



[12] 发明专利申请公开说明书

H04M 11/06 H04M 11/10

[21] 申请号 02115372.8

[43] 公开日 2003 年 12 月 31 日

[11] 公开号 CN1464685A

[22] 申请日 2002.6.13 [21] 申请号 02115372.8

[71] 申请人 优创科技（深圳）有限公司

地址 518057 广东省深圳市南山区高新技术
工业村 R2 -A 三层

[72] 发明人 余 泊

[74] 专利代理机构 深圳市千纳专利代理有限公司

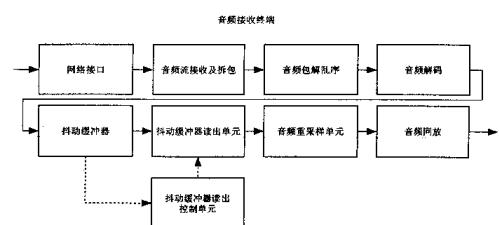
代理人 胡 坚

权利要求书 2 页 说明书 8 页 附图 3 页

[54] 发明名称 一种在网络终端缓冲区中处理音频
流回放的方法

[57] 摘要

一种在网络终端缓冲区中处理音频流回放的方法，解决了在网络上通信易断音、卡音的问题，在包交换网络上(如 IP 网络上)进行音频(如语音)实时通信，在接收端设置一个抖动缓冲区，当接收端收到音频包之后，首先按正常顺序解码、解乱序，之后放入抖动缓冲区；当抖动缓冲区将要充满之时，对音频数据进行降采样率处理，以实现快速回放该音频数据流；当抖动缓冲区将要被读空之时，对音频数据进行升采样率处理，以实现音频流的慢速回放；当抖动缓冲区中的音频数据在正常范围之时，按原采样率进行音频流的回放。采用上述控制方法，可以实现包交换网络上传输来的实时音频流的连续流畅播放，减少卡音、断音等不连续现象。



1、一种在网络终端缓冲区中处理音频流回放的方法，其特征在于：

当包交换网络上传输来音频数据包进入抖动缓冲区时，计算读、写指针之间的距离 D；

在初始化过程中，设置读、写指针之间的距离 D 正常值的范围和一次从抖动缓冲区中读出的数据长度；

在运行过程中，当读、写指针之间的距离 D 正常时，音频流按正常速度回放；当读、写指针之间的距离 D 大于或小于正常值时，通过音频重采样单元对从抖动缓冲区中读出的大于或小于正常播放长度的数据进行重采样处理，使其恢复正常播放长度。

2、根据权利要求 1 所述的一种在网络终端缓冲区中处理音频流回放的方法，其特征在于：所述的抖动缓冲区为环形抖动缓冲区。

3、根据权利要求 1 或 2 所述的一种在网络终端缓冲区中处理音频流回放的方法，其特征在于：所述的读、写指针之间的距离 D 是为进入抖动缓冲区的读指针偏移量与写指针偏移量的差。

4、根据权利要求 1 或 2 所述的一种在网络终端缓冲区中处理音频流回放的方法，其特征在于：当 D 小于正常值范围的下限时，从环形缓冲区读出的数据长度小于正常播放长度，采用升采样处理。

5、根据权利要求 1 或 2 所述的一种在网络终端缓冲区中处理音频流回放的方法，其特征在于：当 D 大于正常值范围的上限时，从环形缓冲区读出的数据长度大于正常播放长度，采用降采样处理。

6、根据权利要求 3 所述的一种在网络终端缓冲区中处理音频流回放的方法，其特征在于：当读指针的偏移量小于写指针的偏移量时，读、写指针之间的距离 D 为写指针的偏移量减读指针的偏移量所得的差。

7、根据权利要求 3 所述的一种在网络终端缓冲区中处理音频流回放的方法，其特征在于：当读指针的偏移量大于写指针的偏移量时，读、写指针之间的距离 D 为环形抖动缓冲区的写指针的偏移量减读指针的偏移量所得的差在加上环形抖动缓冲区的长度所得的值。

8、根据权利要求 1 或 2 所述的一种在网络终端缓冲区中处理音频流回放的方法，其特征在于：该方法在实现方式上既可以通过硬件编码实现也可以通过软件编码实现。

一种在网络终端缓冲区中处理音频流回放的方法

技术领域

本发明涉及一种处理音频流的方法，尤其是指在包交换网络上使通话连续、流畅正常播放的在网络终端缓冲区中处理音频流回放的方法。

背景技术

当前，由于包交换网络的快速发展，原来主要承载在电路交换网络上的话音实时通信，开始大规模以包交换语音（VoIP）的形式转移到IP包交换网络上，最终会形成音频与数据及其它流媒体融合的网络。传统的电路交换网络与包交换网络存在一些不同特性，比如电路交换网络提供端到端的固定带宽并且独占的通信线路，不会出现包交换网络上存在的数据包丢失、乱序、时延抖动等问题；由于其传输带宽利用的高效率，网络融合的需求，以及组网与网络管理、扩容等的灵活性，决定了在包交换网络上传输音频信号的必要性，但是包交换网络上存在的数据包丢失、乱序、时延抖动等问题，会严重影响实时音频通信的质量，造成卡音、断音等不连续现象。

数据包的丢失对音频通信质量的影响主要体现在丢包率上，如果丢包率较低比如1%到2%，则听者不会感觉到明显的音频质量下降，但是随着丢包率的上升，听者会感觉到音频信号的断断续续或不连续，此时听者在感觉到少量断断续续时可能还能听懂对方的说话，但是大量的断续就会使听者听不懂对方在说什么了，从而造成通信终止。少量的数据包丢失可以通过利用音频信号的冗余性，即通过插值对丢失的音频数据包进行弥补，但是大量的数据包丢失，无论如何都会使音

频通信的质量下降。

乱序与时延抖动本质上都会造成需要连续、顺序播放的音频数据包在播放时间上的抖动；由于实时音频通信是一个连续的过程，每个音频数据包要在解码之后按固定时间间隔顺序播放，因此音频数据包在到来时间上的抖动会造成听觉上的不连续；如果抖动太大，则有的数据包由于来的太晚已没有播放的价值而被丢掉。

解决数据包抖动的方法之一是在接收端设置一个抖动缓冲区，在数据包到来时首先放入该缓冲区，在回放音频数据时，本着先进先出的原则，在时间上均匀地取出音频数据包并送往音频回放设备播出；只要该抖动缓冲区不被音频回放程序读空，则可以保证音频流的连续回放。该抖动缓冲区越大，则可平滑的抖动越大，但是过大的抖动缓冲区会造成音频回放时延的加大，过大的时延也是不可取的，会造成实时通讯的困难；抖动缓冲区的大小可根据具体需求来设定，也可以根据网络状况动态调整。但是这样并不能有效保证音频流的连续播放。

发明内容

本发明的目的是提供一种通话质量好、并能维持实时通讯的一种在网络终端缓冲区中处理音频流回放的方法。

本发明是这样实现的：当包交换网络上传输来音频数据包进入抖动缓冲区时，计算读、写指针之间的距离 D；

在初始化过程中，设置读、写指针之间的距离 D 正常值的范围和一次从抖动缓冲区中读出的数据长度；

在运行过程中，当读、写指针之间的距离 D 正常时，音频流按正常速度回放；当读、写指针之间的距离 D 大于或小于正常值时，通过音频重采样单元对从抖动缓冲区中读出的大于或小于正常播放长度的数据进行重采样处理，使其恢复正常

的播放长度。

上述的抖动缓冲区为环形抖动缓冲区。

所述的读、写指针之间的距离 D 是为进入抖动缓冲区的读指针偏移量与写指针偏移量的差。

当 D 小于正常值范围的下限时,从环形缓冲区读出的数据长度小于正常播放长度,采用升采样处理。

当 D 大于正常值范围的上限时,从环形缓冲区读出的数据长度大于正常播放长度,采用降采样处理。

当读指针的偏移量小于写指针的偏移量时,读、写指针之间的距离 D 为写指针的偏移量减读指针的偏移量所得的差。

当读指针的偏移量大于写指针的偏移量时,读、写指针之间的距离 D 为环形抖动缓冲区的写指针的偏移量减读指针的偏移量所得的差在加上环形抖动缓冲区的长度所得的值。

该方法在实现方式上既可以通过硬件编码实现也可以通过软件编码实现。

采用上述方法后,当对读、写指针之间的距离 D 是非正常值时,通过对从抖动缓冲区中读出的大于或小于正常长度的语音数据块进行重采样处理,从而加速或减慢播放速度,最终使环形抖动缓冲区中的读指针、写指针始终保持一定的距离,保证了包交换网络上传输来的实时音频流的连续流畅播放,使通话连续、流畅,减少了卡音、断音等不连续现象。

附图说明

下面结合附图和具体的实施方式对本发明作进一步详述。

图 1 是包交换网络音频通讯示意图;

图 2 是音频通信终端的发送部分示意图;

图 3 是音频通信终端的接收部分示意图；

图 4 是环形抖动缓冲区示意图。

具体实施方式

如图 1 所示，每个终端通过某种接入方式连接到包交换网络之中，每个音频终端可以通过该包交换网络向另一个音频终端发送或接收音频数据包；多个音频终端也可以通过某种形式组成一个可以多方通话的会议网络。

如图 2 所示，音频输入信号首先送到音频采集单元，该音频采集单元完成音频信号从模拟到数字信号的转换，也就是量化过程，该处理过程一般将音频信号量化为 16Bit 精度的有符号数字信号，之后送入音频编码器进行数据压缩以节省网络带宽，音频信号经编码压缩之后，送入打包传送单元，数据打包单元一般将用于实时通讯的音频包按实时通信协议（RTP）标准进行封装，之后再封装为用户数据协议（UDP）包，最后打入互连协议（IP）包传送到网络上。网络接口单元一般是网络层中的物理层及数组链路层，如以太网接口芯片或调制解调器等，经过压缩后的音频数据包最后经过网络接口单元传送到包交换网络上。

在音频数据包到达目标接收终端之前，要通过一系列的网络传输单元；这可能包括多种交换设备和路由设备，不同的路由及网络状况会产生不同的传输时延，从而造成按等时间间隔顺序传送的音频数据包会在时间上非均匀地到达接收端，这就造成数据包接收的抖动；另外按等时间间隔顺序传送的音频数据包有可能在不同的路由上传送，比如顺序发送的数据包 p1,p2,p3, ... 在接收端的顺序可能变成 p1,p3,p2, ... 。

如图 3 所示；音频数据包经过包交换网络传输之后到达接收端的网络接口单元，网络接口单元将收到的数据包拆去物理层地址等信息后还原成互连协议（IP）数据包；该数据包接着被送入音频流接收及拆包单元，在该处理单元中拆去互连协

议 (IP) 包头信息、用户数据协议 (UDP) 包头信息、实时通信协议 (RTP) 包头信息，最后还原为音频数据包；还原后的音频数据包被送入音频包解乱序单元，解出按正常时间顺序排列的数据包；按正常时间顺序排列的数据包接着被送入音频解码单元，解出音频信号的线性码；解出的音频信号的线性码被连续写入环形抖动缓冲区暂时存储；由于音频数据包经过网络传输之后出现的抖动，使音频数据流往环形抖动缓冲区的写入操作在时间上是非均匀的，但音频信号的回放在时间上要求是连续且均匀的，所以从环形抖动缓冲区读出音频数据并回放的操作与往环形抖动缓冲区的写入操作时非同步的。

如下详细描述音频数据在环形缓冲区中的处理及回放过程，包括图 3 中的抖动缓冲器，抖动缓冲器读出单元，抖动缓冲器读出控制单元，音频重采样单元，音频回放单元；图 3 中实线为数据流，虚线位控制流。

如图 4 所示，抖动缓冲区实际上是一个长度为 N 的连续存储空间，该存储空间的起始地址用偏移量 0 表示，结束地址用偏移量 N-1 表示，当前的写入地址指针用偏移量 W 表示，当前的读出地址指针用偏移量 R 表示，当前的写入指针与读出指针之间的距离用 D 表示；对该抖动缓冲区的写入操作为对当前写指针指向的存储位置写入音频数据，之后写指针偏移量加 1 并对抖动缓冲区的长度 N 取余运算，即用 C 语言可以表示为每写入一个数据后 $W=(W+1)\%N$ ；对该抖动缓冲区的读出操作为每读出一个数据单元后读指针偏移量加 1，同样用 C 语言可以表示为每读出一个数据后 $R=(R+1)\%N$ ；这样的读写操作可以保证在读写指针到达抖动缓冲区的顶部时，也就是偏移量 N-1 时，下一次读或写操作会自动翻转到抖动缓冲区的底部，也就是偏移量 0；这样的操作实际上相当于将该抖动缓冲区的顶部与底部相接，组成了一个环形缓冲区，读写指针按着相同的方向（顺时针或逆指针）旋转；为了保证读写指针之间不发生冲突，即读指针要保持始终跟随在写指针

的后方，这是通过读写指针之间的距离 D 来判断的；如果 W 大于 R 则 $D=W-R$ ；如果 W 小于 R 则 $D=N+W-R$ ；由此只要 D 大于 0 则可以保证读指针始终跟随在写指针的后方，同时读写指针之间的距离 D 作为控制信号送入抖动缓冲区读出控制单元，实际上读写指针之间的距离 D 相当于一个闭环控制系统中的误差控制信号，作为对抖动缓冲区读出操作的依据。

抖动缓冲区读出控制单元得到读写指针之间的距离 D 之后，根据 D 的大小来决定每次音频回放设备需要回放数据时读出的音频数据块的大小；音频回放单元一般是按等时间间隔（比如 30 毫秒）需要一定长度的音频数据块进行回放，如果音频信号的采样率为 8000 个采样点每秒（8k/s）则每 30 毫秒需要长度为 240 个采样点的音频数据块；若 D 的大小在正常的范围内，则抖动缓冲区读出控制单元控制抖动缓冲区读出单元每次在音频回放单元需要数据时读出正常长度的一块数据（比如 240 个音频数据），而此时音频重采样单元对此数据不进行任何操作，透明地将该块数据送到音频回放单元播放。

如果读写指针之间的距离 D 超过正常范围，如过大或者过小则需要进行必要的调整，因为如果 D 过小，则可能发生由于数据包到来时间的随机抖动造成该抖动缓冲区被不时读空的情况出现，从而造成 W 小于 R 的情况出现，也就是出现读指针比写指针在环形缓冲区中跑的快的情况，这时就会出现断音与卡音；如果 D 过大，由于上述相同的原因，可能出现 W 绕过 R 一圈的情况出现，这时也会出现断音与卡音。

另外，网络传输上总会出现数据包丢失的情况，从长时统计平均的意义上来讲 D 会因此随着时间的推移越来越小；还有一个因素会影响 D 的变化，如果音频发送端的采样时钟频率不同于音频接收端的播放时钟频率，就会出现如下情况，即当音频发送端的采样时钟频率大于音频接收端的播放时钟频率时， D 会因此随着时间

的推移越来越大，当音频发送端的采样时钟频率小于音频接收端的播放时钟频率时，D 会因此随着 时间的推移越来越小。

假设上述综合原因允许 D 在大于 d1 小于 d2 的范围内变化，也就是说当 $d1 < D < d2$ 时，抖动缓冲区读出单元在相应的控制单元的控制下每次从抖动缓冲区中读出长度为 L 的音频数据块；当 $D < d1$ 时，表示该抖动缓冲区很可能将要被读空，此时控制单元应控制抖动缓冲区读出单元在音频回放单元需要下一帧数据时，读出长度小于正常长度 L 的一帧数据，假定其长度为 L1 ($L1 < L$)；此时音频重采样单元就将长度为 L1 的音频数据块通过符合音频感知特性的插值运算变为长度为 L 的音频数据块，保证音频回放单元有标准长度的音频数据回放；这样相当于减慢了该帧音频数据的回放速度，只要 L 与 L1 的差值不是太大（比如 $((L-L1)/L) < 1\%$ ）则主观听觉上不会有可感知的变化；由于减慢了音频数据的回放速度，可以预知 D 会因此越变越大，当 D 回到正常的范围内时 ($d1 < D < d2$) 就可以按正常的速度进行音频的回放了。

当 $D > d2$ 时，表示该抖动缓冲区很可能将要被写满，此时控制单元应控制抖动缓冲区读出单元在音频回放单元需要下一帧数据时，读出长度大于正常长度 L 的一帧数据，假定其长度为 L2 ($L2 > L$)；此时音频重采样单元就将长度为 L2 的音频数据块通过符合音频感知特性抽取运算变为长度为 L 的音频数据块，保证音频回放单元有标准长度的音频数据回放；这样相当于加速了该帧音频数据的回放速度，同样只要 L 与 L2 的差值不是太大（比如 $((L2-L)/L) < 1\%$ ）则主观听觉上不会有可感知的变化；由于加快了音频数据的回放速度，可以预知 D 会因此越变越小，当 D 回到正常的范围内时 ($d1 < D < d2$) 就可以按正常的速度进行音频的回放了。

如果数据包到来时间的抖动范围太大或者丢包率太大，则有可能造成环形缓冲区被读空或者写满，此时可以做异常处理，在该缓冲区被读空时，可以重放上一帧

语音数据，若紧接着的下一次读操作时缓冲区仍为空，则播放静音信号；如果环形缓冲区被写满，则自动冲掉该缓冲区中的所有未播放数据，并重新开始正常的对该缓冲区的正常读写操作。

该方法在实现方式上既可以通过硬件编码实现也可以通过软件编码实现。

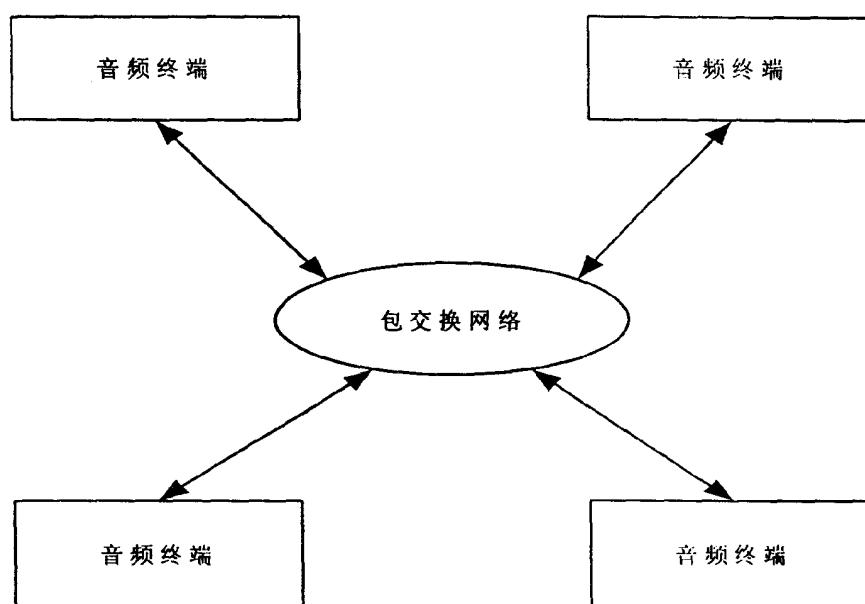


图 1

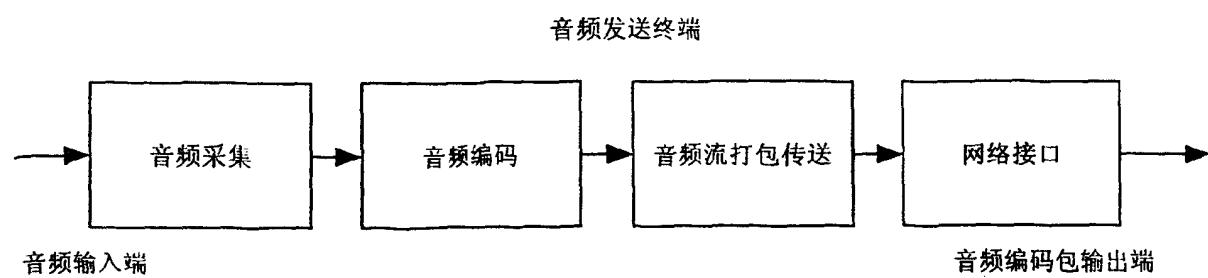


图2

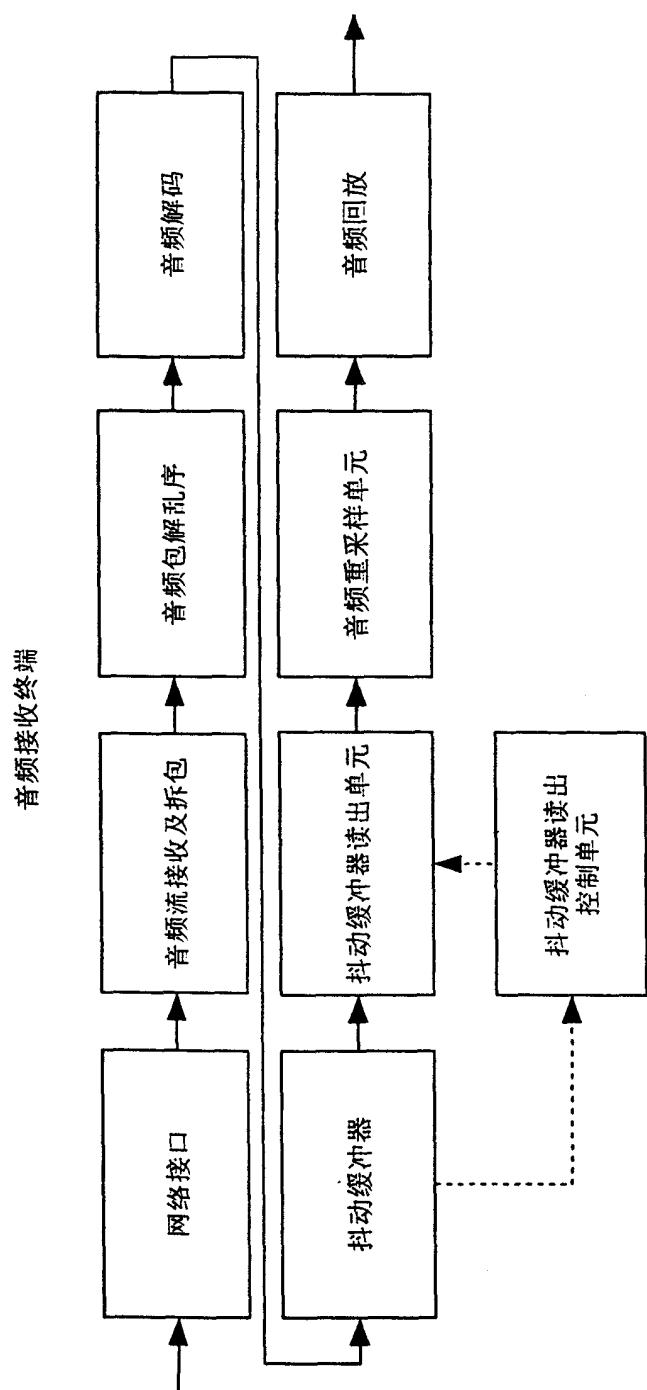


图 3

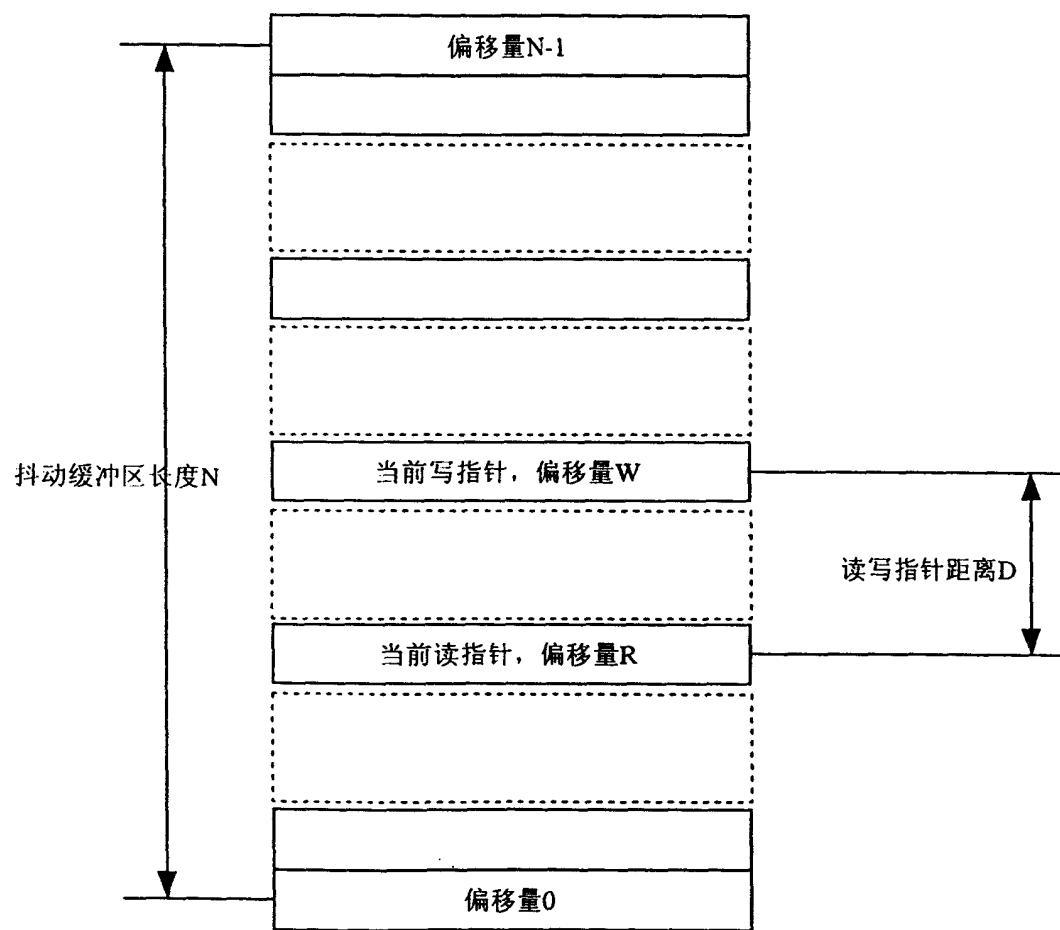


图 4