

一种改进的基音周期提取算法研究

张 雨 高 勇

(四川大学电子信息学院 四川 成都 610065)

摘 要 现有的基音周期提取方法在准确度和稳健性上存在一定的局限性,很难准确地提取出信噪比较低的语音信号的周期。针对这一问题,提出一种基于小波变换的基音周期提取改进算法,将 TEO (Teager Energy Operator)算子与二进小波变换结合起来,首先进行清浊判定,然后只对浊音帧采取多尺度小波分析提取周期,提出用幅度信息和出现概率信息校正二进小波模极大值间距,从而得到准确的周期。计算机仿真结果表明,与传统方法相比,不仅提高了处理效率,而且具有更高的精确度和稳定性。

关键词 基音周期提取 Teager能量算子 二进小波变换 模极大值

RESEARCH ON AN IMPROVED ALGORITHM OF PITCH DETECTION

Zhang Yu Gao Yong

(College of Electronics and Information Engineering, Sichuan University, Chengdu 610065, Sichuan, China)

Abstract Existing methods of pitch detection have their limitation in accuracy and robustness, and is hard to pick up accurate pitches in the condition of low SNR. To solve this problem, an improved algorithm of pitch detection based on dyadic wavelet transform is proposed. This effective method combines Teager Energy Operator with dyadic wavelet transform. Firstly, surd or sonant judgment with each voice frame is done, then the multi-scale wavelet analysis is done only on the sonant frames to get pitches. In order to get accurate pitches, the intervals of local modulus maxima of dyadic wavelet coefficients are corrected according to amplitude information and appearance probability information. The computer simulation results indicate that, the improved method proposed in this paper is more accurate and more robust than the traditional method, and has higher processing efficiency.

Keywords Pitch detection Teager energy operator Dyadic wavelet transform Local modulus maxima

0 引 言

基音周期、清浊信息等是语音信号非常重要的参数,准确地提取基音周期对于高质量的语音分析与合成、识别、语音压缩编码等方面具有重要的意义。混合激励线性预测 (MELP) 声码器算法中采用自相关法,自相关法虽原理简单,但是大量的乘法运算限制了处理效率。倒谱法虽对浊音估计的效果较好,但处理过渡音和含噪语音效果不佳,且算法较复杂。

由于语音信号变化十分复杂,基音周期的动态范围相当大,到目前为止,还没有一个能在任何情况下满意地检测出各种语音基音周期的方法。小波变换在时频域具有超强的分辨能力,是信号分析和处理的一种有力工具。本文将小波变换应用于语音基音周期检测,先通过 TEO 算子对语音信号每帧进行清浊判断,然后对浊音帧进行多尺度小波分析,搜索各尺度下小波系数的模极大值,提出用幅度信息、概率统计方法来修正较低尺度下的模极大值间距,从而准确、快速地提取出最终的基音周期,并通过计算机仿真对真实语音进行了基音周期检测,验证了本文方法的有效性。

1 小波理论基础

小波变换是一种多尺度和多分辨率的分析方法,对应时间

和频率的局部变换,据此可以进行奇异性检测。首先引进信号 $f(t)$ 的卷积型小波变换定义^[1]:

$$Wf(u, s) = f(u) \times_s(u) = \frac{1}{s} \int_{-\infty}^{+\infty} f(t) \left(\frac{u-t}{s}\right) dt \quad (1)$$

其中 $\times_s(t) = \frac{1}{s} \left(\frac{-t}{s}\right)$ 。

若令尺度 $s = 2^j, j \in \mathbb{Z}$, 便得到离散二进小波变换:

$$Wf(u, J) = Wf(u, 2^j) = \frac{1}{2^j} \int_{-\infty}^{+\infty} f(t) \left(\frac{u-t}{2^j}\right) dt \quad (2)$$

所谓信号的奇异性检测是对原信号在不同尺度上进行“磨光”,再检测对磨光后信号的一阶或二阶导数极值点。一般选取高斯函数或样条函数做磨光(平滑)函数,记为 $\phi(t)$, 平滑函数一般要满足:

$$\int_{-\infty}^{+\infty} \phi(t) dt = 1 \text{ 且 } \lim_{|t| \rightarrow \infty} \phi(t) = 0 \quad (3)$$

小波函数 $\psi(t)$ 是一次样条函数 $\phi(t)$ 的一阶导数,也即是二次样条函数。即:

$$\psi(t) = \frac{d}{dt} \phi(t) \quad (4)$$

根据卷积型小波的定义和卷积的性质可得:

收稿日期: 2008-06-12。张雨,硕士生,主研领域:移动通信,语音信号处理等。

$$Wf(u, s) = f(u) \times_s(u) = f(u) \times (s \frac{d}{dt}) = s \frac{d}{dt} (f(u) \times_s(u)) \quad (5)$$

其中 $f(u) \times_s(u)$ 表示的直观几何意义即 $f(t)$ 的尖锐点被 (t) 磨成光滑弧,从而使 $f(t)$ 变成一个光滑函数 $f(u) \times_s(u)$ 。式 (5) 还说明,在某尺度 s 下, $f(t)$ 的小波变换系数 $Wf(u, s)$ 与 $f(u) \times_s(u)$ 的一阶导数成正比,故可以通过检测小波变换的模极大值来确定原始信号的突变位置。也即是说,对 $f(t)$ 作小波变换后,其小波系数的模极大值对应于信号的突变点。将该原理应用到语音信号处理时,即在声门闭合的瞬间,声道受到较强的刺激,在语音信号中引起一个突变。如果能检测出信号的这一突变,那么相邻两次突变的时间间隔就是基音周期。由于信号不连续性在小波变换不同分辨率下具有传递性,可在多个尺度下共同分析基音周期,以确保其准确性。

2 算法原理及实现

本文算法的总流程图如图 1 所示。

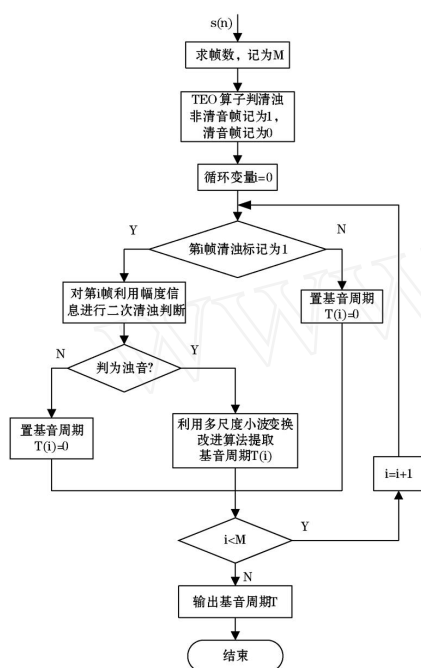


图 1 总流程图

2.1 基于小波变换的清浊判定 (TEO 算子)

传统时域自相关法^[2]先求周期,再进行清浊判定,没有事先排除清音区,不利于提高语音处理速度。TEO是由 Kaiser提出的一种非线性算子。首先计算小波系数的能量,然后用该算子提取能量包络,接着通过选取合适的阈值对信号进行清浊判断,最后对清浊判决结果进行修正。这种方法不仅能提高语音处理效率,而且在信噪比很低的情况下具有很好的鲁棒性。

设 $s(n)$ 为一有限频带的语音信号,则 Teager 能量算子定义为:

$$T[s(n)] = [s(n)]^2 - s(n+1) \times s(n-1) \quad (6)$$

具体实现的步骤:

- 1) 先对 $s(n)$ 进行 4 层小波分解,得到第 4 层小波系数 w 。
- 2) 根据式 (6),计算得到 $e = T[w]$ 。
- 3) 将 e 取模之后,低通滤波得到包络。
- 4) 在小波域,浊音段的小波系数能量比清音段的大,据此选定阈值,然后进行清浊判决。对于 θ 的选取,采用了与文献 [4] 相同的选取方法。

5) 对清浊判决结果进行修正。由于短的清音段不可能出现在连续的浊音段中,比如说判决结果为 ...VVVUVV...,其中 V 表示浊音段, U 表示清音段,那么判决结果应该修正为 VVV-VVV;反之,短的浊音段也不可能出现在连续的清音段中,同理加以修正。

传统 Mallat 塔式算法实现二进小波变换时,每一次分解为高频成分和低频成分,将低频部分数据二倍抽取之后送入下一级再进行分解。虽然二倍抽取时并不丢失有用信息,但随着分解层数的增加,采样点数就越来越少,既不利于观察,而且由于点数的减少会造成判定的清浊区域难以与原语音序列准确对应,故实验中采用冗余二进小波变换。

本文以男音“知识”为例,用改进算法提取清浊音的过程如图 2 所示。为简单起见,在此约定图中及后文出现的 J 均代表二进小波尺度参数。

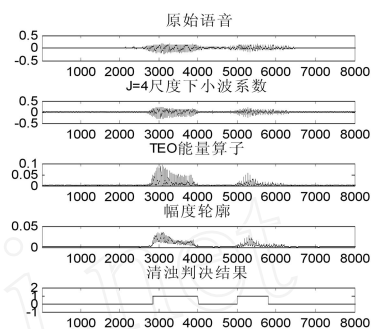


图 2 TEO 算子判定清浊

2.2 基于小波变换基音周期提取原理及实现

经过 TEO 算子的初判清浊,将一段语音加汉明窗每 300 点为一帧,帧间适量重叠,如果初判结果表明该帧为清音,则跳出循环,基音周期直接清零;如果初判结果为浊音或混合区,那么进入二次清浊判决。一般设定阈值 T_0 为整段语音幅度最大值的 40% ~ 60%。若该帧信号幅度均值小于阈值 T_0 ,视为清音,周期值赋零;反之,则为浊音。如果判为浊音即可进行多尺度小波分析以提取基音周期。

本文中采用的子波滤波器和尺度滤波器的参数分别如下:

$$lo_d = [0 \ 0 \ 125 \ 0 \ 375 \ 0 \ 375 \ 0 \ 125 \ 0]$$

$$hi_d = [-0.0061 \ -0.0869 \ -0.5798 \ 0.5798 \ 0.0869 \ 0.0061]$$

多尺度小波分析结果如图 3 所示。

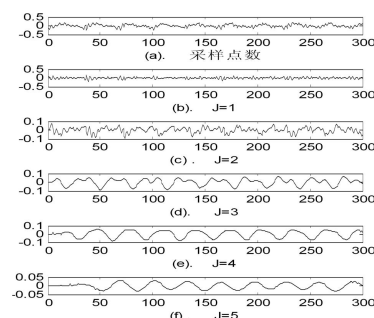


图 3 五个尺度下的小波系数 (J 为尺度参数)

接下来的任务就是从这些小波系数中提取周期。文献 [5] 中提到在尺度参数 J 为 3 时,检测出所有正向局部极大值,依次计算相邻峰值的间隔,比较原则是若本次间隔与前一次间隔差的绝对值小于前一间隔的 10%,则记录该间隔的大小为基音周期候选值,对一帧语音的所有候选值进行平均,得出该尺度下的

基音周期。同理检测 $J=4$ 、 $J=5$ 情况下模极大值的位置, 计算相应尺度下的基音周期, 若三个尺度下周期如果相同, 作为该浊音帧的周期, 若不等判为清音帧。

在实验中, 发现在尺度参数 $J=3$ 时, 按照文献 [5] 的方法, 所有的极大值点均不满足当前间隔与前一次间隔差的绝对值小于前一次间隔 10% 的准则, 提取不出基音周期。

问题在于尺度参数 $J=3$ 的小波系数存在很多虚假极大值点, 在图 4(a) 中用小圆圈加以标明。此时如果按照文献 [5] 的判决原则, 显然不对。

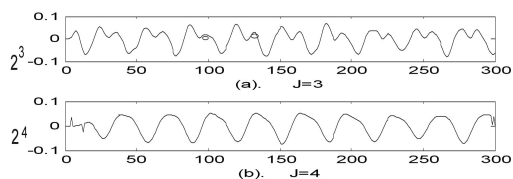


图 4 伪极大值点示例

此外由于加窗造成帧头数据跳变, 对应小波系数产生大量的局部毛刺, 如图 4(b) 所示, 它们的存在导致所求的极大值出现很多虚假点, 进而影响基音周期提取准确性。

2.3 改进方案及其实现

下面将对各个尺度下基音周期提取的改进方案进行详细的叙述。

不难发现, 第 3 层小波系数伪极大值虽然很多, 但伪极小值却很少。可先对其类滤波处理 (保留幅度为负值的小波系数), 然后搜索该尺度下所有的负向极小值, 再计算相邻极小值点的间隔存入 $interval$ 数组, 将其修正之后可确定该尺度下的周期。具体修正算法在后文将被提到。

对于尺度参数 $J=4$ 、 $J=5$ 时, 从图 3(e) - (f) 和图 4(a) 对比可看出, 小波系数伪极大值点较少, 可直接按照改进算法修正, 求出各自尺度下的基音周期。流程如图 5 所示。

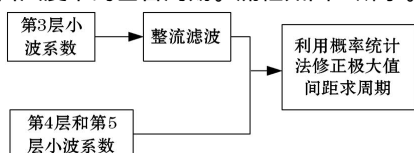


图 5 改进方案流程

此外, 对于因加窗造成小波系数起始位置出现大量局部毛刺的问题, 可以采取与每帧起始位置间隔一定距离的位置开始对小波系数的极大值进行统计的方法加以解决。

理想情况下, $interval$ 数组存放的数据应为基音周期, 但事实上由于伪极大值点的影响必须进行修正。现在以某帧的 $interval$ 数组数据为例说明修正的方法:

$$interval = [5, 7, 2, 31, 6, 25, 30, 32, 29, 32, 32, 33, 31, 8, 23, 31]$$

不难看出, 基音周期大体应在 30 左右, 其中 5, 7, 2 是由帧头毛刺造成的, 而 6 和 8 则是混入的少量伪极大值点造成的。从图 4 看出, 一个伪极大值点混入, 对原相邻极值点间隔产生影响, 将改变 $interval$ 的两个数据, 修正时将这两个数据相加即可, $6 + 25 = 31$; $8 + 23 = 31$, 同时应该注意合并时, 数组下标也对应改变。

经过举例说明和分析结果, 本文从统计学的角度出发, 给出对其进行修正的具体实现步骤如下:

- 1) 设置区间长度 L , 初始值为 5。
- 2) 将帧长等分成区间长度为 L 的段, 结果为: $[0, L-1], [L, 2L-1], \dots$, 统计 $interval[i]$ 在各区间范围内取值出现的

次数, 选取次数最大者对应的区间为基音周期的候选区间。

3) 若上述过程出现多个区间的统计结果相同或相近, 则置 $L = L + 1$, 重复第 2) 步。最终得到候选区间记为:

$$[low_lim, up_lim]$$

4) 对 $\forall i$, 若 $interval[i] < low_lim$, 则 $interval[i] = interval[i] + interval[i+1]$, 直到 $interval[i]$ 值落在候选区间为止。若 $low_lim < interval[i] < up_lim$, 则标记为候选基音周期。对所有候选基音周期求算术平均, 作为当前尺度的基音周期。

3 实验结果与评价

3.1 算法复杂度及效率分析

自相关函数算法的运算单元是乘法运算, 倒谱法的基本运算单元是快速傅里叶变换, 小波变换的计算复杂性介于自相关法和倒谱法之间。由于采取 TEO 算子进行语音清浊判断, 接下来只对浊音进行基音周期提取, 大大减少了运算量, 与传统的时域自相关法相比, 提高了语音信号的处理效率。

3.2 鲁棒性分析

由于浊音信号可认为是由低频正弦信号及其高次谐波叠加而成, 为了验证本算法的鲁棒性, 构造如下信号:

$$s(t) = 5 \sin(2\pi/60 \times t) + \sin(2\pi \times 6/60 \times t) + \sin(2\pi \times 10/60 \times t) \quad (7)$$

它由已知频率为 1/60Hz (周期为 60) 的正弦信号和高次谐波 (频率分别为 1/10 和 1/6) 叠加而成。在式 (7) 基础上加入不同程度的高斯白噪声, 周期提取结果如图 6 所示。

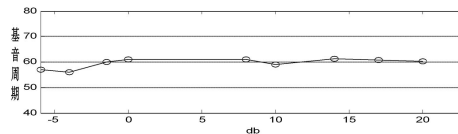


图 6 噪声干扰下的周期提取

从图 6 可以看出, 在噪声干扰下, 提取的周期值均在 60 附近, 新方法仍具有较强的稳定性。

3.3 准确性分析

在男声“知识”这段语音中, 选取 10 帧浊音做准确性分析实验。以 MELP (2.4Kb/s) 声码器算法提取到的基音周期结果作为标准, 将本文方法提取的周期与其进行对比, 结果如图 7 所示。由图可见, 本文提出的改进算法提取的周期与美国联邦标准声码器中提取周期的结果基本吻合, 相对误差较小, 具有很高的准确度。

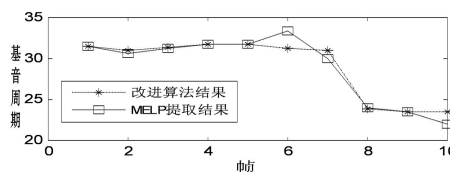


图 7 准确性分析

4 结论

该算法针对传统的小波检测周期准确度低的弊端, 并在其基础上进行改进, 将 TEO 算子判定清浊与二进小波变换提取周期结合起来, 清浊分区之后只对浊音提取基音周期, 提高了处理效率, 并且还使用改进算法对周期值加以修正。从仿真结果看,

取得比较好的效果,具有较高的准确性和较好的稳定性。

参 考 文 献

- [1] 刘涛,曾祥利,等.实用小波分析入门[M].北京:国防工业出版社,2006:59-61.
- [2] 王洪,唐凯.低速率语音编码[M].北京:国防工业出版社,2006:122-125.
- [3] Kaiser J F. On a simple algorithm to calculate the energy of a signal [C]//Albuquerque, USA: Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP '90), 1990: 381-384.
- [4] 陈宁,胡瑛.基于 Teager 能量算子和空域相关函数的基音检测算法[J].微电子学与计算机,2007,24(11):21-23.
- [5] 董婧,赵晓晖,应娜.基于二进小波变换的基音检测算法[J].吉林大学学报:工学版,2006,36(6):978-982.
- [6] 罗雪松.小波分析用于语音信号的基音检测[D].重庆:重庆大学,2002.
- [7] Ugveje O C. Selective noise filtration of image signals using wavelet transform[J]. Measurement, 2004, 36: 279-287.
- [8] Xu Yansun. Wavelet transform domain filters: a spatially selective noise filtration technique[J]. IEEE Transactions on Image Processing, 1994, 3(6): 747-758.
- [9] 王长富,戴蓓倩,李辉,等.基于声门闭合时刻的语音基音周期的提取[J].中国科学技术大学学报,1998,28(3):362-367.

(上接第 259 页)

(4) 1/4 剪切攻击

对含水印图像在 1/4 剪切攻击情况下,NMF 算法与 DCT 算法提取水印及水印相关值的效果比对。NMF 算法提取水印依然清晰且有很高的相关值,而 DCT 算法使数据出现负值,已无法正确显示和辨析提取的水印,如图 6 所示。

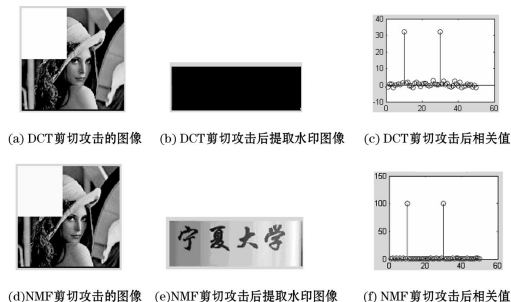


图 6 剪切攻击

(5) JPEG 压缩攻击

对含水印图像在受到 10% 的 JPEG 压缩情况下,NMF 算法与 DCT 算法提取水印及水印相关值的效果比对。两者实验结果基本一致,如图 7 所示。

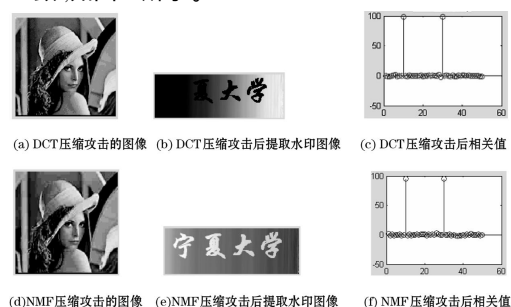


图 7 JPEG 压缩攻击

(6) 45 度旋转攻击

对含水印图像在旋转 45 度情况下,NMF 算法与 DCT 算法提取水印及水印相关值的效果比对。两者实验结果基本一致,如图 8 所示。

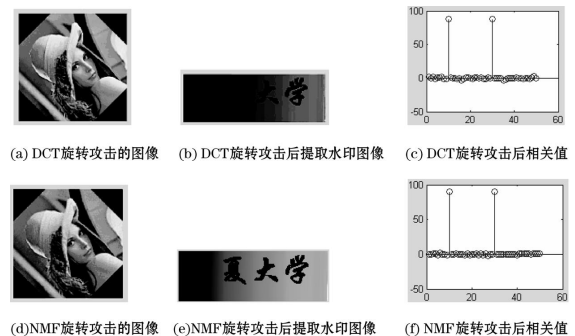


图 8 旋转攻击

由上述客观和主观实验结果可见,DCT 算法与 NMF 算法在随机噪声、JPEG 压缩、图像旋转攻击情况下提取的水印和水印相关值基本相同。而在滤波和剪切攻击情况下,DCT 算法抵抗性略弱,NMF 算法提取水印的清晰度和水印相关值都高于 DCT 算法,表明 NMF 算法具有更好的鲁棒性。

我们认为将 NMF 应用于数字水印算法可以得到两个好处:一是 NMF 的非负性能够克服了其他分解方法在特征空间上投影得到的投影系数可能出现正负相互抵消的情况,增强了算法对数据特征的准确表达;二是 NMF 具有基于部分表示整体的能力,对于一定程度的遮挡和信息丢失比较鲁棒。

4 结 论

本文提出的一种基于非负矩阵分解的水印算法,利用 NMF 表现出的数据非负特性和基于部分表示整体的能力,该算法生成的含水印图像在经过图像剪裁、高斯低通滤波攻击后仍能清晰地从中提取出嵌入的水印图像,体现出独特的优势。总体效果上看,NMF 算法比 DCT 算法有更好的性能,是一项有着发展前景的数字水印嵌入算法。

参 考 文 献

- [1] 周庆松,何礼富.数字水印技术研究[J].电脑知识与技术,2007(20):517-518.
- [2] 曹丽娜,陈亚军,任林政.一种基于 DCT 域的鲁棒水印算法及其 Matlab 实现[J].电脑知识与技术,2007(9):813-814.
- [3] 张勇.数字水印算法研究与实现[D].吉林大学,2005.
- [4] Paatero P, Tapper U. Positive matrix factorization: A non-negative factor model with optimal utilization of error estimates of data values. Environmetrics, 1994(5): 111-126.
- [5] 艾竹君,屈志毅,赵玲,等. NMF 在人脸表情识别中的应用[J].甘肃科技,2006,22(3):127-128.
- [6] 刘维湘,郑南宁,游屈波.非负矩阵分解及其在模式识别中的应用[J].科学通报,2006,51(3):241-250.
- [7] Lee D D, Seung H S. Algorithms for non-negative matrix factorization[J]. Advances in Neural Information Processing Systems, 2001, 13: 556-562.
- [8] Macq B, Quisquater J J. Cryptology for digital TV broadcasting Proc. of the IEEE, 1995, 83(6): 944-957.