文章编号:1674 - 2869(2018)02 - 0208 - 06

基于自相关函数的钢琴乐音改进识别算法

刘 莹[1,2],赵彤洲*[1,2],江逸琪[1,2],柴 悦[1,2],李 翔[3]
1. 智能机器人湖北省重点实验室(武汉工程大学),湖北 武汉 430205;
2. 武汉工程大学计算机科学与工程学院,湖北 武汉 430205;
3. 武汉天喻信息产业股份有限公司,湖北 武汉 430223

摘 要:在传统三电平削波结合自相关函数识别算法的基础上,经过准确的音频分割后,提出了帧移法提取乐音基音信号。该算法能在更精细尺度上寻找最大自相关函数,进而准确定位基音位置,较好地解决了传统算法中当乐音节奏较快时,无法区分半频和倍频对基音的影响,从而导致的识别率低的问题。实验表明,本算法对于节奏快慢不同的钢琴乐音的平均识别率约为83.0%,且快节奏乐音的识别率较传统算法高出20.2%,因此该方法对乐音识别尤其对快节奏乐音识别有显著效果。

关键词:基音周期; 三电平中心削波; 自相关函数; 帧移; 乐音识别
中图分类号:TP391.4 **文献标识码**:A **doi**:10. 3969/j. issn. 1674-2869. 2018. 02. 017

Improved Piano Audio Recognition Algorithm Based on Autocorrelation Function

LIU Ying[1,2], *ZHAO Tongzhou**[1,2], *JIANG Yiqi*[1,2], *CHAI Yue*[1,2], *LI Xiang*[3]
1. Hubei Key Laboratory of Intelligent Robot(Wuhan Institute of Technology), Wuhan 430205, China;
2. School of Computer Science & Engineering, Wuhan Institute of Technology, Wuhan 430205, China;
3. Wuhan Tianyu Chengdu Westone Information Industry Inc., Wuhan 430223, China

Abstract: Combined with traditional three-level center clipping method and autocorrelation function recognition algorithm, an improved frame-shift algorithm to extract precisely the pitch signal was presented, which could search the maximum autocorrelation function at a finer scale to accurately locate the pitch position after accurate audio segmentation. This algorithm solved the problem that the traditional algorithms could not distinguish the influence of half-frequency and double-frequency on the pitch with fast rhythm, which degraded the recognition rate. Experiments showed that the improved algorithm had an average recognition rate of 83.0% for piano music with different rhythms, and the recognition rate with fast-paced music was 20.2% higher than that of traditional method. Therefore, the proposed algorithm has a significant improvement on music recognition, especially for fast-tempo music.

Keywords: pitch period; three-level central clipping filter; autocorrelation function; frame-shift; music recognition

钢琴乐音信号是由基音及泛音共同组成的,而决定其音高的是基音,因此基音周期的检测是钢琴音符识别的关键所在[1-2]。基音周期的检测的方法主要包括频域识别和时域识别,短时自相关法是一种经典的时域检测算法,它计算简单,应用广泛,但是该算法会发生基音倍频或半频错误。

收稿日期:2016-12-03
基金项目:国家自然科学基金(61103136);武汉工程大学研究生创新基金(CX2017076)
作者简介:刘 莹,硕士研究生。E-mail:yingtpr@foxmail.com
***通讯作者**:赵彤洲,硕士,副教授。E-mail:zhao_tongzhou@126.com
引文格式:刘莹,赵彤洲,江逸琪,等. 基于自相关函数的钢琴乐音改进识别算法[J]. 武汉工程大学学报,2018,40(2):208-213.

在此基础上，在计算自相关函数前进行三电平中心削波运算是一种经典的改进算法[3-5]。由于该运算去除了各个音符能量相对集中在中心区域的部分，保留了在峰值附近的能量，因而可以减少计算量，加快运算速度，同时，在一定程度上避免上述错误的发生，进而提高识别率，但是，这种算法仍然有一定的局限性。为抑制高次谐波的干扰，文献[6]提出进行两次三电平中心削波和自相关处理，但这种方法增加了计算量，不适于快速计算的应用场景。此外，在用自相关法估计基音周期时，会发生帧间基音周期跳跃的现象，并且识别过程会受到半频点、倍频点和随机错误点的干扰，对于这些问题前期工作者提出了各种平滑滤波算法[7-9]，其目的就是过滤掉各种干扰点。文献[10]提出将插零算法以及相应的低通滤波器应用于三电平削波的自相关法。文献[11]提出将三电平中心削波自相关函数与循环均值幅度差分函数相结合。上述算法在处理节奏较为缓和的乐曲时，可以达到较为满意的识别率，但是在处理节奏较快的乐曲时，识别率会迅速下降。本文提出了帧移自相关函数法，目的是在更小尺度上寻找最大自相关函数，以适应快节奏乐曲，因而一定程度上避免了传统算法对快节奏乐曲的漏检、误检或识别错误等情况发生，进而可以显著提高识别准确率。

1　三电平中心削波的自相关函数

假设 $z_i(x)$ 是乐音信号的时间序列 $w(t)$ 加窗分帧后的第 i 帧信号，其中下标 i 表示第 i 帧，设每帧帧长为 N。$z_i(x)$ 的短时自相关函数定义为：

$$R_i(k)=\sum_{m=1}^{N-m} y_i(m)y'_i(m+k)$$

式中，k 是时间的延迟量。

短时自相关函数具有如下性质：

1）如果 $z_i(x)$ 是周期信号，周期是 P，则 $R_i(k)$ 也是周期信号，且周期相同，即有

$$R_i(R)=R_i(R+P)$$

2）当 $k=0$ 时，短时自相关函数具有最大值，即在延迟量为 0，$\pm P$，$\pm 2P$，… 时，周期信号的自相关函数也达到最大值。

3）短时自相关函数是偶函数，即 $R_i(k)=R_i(-k)$。

短时自相关函数法基音检测的主要原理都是利用短时自相关函数的这些性质，通过比较原始信号与它延迟后的信号之间的类似性来确定基音周期的。如果延迟量等于基音周期，两个信号就具有最大类似性；或是直接找出短时自相关函数的两个最大值间的距离，作为基音周期的初估值。

C_L 是削波电平，中心削波函数 $C[z_i(x)]$ 的数学关系式为：

$$C[z_i(x)]=\begin{cases} z_i(x)-C_L & z_i(x)>C_L \\ 0 & |z_i(x)|\leqslant C_L \\ z_i(x)+C_L & z_i(x)<-C_L \end{cases} \quad (1)$$

三电平中心削波法的输入输出函数为：

$$y'_i(x)=C'[z_i(x)]=\begin{cases} 1 & z_i(x)>C_L \\ 0 & |z_i(x)|\leqslant C_L \\ -1 & z_i(x)<-C_L \end{cases} \quad (2)$$

即削波器的输出 $y'_i(x)$ 在 $z_i(x)>C_L$ 时为 1，$z_i(x)<-C_L$ 时为 −1，其余为 0。

按文献[7]的介绍，C_L 的取法是取 $z_i(x)$ 前部 100 个样点和后部 100 个样点的最大幅度，从中取较小者，并乘以 0.68 作为门限电平 C_L。按式（1）得到的中心削波输出 $y_i(x)$：

$$y_i(x)=C[z_i(x)]=\begin{cases} z_i(x)-C_L & z_i(x)>C_L \\ 0 & |z_i(x)|\leqslant C_L \\ z_i(x)+C_L & z_i(x)<-C_L \end{cases}$$

与按式（2）得到三电平削波输出 $y'_i(x)$ 求自相关函数：

$$R_i(k)=\sum_{m=1}^{N-m} y_i(m)y'_i(m+k) \quad (3)$$

其中，$y'_i(x)$ 的取值只有 1，0 和 −1 三种，所以式（3）中只有加（减）法，在实际运算中节省了大量时间，为实时运算创造了条件。

2　改进的自相关函数基音提取算法

若对音频序列 $w(t)$ 端点检测，每个端点在原序列的起始位置记为 $S(i),(i=1,2,\cdots,n)$。经过准确的音符分割[12-13]后，认为一个端点对应一个基本音符的起始点，设 $T(i)$ 为原音频序列中第 i 个音符的基音周期，由音乐的短时平稳性，将区间 $s(i)=w[S(i),S(i)+l]$ 按定长窗计算自相关函数，其中窗长 l =409 6。理论上认为，第一个最大自相关函数对应的位置，即为基音的周期。由式（3）可得 seg(i) 的自相关函数为：

$$R_i(x),\mathrm{corr}(s(i)),(x=1,2,\cdots,l)$$

取 $R_i(x)$ 最大值 $R_{i,\max}(x)=\max(R_i(x))$，理想状况下，约 70% 音符的数据帧在经过三电平削波和自相关函数计算后的波形图符合如下规律：$R_{i,\max}(x)$ 所在的点 $P_{i,\max}$ 与首个峰值点 $P_i(1)$ 重合，如图 1 所示，首个峰值点 1 与最大峰值点 3 为同一个点，在此情形下可得基音周期 $T(i)=I_i(1)$。

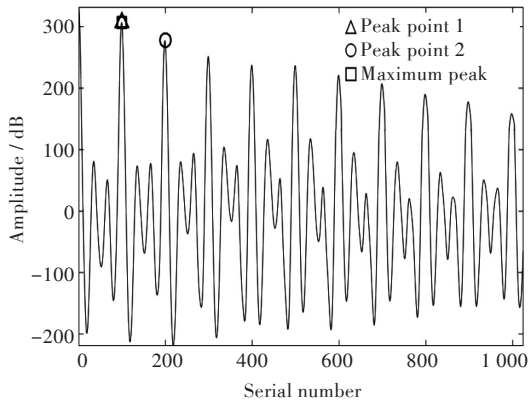

图1 理想状况下自相关函数波形图

Fig. 1 Waveform diagram of autocorrelation function in ideal condition

少数情况下，信号受到共振峰的影响，会出现倍频波干扰，即出现 $P_{i,\max}$ 与 $P_i(1)$ 分离的现象。

为消除此影响，首先需要选取合适的峰值点。图2显示了钢琴曲“致爱丽丝”(For Elise)的第11个音符E4的数据帧在经过三电平削波和自相关函数计算后的波形图，第二个峰值点2与最大峰值点3是同一个点。

设阈值 $H_{i,\min}=R_{i,\max}(x)/k_1$，其中 k_1 为一常量。记录满足条件 $R_i(x)>H_{i,\min}$ 的峰值序列 $P_i(j)$ 与对应于 $R_i(x)$ 的序号 $I_i(j)$，即 $R_i(I_i(j))=P_i(j)$。

k_1 的取值需要保证 $P_i(j)$ 不包含图2中幅值过小的峰值点1，同时也要包含幅值较大且可能正确的峰值点2。因此，峰值点2的幅值为 $P_i(1)$，序号为 $I_i(j)$；峰值点3的幅值为 $P_{i,\max}$，序号为 $I_{i,\max}$。在本文中阈值 $k_1=2$。

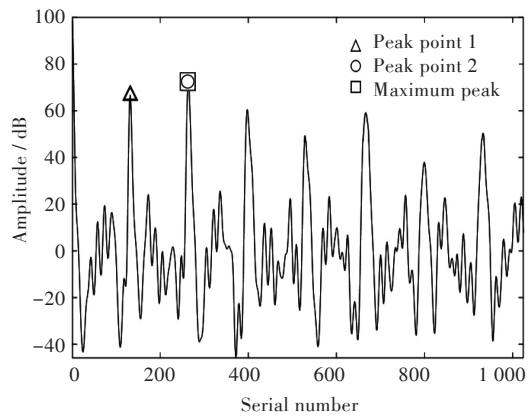

图2 音符E4自相关函数波形图

Fig. 2 Waveform diagram of autocorrelation function of note E4

为选出正确的峰值点，还需要进行进一步阈值判断。取最大峰值点与首个峰值点的幅值比 $C_R=P_{i,\max}/P_i(1)$。图3显示了钢琴曲“梦中的婚礼”(MARIAGE D´AMOUR)的第35个音符D5的数据帧在经过三电平削波和自相关函数计算后，数据帧平移前后波形对比图。

经过计算可得图3(a)、3(b)中的幅值比分别为 $C_{R,1}=1.66$，$C_{R,2}=1.36$，由此可见，当数据帧进行平移后，C_R 的值会发生变化。在本文中，数据帧平移指的是使选取的信号区间上界与下界都增加64，即平移后的信号区间 $s'(i)=w[S(i)+64, S(i)+l+64]$。

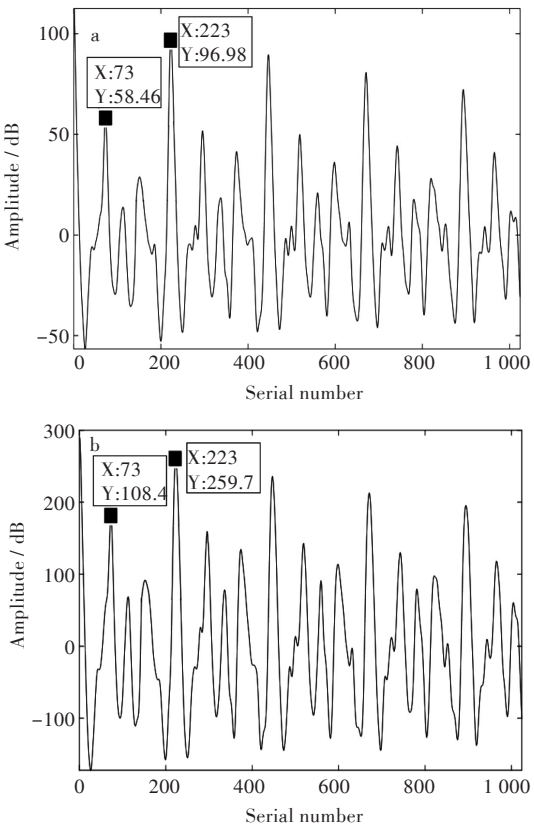

图3 音符D5波形(a)帧移前和(b)帧移后对比图

Fig. 3 Contrast diagrams of (a) before and (b) after frame-shift of note D5 waveform

对数据帧进行多次平移后，可以发现其幅值比在一定范围内波动，如图4所示，将上述数据帧进行8次平移得到幅值比序列 $C_R(b)$。

设阈值 k_2 为一常量，分别统计 $C_R(b)>k_2$ 的个数 n_1 与 $C_R(b)<k_2$ 的个数 n_2，若 $n_2>n_1$，认定 $T(i)=I_i(1)$；若 $n_1>n_2$，则认定 $T(i)=I_{i,\max}$。k_2 的值对统计结果有直接影响，经过多首乐曲的调整，取 $k_2=1.43$ 结果较为理想。

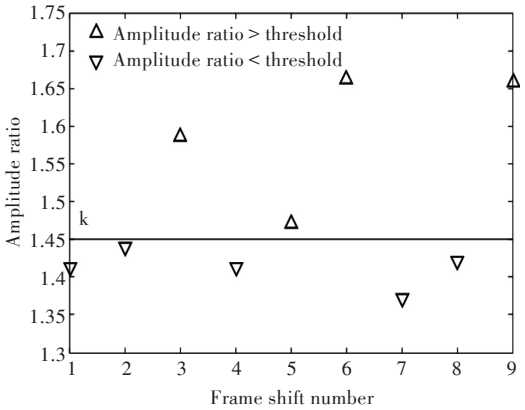

图4　多次帧移的幅值比变化情况

Fig. 4　Amplitude ratio change of multiple frame-shift

3　结果与讨论

所用乐音数据文件由软件EveryonePiano根据曲谱合成，并通过立体声混音内录钢琴曲谱的右手演奏部分得到，其中软件所用音源为mdaPiano。

由音乐基础理论可知，音符 i 的标准频率[14] $F(i)=f_{a^1}*2^{\frac{n}{12}}$，其中 $f_{a^1}=440$ 为第一国际高度，n 为音 i 到音 a^1 间隔的半音数目，当音 i 比音 a^1 低时 n 取负数。若经第二节算法计算得出基音周期为 $T(i)$，则相应的基音频率为 $F'(i)=f_s/T(i)$，其中 f_s 表示乐曲采样频率。音分[15]偏差定义为 $O(i)=\log_k(F'(i)/F(i))$，其中 $k=\sqrt[1200]{2}$。令集合 $U=\{x|-50<x<50\}$，当音分偏差 $O(i)\in U$ 时，认为音符 i 识别正确。

图5为乐曲“致爱丽丝”(For Elise)的前35个音符，用传统三电平削波自相关函数法和改进的自相关函数法识别结果对比图。

以音分偏差作为判定条件，传统识别算法正确率只有77.1%，其中错误主要体现在识别结果为标准频率的一半，如图5(a)所示；而本文提出的帧移法可达到100%，如图5(b)所示。

为检验本文算法的有效性，在根据曲谱合成音乐时，有意识地尽量将一个曲谱按照不同演奏频率合成为变速音乐，目的是检验该算法在低、中、高三种速率条件下的识别率。但是由于合成时EveryonePiano软件本身的快倍速模式只能达到原曲谱速率的2倍，所以，并不是每首音乐都能由慢速合成为快速音乐。

表1~3列出了用传统三电平削波自相关函数法和改进自相关法，作用在更多样本上对识别结果进行对比，对所有乐曲及其变速版本按照速率(音符数/s)分成慢速、中速和快速三组样本。将乐曲的每秒音符数视为其平均速率 v，设定当 $v<3$ 时，

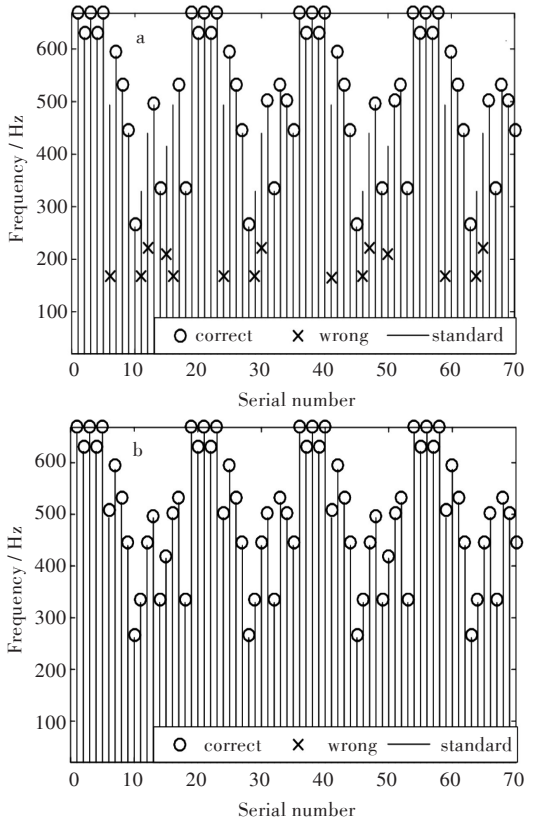

图5　传统自相关法(a)和改进自相关法(b)对“致爱丽丝”前35个音符识别结果对比

Fig. 5　Comparison between (a) traditional autocorrelation and (b)improved autocorrelation method for the recognition rate of the first 35 notes from “For Elise”

乐曲是“慢速”的；当 $3\leqslant v<4$ 时，乐曲是“中速”的；当 $v\geqslant 4$ 时，乐曲是“快速”的。最后，表4对比了慢速、中速和快速三组乐曲的识别结果。

表1~3中的最后一列是本文算法与三电平削波法的相对误差，由表1可知，当乐曲节奏较慢时，两种方法的相对误差率仅在5.1%以内，说明传统三电平削波方法与本文方法识别率接近，但从表2可知，当乐曲节奏较快，两种方法的平均相对误差率在20.6%，改进算法的准确率显然高于传统算法。从表3中可以看出，当乐曲节奏进一步加快时，两种方法的相对误差率更大(平均相对误差率为64%)，尽管本算法在快节奏条件下识别率有所降低，但识别率仍然显著高于传统算法。

此外，同一首曲子的不同速度也会对识别结果有所影响，如：“Faded”、“Faded(1.5倍速)”和“Faded(2倍速)”这三首乐曲在两种方法的识别结果随乐曲速度增大而减小，并且当曲子速度分别为1.9(音符数/s)、3.2(音符数/s)和4.3(音符数/s)时，传统方法识别正确率的下降幅度要大于本文方法的正确率，说明同一首曲子的速度改变对传统方法的影响较大。

表1 慢速乐曲识别结果对比

Tab. 1 Comparisons of recognition results of low speed music

乐曲(节选)	音符数	乐曲时长/s	乐曲速率v/(音符数/秒)	三电平中心削波方法正确率R_1/%	本文方法正确率R_2/%	相对误差率$\lvert R_1-R_2\rvert R_1^{-1}$/%
小星星	42	30	1.4	95.2	100	5
天空之城	56	33	1.7	98.2	100	1.8
Faded	51	26	1.9	96.3	100	3.8
爱的罗曼史	46	21	2.2	95.7	100	4.5
Maps(0.75倍速)	41	19	2.2	95.1	100	5.1
致爱丽丝	70	26	2.7	87.1	91.2	4.7
卡农(0.75倍速)	74	27	2.7	85.5	89	4.1
平均值	54.3	26	2.1	93.3	97.1	4.1

表2 中速乐曲识别结果对比

Tab. 2 Comparisons of recognition results of medium speed music

乐曲(节选)	音符数	乐曲时长/s	乐曲速率v/(音符数/秒)	三电平中心削波方法正确率R_1/%	本文方法正确率R_2/%	相对误差率$\lvert R_1-R_2\rvert R_1^{-1}$/%
梦中的婚礼	106	34	3.1	85.5	100	16.9
野蜂飞舞(0.25倍速)	64	20	3.2	75	87.5	16.7
Faded(1.5倍速)	51	17	3.2	75.5	86.3	14.3
致爱丽丝(1.25倍速)	125	22	3.2	68.5	80	16.8
爱的罗曼史(2倍速)	46	14	3.3	86.9	95.7	10.1
快乐的农夫	44	13	3.3	53.1	79.6	49.9
Maps	41	12	3.4	92.1	97.6	6
卡农	74	20	3.7	64.6	85.9	33
小星星(2倍速)	42	11	3.8	70.5	85.7	21.6
平均值	65.9	18.1	3.4	74.6	88.7	20.6

表3 快速乐曲识别结果对比

Tab. 3 Comparisons of recognition results of high speed music

乐曲(节选)	音符数	乐曲时长/s	乐曲速率v/(音符数/秒)	三电平中心削波方法正确率R_1/%	本文方法正确率R_2/%	相对误差率$\lvert R_1-R_2\rvert R_1^{-1}$/%
快乐的农夫(1.25倍速)	44	11	4	27.2	59.1	117
梦中的婚礼(1.25倍速)	106	25	4.2	68.8	85.8	24.7
Faded(2倍速)	51	12	4.3	56.8	78.8	38.7
Maps(1.5倍速)	41	9	4.5	51.5	65.6	27.3
克罗地亚狂想曲(0.75倍速)	122	27	4.5	25.4	59	132
卡农(1.25倍速)	74	16	4.6	48.2	57.5	19.2
野蜂飞舞	64	11	5.9	45.7	51.3	12.2
克罗地亚狂想曲	122	20	6.1	19.7	47.5	141
平均值	78	16.4	4.8	42.9	63.1	64

而对于相同速度的不同曲子,两种方法在识别结果上均有差异,如表1中每秒音符为2.7的曲子:“致爱丽丝”和“卡农(0.75倍速)”,两种方法的识别正确率不相同,局部节奏较快的乐曲,即“卡农(0.75倍速)”识别正确率较低;甚至有些慢速乐曲的识别结果要比快速乐曲的差,如表3中“梦中的婚礼(1.25倍速)”的每秒音符数为4.2,两种方法的正确率分别为中68.8%和85.8%,而表2中“快乐的农夫”每秒音符数为3.3,但两种方法的识别率仅为53.1%和79.6%,可能原因在于乐曲本身节奏

表4　慢速和中速与快速乐曲识别结果对比

Tab. 4　Comparisons of recognition results of low, medium and high speed music　%

乐曲类型	三电平中心削波方法正确率 R_1	本文方法正确率 R_2	绝对误差率 $\lvert R_1-R_2\rvert$
慢速	93.3	97.1	3.8
中速	74.6	88.7	14.1
快速	42.9	63.1	20.2
平均值	70.3	83.0	12.7

不均匀。如，“梦中的婚礼(1.25倍速)”虽然每秒音符数较高(平均速度高)，但乐曲节奏均匀，节奏最快部分的相邻音符间隔时间为0.206 s，而“快乐的农夫”虽然每秒音符数比较低(平均速度低)，但乐曲节奏不均匀，整个乐曲有快有慢，使得在最快节奏部分的相邻音符间隔时间仅为0.193 s。由于较短的时间间隔会导致前一个音符的谐波尚未充分衰弱，从而对后一个音符的谐波造成干扰。因此对于存在局部节奏快、音符密集的乐曲，不管其平均速度快慢与否，都会影响两种方法的识别正确率。

4　结　语

提出了一种改进的自相关基音周期提取算法，该算法能较好地解决传统识别方法中因为无法明确区分半频或倍频对基频的影响而造成的识别误差，当钢琴乐音节奏较快时($v\geqslant4$)，本文算法平均准确率为63.1%，比三电平削波算法高出20.2%；当钢琴乐音节奏适中时($3\leqslant v<4$)，本文算法平均准确率为88.7%，比三电平削波算法高出14.1%；当乐音节奏较慢时($v<3$)，文本算法平均准确率为97.1%，比三电平削波算法高出3.8%，综合来看，本文所用的算法对以上3组慢、中和快速乐曲的平均识别准确率为83.0%，比传统三电平削波算法高出12.7%。因此本文的算法在快慢节奏不同的钢琴乐音识别中取得了较高的识别准确率，并且对快节奏钢琴乐音的识别准确率有明显的提升。

考虑到周围环境声音以及钢琴弹奏者触键方式的差异，如力度、速度、角度等，这些个体差异会对钢琴音色有影响，进而对识别准确率有一定影响，因此，本算法仅验证了钢琴单键识别算法的有效性和可靠性，没有考虑降噪、双键乐音识别等情况。如果要使本算法有更广泛的适应度，环境噪声的降噪、双键音频分离及自适应阈值的训练是今后需要研究的方向。

参考文献：

[1] 徐国庆，张彦铎，王海晖，等. 乐音旋律识别研究[J]. 武汉工程大学学报，2007，29(2)：60-62.

[2] 徐国庆，张彦铎，王海晖. 基于多分辨分解的乐音水印算法实现[J]. 武汉工程大学学报，2008，30(2)：91-93.

[3] 易克初，田斌，付强. 语音信号处理[M]. 北京：国防工业出版社，2000：62-63.

[4] 吴兴铨，周金治. 基于改进小波变换的语音基音周期检测[J]. 自动化仪表，2017，38(6)：67-70.

[5] 李嘉安娜. 噪声环境下的语音端点检测方法研究[D]. 广州：华南理工大学，2015.

[6] 何晓亮，贾亮，秦文健. 舞蹈机器人中音乐基音频率的提取[J]. 电子设计工程，2011，19(13)：39-45.

[7] 翟景瞳，王玲，杜秀伟. 改进的音高识别算法[J]. 计算机工程与应用，2009，45(20)：228-230.

[8] 马效敏，郑文思，陈琪. 自相关基频提取算法的MATLAB实现[J]. 西北民族大学学报(自然科学版)，2010，31(4)：54-63.

[9] 沈瑜，党建武，王阳萍，等. 加权短时自相关函数的基音周期估计算法[J]. 计算机工程与应用，2012，48(35)：1-6.

[10] 栾极，马太，王飞，等. 插值采样增强钢琴音高识别能力的方法[J]. 数字技术与应用，2014(6)：73-75.

[11] 李嘉安娜. 噪声环境下的语音端点检测方法研究[D]. 广州：华南理工大学，2015.

[12] 冷娇娇，赵彤洲，方晖，等. 基于方差稳定性度量的乐器音频分割算法[J]. 计算机工程与设计，2016，37(3)：768-772.

[13] CARDINAL J，FIORINI S，JORET G. Minimum entropy combinatorial optimization problems [M]. New York：Springer，2012：4-21.

[14] 吴晶晶. 钢琴音乐信号的特征识别[D]. 秦皇岛：燕山大学，2009.

[15] 杨帆，杨杰朝. 基于LabVIEW的频率-音分转换设计[J]. 应用声学，2014(6)：554-559.

本文编辑：陈小平