

乐音音符基频识别

吴晶晶 孟玲玲

(燕山大学信息科学与工程学院实验室 秦皇岛 066004)

摘要: 音乐识别的研究在很早就开始了,但识别系统的性能仍未达到满意的效果。通过研究乐音的声音和物理特性,该文提出一种短时自相关基音检测算法。实验证明,利用该文提出的方案,对于传统的自相关基音检测算法,可以提高基音频率提取的准确度,取得了较好的识别效果。

关键词: 音乐识别;基音频率;自相关

中图分类号: TP39 **文献标识码:** A

The recognition of the musical pitch

Wu Jingjing Meng Lingling

(Laboratory, Department of information science and engineering, Yanshan University, Qinhuangdao 066004)

Abstract: The research for music recognition has started long ago, yet recently the performance has not been satisfied. The thesis investigates the musical and physical characters of the musical sound, and advance an short-time auto-correlation pitch detection algorithms. The performance of this system shows that this solution greatly improves the precision of pitch detection toward traditionary auto-correlation pitch detection algorithms, and acquires better results.

Keywords: music recognition; pitch; auto - correlation

0 引言

音乐识别的研究早在20世纪70年代就开始了。第一个识别系统出现在1975年。1996年的系统已经能够处理较复杂的钢琴作品。计算机的出现推动了现代电子音乐技术的发展。计算机多媒体技术和信号处理技术的日新月异和其在音乐欣赏与创作领域的渗入,使以电子音乐为代表的现代音乐技术飞速发展,而且技术革新的领域也在逐渐扩大。基于计算机、电子技术的音乐识别、分类、特征提取等一系列问题,被越来越多研究人员关注。

1 乐音的音乐特性

1.1 预处理

钢琴乐音由88个音符($A_2 \sim c^5$)构成,其中包含7个完整的八度音程^[1]($C_1 \sim b_4$),另外在最低频段有3个音($A_2 \sim B_2$),最高频有1个音 c^5 。

钢琴是用十二平均律制成的,平均律音程是音律测定与计算中的数学框架。十二平均律是按照 $\frac{1}{12}$ 倍频程来划分音阶的。每八度音为一个倍频程,将一个倍频程划分为12个半度音阶,每两个相邻的半度音阶的频率之比为2的

$\frac{1}{12}$ 次方;按照十二平均律,钢琴的个音符对应基音频率范围为:(27.5 Hz(A_2)~4186.00 Hz(c^5)).其基音频率数值决定了音符的音高。

原始乐音中存在两种性质的噪声:工频噪声和高频白噪声。使用一阶高通滤波器工频干扰,并提升高频部分;白噪声则可以直接通过低通滤波去除。在实际乐音音频采集(录制)中可能还会混入话筒撞击等短时高幅噪声,这可以通过文中提出的尖噪声消除方法去除。然后进行端点检测和单音分割,设置帧长为 m ,分成 k 帧,逐帧求出其

$$\text{短时能量: } E(i) = \sum_{n=1}^m x_i^2(n)$$

考虑到短时能量对高电平(乐音)信号敏感,为了区分静音和乐音段可以采用短时平均幅度: $E(i) = \sum_{n=1}^m |x_i(n)|$

短时过零率反映了乐音信号的频率情况,以采样频率 f_s 对频率为 f_0 的正弦信号采样,其过零率为 $\frac{2f_0}{f_s}$ 。短时过

$$\text{零率定义为: } ZCR(i) = \sum_{n=1}^m |x_i(n) - x_i(n+1)|$$

实际应用中,为了判断方便,可以设置门限值 δ ,当前后两个采样信号值之差大于门限值或改变符号时,分别将短时能量和过零率进行归一处理:

$$E(i) = \begin{cases} 1, \sum_{n=1}^m |x_i(n)| \geq \delta \\ 0, \sum_{n=1}^m |x_i(n)| < \delta \end{cases}$$

$$ZCR[i] = \begin{cases} 1, x(i) * x(i+1) \leq 0 \\ 0, x(i) * x(i+1) > 0 \end{cases}$$

在静音段有可能出现尖噪声,所以要设法防止将尖峰噪声误判为端点,此时必须再检测本帧后连续 t 帧的情况,如连续 t 帧均有乐音,则可确定本帧为端点开始,否则是短时尖峰噪声;同样在乐音段中出现短时极小值则必须再持续检测本帧后连续 t 帧的情况,如连续 t 帧均无乐音,方可确定本帧为端点结束。

将确定的端点位置数据映射到乐音时间序列中就是单音端点位置。用这种方法检测端点计算量不大,检测的效果很明显。

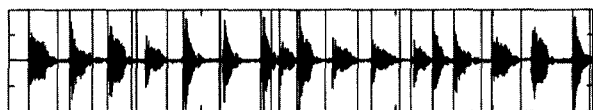


图1 单音符端点检测

1.2 短时自相关

一个时间离散确定信号的自相关函数定义为:

$$\varphi(k) = \sum_{m=-\infty}^{\infty} x(m)x(m+k) \quad (1)$$

如果信号是随机的或周期性的,这时的定义是:

$$\varphi(k) = \lim_{N \rightarrow \infty} \frac{1}{2N+1} \sum_{m=-N}^N x(m)x(m+k) \quad (2)$$

在任一种情况下,信号的自相关函数表示法都是描述信号一定特征的一种方便的方法。检测流程如图2所示。

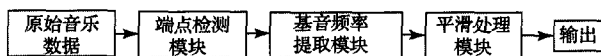


图2 检测流程图

根据周期信号的自相关函数也以同样的周期为周期,自相关函数的其他重要特性:

- 1) 它是一个偶函数,即 $R(k) = R(-k)$ 。
- 2) 在 $k=0$ 达到它的最大值,即所有的 k 来说 $|R(k)| \leq (0)$ 。
- 3) 对确定信号,数量 $R(0)$ 等于能量(式(1)),对随机或周期信号,它等于平均功率。

因此,自相关函数相当于一个特殊情况能量。而更为重要的是它是表示周期性的一个方便方法。如果我们将特性1)和特性2)结合起来一起来考虑式(3),将会看到,样本 $0, \pm P, \pm 2P, \dots$ 周期信号的自相关函数达到最大值。也就是,可以不用考虑信号的起始时间,而借确定自相关函数中的第一个最大值的位置来估计其周期。另外,在音乐信号处理中,计算自相关函数所用的窗口长度与平均能量等情况略有不同。这里 N 至少要大于基音周期的

二倍,否则将找不到第二个最大值点(除 $R(0)$ 以外的最近的一个最大值点)。另一方面, N 值也要尽可能地小,因为语音信号的特性是变化的,如 N 过大将影响短时性。理论上,音符的频率的范围大约在 $20 \sim 20\,000$ Hz。而在检测语音基音周期时,一帧音乐中应包含有2个以上的周期才可检测出基音周期。由于基音周期的范围很宽,所以应使窗宽匹配于预期的基音周期。长基音周期用窄的窗,将得不到预期的基音周期;而短基音周期用宽的窗,自相关函数将对许多个基音周期作平均计算,这是不必要的。

对中音区完整音阶($c1-c2$)采用上述的音符识别方法,对其基音频率如识别结果表2中音阶识别的误差绝对值在0.5%以内,根据十二平均律,连续两个半音之间的频率比为1.05946,即频率偏差率为5.946%,远高于识别的平均误差率,在这个偏差范围以内很容易判断识别单音符的标准频率。准确识别出基频以后,根据系统中的Tones;向量中保存的基频和音名映射表,就可以得到单音符的音名。

表2 中音音阶基音频率识别及误差

音符	标准频率	识别频率	误差率(%)
c^1	261.63	261.30	-0.13
$\#c^1$	277.18	277.33	0.05
d^1	293.66	293.26	-0.14
$\#d^1$	311.13	310.4	-0.23
e^1	329.63	329.94	0.09
$\#f^1$	349.23	348.37	-0.27
f^1	369.99	367.46	-0.68
g^1	392.00	392.43	0.11
$\#g^1$	415.30	415.92	0.15
a^1	440.00	438.67	-0.30
$\#a^1$	466.16	466.10	-0.01
b^1	493.88	494.83	0.19

表3是经过端点检测和单音符识别所得到的波形乐音为乐曲“洋娃娃和小熊跳舞”开始的17个单音符的一段乐音的频率、音名。结果表明,系统正确识别出了乐音中所有单音符的音名,效果比较理想。

表3 乐音音符识别结果

音符序号	识别频率(Hz)	识别音名
1	261.566	e^1
2	293.289	d^1
3	330.756	e^1
4	349.256	f^1
5	391.594	g^1
6	391.054	g^1
7	393.449	g^1
8	350.896	f^1
9	330.194	e^1
10	348.833	f^1
11	348.403	f^1
12	348.451	f^1
13	329.352	e^1
14	291.394	d^1
15	261.047	e^1
16	330.384	e^1
17	392.010	g^1

1.3 中值滤波平滑

中值平滑处理是一种采用滑动窗的直方图统计处理的方法。其基本原理是:设 $\{x(n)\}$ 为输入信号, $\{y(n)\}$ 为中值滤波器的输出窗长为 $2L+1$, 那么, n_0 处的输出值 $y_0(n)$ 就是将窗口的中心移到 n_0 处时窗内输入样点的中值。所谓中值就是将窗内的 $2L+1$ 输入样本 $x(n_0-L)$, $x(n_0-L+1)$, \dots , $x(n_0)$, $x(n_0+1)$, $x(n_0+2)$, \dots , $x(n_0+L)$, 进行统计, 求出一个直方图, 其中 $1/2$ 分位数就是中值。中值滤波可以纠正个别奇异点而不影响周围的样点的值。

2 结 论

本文先通过语音信号的短时能量和过零率来确定语音的端点以及分割音节, 进而通过短时自相关函数法来提取基音周期。在对信号的处理中, 本文运用了低通滤波器进行预处理, 还进行了平滑的后处理。通过对十首曲子进行实验, 获得较好的效果, 但由于用 MATLAB 编程, 运行速度较慢。

参 考 文 献

- [1] 缪天瑞. 律学[M]. 北京: 人民音乐出版社, 2002.
- [2] DAUBECHIES I. Ten Lectures on Wavelets [M]. SIAM Philadelphia, 1992.
- [3] 刘丹, 张乃尧, 朱汉城. 音乐特征识别的研究综述[J]. 计算机工程与应用. 2004.

- [4] DAVID C. Recombinant music using the computer to explore musical style[J]. Computer, 1991; 24(7): 22-27.
- [5] NAGASHIMA K, KAWASHIMA J. Experimental study on arranging music by chaotic neural network [J]. International Journal of Intelligent Systems, 1997; 12: 323-339.
- [6] JOHNSON M L. Toward an expert system for expressive musical performance[J]. Computer, 1991; 24(7): 30-34.
- [7] BOZENA K. Computer-based recognition of musical phrase using the rough-set approse [J]. Intelligent Systems, 1998; 104(1-2): 15-30.
- [8] 李功宇. 弦乐和弦声信号时频特征的小波包分析[J]. 昆明理工大学学报(理工版), 2003.
- [9] 赵瑞珍, 宋国乡. 基音检测的小波快速算法[J]. 电子科技, 1998(1): 16-19.

作 者 简 介

吴晶晶, 1983 年出生, 毕业于燕山大学。

E-mail: jing-2008-jing@163.com



安捷伦科技为其 PXI 数字化仪新增高阻输入模块

(北京, 2009 年 3 月 12 日), 安捷伦科技公司(NYSE: A)日前宣布推出具有增强功能的 PXI 数字化仪, 以扩充其 Acqiris 高速数字化仪产品线。其 8 位和 10 位数字化仪, U1061A 和 U1062A, 通过 配备可选高阻输入模块, 为通信和半导体器件测试等 ATE 应用, 提供理想的信号探头组合。

U1062A 是全球首款 10 位、4 GS/s 采样、3U PXI 数字化仪, 其输入带宽可高达 2 GHz, 采集内存达每通道 256 MS(样点)。U1062A 卓越的性能极大地缩短了数据采集和测试时间, 使其更适合高速 ATE 应用, 而在此类应用中测试时间的瓶颈仅仅在于被测器件(DUT)的速度限制。具有高达 1 GHz 的瞬时带宽的 U1061A, 8 位双通道数字化仪, 可为每通道提供 8 MS(样点)的采集内存, 并在两个通道上实现 1 GS/s 同步采样。在单通道应用中, 这两款数字化仪能够实现通道交错采集: 即可在指定的任何一个通道上通过 ADC 进行交错采集, 并实现高达 2 GS/s 的采样率和 16 MS 存储(U1061A), 以及 4 GS/s 的采样率和 512 MS 存储(U1062A)。

在这两款产品上应用全新高阻前端可进行高压测量。它是通过可编程的前端电子器件实现了一整套输入电压范围: 50 Ω 阻抗端为 50 mV 至 5 V 全量程; 1 M Ω 阻抗端为 50 V 全量程。阻抗端为 50 Ω 时, U1061A 的带宽是 1 GHz, 而 U1062A 的带宽则为 1.4 GHz; 当应用 1 M Ω 阻抗端时, 带宽为 300 MHz。通过应用可切换的滤波器以及超量程信号快速恢复功能, 放大器的响应实现了优化。

U1061A 和 U1062A 数字化仪完全符合 PXI 和 CompactPCI 标准。配合 AcqirisLive 和 AcqirisMAQS 软件以及 Windows®、Linux 和 VxWorks 驱动程序, 使其能够轻松地集成到生产测试设备和自动测试设备(ATE)中。

如欲了解与 Agilent U1061A 和 U1062A PXI 高速数字化仪有关的更多信息, 敬请访问: www.agilent.com/find/high-speed-digitizers。