基于小波变换的语音基频提取新算法 [1]

钟金宏 杨善林 林逸榕* 鲁 奎

(合肥工业大学计算机网络系统研究所 合肥 230009)
*(合肥工业大学电气工程学院 合肥 230009)

摘 要 该文将小波变换应用于具有连续语音特征的三字词语音的基频提取，并针对实验中出现的问题对算法进行了改进，提出了一种新的基于小波变换的语音基频检测算法。该算法主要包括：离散小波变换计算、基于投票策略的基频点选择和基频起点确定、基频检查、异常点修正、头尾漏点处理以及基于投票策略的基频点精确定位。实验表明，该算法较好地克服了基于小波变换传统算法的不足，更适合于连续语音的基频提取，缺陷是需要较多的计算时间，不太适合于实时性要求较高的系统。

关键词 小波变换，基频检测，投票策略，语音信号
中图号 TN912.3, O177.6

1 引 言

基频是语音信号中运载的重要信息，准确地检测它具有非常重要的意义。目前已有很多基频提取方法，但由于语音信号自身的复杂性，这个问题一直未能得到很好解决。小波变换具有良好的时频局部分析能力，非常适合于探测正常信号中的突变。据此 S. Kadambe 等将小波变换应用到语音基频提取中，并演示了其相对于自相关法和倒谱法的优点 [1]。我国学者将小波变换引入到汉语基频检测 [2,3]，给出了一系列实用的算法。本文在他们工作的基础上，将小波变换应用到具有连续语音特征的三字词语音的基频提取中，并根据实验中出现的问题，对传统算法进行了改进，给出了一种基于小波变换的语音基频检测新算法，获得了较好的实验结果。

2 二进小波变换

设 $\psi(t) \in L^2(R)(L^2(R)$ 表示均方可积一维函数的 Hilbert 空间)，其傅里叶变换为 $\hat{\psi}(\omega)$。当 $\hat{\psi}(\omega)$ 满足完全重构条件

$$C_\psi = \int_R \frac{|\hat{\psi}(\omega)|^2}{|\omega|} \mathrm{d}\omega < \infty \tag{1}$$

称 $\psi(t)$ 为母小波。将母小波 $\psi(t)$ 经二进伸缩和平移后，就可以得到一个小波序列：

$$\psi_{j,k}(t) = 2^{-j/2}\psi(2^{-j}t - k), \qquad j,k \in Z \tag{2}$$

显然 $\psi_{j,k}(t) \in L^2(R)$，若存在 $0 < A < B < \infty$，使 $A \le \sum_{j\in Z} |\hat{\psi}(\omega)|^2 \le B$，则称 $\psi_{j,k}(t)$ 为一个二进小波。对任意函数 $f(t) \in L^2(R)$ 的二进小波变换为

$$W_\psi f(j,k) = < f, \psi_{j,k} > = 2^{-j/2} \int_{-\infty}^{+\infty} f(t)\psi^*(2^{-j}t - k)\mathrm{d}t \tag{3}$$

[1] 2001-02-26 收到，2002-01-08 定稿
博士点基金资助 (98035904)

小波变换一般由符合条件的有限长脉冲响应滤波器 (FIR) 实现，其实现算法为 [4]

$$S_{2^j}f(n)=\sum_{l\in Z}h_l S_{2^{j-1}}f(n-2^{j-1}l) \tag{4}$$

$$W_{2^j}f(n)=\sum_{l\in Z}g_l S_{2^{j-1}}f(n-2^{j-1}l),\ j=1,2,\cdots,J \tag{5}$$

本文采用具有紧支集且有线性相位的正交镜像 FIR 滤波器，其传递函数为 [3]

$$H(\omega)=e^{i\omega/2}[\cos(\omega/2)]^{2n+1} \tag{6}$$

$$G(\omega)=4ie^{i\omega/2}\sin(\omega/2) \tag{7}$$

3 基频检测原理及其存在的问题

当信号中有突变点时，其小波变换将在该点附近表现为局部最大，且对于信号中真正的突变，小波变换时信号的不连续性在不同分辨率层有传递性．人在发声时，由于声门瞬时闭合，对声道形成较强冲击，从而在语音信号中引起一次锐变．通过小波变换可检测出语音信号的这一锐变，即相当于检测出声门闭合时刻，而相邻两次的声门闭合之间的时间长度的倒数即为该处基频，因此求相邻两次突变的时间间隔的倒数就可得到基频 [1] ．

根据文献 [1] 提供的算法，对三字词语音进行基频提取实验．该算法对部分语音得到了较好的结果 (见图 1)，但对大部分语音提取的基频含有较多的错误点，甚至是一堆乱点 (见图 2)．这主要是由于三字词语音受到了声道响应、音联、协同发音和变调规律等的影响．其中受影响较小的语音基频提取效果较好，而受到较大影响的语音基频提取效果较差．因此基于小波变换传统算法无法解决三字词的基频提取问题，更别提连续语音．

通过对出错语音的分帧分析发现：(1) 对于每帧语音的小波变换，一般会在某一尺度上表现出比其它尺度上更好的周期性，即更少的错误点，(见图 3)；(2) 真正的基频点就隐藏在几个尺度上小波变换的局部极大值点中，只是算法没有将其检测出来．因此直观上可以认为是不适合的基频点选择方法造成了基频检测错误．

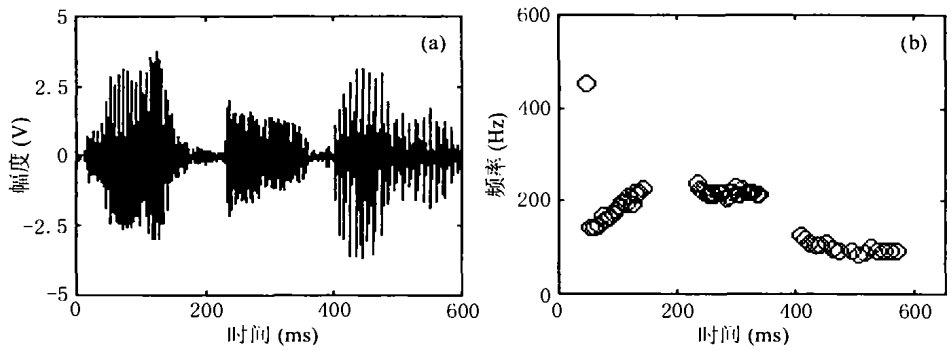

图 1　三字词“卢森堡”的基频曲线

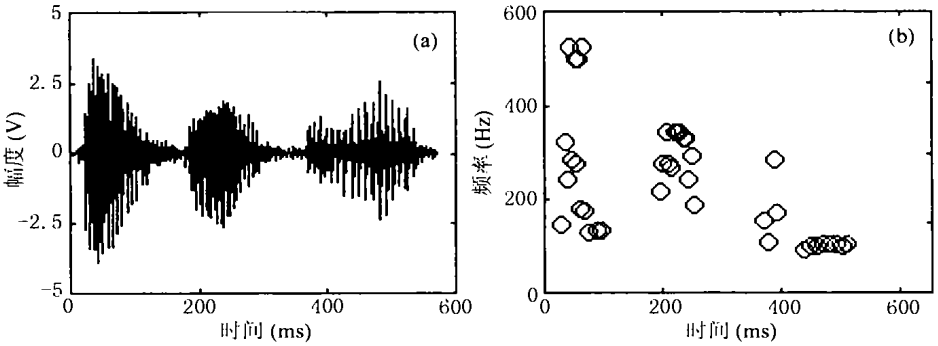

图 2 三字词“党代表”的基频曲线

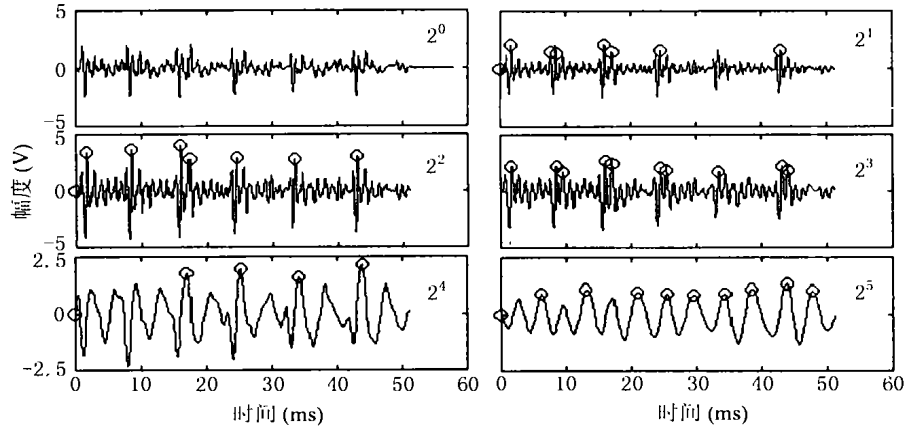

图 3 一帧语音信号的小波变换结果

4 基于小波变换的基频检测新算法

新算法相对于传统算法的改进主要在以下两个方面：(1) 基于投票策略的基频点选择、基频起点确定和基频点的精确定位；(2) 对每帧语音基频增加基频检查和推算过程，以去除错误点、误差点和头尾的漏点现象.

4.1 基频点选择

基于投票策略的基频点选择方法的主要依据是：真正的突变点在小波变换的不同分辨率层具有传递性. 实验中也发现了这一现象，且其在不同尺度上的局部极值点位置相差不大，一般小于 5，如表 1 所示. 因此可将不同尺度上的基频点位置相差小于 5 的基频点认为代表的是同一突变 (基频点). 实验中还观察到真正的基频点一般会在 5 个尺度上至少出现三次. 因此可通过投票策略进行基频点选择，将在 3 个或 3 个以上尺度上都出现的基频点作为真正的基频点.

定义 1 设有两个向量 A 和 B，其中 $A=\{a_1,a_2,\cdots,a_i,\cdots\}$, $B=\{b_1,b_2,\cdots,b_j,\cdot\}$，对于向量 A 中某一元素 a_i，令向量 $E=|a_i-B|=\{|a_i-b_1|,|a_i-b_2|,\cdots,|a_i-b_j|,\cdots\}$, $\{e,k\}=min(E)$，式中 $\min(E)$ 是求取向量中最小元素运算，e 为所求的最小元素的值，k 为该最小元素在向量中的标号，如 $e<5$，则称向量 A 中元素 a_i 与向量 B 中元素 b_k 相同.

在定义 1 中如向量 A 和 B 中元素均为基频点位置，则该定义可作为判断基频是否相同的准则. 根据文献 [2, 3] 和实验观察，在尺度 2，3 和 4 下的小波变换能较好地反映三字词语音

表 1　一帧语音在五个尺度上的基频点

尺度	基频点位置										
1	8	68	126	180		234					
2		68	124	181	202	235	288				
3		67	124	180		234	288			443	
4		65	123	179		232	286	338	390	442	495
5		65	122	176		231	284	336	388	441	493

的基音周期 (见图 3)，因此可设定尺度 2， 3， 4 下检测到的基频点为基频侯选点 (候选人)，并以它们的位置值作为各自的标识，令尺度 1， 2， 3， 4， 5 下检测到的基频点为投票人，以它们的位置值作为各自的投票。

定义 2　设定向量 A 表示尺度 2， 3， 4 下得到的基频点集合，向量 B 为某一尺度下检测到的基频点集合，对于向量 A 中元素 a_i，由定义 1 如向量 B 中存在与 a_i 相同的元素，则称向量 A 中元素 a_i 得到一票，可视为对应基频侯选点得到一票。

定义 3　由定义 2 计算向量 A 中每一元素在各阶尺度上所得的票数，如最后所得总票数大于或等于 3，则称该基频侯选点为真正的基频点。

在定义 2 和 3 中为避免相同或相近的基频侯选点多次参加选举，某一基频侯选点在参加选举前，先与已选出的基频点相比较，如位置值相差小于 18，则认为是同一点，不再参加选举。这主要是考虑三字词基频在 10k 采样时的变化范围为 18~125 采样点 [5]。

4.2 基频检查

基频检查的基本思想是先确定基频的起点，即本帧语音基频的开始点，然后根据语音信号短时稳定的特点，即相邻基频间通常不发生突变，进行基频推算，根据定义 4 进行异常判断，如出现异常，修正异常基频点。

定义 4　设向量 A 中第 $i-1$ 个元素和第 i 个元素间的差值为 a，令 $z=a_i$, $E=|z+a-A|$, $\{e,k\}=\min(E)$，式中 $\min(E)$, e,k 的含义与定义 1 中相同，如 DT< $(a_k-z)/a$ <UT，其中 DT 和 UT 为上下界，则称为是正常变化，反之则称为是异常变化。

4.2.1 *基频起点确定*　基频起点确定也是通过投票评比得到的，主要包括：计算本帧语音的基音周期、确定参加评比的基音周期、基频评比和倍频分析。

在确定参加评比的基音周期的过程里有两条限制：基音周期基本相同的，不参加基频检查，直接进入头尾漏点处理和精确定位；基音周期应大于 125 采样点的，不参加基频检查，认为是漏点或分频。

基频评比的具体过程如下：(1) 对每一基音周期从其自身位置开始向前和向后推算，根据定义 4 进行推算和判断，如是正常变化，给该基音周期对应的计数器加 1，并令 $a=a_k-z$, $z=a_k$，如是异常变化，令 $z=z+a$, a 值不变，再次按定义 4 进行推算直至本帧语音结束；(2) 计算每个基音周期在本帧语音上应有的基频点数，方法为基频点位置的最大值除以该基音周期，再将每个基音周期对应的计数器值与应有的基频点数相除，商值最大的为本帧语音起始基音周期，其起点为本帧语音的基频起点。基频评比过程中有两条限制：基音周期对应的计数器值的最大值为 1 或相除后结果的最大值小于等于 0.5，表明它们都不具备成为基频的资格，进入倍频分析；实验中观察到大于 100 采样点的基音周期参加评比会造成分频现象，为此对它实行了限制，如它是检测到的大多数基音周期的两倍，则取消它的参评资格。

倍频分析是在严格的前提条件下才进行的，因此为提高算法的速度仅推算一次，设向量 A 中第 $i+2$ 个元素和第 i 个元素间的差值为 a，取 $z=a_i$，按定义 4 进行一次推算，如属正常变化，则以第 i 个元素为本帧语音的基频起点，差值 a 为本帧语音的起始基音周期。如属异常变化，令 $i=i+1$, 继续推算。如找不到起点，以向量 A 中第 1 个元素作本帧语音的基频起点。

4.2.2 基频检查 这个过程主要检查基频中是否包含非基频点 (异常点) . 根据异常点出现的位置, 可将异常情况分为两种. 一种是异常点处于两基频点间的 0~0.85 之间, 如图 4(b) 所示的 $i+2$ 点. 在 0.85~1 之间为另一种异常情形, 如图 4(a) 所示的 $i+1$ 点.

基频检查时如出现异常, 按这两种情况处理, 具体方法为: 将基频点集合表示为向量 A, 基频点表示为 a_i, 即 $A=\{a_1,a_2,\cdots,a_i,\cdots\}$, 设第 $i+1$ 点和第 i 点间的差值为 e_1, 第 $i+2$ 点和第 $i+1$ 点间的差值为 e_2, 第 i 点和第 $i-1$ 点间的差值为 e_3, 第 i 点和第 $i+2$ 点间的差值为 e_4, 如 DT< e_2/e_1 <UT 成立, 其中 DT 和 UT 与定义 4 中相同, 则称第 i 点是正常变化, 反之称第 i 点是异常变化. 假设在第 i 点出现异常, 需判断上述两种情形, 即判断第 $i+1$ 点错误还是第 $i+2$ 点错误, 如 DT< e_4/e_3 <UT 成立, 则第 $i+1$ 点错误, 反之为第 $i+2$ 点错误. 这里 DT 和 UT 的取值比较关键, 实验中 DT= 0.85, UT= 1.15 .

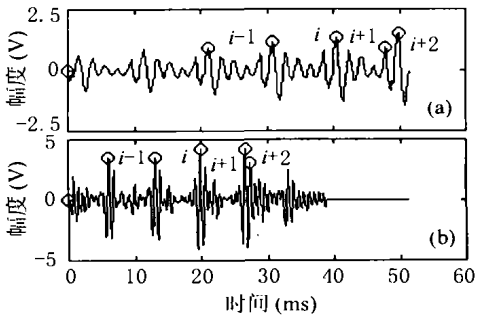

图 4 两种异常情形的图形演示

4.2.3 异常点修正 以尺度 1, 2, 3, 4 和 5 下检测到的基频组成向量 H, 对于第 $i+1$ 点错误 (图 4(a)), 令 $z=a_i, y=e_3$, 记 $E=|z+y-A|$, $\{e,k\}=\min(E)$, 如 DT< $(a_k-a_i)/e_3$ <UT, 则第 $i+1$ 点应为 a_k; 否则令 $E=|z+y-H|$, $\{e,k\}=\min(E)$, DT< $(H_k-a_i)/e_3$ <UT, 则第 $i+1$ 点应为 H_k; 否则第 $i+1$ 点应为 a_i+e_3. 对于第 $i+2$ 点错误 (图 4(b)), 令 $z=a_{i+1}$, $y=e_1$, 可进行类似处理.

4.3 头尾漏点处理

实验中发现一帧语音的开头和结尾存在漏点现象, 需进行漏点处理. 本阶段是在基频点选择和基频检查后进行的. 具体方法为: 设基频点组成的向量为 A, a_i 表示向量 A 中的元素, 基频点数为 n, 帧长为 L 点, 则开头最大可能的漏点数为 $\mathrm{int}(a_1/(a_2-a_1))+1$, 结尾最大可能的漏点数为: $\mathrm{int}((L-a_n)/(a_n-a_{n-1}))+1$, 式中 $\mathrm{int}(F)$ 为取整运算, 以它们作为循环结束条件; 设 $e_i=a_{i+1}-a_i$, 记 $E=|a_i-e_i-H|$, $\{e,k\}=\min(E)$, 如 DT< $(H_k-a_i)/e_i$ <UT, 将 H_k 添加到基频中, 否则停止漏点检查. 结尾漏点检查与上述过程一致.

4.4 基频点精确定位

由表 1 可看出, 在不同尺度上, 基频位置存在偏移, 而基于投票策略方法得到的基频来自不同尺度, 这势必造成不必要的误差, 因此需进一步进行定位. 本文所采用的方法是基于投票策略选择用来定位的分辨率层. 以 2, 3 和 4 阶尺度作为候选尺度, 投票选出的基频点组成向量 A, 分别依次以尺度 2, 3 和 4 下检测到的基频点组成向量 B. 根据定义 1 分别计算每个尺度上支持向量 A 中基频的票数, 选择得票数最高的尺度, 将向量 A 中基频与所选尺度上的基频相比较, 不同的予以保留, 相同的取所选尺度上的基频.

4.5 算法步骤

综上所述, 基于小波变换的基频提取新算法如下:

(1) 初始化 设定帧长 $L=51.2$ms, 帧移 $S=41.2$ms, 交叠为 10ms, 清音和浊音区别门限的系数取为 0.05, 判断基音点门限系数取为 0.68 .

(2) 取一帧语音信号, 计算其在压扩因子 $\alpha=2^j, j=1,2,3,4,5$ 上的离散二进小波变换.

(3) 定位在各阶尺度上的极大值点. 进行清浊音判别, 如该段语音的小波变换的最大值小于清浊音区别门限, 则该段为清音, 没有基音周期; 进行基音点判别, 当极大值点的值大于基音点门限, 该极大值点为基音点.

(4) 基于投票策略进行基频点选择和基频起点确定。

(5) 进行基频检查和异常点修正，处理起始和结尾的漏点。

(6) 基于投票策略进行基频点精确定位，计算本帧语音的基音频率。

(7) 若语音信号没有结束，帧移 S 点，转 (2)。

(8) 去除帧间交叠，进行插值平滑。

表 2 为表 1 所示的一帧语音经上述处理后的分步结果。图 5 为经上述算法处理后得到的三字词“党代表”的基频曲线 (未经平滑)，与图 1 相比基频曲线得到了很大改善。

表 2　表 1 所示语音经上述算法处理后的分步结果

说明	基频点位置									
投票选出的基频点	68	124	181	235	288	443				
确定起点后的基频点	181	235	288	443						
基频检查后的基频点	181	235	288	388						
补充起始和结尾漏点	8	68	126	181	235	288	338	388	441	495
精确定位后的基频点	8	65	123	179	232	286	338	390	442	495

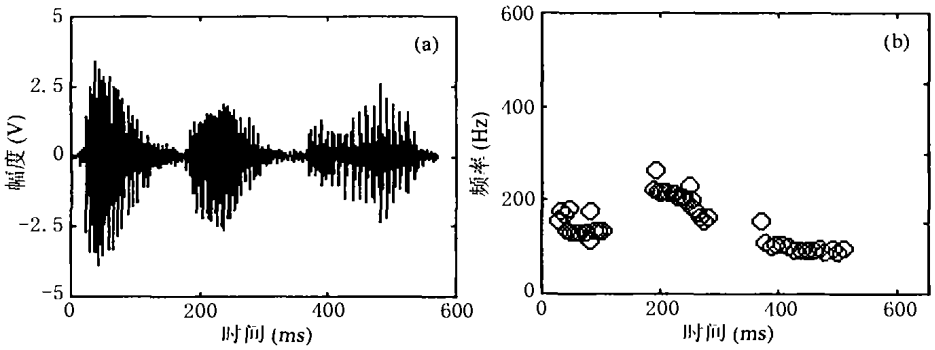

图 5　三字词“党代表”的基频曲线

5 实验与结论

语音信号由采样系统采集，为 10k 采样，12 位量化。实验的三字词词汇有 192 个，语音材料由 5 男 5 女发音，10 人的语音样本为 1920 个。在实验中发现，小波变换方法具有很好的抗噪性，因此取消了基频提取前的滤波处理。实验中用能量过零率方法实现有效语音段截取，用本文提出的算法检测三字词的基频。三字词基频提取成功率高于 90%，与自相关法的基频提取结果相当 [5]。自相关法需要经常调节削波电平阀值 [5]，为得到一个三字词的基频，往往要经过多次尝试。相比而言，小波变换方法要更方便。基于小波变换传统算法的基频检测结果总体上有较多的错误点和误差点，仅对少部分语音得到了较好的结果。经基频提取，得到的三字词基本调型与文献 [6] 所述的基本相符。男女声的声调变化主要在频率绝对值的高低，而声调模式基本不变。如音节中有清声母，声调可提供精确的音节分割和声韵分离信息；当浊音相连时声调是连续的，需借助于其它特征才能进行音节分割。提取的三字词基频曲线中含有孤立误差点，一般可用插值和平滑来消除。基频提取错误主要有两种情况，一种是语音本身读错了；另一种是声道响应非常强烈，无法形成规则的声调曲线。

实验表明，本文提出的基频检测算法较好地克服了基于小波变换传统算法的不足，解决了具有连续语音特征的三字词语音的基频提取问题，具有提取的基频更准确、对噪声不敏感和更

适于处理连续语音等特点．但花费的计算时间较多，为更好地进行实时语音处理，这一点还有待进一步改进．

参 考 文 献

[1] S. Kadambe, G. Faye Boudreaux-Bartels, Application of the wavelet transform for pitch detection of speech signal, IEEE Trans. on IT., 1992, IT-38(2), 917-924.

[2] 王长富，林志刚，戴蓓倩，张劲松，基于小波变换的语音基音周期检测，合肥，中国科学技术大学学报，1995, 25(1), 47-52.

[3] 程俊，张璞，戴善荣，易克初，小波变换用于信号突变的检测，通信学报，1995, 16(3), 96-104.

[4] S. Mallat, A theory for multiresolutions signal decomposition: wavelet representation, IEEE Trans. on PAMI., 1989, PAMI-11(7), 674-692.

[5] 钟金宏，杨善林，张学应，汉语连续语音三字词声调提取方法研究，合肥，合肥工业大学学报（自然科学版），2000, 23(5), 710-714.

[6] 陶维青，徐士林，钟金宏，汉语三字词声调的模式分析，中文信息学报，1998, 12(3), 29-35.

A NEW ALGORITHM FOR PITCH DETECTION OF SPEECH SIGNALS USING WAVELET TRANSFORM

Zhong Jinhong Yang Shanlin Lin Yirong* Lu Kui

(*Institute of Computer Networks System, Hefei University of Technology, Hefei* 230009, *China*)

*(*School of Electrical Engineering, Hefei University of Technology, Hefei* 230009, *China*)

Abstract In this paper wavelet transform is applied to pitch detection of Chinese trisyllabic word with characteristic of continuous speech, a new algorithm for pitch detection of speech signals using wavelet transform is proposed for the question appearing in pitch extraction. It is made up of the following processes: calculating discrete wavelet transform, selecting pitch points and determining the first pitch point based on voting strategy, checking pitch, modifying abnormal pitch point, processing pitch miss point at the beginning and end of a frame of speech signals, relocation of pitch point based on voting strategy. It has been found that this algorithm is valid, and very suits to process continuous speech. But it needs much time for computation, so it is not suit for real-time system.

Key words Wavelet transform, Pitch detection, Voting strategy, Speech signals

钟金宏： 男，1971 年生，博士生，研究方向为语音信号处理和人工智能．

杨善林： 男，1948 年生，教授，博士生导师，主要从事人工智能、计算机控制和决策支持系统等方面研究．

林逸榕： 女，1971 年生，硕士，主要研究方向为信号处理．

鲁 奎： 男，硕士生，研究方向为管理信息系统．