

# 一种改进的信号灵活分割算法及 语音信号清-浊音的自动分割

董恩清, 刘贵忠, 周亚同, 顿玉洁

(西安交通大学电信学院信息与通信工程系, 陕西西安 710049)

**摘 要:** 文中主要对王永忠等提出的灵活分割算法存在的问题做了相应的改进, 并做了比较分析, 然后将改进后的分割算法应用于语音信号的清-浊音自动分割中. 经过大量的理论模型与实际语音信号验证该改进后的算法确实解决了二进分割算法及王永忠方法存在的问题, 达到了对信号自适应有效分割. 仍然采用 Wesfreid 等提出的清-浊音识别准则, 将新的分割方法应用到实际语音信号的清-浊音自动分割中, 不仅同样产生较好划分结果, 而且在时间上没有过多的冗余分割.

**关键词:** 分割; 二叉树; 最佳基; 局部余弦变换; 语音信号; 清-浊音

**中图分类号:** TN912.3 **文献标识码:** A **文章编号:** 0372-2112(2001)10-1364-04

## An Improved Flexible Signal Segmentation Algorithm and Automatic Speech Voiced-unvoiced Segmentation

DONG En-qing, LIU Gui-zhong, ZHOU Ya-tong, DUN Yu-jie

(School of Electronic and Information Engineering, Xi'an Jiaotong University, Xi'an, Shanxi 710049, China)

**Abstract:** To the existing problems of the flexible segmentation algorithm proposed by Wang Yongzhong, et al., in this paper, we make some improvements, and apply the new improved algorithm to automatic speech voiced-unvoiced segmentation. A great deal of numerical simulations and practical data processing show that this new algorithm assuredly solves several problems of the binary based segment algorithm and the segment algorithm proposed by Wang Yongzhong, and achieves adapted effective segmentation for speech signal. The criterion of distinguishing voiced from unvoiced speech segments proposed by Wesfreid, et al. is adopted, and the new improved algorithm is also applied to automatic speech voiced-unvoiced segmentation. It not only obtains ideal segmentation result, but also there are not much more redundancy segmentations.

**Key words:** segmentation; binary tree; best basis; local cosine transform; speech signal; voiced-unvoiced

## 1 引言

由 Coifman 和 Meyer(1991)<sup>[1]</sup> 及 Auscher 等(1992)<sup>[2]</sup> 先后构造的局部余弦基是由平滑、紧支撑钟函数与余弦函数乘积构成的. 这些局部化的余弦函数仍保留着正交性, 并且具有较小的 Heisenberg 乘积. 可是在自适应局部余弦变换中, 进行基于最佳基搜索算法<sup>[3]</sup> 的二进分割具有其所固有的缺陷. 首先二进分割对于沿轴分布的一组信号不能灵活地分割, 且具有相当多的冗余分割; 其次对原始信号在时间方向的时移相当敏感, 以致于若信号时移一些样点后最佳基将产生较大的改变. 在文献[4]中, 王永忠等针对二进分割以上存在的问题, 提出一种灵活的分割算法, 基本上解决了以上问题, 但也存在一定问题. 举一个简单的例子, 如对于有值的分割段之外存在无值的分割段时, 该方法存在将有值分割段与无值分割段合并为一个大的分割, 这显然是不合理的. 所以本文在文献[4]的基

础上, 针对这一问题提出了相应的解决方法. 此外, 文献[4]中的合并始终是从左至右进行单方向合并, 这样可能导致不是全局相邻最优合并. 为了解决这一问题, 达到全局搜索最优合并的目的, 本文也提出了相应的解决办法. 并将改进后的新的灵活的分割算法应用于语音信号的清-浊音自动分割中.

## 2 自适应局部余弦变换

### 2.1 局部余弦变换<sup>[5,6]</sup>

我们首先简要介绍局部余弦变换(LCT), 假设时间轴被划分若干段, 即  $R = \bigcup_{j \in Z} I_j$ , 这里  $I_j = [a_j, a_{j+1}]$ , 保证间隔宽度不小于某一固定正数, 即对于  $j \in Z$ ,  $a_{j+1} - a_j \geq \varepsilon > 0$ . 定义截断函数  $b_j(t)$  如下式, 其中  $\beta(t) = \sin[(\pi/4)(1 + \sin(\pi/2)t)]$ ,  $0 < r \leq \varepsilon$ . 函数  $b_j(t)$  在区间  $(a_j - r, a_{j+1} + r)$  内是平滑紧支撑函数, 这一间隔包含的区间  $[a_j, a_{j+1}]$  称为它的实际支撑.

收稿日期: 2000-06-01; 修回日期: 2000-09-25

基金项目: 国家自然科学基金(No. 69872030); 陕西省自然科学基金(No. 98x08)

$$b_j(t) = \begin{cases} \beta \left( \frac{t - a_j}{r} \right), & t \in [a_j - r, a_j + r) \\ 1, & t \in [a_j + r, a_{j+1} - r) \\ \beta \left( \frac{a_{j+1} - t}{r} \right), & t \in [a_{j+1} - r, a_{j+1} + r) \\ 0, & t \in (-\infty, a_j - r] \cup [a_{j+1} + r, \infty) \end{cases}$$

那么如下函数集就是在  $L^2(R)$  上的正交基

$$\psi_k(t) = b_j(t) \frac{\sqrt{2}}{\sqrt{|I_j|}} \cos \frac{\pi}{|I_j|} \left( k + \frac{1}{2} \right) (t - a_j), j \in Z, k \in N$$

我们可以通过调整控制窗宽度  $I_j$ , 表示实际支撑的左端点位置  $a_j$  及表示频率(波数)指标  $k$  来改变基函数  $\psi_k$ . 因此, 信号  $S(t) \in L^2(R)$  可以表示成  $\psi_k$  函数形式

$$S(t) = \sum_{\substack{j \in Z \\ k \in N}} c_k \psi_k(t)$$

其中

$$c_k = \langle S(t), \psi_k(t) \rangle = \frac{\sqrt{2}}{\sqrt{|I_j|}} \int S(t) b_j(t) \cdot \cos \frac{\pi}{|I_j|} \left( k + \frac{1}{2} \right) (t - a_j) dt \quad (1)$$

在进行折叠预处理步骤后(有关折叠的详细论述参见文献[6]), 式(1)的内积可以应用标准的快速离散余弦变换计算. 另外, 有关二叉树算法的合成过程参见文献[5、6].

## 2.2 最佳基的搜索<sup>[3]</sup>

基于二进的信号分解树由不同级上的基构成. 然而, 并不是所有的基对匹配一给定的信号是有效的. 所以, 我们必须基于一个代价函数, 从所有可能的局部余弦包基中搜索最佳基.

定义一代价函数求取分解信号的熵来寻找局部余弦最佳基, 也即是获得对某一信号的最佳匹配的自适应局部余弦基. 在文献[7]中给出了几种代价函数的类型, 这里, 我们使用 Shannon 熵作为代价函数.

在实现中, 采用了 Coifman 和 Wickerhauser<sup>[3]</sup> 提出的基于 Shannon 熵搜索最佳基的快速算法, 最后得到对于一给定信号的最优截断树, 也就得到一个自适应局部余弦基.

## 3 改进后的灵活自适应分割算法

### 3.1 文献[4]的分割算法

二叉分解树的局限性是它的二进分解, 为了克服二进分解的约束, 文献[4]中提出了一个灵活的分割算法. 令  $L$  表示一维信号的时间分辨率, 即是对于长度为  $N$  的信号想得到的最小时间分割, 通常假设  $N = KL$ , 这里  $K$  是最细分割数. 在这一分割算法中, 采用 Shannon 熵作为代价函数.

首先对一信号进行均匀分割, 然后采用从左到右的合并处理, 以便得到不受二叉树限制的最优分割. 对于每一个可能的合并, 比较合并后的分割所具有的代价与合并前两个独立分割的代价和. 如果合并后的分割的代价小于两个独立分割代价的和, 那么两个独立的分割合并为一个分割, 这一合并后的分割在下一次的可能合并中将作为一个整体分割对待; 否则, 拒绝合并. 由于合并仅在相邻的分割之间进行, 拒绝合并操作的左边分割将被从合并候选分割中剔除, 新的合并过程

将把前面合并成一个整体分割作为下一次合并的左边分割部分, 并从候选分割中选出一个作为右边的分割. 这种合并及合并中右边的候选分割的选取都是始终从左至右进行的.

需要说明的是在分割中, 分割的最小长度是预定的时间分辨率, 该方法使二进分割处理的很多缺陷得到了解决.

### 3.2 改进后的分割算法

**3.2.1 处理单一分割段的 LCT 方法** 虽然文[4]中的分割算法基本上解决了二进分割存在的问题, 但该算法也存在着一定问题. 所以本文在文献[4]的基础上, 针对存在的问题提出了解决该问题的方法. 解决的途径主要体现在对处理单一分割段的 LCT 方法上. 也就是在对每一小分割段做 LCT 时, 不是采用相邻数据进行折叠处理, 而是将每一个小的分割段首先将其做 0 延拓后, 再进行折叠处理. 这样使得数据点的值皆为 0 的分割段在经过单独的 LCT 后其值仍为 0, 这样与相邻的数据中有非零值的分割段在相邻合并时不会产生合并. 当两个相邻独立小段合并后, 计算合并后的 LCT 系数及熵是采用将原始信号对应段的数据同上单独做 LCT.

**3.2.2 全局最优搜索合并方法** 文献[4]中的合并始终是从左至右进行单方向合并, 这样可能导致不是全局相邻最佳合并, 在此提出了如下相应的解决方法. 即是采取同时计算所有相邻分割段间的熵, 如果两个相邻分割段的熵之和大于两个相邻段作为一个分割段时的熵, 则这两个相邻分割段作为待合并的分割段, 否则这两个分割段将拒绝合并. 每次合并只将所有待合并的分割段中最佳的相邻两个分割段合并. 在这里, 待合并的两个相邻最佳分割段的选取方法是从所有相邻两个分割段能够合并为一个分割段的对象中选取其中的熵最小者; 然后再计算合并后的分割段与相邻的分割段间的熵, 确定是否为待合并段. 接下来再进行下一次全局搜索合并, 直到没有可合并的分割为止. 这样避免了单向合并方法可能造成合并不是最佳合并的问题. 这一过程可以通过图 1 简单加以说明, 其中  $E(2, 3)$  表示第二次合并处理时从左至右第三组相邻段组合可作为一个待合并的分割段的熵值,  $E(1, 1) = N$  表示第一次合并处理时从左至右第一组相邻段组合不能作为一个待合并分割段.

## 4 模拟实验的分析比较

在模拟实验中我们选择信号的长度为  $N = 1024$ , 虽然任意的时间分辨率在这一新方法中均可以采用, 在这里把与时间分辨率相对应的最小时间分割设为  $L = 32$ , 则最小分割数  $K = 32$ . 为了方便比较, 在二叉树的算法中, 对信号最大分解级数设为 5 级, 其最小时间分割同样也是 32.

### 4.1 检验有无二进分割的约束

为了解决二叉树算法的二进约束, 并能够说明改进后的灵活分割算法的优越性, 我们选择一支撑长度为 17 ( $t = 504 \sim 520$ ) Ricker 子波, 这一子波的峰值恰好位于整个信号的中点, 这一点皆是二叉树算法和改进后的灵活分割算法的可能分割点. 在该信号中  $t = 1 \sim 128$  处, 还有一个两个周期的正弦信号. 图 2(a) 为应用二叉树算法的最佳分割, 从中我们看到虽然正弦信号被正确地分割为一个区间, 但完整 Ricker 子波信

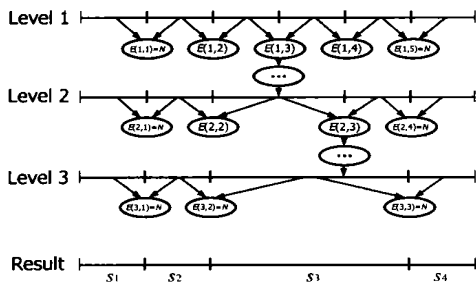


图 1 全局最优搜索合并示意图

号却被毫无道理地在中间峰值点分割开来。其它没有信号的区域也被分割为若干个分割, 形成分割冗余, 这显然是不合理的。如果不把 Ricker 子波从中间分开, 那么只能把整个信号作为一个分割了, 这样一来, 前面与之不同的正弦信号与 Ricker 子波作为一体看待显然无道理, 这些正是二叉树算法所固有的局限性。图 2(b) 为由文献[4]的分割算法得到的分割结果, 从图中明显地看出, 对于正弦信号的分割点应该在 128 的位置, 而其却在 160 点位置, 这正是由于以上论述的原因使文献[4]的算法出现这一问题。与之相反, 图 2(c) 为应用改进后的灵活分割算法的分割结果, 并没有把 Ricker 子波一分为二, 两种类型的信号皆合理地得到分割, 同时也并没有产生多余的分割。

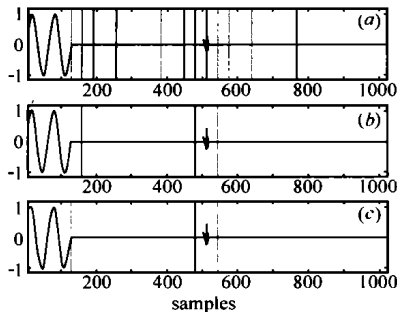


图 2 没有时移的理论记录的分割对比图

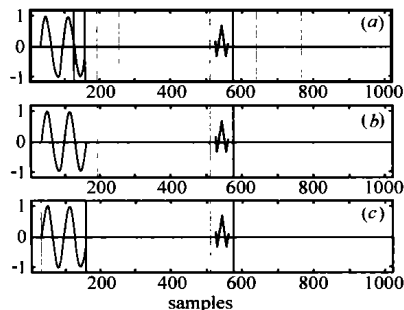


图 3 有时移的理论记录的分割对比图

#### 4.2 时移灵敏性的检验

假设让该合成信号向后时移 32 个样点, 图 3(a) 得到了二叉树算法的分割结果, 其与图 2(a) 相比有较大的变化。虽然 Ricker 子波被正确划分, 但正弦波形却被一分为二, 这清楚地说明了二叉树算法的分割结果并不是时移不变的。图 3(b) 为经时移后由文献[4]的算法得到的分割结果, 对于正弦波的

分割点位置与图 3(c) 相比向两侧移一个最小分割段的距离, 显然具有一定的时移敏感性, 这表明文献[4]的分割算法的问题所在。然而, 图 3(c) 为经时移后由改进后算法的分割结果, 与图 2(c) 相比显然是时移不变的。

### 5 清音-浊音自动分割算法<sup>[5]</sup>

本文提出的新的灵活分割算法, 产生自适应正交分解, 得到较好适应信号谱的窗尺寸来对信号进行加窗余弦变换, 其是对于信号的局部谱表示。与之相联系的时间划分, 或窗选择, 显然根据波峰数对语音信号进行清音和浊音划分是可行的。本文同文献[5]一样把新提出的方法应用于清音-浊音自动分割中。清音-浊音自动分割的准则简述如下:

这里假设信号  $S = S(t)$  是在整个时间间隔  $[0, T]$  上, 那么我们可以计算它的自适应时间划分

$$[0, T] = \bigcup_{0 \leq j < N} I_j$$

$S$  在相邻子间隔边界的折叠可以看成该函数的一个划分  $S(t) \rightarrow (S_0(t), S_1(t), \dots, S_{N-1}(t))$ , 应用离散局部余弦变换, 每一  $S_j(t)$  可以被分解正交基本波形的集合

$$S_j(t) = \sum_{0 \leq k < n_j} c_k^j \phi_k(t)$$

其中  $n_j$  为  $I_j$  中的采样点数,  $c_k^j$  为经过计算标准 DCF IV 变换的离散局部余弦变换系数。

每一个系数  $c_k^j$  给出了基本正交波形成分的振幅, 这基本波形的周期为  $T_k = (2\pi / \omega_k)$ , 其中  $\omega_k = (\pi / |I_j|) (k + \frac{1}{2})$ , 所以其频率为

$$F_k = \frac{\omega_k}{2\pi} = \frac{k + (1/2)}{2|I_j|} \quad (2)$$

浊音是由规则的全程激励产生的, 其典型基本频率对于女音一般在 140~250Hz, 而男音一般在 100~150Hz。频率  $F_k = 250\text{Hz}$  相应  $k$  大约等于  $n/16$ , 使用这一估计, 将引入在语音分割中区分清浊音的准则。首先我们定义最强谱成分中的频率索引  $k_0$

$$|c_{k_0}^j| = \max_{0 \leq k < n} |c_k^j| \quad (3)$$

清-浊音划分准则<sup>[5]</sup>: 我们称  $F_{k_0}$  为分割的第一个基本频率, 如果  $k_0 < n/16$ , 在  $I_j$  上的信号分割  $S_j(t)$  为浊音; 如果  $k_0 > n/16$ , 则称清音。对于更复杂的识别问题, 我们可以利用多个频率描述一个分割中的谱。在找到第一个基本频率后, 我们将其相邻的频率系数去掉, 然后在剩余的系数中寻找次最强的系数。

### 6 实际语音信号的清-浊音自动分割

#### 6.1 实际语音信号的时间分割

无论何种语音识别系统, 在系统的最初阶段都是需要把语音分解为可识别的最小语音单元, 以便于后期的处理。所以能够把语音信号正确地进行时间分割是很有意义的。据此, 这里我们采用汉语数字“721”语音信号来作为时间分割对象, 分别应用二叉树算法、文献[4]的分割算法以及改进后的新算法对其进行分割。在分割中采用了对整个信号做二叉树深度为

5 级分解, 相当于最小时间分辨率  $L/2^5$ , 其中  $L$  为信号的总长度. 图 4(a) 为应用二叉树算法进行 5 级的最佳分割结果, 图 4(b) 为文献[4]的分割算法结果, 而图 4(c) 为采用改进后的灵活分割算法的分割结果. 从中看出改进后的算法比二叉树算法分割的结果以及文献[4]的分割算法结果更合理, 没有它们中那么多冗余分割.

## 6.2 语音信号清·浊音的自动分割

在语音识别中, 能够正确地区分清音·浊音是非常重要的先期基础工作. 本文同样采用文献[5]的方法来进行清·浊音的自动分割, 以便来验证改进后的新方法在语音的清·浊音划分中的有效性. 在这里采用的是较复杂的识别模式, 图 5 是上面的连续汉语语音数字“721”的清·浊音分割结果, 在图的上方 0.15、0.20、0.25 的水平线分别表示语音信号为无声、清音和浊音. 从图中可以看出基本上正确地对清浊音进行了正确的分割.

需要说明的是在 0.6~0.8 之间出现无声和浊音两次的交替情况是不能单靠此分割后的清·浊音识别来把这两个转变点分开, 也就是说利用该分割算法进行的清·浊音识别不能同时进行有声和无声的判别. 所以在利用该分割算法进行清·浊音识别必须与有声和无声判别结果结合起来. 另外虽然在 0.4~0.6 范围内有几个冗余分割, 但较二进分割算法还是减少了许多.

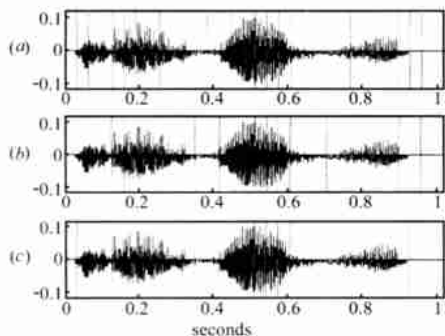


图 4 汉语数字“721”语音分割对比图( $L/2^5$ )

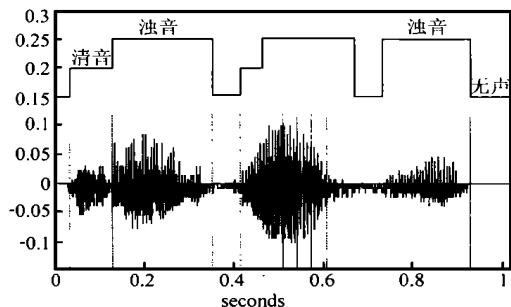


图 5 基于新的灵活分割语音清浊音自动识别

## 7 结论

文中主要对文献[4]中提出的一种灵活分割算法存在的问题做了相应的改进, 经过大量的理论模型验证该改进后的算法的有效性. 并将改进后的算法应用于语音信号的清·浊音自动分割中, 不仅同样产生较好划分结果, 而且在时间上没有过多的冗余分割.

## 参考文献:

- [1] Coifman R R, Meyer Y. Remarques sur l'analyse de Fourier à fenêtre [J]. Comptes Rendus de l'Académie des Sciences, 1991, 312: 259-261.
- [2] Auscher P, Weiss G, Wickerhauser M V. Local Sine and Cosine Bases of Coifman and Meyer and the Construction of Smooth Wavelets [M]. in C. K. Chui, Ed, Wavelets: A Tutorial in Theory and Applications, Boston, MA: Academic Press, 1992: 237-256.
- [3] Coifman R R, Wickerhauser M V. Entropy Based algorithms for best basis selection [J]. IEEE Trans. on Information Theory, 1992, 38(2): 713-718.
- [4] Wang YongZhong, Wu RuShan. New flexible segmentation technique in seismic data compression using local cosine transform [A]. Wavelet Applications in Signal and Image Processing VII, Proc. of SPIE [C], 1999, 3813: 784-794.
- [5] Wesfreid E, Wickerhauser M V. Adapted local trigonometric transforms and speech processing [J]. IEEE Trans. on Signal Processing, 1993, 41(12): 3596-3600.
- [6] Wickerhauser M V. Adaptive wavelet analysis from theory to software [M]. A. K. Peters, Wellesley, mass, 1994: 103-152.
- [7] Misiti M, Misiti Y, Oppenheim G, Poggi J M. Wavelet Toolbox for Use with MATLAB [M]. THE MATH WORKS Inc., 1996.

## 作者简介:



董恩清 男, 1965 年 7 月生于辽宁省营口市. 讲师, 1987 年于中国矿业大学获工学学士学位, 1993 年于长安大学获理学硕士学位, 现为西安交通大学电信学院通信与信息系统专业在读博士生, 主要研究兴趣包括小波分析、时频分析、非平稳信号处理和语音数据压缩等.



刘贵忠 男, 1962 年 9 月生于陕西省吴旗县. 1989 年 6 月获荷兰 Eindhoven 大学博士学位, 现为西安交通大学电子与信息工程学院教授, 博士生导师. 主要从事非平稳信号的分析、处理、模式识别、反演等方面的理论、方法与应用研究.