序列及子带能量方差作为特征,通过 SVM 和神经网络等机器学习算法分辨有用信号与噪声信号,从而实现声音信号的端点检测。显然,如果信号主要由白噪声组成,那么,由于白噪声的性质,小波子带能量序列应该近似为一个常数序列,其方差近似为 0。所以,小波子带平均能量提取的特征向量能够在信噪比很低的情况下区分声音信号中的高斯白噪声,另外,对于突发性和有色噪声,在信噪比较高的情况下,噪声信号的能量相比于有用信号比较小,通过能量大小也能得到较好的结果。

当遇到其他突发性的噪声或有色噪声时,信噪比较高时检测率还是比较高的,当信噪比较低时检测率就下降明显,甚至失效。这是由于,在信噪比较高时,声音信号和噪声信号还可以通过能量的大小来做出区分。当信噪比较小时,由于噪声信号不是白噪声,能量序列并不能近似成一个常数序列,在能量与声音信号能量相当的情况下,这种方法就不能很好的区分噪声和语音信号,因此也就不能获得好的端点检测结果。

3 基于 MFCC 的声信号特征提取

3.1 MFCC 原理

人类的听觉系统是一个特殊的非线性系统,它对不同频率信号响应的灵敏度是不同的,基本上呈现对数模型^[10]。MFCC(Mel 频率倒谱系数)是根据人耳的听觉特性提出的一种特征参数,与根据实际频率倒谱提取的特征参数不同。人耳对声音频率的响应与声音信号的实际频率呈非线性关系,而 Mel 频率倒谱系数的设计就是基于人耳的听觉特性的。Mel 频率与线性频率的关系如图 1 所示。

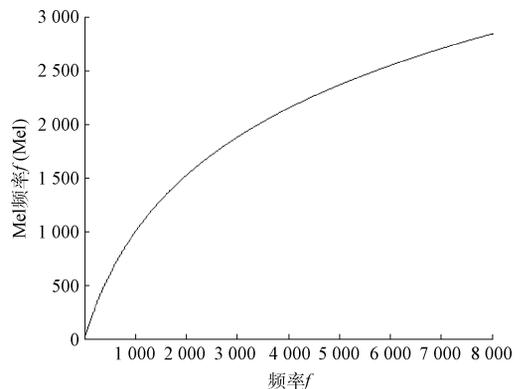


图 1 Mel 频率与线性频率关系曲线

Mel 倒谱参数特征对于信号性质的依赖性不强,对输入的声音信号没有任何的限制,该特征参数的设计利用了听觉模型的研究成果。所以,该特征参数与利用声道模型设计的 LPCC 参数相比具有更好的鲁棒性和适应性,与人耳的听觉特性更加符合。在信噪比较低的情况下,MFCC 特征参数仍然具有较好的检测性能。

3.2 MFCC 特征提取步骤

MFCC^[11]的计算步骤如下:

- 1) 利用信号 FFT 计算语音信号的频谱 $F(k)$ 。
- 2) 实际频率到 Mel 频率的映射,如式(5)所示:

$$\text{Mel}(f) = 2595 \lg(1 + f/700) \quad (5)$$

式中:参数 f 表示实际频率,单位为 Hz。

3) 定义一个含有 L 个三角滤波器的滤波器组,滤波器的个数取决于信号的截止频率。在 Mel 频率尺度上,每个滤波器的中心频率间隔相等。设第 l 个三角滤波器的下限频率、中心频率和上限频率分别是 $o(l)$ 、 $c(l)$ 、 $h(l)$,由于相邻的三角滤波器有重叠,因此式(6)表示它们的关系:

$$c(l) = h(l-1) = o(l+1) \quad (6)$$

4) 将信号频谱输入三角滤波器组,得到滤波器的输出:

$$m(l) = \sum_{k=o(l)}^{h(l)} W_l(k) |F(k)|, l = 1, 2, \dots, L \quad (7)$$

5) 计算每个滤波器输出的对数能量,然后经过离散余弦变换(DCT)可得 MFCC 参数:

$$c_{\text{MFCC}}(i) = \sqrt{\frac{2}{N}} \sum_{l=1}^L \lg[m(l)] \cos \left[\left(l - \frac{1}{2} \right) \frac{i\pi}{L} \right] \quad (8)$$

3.3 MFCC 特征存在问题

MFCC 参数很好地描述了人耳的听觉机制,能够很好地区分语音和噪声,其具有较强的稳定性以及抗噪性。然而基于 MFCC 相似度的端点检测准确率并不高,特别是当信噪比提高后未有显著提升。

4 基于 DWT-MFCC 特征参数的提取

小波变换的基本思想是多分辨率分析,是对短时傅里叶变换的思想的继承和发展,它克服了短时傅里叶变换中窗口大小固定等缺点,提供的时频窗口随频率的变化而改变,使得信号的时域和频域部分都获得合适的分辨率。将声音信号进行小波分解,得到的小波系数分属不同的频段,高频段的小波系数表示声音信号的快变分量,低频段的小波系数表示信号的慢变分量。对不同频段的小波系数,采用分层 FFT(即不同频段分别做 FFT 变换),根据系数所属的频段将 FFT 结果拼接成完整的频谱,再利用 Mel 滤波器将得到的频谱转换成 Mel 倒谱系数,从而得到一种新的参数 DWT-MFCC。如图 2 所示,DWT-MFCC 特征参数的提取过程与 MFCC 相比,主要的不同在于对信号的处理,特点是该参数采用小波变换和分层 FFT 的方法代替原来的 FFT。

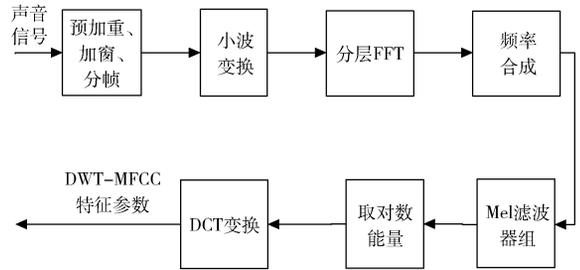


图 2 DWT-MFCC 特征参数提取框图

5 基于 SVM 的声信号端点检测算法

20 世纪 90 年代,Vapink 等人提出了一种机器学习理论,支持向量机(SVM)。SVM 的主要思想是找到一个最优的超平面,通过这个超平面可以实现数据分类,使得待分类数据集中的点尽可能地与该超平面距离最远^[12]。SVM 是一种广泛应用的机器学习理论,它成功的运用于很多现实生活中的分类和回归问题。

基于 DWT-MFCC 特征和 SVM 的声音信号端点检测算法步骤如下:选择一段含噪声的声音信号作为 SVM 的训练样本,首先人工标识每一帧信号,有用信号标示为 1,噪声则表示为 -1。然后对每一帧信号进行处理,提取每帧信号的 DWT-MFCC 特征量,将特征量作为输入向量输入到 SVM 中。将样本信号提取的特征量作为数据输入 SVM,产生输出结果(输出结果值为 -1 或 1),人工标示信号的输出作为 SVM 的期望输出,对 SVM 进行训练,得到训练后的 SVM 模型。训练完后,取一段声音信号作为输入信号,提取每一帧的 DWT-MFCC 特征参数,将其作为 SVM 的输入,SVM 的输出结果即为该帧信号的检测结果,找到声音信号的起止点即为端点检测的结果。

6 端点检测实验及结果分析

6.1 仿真实验数据来源

纯净的声音信号来自无噪声实验室内利用敲击法采集的声音信号,在计算机中的保存格式为 .wav。实验中首先对信号进行分帧处理,并利用 Cool Edit Pro 软件对声音信号样本的起点和终点进行手工标识,此时标示的起止点就是端点位置,后面的端点检测以此为标准起止点。噪声库采用 Noisex-92 中的噪声信号,其中包含了 White、Pink、Babble 等噪声,在 MATLAB 平台上,采用人工加入的方法为纯净的声音信号添加 Noisex-92 噪声库中的噪声,信噪比(SNR)分别为 -5 dB, 0 dB, 5 dB, 10 dB。纯净声信号和各类噪声在 0 dB 的信噪比下叠加声信号波形图如图 3 所示。

6.2 结果和分析

实验通过对加载不同噪声的信号在 MATLAB 环境下进行仿真。图 4~6 分别为加入 White、Pink、Babble 噪声的声音信号、经过本文方法处理后的输出结果。

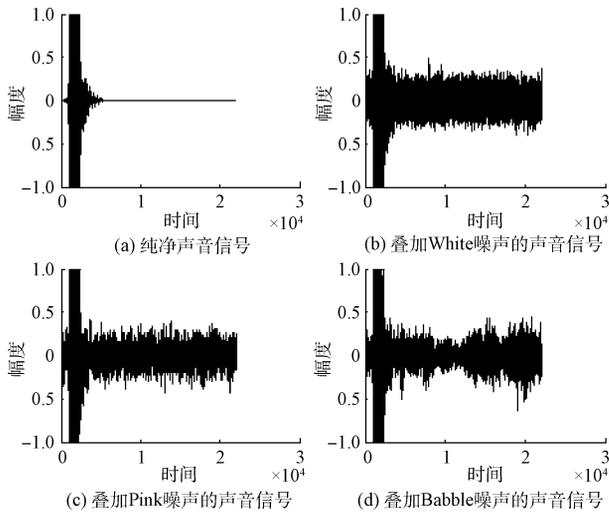


图 3 声音信号波形

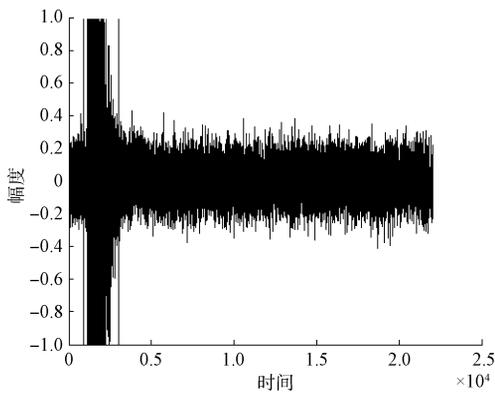


图 4 含高斯白噪声的声音信号端点检测结果

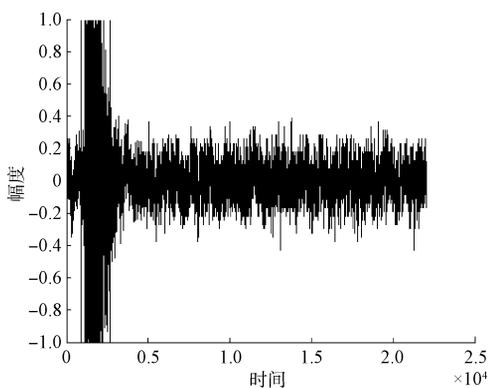


图 5 含 Pink 噪声的声音信号端点检测结果

为了更好地分析 DWT-MFCC 特征参数在声音信号端点检测中的性能,同时对小波系数能量、MFCC 相似度和短时能量法进行实验,作为 DWT-MFCC 参数的对比实验,在不同的信噪比下,比较各种算法的端点检测识别率。表 1 列出了 3 种端点检测算法的检测结果。

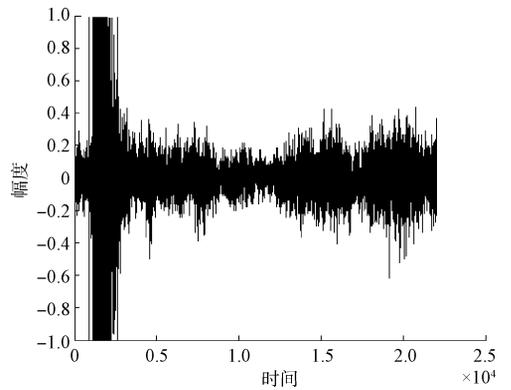


图 6 含 Babble 噪声的声音信号端点检测结果

表 1 几种算法对声音信号端点检测识别率(%)

SNR/dB		-5	0	5	10	无噪
改进方法	White	88.3	90.1	90.8	93.1	
	Pink	86.8	89.4	90.5	92.1	95.9
	Babble	82.4	87.5	89.7	91.6	
小波能量	White	88.7	92.6	93.1	94.5	
	Pink	72.7	86.5	92.6	93.5	94.5
	Babble	64.7	72.7	89.3	92.8	
MFCC	White	82.8	82.8	90.3	89.4	
	Pink	87.2	83.2	90.5	91.2	91.8
	Babble	79.1	76.6	83.9	87.5	
短时能量	White	17.6	17.6	17.6	92.8	
	Pink	17.6	17.6	17.6	91.1	93.4
	Babble	17.6	17.6	79.3	91.4	

如表 1 端点检测结果所示,信噪比较高时,3 种端点检测算法都具有良好的检测性能,但当噪声不断加大时,几种不同的端点检测算法的性能下降也有所不同,短时能量法已经不能达到实际运用的要求了。对于白噪声,小波能量算法在信噪比较低的情况下也能得到较好的检测率,但对于非平稳的噪声 Pink、Babble 则在信噪比较低的情况下就无法达到较好的识别率。MFCC 算法对于各种噪声在低信噪比的情况下都能得到较好的效果,特别是对于 Pink 和 Babble 这种突发性和有色噪声情况下都能达到 80% 左右的识别率。MFCC 虽然具有较强的稳定性以及抗噪性,然而其识别率并不高,特别是当信噪比提高后未有显著提升。DWT-MFCC 算法在 MFCC 算法上加入了小波变换,相比 MFCC 算法,识别率上有了比较大的提升,而在对 pink 和 babble 这种突发性和有色噪声的识别上比单独做小波变换的识别率高,从而说明了 DWT-MFCC 算法有更好的识别效果。

7 结 论

本文研究了小波变换和 MFCC 在端点检测上的应用,

在 MFCC 特征提取的基础上,加上了小波变换,提出了基于小波变换的 MFCC 特征参数(即 DWT-MFCC)。这种参数与 MFCC 及小波子带能量这两种参数相比具有更好的鲁棒性和适应性,对于不同的噪声,在低信噪比的条件下 DWT-MFCC 都能维持很高的识别率,这一特性使得该端点检测算法在强噪声环境下应用成为了可能,并可很好地应用到风机叶片声振法检测中。

参考文献

- [1] 邬冠华,林俊明,任吉林,等. 声振检测方法的发展[J]. 无损检测,2011(2):35-41.
- [2] 黎林,朱军. 基于小波分析与神经网络的语音端点检测研究[J]. 电子测量与仪器学报,2013,27(6):528-534.
- [3] 张德祥,吴小培,吕钊,等. 基于经验模态分解和 Teager 峭度的语音端点检测[J]. 仪器仪表学报,2010,31(3):493-499.
- [4] 董胡. 倒谱距离和短时能量的语音端点检测方法研究[J]. 计算机技术与发展,2014(7):77-79.
- [5] 李战明,尚丰. 一种基于谱熵的语音端点检测方法[J]. 电子技术与软件工程,2015(1):200-202.
- [6] 丁宁,沈二波,杨红卫. 小麦音频信号预处理方法的研究[J]. 农机化研究,2010(8):158-161.
- [7] 李荣荣,胡昌奎,余娟. 基于谱熵的语音端点检测算法改进研究[J]. 武汉理工大学学报,2013,35(7):134-139.
- [8] DI W U, ZHAO H, HUANG C, et al. Speech endpoint detection in low-SNRs environment based on perception spectrogram structure boundary parameter[J]. Chinese Journal of Acoustics, 2014(4).
- [9] 朱恒军,于泓博,王发智. 小波分析和支持向量机相融合的语音端点检测算法[J]. 计算机科学,2012,39(6):244-246.
- [10] 朱坚民,张雷,翟东婷,等. 基于声音多特征贝叶斯网络融合的话者识别研究[J]. 仪器仪表学报,2013,34(9):2058-2067.
- [11] 王宏志,徐玉超,李美静. 基于 Mel 频率倒谱参数相似度的语音端点检测算法[J]. 吉林大学学报:工学版,2012,42(5):1331-1335.
- [12] ZHU L Q, ZHANG Z. Automatic recognition of insect sounds using MFCC and GMM [J]. Acta Entomologica Sinica, 2012, 55(4):466-471.
- [13] 丁世飞,齐丙娟,谭红艳. 支持向量机理论与算法研究综述[J]. 电子科技大学学报,2011(1):2-10.

作者简介

张芝旖,工学硕士,主要研究方向为信号处理。

E-mail: zhiyi199201@hotmail.com

姚恩涛,教授,主要研究方向为传感器技术。

石玉,讲师,主要研究方向为信号处理、优化算法。