改进的音高识别算法

翟景曈,王 玲,杜秀伟
ZHAI Jing-tong,WANG Ling,DU Xiu-wei

湖南大学 电气与信息工程学院,长沙 410082
College of Electrical and Information Engineering,Hunan University,Changsha 410082,China
E-mail:zjt19840207@163.com

ZHAI Jing-tong,WANG Ling,DU Xiu-wei.Improved method of pitch recognition.Computer Engineering and Applications,2009,45(20):228-230.

Abstract: An improved pitch recognition algorithm based on the relevant processing and Fast Fourier Transform(FFT) is proposed.The music signal is preprocessed by the use of the three-level center clipping function.Then the pitch is estimated by using the related processing method.A filter is designed by using the parameters which are aroused from the estimated pitch value.After filtering the music signal the precise frequency result can be got by using FFT method.The algorithm has higher recognition accuracy than the traditional time domain processing method,and more effectively recognizes the pitch which is rich of harmonic than the traditional harmonic peaks method,and has less amount of computing than the wavelet algorithm.The experimental simulation proves that the method is feasible,fast and reliable.

Key words: recognition of pitch;fundamental tone detection;Fast Fourier Transform(FFT);endpoint detection

摘 要:提出了一种基于自相关处理和快速傅里叶变换(FFT)的改进的单音音高识别算法。利用修正的三电平中心削波函数对音乐信号进行预处理,再自相关处理估计基音周期,以估计周期为参数设计滤波器,音乐信号滤波后用FFT实现频域的准确定位。该算法比传统的时域处理法具有更高的识别精度,能比谐波峰值法更有效地解决谐波丰富、基频分量小的信号的识别,且运算量比小波算法小。经实验仿真验证该方法可行、快速可靠。

关键词:音高识别;基音检测;快速傅里叶变换;端点检测
DOI:10.3778/j.issn.1002-8331.2009.20.066 **文章编号:**1002-8331(2009)20-0228-03 **文献标识码:**A **中图分类号:**TP391.42

1 引言

音乐识别在音乐数据库检索技术、乐器调音技术和自动记谱技术中有很重要的应用价值,其重点是音高识别。目前有很多音高识别算法,例如并行处理法、谐波峰值法和小波分析法。但并行处理法的识别精度一般不高,谐波峰值法对于基频分量小、谐波非常丰富的信号识别比较困难,而小波分析法的计算量一般很大。本文通过深入研究音乐信号的物理特性和音乐特性,提出了一种改进的音高识别方法。

2 音高识别原理及常用识别算法

2.1 音高识别原理

音乐的音高(音调)是指声音的高低,主要是由信号的基频决定,信号基频越小,音调越低;信号基频越大,音调越高。对音乐信号的音高的识别就是检测信号的基音[1]。实际上音高与声音的频率并不成正比关系而是近似成对数关系,但音高与频率有一一对应的关系,如表1所示为小字一组各音高的音名与频率的对应关系。音乐信号的音高识别,其实质就是音乐信号的基音频率的检测。

表1 小字一组音阶与频率的对应关系

音阶	简谱	频率/Hz
C	1	261.6
D	2	293.7
E	3	329.6
F	4	349.2
G	5	392.0
A	6	440.0
B	7	493.9

2.2 几种常用的音高识别算法

基音检测的方法大致可分为:时域处理法和变换域处理法。时域处理法是直接由语音波形来估计,分析出波形上的周期峰值。这类方法包括并行处理法(PPROC)、数据减少法(DARD)和短时自相关法等。变换域处理法是将语音信号变换到其他空间域来分析提取基音。这类方法包括谐波峰值法、倒谱法和小波分析法。常用的单音音高识别算法有:并行处理法、谐波峰值法和小波分析法。

2.2.1 并行处理法

并行处理法的实现框图[2]如图1所示。

作者简介:翟景曈(1984-),男,硕士研究生,主要研究方向:图像、声音信息处理与传输技术;王玲(1962-),女,教授,主要研究方向:现代网络与通信技术;杜秀伟(1982-),男,硕士研究生,主要研究方向:模式识别。

收稿日期:2008-04-17 **修回日期:**2008-07-23

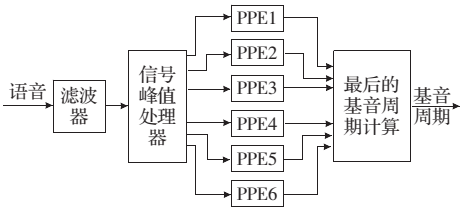


图 1　并行处理法的基音检测框图

音乐信号在经过预处理后，由峰值处理器找出峰点和谷点,再根据位置和幅度产生 6 个脉序列,然后由 6 个并行的基音周期估计器(PPE)估计基音周期。最后,对这些基音估计器的输出作逻辑组合,得出估计值。

并行处理法是时域处理法的典型代表，其优点是运算简单、硬件容易实现,但很多情况下识别误差会比较大,甚至会大于 100 音分。音分是音乐领域用来表征音高差值的概念。一个半音的差值用 100 音分度量。当识别误差大于 100 音分时,识别结果即为错误结果。音分的计算公式为:d=3 986*$\log\frac{f_0}{f}$,其中 f_0 为识别音高,f 为待测音高。

2.2.2　谐波峰值法

谐波峰值法是基于快速傅里叶变换(FFT)的分析法,将信号通过 FFT 变换得到离散的频率谱，最大峰值对应于基音频率。但在实际中,基音峰值的幅度不一定是最大的,因此需要提取峰值频率的最大公因子作为基频。文献[3]中提出了一种方便的基于置信度的方法:由于基音的幅度不一定最大,但幅度大的分量一定是基音的某次谐波(一般不大于 5)或基音本身,因此可对最大峰值频率求出 1 到 5 倍因子作为候选基音,然后对每个候选基音的 n 个谐波的幅度求和,和值最大的候选基音的置信度最大,继而是基音的可能性最大。公式为:

$$L(N)=f_p/N,1\leqslant N\leqslant 5 \tag{1}$$

$$B(N)=\sum_{i=1}^{n}P(i) \tag{2}$$

其中 $L(N)$ 为候选基音,f_p 为最大峰值频率,$B(N)$ 为置信度,$P(i)$ 为某次谐波的幅度,n 为谐波的个数。

由于音乐信号的频宽较大，对于音高跨度较大的乐曲,如果乐器谐波比较丰富,采用谐波峰值法就很有可能把二次甚至三次谐波误定为音高。

2.2.3　小波分析法

小波具有良好的时频特性,能很好地调节时域和频域的分辨率[4]。做基音检测时,小波变换相当于一个中心频率和带宽可调的滤波器,压扩因子 j 每增加一次,小波函数的中心频率和带宽便缩小一半,每经一次变换,高频谐波部分就被滤去一半,而基音部分被保存下来,变换后的波形也越来越“纯”,最终可确定基音。

小波分析法能够很好地提取基音,但小波分解的运算量很大,尤其当需要较大的分解尺度时,所需运算量更加庞大[5]。

3　改进的音高识别算法

本文的改进算法是:先用时域处理法对音乐信号进行基音频率估计，然后以得到的基音频率为参数设计数字低通滤波器。音乐信号经滤波器滤波后,滤除了高频分量,相当于减小了音乐信号的频宽，排除了谐波成分的干扰。在此基础上再做 FFT 就可得到准确的音高。算法流程如图 2 所示。

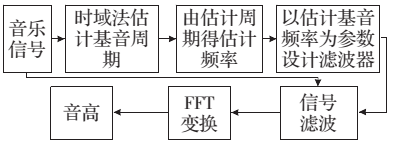


图 2　算法过程

(1)时域法估计基音周期采用自相关处理法。浊音信号的自相关函数在基音周期的整数倍位置上出现峰值,因此检测峰值的位置就可提取基音周期值。由于短时自相关函数中保留的幅度太多,有许多峰值,可能被检测出来的峰值就会偏离原来峰值的真实位置。因此需要采取中心削波对音乐信号进行预处理来改善自相关函数的性能。计算自相关函数要进行乘法运算,其运算量还是比较大的,为进一步减少运算量可对中心削波函数进行修正,采用三电平中心削波的方法,其输入输出函数为[2]:

$$y(n)=C'[x(n)]=\begin{cases}1, x(n)>C_L\\ 0, |x(n)|\leqslant C_L\\ -1, x(n)<-C_L\end{cases} \tag{3}$$

即削波器的输出在 $x(n)>C_L$ 时为 1,$x(n)<-C_L$ 时为-1,除此以外均为零。这使得大多数次要的峰被滤除掉了,而只保留了明显显示周期性的峰。

三电平中心削波的自相关函数的计算很简单,设 $y(n)$ 表示削波器的输出,则由自相关函数直接计算的公式为:

$$R_n(k)=\sum_{m=0}^{N-1-k}[y(n+m)w'(m)][y(n+m+k)w'(m+k)] \tag{4}$$

上式中 $w'(m)$ 为窗函数经削波器的输出。如果窗口采用矩形窗,则上式变为:

$$R_n(k)=\sum_{m=0}^{N-1-k}[y(n+m)][y(n+m+k)] \tag{5}$$

上式中 $y(n+m)$、$y(n+m+k)$ 的取值只有-1、0、1 三种情况,因而不需作乘法运算而只需要简单的组合逻辑即可，运算量大大减小。

削波电平由音乐信号的峰值幅度来决定,一般取音符音乐段最大幅度 $A_{\max}$ 的一个固定百分数。但如果 $A_{\max}$ 为杂波干扰成分,这样选取削波电平会损失掉基音信息。可以采用下式计算削波电平:

$$threshold=\partial(\sum_{i=1}^{n}A_i)/n \tag{6}$$

即前 n 个大的幅值的平均值的一个固定百分数。上式中 A 为前 n 个大的峰值的幅度,∂ 为一个固定的百分数,是一经验值。

一般当 $R_n(k)$ 为最大值时的 k 值就是基音周期,但由于谐波峰值的干扰,很有可能 $R_n(k)$ 为最大值时的 k 值为谐波周期,这样就会将谐波频率误认为基音频率。为了防止这种基音误取情况的发生,可对自相关系数 $R_n(k)$ 再进行一次三电平削波、自相关处理,此时得到的自相关系数 $R_n'(k')$ 为最大值时的 k' 值即为基音周期。这是因为第一次自相关处理得到的 $R_n(k)$ 也是以基音周期为周期的序列,在基音及各谐波位置上的幅值比较大,对 $R_n(k)$ 做一次自相关处理可进一步排除谐波的干扰,得到准确的基音周期。基音频率估值可由 $f_0'=f_s/k'$ 计算,f_s 为采样频率。

(2)当得到基音频率估值 f_0' 后,就需要以 f_0' 为参数来设计数字低通滤波器。设计滤波器的目的是为了在排除高频谐波分

量的干扰和减小音乐信号的频宽的基础上用 FFT 来准确定位基音频率,因此可以 f_0' 为通带频率、$2f_0'$ 为阻带频率、通带衰减 3 dB、阻带衰减 20 dB 为技术指标采用双线性变换法设计数字低通滤波器。

(3)音乐信号经滤波后,通过快速傅里叶变换 FFT 后就可得到无谐波干扰的理想频谱,最大峰值即为基音,由 $f_0=f_s/N\cdot K$ 即可计算出准确的基音频率 f_0。N 为 FFT 变换点数,f_s 为采样频率,K 为最大峰值的位置。需要注意的是,当时域估计的基音值较小,即待测音符为低音时,需要通过对原始数据进行抽取来降低采样频率 f_s,进而改善频域分辨率 f_s/N 以达到识别精度要求。

(4)对连续音乐片段进行识别时,需要分割音符,将连续的音乐片段分割成单个音符,映射到原始数据中再利用上面的算法实现对单个音符的识别。分割音符可采用端点检测的方法:首先将音乐片段信号进行分帧,设置帧长为 m,帧移为 n,然后计算每帧的短时能量。假定音乐信号开始前 N 帧信号为纯背景噪声,初始门限值 H 可以定为该 N 帧信号短时能量值的几何平均值的 a 倍,a 取 1 至 1.5 之间的数。如果当前帧判定为噪声,那么门限的更新采用下述方法:

$$H_i=\beta H_{i-1}+(1-\beta)aN_{i-buffer} \tag{7}$$

式中 β 的是一个经验值取 0.9,H_{i-1} 为前一音符的门限值,$aN_{i-buffer}$ 当前 N 帧信号短时能量值的几何平均值的 a 倍。

音符分割的流程为:

(1)计算前 10 帧的短时能量的平均值得到初始门限值。

(2)若当前帧的短时能量超过了前面计算出的门限值,且此后连续 5 帧都超过了该门限,即认为该位置为音符的起始点,否则认为处于过渡带。

(3)若当前帧的短时能量低于门限值,且此后连续 5 帧都低于该门限,即认为该位置为音符的终止点,否则认为处于有音区。

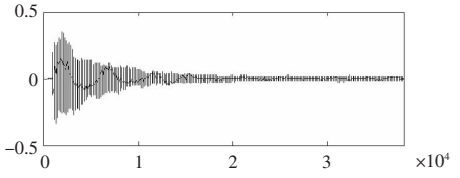
(4)由式(7)重新计算门限。

(5)回到第(2)步,直到整个音乐信号结束。

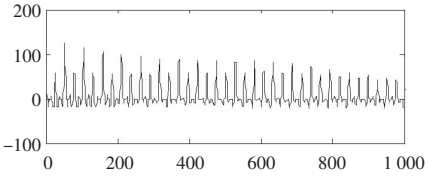
4　实验测试

4.1　单个音符识别

(1)图 3(a)是标准频率为 207.7 Hz 的小字组钢琴 #G 音符的原始波形,经过三电平削波、自相关处理后的波形如图 3(b)所示。



(a)原始信号波形



(b)削波、自相关处理后波形

图 3　三电平削波、自相关处理

(2)由上步得到的基音估计为 211 Hz,所以用 211 Hz 为通带频率和 422 Hz 为阻带频率设计数字低通滤波器,频率响应如图 4。

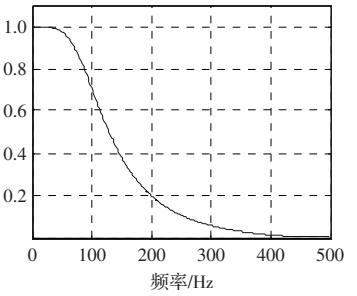
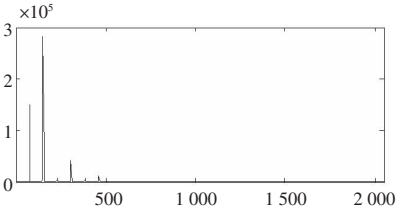
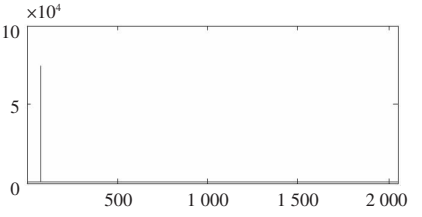


图 4　滤波器的频率响应

(3)#G 音符信号直接 FFT 变换的频谱如图 5(a)所示,明显看出基频分量小于二次谐波分量。当经(2)设计的滤波器滤波后再做 FFT 得到的频谱则为理想频谱,如图 5(b)所示,频谱中只剩基频分量。最终得到音高值为 207.26 Hz。



(a)直接 FFT 的频谱

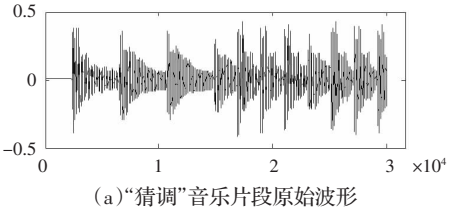


(b)先滤波后再 FFT 的频谱

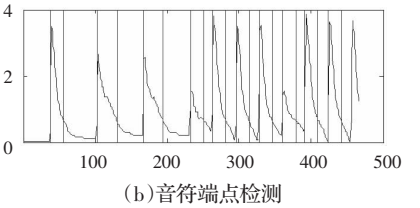
图 5　FFT 频谱图对比

4.2　音乐片段识别

由 MIDI 音频器产生的乐曲“猜调”开始的 10 个单音符的音乐片段的原始波形如图 6(a)所示,音符分割结果如图 6(b)所示,用本文提出的音高识别算法识别的结果如表 2 所示。



(a)“猜调”音乐片段原始波形



(b)音符端点检测

图 6　音符分割

表 2 中第 1 栏为测试音乐片段中前 10 个音符的序号,第 2 栏为各音符对应的频率值,第 3 栏为识别结果,第 4 栏为识

(下转 242 页)