

一种基于声音信号形态特征参数的压缩算法及其分析^{*}

A Compression Algorithm based on Sound Signals Characteristic Parameters and it's Analysis

王昆仑

(新疆师范大学数理信息学院计算机科学与技术系 乌鲁木齐830054)

Abstract This paper refers to a new compressed algorithm based on SSCPA (Sound Signals Characteristic Parameters Analysis) and quadratic curve simulative method. The algorithm use short-time energy to divide sound signals wave form ton on-sound band and sound bands, and it can scan original sound signals and draw out two types of wave form characteristic parameter of sound signals in onetime during the coding stage. It is used recursive formula structured by the quadratic Bézier curve simulative way to recover the sound signals waveform in decoding. MOS amount accomplish 4.5, In information by calculated and analysis, the interactive relevance coefficient of original signal and reinstate signal is 0.9853 and them power spectral desity is analogy. This algorithm have traits such as: calculating simply, much less lack-rate, lower coding-rate, non-uniform velocity, lower complexity in algorithm space and powerful practicable.

Keywords Sound signals, Characteristic parameters analysis, Compressed algorithm, Data compressed, Capability analysis

1. 引言

声音信息能够压缩的基本依据是语音信号的冗余度及人类的听觉感知机理^[1]。声音信号编码方法可大致分为两类:一类方法要求重建的信号听起来与输入语音一样,但其波形可以不同,此类编码方法的缺点是编码方法所需的运算量大,对于环境噪声较敏感。第二类方法则使重建的语音信号基本上与原始语音信号波形相同,对声音信号的波形直接进行编码。实际上,目前中低码率编码的发展方向是这两类方法之结合。当前低速率语音编码算法^[2]如:混合激励声编码器;插值编码的原型波形插值、时频插值;谐波编码的多带激励、正弦变换等,都需要精确估计浊音段的基音周期。文[3,4]提出了一种新的语音信号波形编码方法,将语音信号分为无音段,清音段和浊音段三类,然后对不同的语音信号采用矢量和标量量化的方法生成相应的码本进行编码,在低码率下取得较好的话音质量。本文提出了一种新的基于声音信号形态特征参数分析 SSCPA (Sound Signals Characteristic Parameters Analysis)和二次曲线拟合方法的压缩算法。算法具有计算简便、声音失真率很小、低码率、非均匀速率、算法时空复杂度小和实用性强的特点。

2. 信号波形的形态特征分析

2.1 声音信号波形分析

声音经过采样和量化得到数字声音信号,我们把声音信号波形分为两种:无音段类型和有音段类型。

无音段除了数字声音信号分析中所描述的通常意义下的无音段之外,还包括清音段。无音段所包含的信息量基本上为零,而清音段在本质上就是噪音。在数字声音信号中,真正的无音段一般多表现为噪音。对于无音段,选择合适的短时能量阈值,可以有效地把无音段分离出来并赋予其编码。解码时只要保持波形长度,以静音形式恢复。

有音段即是数字声音信号分析中所描述的通常意义下浊

音段。我们知道,浊音的基频以及低频主要由过零率来决定。只要算法编解码过程中使有音段过零率不变,就可以保持声音信号基频及低频性能。浊音段的信号波形的变化是比较复杂的,但从信号变化总体看,其变化是有规律的。要使解码后的声音波形与编码以前的声音波形保持基本一致,也就是保证其形态基本不变。所以,本算法中编码的主要功能就是提取信号波形的形态特征参数,解码的主要功能就是恢复声音波形。

2.2 声音信号波形的形态特征分析

对声音信号有音段波形分析,其形态特征参数主要有以下两类。

声音信号取样值的幅度分布是非均匀的,声音中的小幅度样本比大幅度样本出现的概率要高。但无论是小幅度样本还是大幅度样本,其波形可以由两类控制样点确定其变化。第一类控制样点是声音信号的局部最大(小)值特征样点,这一类控制样点在一帧内有多个。它不仅描述了声音信号的幅度,而且描述了声音信号的变化趋势。另外,声音信号的过零率也完全包含于其中。第二类控制样点是声音信号‘拐点’附近的局部特征样点,这类控制样点存在于相邻局部最大(小)值控制样点之间。它描述了声音信号的局部的细节变化。

只要取得这两类控制样点,声音信号波形的恢复就可以比较准确地实现。

声音信号波形样本间存在着相关性,尤其是邻近样本之间。但是对这种邻近样本间相关性的描述和提取是要花费大量代价的,而且也是不精确的^[5]。如预测编码方法,提取了上述两类控制样点,声音信号样本间存在的相关性就可以表示为控制样点间的长度,也就是样本间的时长关系。样本间的形态变化可以在第一类控制样点和第二类控制样点控制之下,由二次曲线拟合来描述。

3. SSCPA 算法基本原理和编解码过程

3.1 SSCPA 编码原理和过程

^{*} 本文得到国家自然科学基金(编号:69562001)和新疆维吾尔自治区自然科学基金(编号:95901)资助。

我们首先将声音信号 $S(t)$ 按一定长度分帧, 帧长取 512 字节, 然后逐帧按以下步骤进行。

3.1.1 无音段和有音段的判定 本算法只要求简单切分无音段和有音段即可。如果每帧信号短时平均能量 E_n 较小, 则表明这一帧的数据变化幅度很小, 且在零点处上下浮动。我们采用短时能量作为判别判据, 当区分无音段和有音段阈值 K 设定得较小时, 可以满足本算法对无音段和有音段切分的要求。经实验统计, 我们取短时平均能量

$$E_n = \frac{1}{N} \sum_{n=0}^N S_n(t) \quad (1)$$

当 E_n 的阈值 K 为 200 时, 可以有效地区别无音段和有音段, 满足本算法对无音段和有音段进行的切分判定要求。或者说, 对于 $K > 200$ 的帧均作为有音段进行压缩处理。

3.1.2 编码码字设计和使用 无论是对一帧无音段, 或者是对有音段中存在的局部最大(小)值控制样点和局部控制样点, 我们都分别用一个码字来表征。该码字由两部分组成, 第一部分描述其幅度值, 第二部分描述其相对长度。

在无音段, 一帧用一个码字来描述。码字第一部分代表其幅度值(用定值表示), 码字第二部分也就是帧长。在有音段, 码字的数量是该段中局部最大(小)值控制样点和局部控制样点的个数。控制样点码字的第一部分代表其幅度值, 码字的第二部分是每一个局部最大(小)值控制样点或局部控制样点相对与前一个局部最大(小)值控制样点或局部控制样点的长度值。

3.1.3 无音段编码 其编码过程很简单, 码字的第一部分为定值, 第二部分为这段无音段帧的长度。

3.1.4 有音段编码 首先我们对信号流中的样本 $P_i(x_i, y_i), P_{i+1}(x_{i+1}, y_{i+1}), P_{i+2}(x_{i+2}, y_{i+2})$ 用 $x_{i-1} \leq x_i \geq x_{i+1}$ 或 $x_{i-1} \geq x_i \leq x_{i+1}$ 为判别局部最大(小)值控制样点判据, 逐个样本地检索得到声音信号波形的相邻局部最大(小)值控制样点 $P_a(x_a, y_a)$ 和 $P_b(x_b, y_b)$ 。其后在 $P_a(x_a, y_a)$ 和 $P_b(x_b, y_b)$ 之间用下列算法计算介于其间可能存在的拐点。

1) 构造相邻局部最大(小)值控制样点 $P_a(x_a, y_a)$ 和 $P_b(x_b, y_b)$ 的直线方程

$$Ax + By + C = 0 \quad (2)$$

其中: $A = y_b - y_a, B = x_a - x_b, C = x_a y_b - x_b y_a$, x_b 是相对与 x_a 的样点长度值。

2) 计算介于相邻局部最大(小)值控制样点 P_a, P_b 之间的样点 $S_i(x_i, y_i)$ 与直线(2)的距离 $D(i)$ 。

$$D(i) = \frac{Ax_i + By_i + C}{\sqrt{A^2 + B^2}}, i = 1, 2, \dots, n \quad (3)$$

n 是介于 P_a, P_b 之间样点的个数。

3) 求介于相邻局部最大(小)值控制样点 P_a, P_b 的“拐点” Q_c 。

$$Q_c = \begin{cases} \text{不存在;} & \text{当 } D(i) \text{ 符号无变化时。} \\ \text{取到 } \text{Min}\{D(i)\} \text{ 处。} & \text{当 } D(i) \text{ 符号有变化时。} \end{cases} \quad (4)$$

当 Q_c 不存在时, 局部控制样点 $P_c(x_c, y_c)$ 取 $x_c = \text{Max}\{D(i)\}$;

当 Q_c 存在时, 局部控制样点有两个: $P_{c1}(x_{c1}, y_{c1})$ 和 $P_{c2}(x_{c2}, y_{c2})$, 取 x_{c1} 和 x_{c2} 如下:

$$\begin{cases} x_{c1} = \text{Max}\{|D(i)|\}, x_{c1} \in [x_a, Q_c] \\ x_{c2} = \text{Max}\{|D(i)|\}, x_{c2} \in [Q_c, x_b] \end{cases} \quad (5)$$

4) 最后, 将相邻局部最大(小)值控制样点和局部控制样点 P_a, P_c, P_b 的码字(或 P_a, P_{c1}, P_{c2}, P_b 的码字)依次输出。

3.2 SSCPA 解码原理和过程

首先从传送过来的码字流中读入码字, 然后进入解码阶段。

3.2.1 无音段解码 如果该码字是无音段码字, 则解码为帧长度之个零值。

3.2.2 有音段解码 如果该码字是有音段码字, 其解码按二次 Bézier 曲线拟合算法实现。

对于取得的 N 个码字 P_1, P_2, \dots, P_n , 对其相邻的三个码字 $P_i(x_i, y_i), P_{i+1}(x_{i+1}, y_{i+1}), P_{i+2}(x_{i+2}, y_{i+2})$ 用 Bézier 曲线算法拟合, 二次 Bézier 曲线 S_i 可以表示为($n=2$):

$$Q(t) = \sum_{i=0}^N P_i B_{i,n}(t), t \in [0, 1] \quad (6)$$

其中:

$$B_{i,n}(t) = \frac{n!}{i!(n-i)!} t^i (1-t)^{n-i}, \quad (7)$$

$i = 1, 2, \dots, n, P_i (i = 1, 2, \dots, n)$ 表示特征多边形顶点的位置向量。

二次 Bézier 曲线 S_i 的调和函数为:

$$\begin{cases} B_{0,2}(t) = (1-t)^2 \\ B_{1,2}(t) = 2t(1-t) \\ B_{2,2}(t) = t^2 \end{cases} \quad (8)$$

$$\begin{aligned} \text{即: } Q(t) &= B_{0,2}(t)P_i + B_{1,2}(t)P_{i+1} + B_{2,2}(t)P_{i+2} \\ &= (P_{i+2} - 2P_{i+1} + P_i)t^2 + (P_{i+1} + P_i)t + P_i \end{aligned} \quad (9)$$

若式(6)对应曲线 S_i 的参数方程($n=2$)为:

$$\begin{cases} x(t) = a_x t^2 + b_x t + c_x \\ y(t) = a_y t^2 + b_y t + c_y \end{cases} \quad (10)$$

其中: 系数 $a_x, b_x, c_x, a_y, b_y, c_y$ 由式(9)可以得到, 如式(11)。

$$\begin{cases} a_x = x_{i+2} - 2x_{i+1} + x_i \\ a_y = y_{i+2} - 2y_{i+1} + y_i \\ b_x = 2(x_{i+1} - x_i) \\ b_y = 2(y_{i+1} - y_i) \\ c_x = x_i \\ c_y = y_i \end{cases} \quad (11)$$

进一步由式(10)可以得到离散化公式为:

$$\begin{cases} x_i = a_x t_i^2 + b_x t_i + c_x \\ y_i = a_y t_i^2 + b_y t_i + c_y \end{cases} \quad (12)$$

其中 $t_i = i \cdot dt, i = 1, 2, \dots, N$ 是离散化后所取的样点数, dt 为点数步长。二次 Bézier 曲线曲率

$$D = \frac{1}{4} \sqrt{a_x^2 + a_y^2}$$

我们由 $(P_{i+1} - P_i) - (P_i - P_{i-1}) = 2adt^2$, 即: $P_{i+1} = P_i + P_{i-1} - 2adt^2$, 实际上, 由于取: $dt = t_{i+1} - t_i = t_i - t_{i-1} = 1$ 。故由离散化公式(12)得到的 x_i 和 y_i 可以由下面递推公式求得:

$$\begin{cases} x_{-1} = a_x - b_x + c_x \\ x_0 = c_x \\ x_{i+1} = x_i + x_i - x_{i-1} + 2a_x \end{cases} \quad i = 1, 2, \dots, N \quad (13)$$

$$\begin{cases} y_{-1} = a_y - b_y + c_y \\ y_0 = c_y \\ y_{i+1} = y_i + y_i - y_{i-1} + 2a_y \end{cases} \quad i = 1, 2, \dots, N \quad (14)$$

相邻的三个码字 P_i, P_{i+1}, P_{i+2} 和 $P_{i+1}, P_{i+2}, P_{i+3}$ 用二次 Bézier 曲线拟合得到 S_i 和 S_{i+1} , 其拟合时应达到 $C^{(1)}$ 连续。可以证明^[6]: 由式(11)确定得二次拟合曲线(式(10))在 $t = 1/2$ 处的切线平行于直线 $P_i P_{i+2}$ 。

3.3 SSCPA 算法结构

编解码系统结构如图1所示。编码和解码分别进行。在编

码器中主要是提取声音信号的波形形态特征参数,在解码器

中主要是对有音段的波形拟合过程。

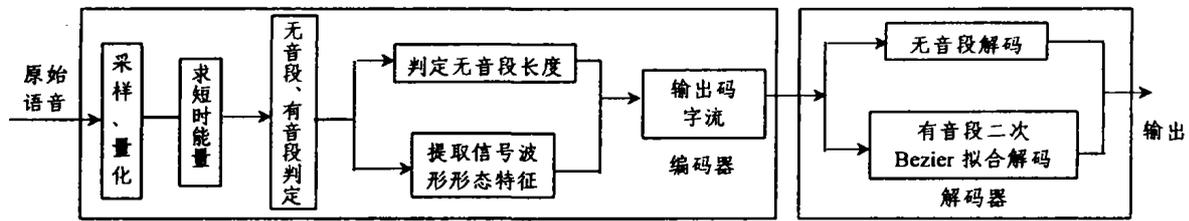


图1 SSCP A 编解码系统框图

4. SSCP A 算法实验及性能分析

4.1 压缩率测试

经对十余种不同素材的声音文件(其中包括语音、歌曲、乐曲等,共2.4GB)进行了压缩测试。测试结果见表1。从压缩测试结果看,平均压缩率达到44%左右。

表1 声音文件压缩测试实验

文件类型	压缩前文件大小	压缩后文件大小	压缩率(%)
语音文件	3,356,176	1,698,733	49.38
歌曲文件	11,716,780	7,357,403	37.21
乐曲文件	9,163,294	5,001,010	45.42
总计	24,236,250	14,057,146	(平均)44.00

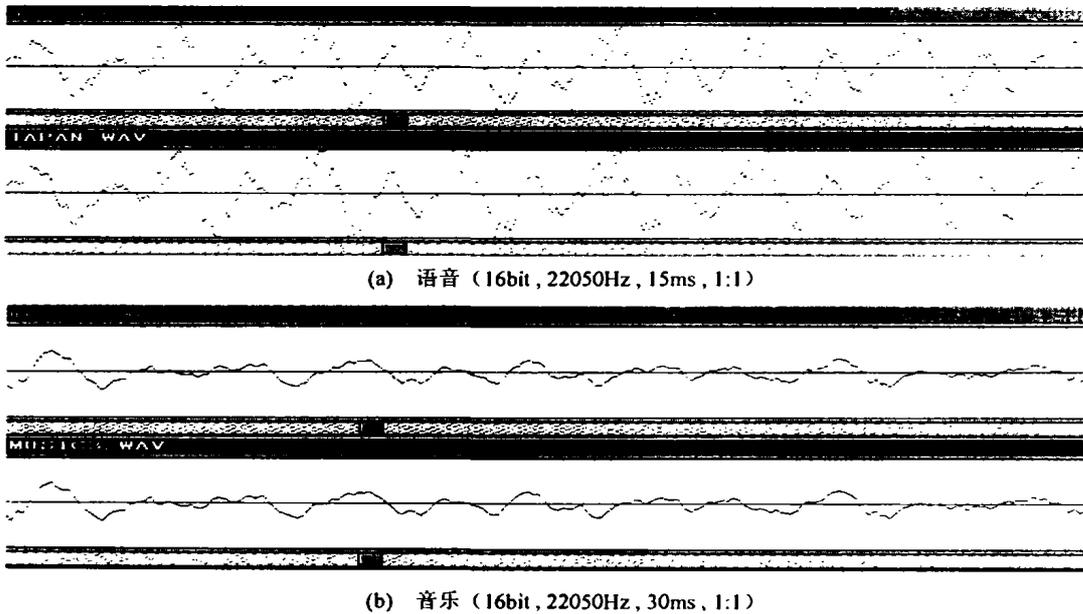


图2 声音信号波形与恢复后的声音信号波形的对比图

4.2 性能分析

经 SSCP A 对声音文件的压缩和解压缩,其原始信号波形与恢复后的信号波形的对比可以参见图2。图2(a)是语音信号波形,图2(b)为音乐信号波形。图2(a)和图2(b)中的上图和下图分别为原始信号波形与恢复后的信号波形。

从图2中可以看出,原始声音信号与恢复后的声音信号变化极小;图2(a)中当信号幅度很大以及信号噪声较强时也能使信号得到很好的恢复,可以看到本算法的失真率很小,编解码精准度是很高的。经一组(7人)测听,MOS(Mean Opinion Score)打分为4.5分(五分制),所恢复的声音质量很好,完全保持了说话人的个人特性,而且音乐的失真率也很小,达到了高保真的效果。

从测试声音信号中,提取原始有音段随机信号 X 以及还原后的同段随机信号 Y,其长度同为318504字节。经 MATLAB^[7]计算,随机信号 X 和 Y 的标准差分别是0.0474和0.0474,方差分别为0.0022和0.0023;协方差分别是0.0022,随机信号 X 和 Y 的互相关系数为0.9853,(当互相关系数为1时,称随机信号 X 与 Y 是完全正相关的)。从中可以看出,随

机信号 X 的和 Y 的相似程度是极高的。

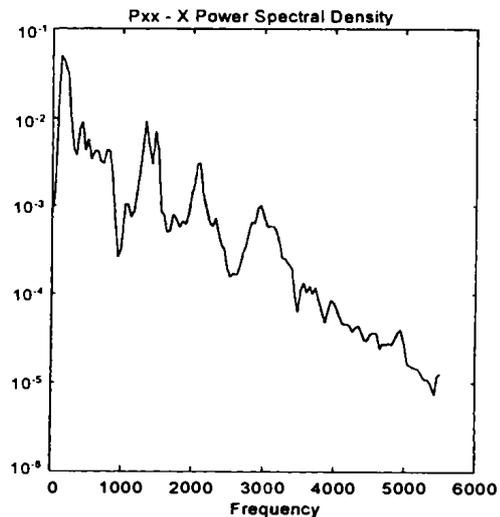


图3 随机信号 X 的功率谱密度图

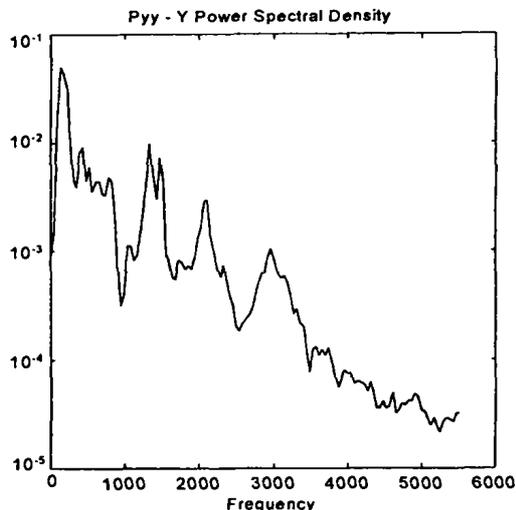


图4 随机信号 Y 的功率谱密度图

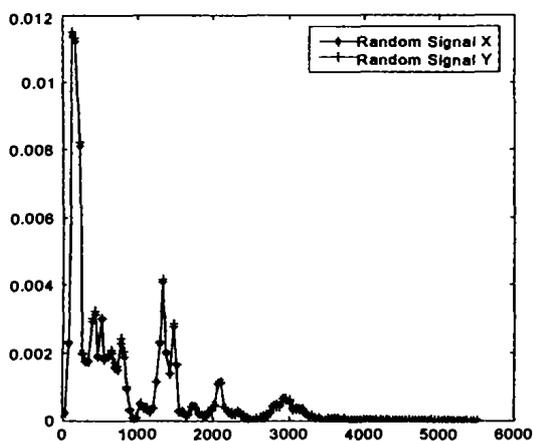


图5 随机信号 X、Y 的功率谱估计

用平均周期图法^[3](置信度为1)对随机信号 X 和 Y 进行功率谱估计,随机信号 X 和 Y 的帧长取256个信号样点,帧移取128,窗函数取 Hanning 窗,得到随机信号 X 的功率谱密度

如图3所示,随机信号 Y 的功率谱密度如图4所示。图5是随机信号 X 和 Y 是功率谱估计的,可以看出经 SSCPA 恢复的信号 Y 的频率特征被基本保持。

结束语 本文提出的方法可以对声音信号进行压缩,平均压缩率达到44%左右。提出的算法运算量很小,编码阶段只要用短时能量的计算就可以区分无音段和有音段,声音信号波形的形态特征只要对输入原始信号一次扫描即可得到,所以编码速度很快。解码阶段用二次 Bézier 曲线拟合法,从式(11)、式(13)和式(14)可以看出,递推公式减少了计算工作量,算法十分简便,由7次乘法和6N次加法完成 S_i 拟合,对输入码字一次扫描即可实现解码,拟合时达到 $C^{(1)}$ 连续。本算法的编解码声音失真率很小,所恢复的声音质量很好。完全保持了说话人的个人特性,达到了高保真的效果,解码后的声音信号与原始声音信号的互相关系数达到0.9853。由于本算法没有使用其他的数据结构(如码本)等,其码字的码率由声音信号本身决定。所以本算法具有计算简便、声音失真率很小、低码率、非均匀速率、算法时空复杂度小和实用性强的特点,具有较好的发展前景。

参考文献

- 1 杨行峻,迟惠生,等. 语音信号数字处理. 北京:电子工业出版社,1995
- 2 Lafiamme C, et al. Harmonic-stochastic excitation (HSX) speech coding below 4 kbit/s. ICASSP-96, 1:204~207
- 3 宋岩涛,余铁城,周键来. 一种新的语音波形压缩编码算法. 见:王承发主编,第五届全国人机语音通讯学术会议论文集,哈尔滨:哈尔滨工业大学,1998. 362~365
- 4 涂哲明. 基于半波波形的语音信号编码算法研究. [硕士学位论文]. 中国科学院声学研究所,1997
- 5 高文. 多媒体数据压缩技术. 北京:电子工业出版社,1994
- 6 唐泽圣,周嘉玉,李新友. 计算机图形学基础. 北京:清华大学出版社,2000
- 7 陈桂明,张明照,戚红雨. 应用 MATLAB 语言处理数字信号与数字图像. 北京:科学出版社,2001年
- 8 肖先赐. 现代谱估计原理与应用. 北京:科学出版社,1991

(上接第95页)

广播环境中高性能并发控制机制的研究才刚刚开始。考虑到广播环境中通讯的不对称性,传统的并发控制协议显得过于严格、不必要、不可行,本文提出了一种新的基于弱可串行化的并发控制协议 BCC-WSR。实验结果表明,新协议可显著提高系统性能。

表1 模拟实验参数设置

参数项	默认设定值
客户端事务长度	5
服务器端事务长度	10
服务器端事务到达率	1 in 2.5-105 bit-units
数据库中数据对象数	500
数据对象大小	1kB
服务器端读操作概率	0.5
客户端操作间延迟	64k bit-unist
客户端事务间延迟	128k bit-unist
客户端重启延迟	0 bit-unist

参考文献

- 1 Pitoura E, Bhargava B. Maintaining Consistency of Data in Mobile Distributed Environments. In: Proc. of the 15th Intl. Conf. on Distributed Computing Systems. 1998. 404~413
- 2 Pitoura E. Supporting Read-Only Transactions in Wireless Broadcasting. In: Proc. of the DEXA'99 Workshop on Mobility in Databases and Distributed Systems, August 2000
- 3 Acharya S, et al. Broadcast Disks: Data Management for Asymmetric Communications Environments. In: Proc. of ACM SIGMOD, 1997
- 4 Datta A, et al. Adaptive Broadcast Protocol to Support Power Conservant Retrieval by Mobile Users. In: Proc. of 13th Intl. Conf. on Data Engineering, 1997
- 5 Leong H V, Si A. Database Caching over the Air-Storage. The Computer Journal, 1997, 40(7): 401~405
- 6 Acharya S, Franklin M, Zdonik S. Disseminating Updates on Broadcast Disks. In: Proc. of the VLDB Conf., Mumbai (Bombay), India, 1999