

基于 AMR 的抗分组丢失主次型语音编解码器

贾龙涛, 鲍长春

(北京工业大学 电子信息与控制工程学院, 北京 100022)

摘 要: 在基于分组传输的实时网络语音通信中, 分组丢失不可避免, 这也造成了语音质量的恶化. 本文提出了一种基于 AMR 的主次型语音编解码方法, 可以有效减少分组丢失对语音质量的影响. 仿真实验证明, 这种方法明显改善了 IP 网络语音通信的质量.

关键词: 语音编码; 网络通信; 分组丢失; 通信质量

中图分类号: TN912. 3 **文献标识码:** A **文章编号:** 0372-2112 (2005) 04-0697-04

Main-Sub Speech Codec for Preventing Packet Loss Based on AMR

JIA Long-tao, BAO Chang-chun

(School of Electronic Information and Control Engineering Beijing University of Technology, Beijing 100022, China)

Abstract: In packet-based real-time network, packet loss can not be avoided, which may greatly damage the speech quality. This paper proposed a main-sub speech codec method based on AMR for reducing influence of the packet loss. The simulation results show that the speech quality over the IP network was improved efficiently with the proposed method.

Key words: speech coding; network telecommunication; packet loss; communication quality

1 引言

自适应多速率语音编码 (AMR) 包括窄带自适应多速率编解码 (AMR-NB)^[1] 和宽带自适应多速率编解码 (AMR-WB)^[2]. 作为 ETSI 和 3GPP 的标准, AMR 语音编解码算法是第三代移动通信的核心编码算法, 主要应用于移动通信. 2002 年, AMR-WB 被 ITU-T 采纳为标准 G. 722. 2, 这样无线通信和有线通信就可以采用同一种编解码标准. 同一标准的应用, 使得无线通信和有线通信接口处不用再进行码流的转换, 总的延迟和复杂度大大降低, 语音通信质量得到提升. 因此, 将 AMR 语音编解码器应用在 IP 网络传输系统中, 是非常具有实用性的. 系统可以根据网络环境自适应地改变发送速率, 能够有效地减少连续分组丢失, 但是针对每一种速率的单一分组丢失却没有很好的解决方法.

在文献[3]中介绍了很多分组丢失补偿的方法, 静音替代和波形替代等方法, 在应用中比较简单, 复杂度很小, 但是当分组丢失率比较高时, 重建语音质量不高. 增加冗余度是一种比较好的补偿技术, 也是现在这一领域研究的热点问题, 出现了很多基于增加冗余度的分组丢失补偿方法. 这种方法可以大幅度提升分组丢失重建语音的质量, 但是同时也大大增加了传输数据量和计算复杂度, 对网络带宽和延时要求较高.

本文基于 AMR-NB 的 4. 75kbit/s 模式, 设计了一种主次型

编解码器. 编码器由三部分组成: 主编码器、次编码器和选择器. 解码器由两个部分组成: 主解码器和次解码器. 这种主次型编解码器能够较好的适应 AMR 编码特点, 在分组丢失补偿方面表现出很好的鲁棒性, 计算复杂度也相对较低.

表 1 4. 75kbit/s 模式编
码参数表

Parameter
LPC 1
LPC 2
LPC 3
LTP_LAG 1
POS 1 ... 1 ... 2
SIGN ... 1 ... 1 ... 2
GAIN_VQ 1
LTP_LAG 2
POS 2 ... 1 ... 2
SIGN ... 2 ... 1 ... 2
LTP_LAG 3
POS 3 ... 1 ... 2
SIGN ... 3 ... 1 ... 2
GAIN_VQ 3
LTP_LAG 4
POS 4 ... 1 ... 2
SIGN ... 4 ... 1 ... 2

表 2 速率模式对应表

MR475	4. 75kbit/s
MR515	5. 15kbit/s
MR59	5. 9kbit/s
MR67	6. 7kbit/s
MR74	7. 4kbit/s
MR795	7. 95kbit/s
MR102	10. 2kbit/s
MR122	12. 2kbit/s

2 AMR 编码器传输的参数

(1) 编码参数

在 AMR 中, 因为速率的改变, 编码参数的数目也不同, 其中最低速率 4. 75kbit/s 编码参数有 17 个; 最高速率 12. 2kbit/s 编码参数有 57 个. 其他速率的编码参数个数在 17—57 之间. 表 1 为 4. 75kbit/s 编码参数表.

(2) 帧类型信息

如果应用非连续传输技术(DTX),语音帧在编码端有下面四种类型:TX. SPEECH 表示正常的语音帧;TX. SID. FIRST 表示正常语音帧后的第一个静音帧;TX. SID. UPDATE 表示 DTX 系统中的静音更新帧;TX. NO. DATA 表示 DTX 系统中的空帧.

(3) 模式信息

不同的编码速率表示不同的模式,AMR 应用中需要传输这些信息,模式对应情况如表 2 所示.

3 主次型编解码算法

Toru Morinagan 等在文[4]中给出了一种带有子编码器(subr_codec)的语音分组丢失补偿方法,取得了较为理想的效果,但是针对 AMR 编码的特点,这种方法已经不适用,需要进行改进,特别是在分类器部分.

3.1 主次型编码器

主次型编码器的设计思路是:系统编码器包括两个编码器,一个是主编码器,另一个是次编码器.主编码器采用一种低压缩高速率的编码方式(在本文中应用 AMR 中的 MR122 模式),次编码器采用一种高压压缩低速率的编码方式(本文采

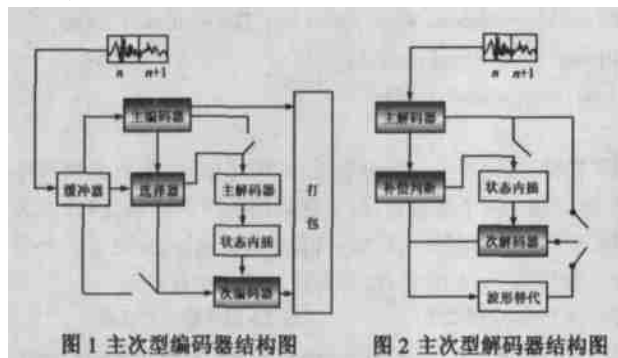


图1 主次型编码器结构图

图2 主次型解码器结构图

用 AMR 中的 MR475 模式).主次型编码器原理如图 1 所示.缓冲器首先缓存两帧(N 和 $N+1$) 的语音信号,第 N 帧语音信号送到主编码器进行编码,然后启动编码选择器,根据阈值判断是否需要附加冗余信息,如果不需要冗余信息,直接将主编码器参数打包输出,如果需要冗余信息,将第 $(N+1)$ 帧语音信号送到次编码器进行编码,然后把主编码器参数、次编码器参数联合打包输出.

本文给出的主次型编码器与传统的分组丢失冗余信息补偿方法有较大的不同.传统方法没有编码选择器,直接对第 N 和第 $N+1$ 帧编码,除了编码信息打包外,还要将编码器的内部状态信息(滤波器状态信息)进行保存,因此传统方法的平均速率较大,不利于传输应用.主次型编码器根据每帧的语音特性首先通过编码选择器判断是否需要附加下一帧的语音信息(下面一节详细介绍选择器),如果需要,就应用本地主编码器合成第 N 帧语音,利用合成的语音对第 $N+1$ 帧进行编码,这样就不需要传输编码器状态信息,在解码端如果分组丢失,可以根据第 N 帧的解码语音来对 $N+1$ 帧进行解码.因为主编码器采用低压缩高质量的编码器,解码语音质量与原始语音质量差别不大,因此对次编码器编解码语音质量影响不是很大.根据实验:如果主编码器采用 AMR 中的 MR122 模式,次

编码器采用 AMR 中的 MR475 模式, MOS 分满分为 5 分,次编码器应用原始语音编解码后 MOS 分约为 3.775,如果次编码器利用主编码器的解码语音编码,解码后语音 MOS 分约为 3.589.从实验也可看出,两种形式恢复语音质量差别不大,听觉上几乎察觉不出来,但是主次型编码器不用传输滤波器的状态信息,这可以很大程度上降低平均编码速率.

在开环基音搜索中语音信号 $s(n)$ 通过加权滤波器 $W(z) = A(z/Y_1)/A(z/Y_2)$ 后,得到加权滤波信号 $s_w(n)$:

$$s_w(n) = s(n) + \sum_{i=1}^{10} a_i Y_1^i s(n-i) - \sum_{i=1}^{10} a_i Y_2^i s_w(n-i) \quad (1)$$

如果按照传统的冗余补偿方法,在分组中需要保存 $W(z)$ 的状态信息,在式(1)中就要保留 $s_w(n)$, $n = -9, -8, K, 0$,也就是上一帧加权语音的最后 10 个值.在主次型编码器中,如果分组 $(N+1)$ 丢失,在补偿过程中,可以直接利用解码器第 N 帧的加权语音,大大降低了传输速率.

3.2 主次型解码器

语音信号经过主次型编码器压缩编码,然后打包在网络中进行传输.在接受端,主次型解码器负责语音解码.这种解码器主要有三个部分组成:主解码器、次解码器、补偿判断器.如图 2 所示.

语音流到达解码端后,如果没有分组丢失现象,则语音流直接送到主解码器进行解码,输出解码语音.如果第 $N+1$ 帧语音分组丢失,则语音流同时也进入补偿判断器,根据编码端加载在语音流最后的补偿判决比特位信息进行判断,如果为 1 就把次要信息流送到次解码器,应用冗余编码参数进行解码,同时应用主解码器的信息更新次解码器中的滤波器状态.如果补偿判断器的输出为 0 值,则不应用次解码器,直接应用简单的前一帧语音波形替代法对丢失的分组进行补偿,最后,输出补偿后的波形.原理如图 3 所示,主次型解码器不存在解码延时现象.

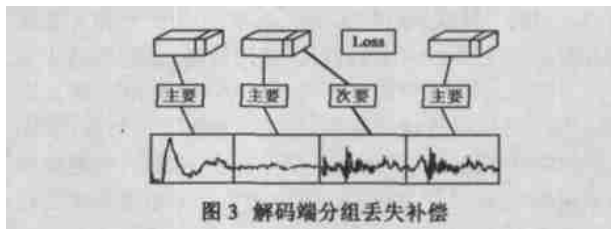


图3 解码端分组丢失补偿

3.3 冗余信息选择器

在主次型编码器(图 1)中,需要用到一个冗余信息选择器.选择器在本文介绍的分组丢失补偿方法中起着很重要的作用.冗余信息法可以很好的解决分组丢失语音补偿问题,但同时也造成平均传输速率比较高,有悖于压缩编码的基本原则.按照传统方法,冗余信息法会大大增加传输速率.例如主编码器采用 MR122,次编码器采用 MR475,采用传统方法后比特率比没有采用这种方法增加 38.9%.因此,如何根据语音和不同编码算法特性,开发一种高效的冗余信息选择器来决定语音帧是否需要加载冗余信息,是冗余信息补偿法能够应用的关键所在.

(1) MR 编码算法特性

AMR 编码算法每帧长度为 20ms,分四个子帧,每个子帧

长 5ms. 选择器就是判定帧 $N+1$ 跟帧 N 的相似度,同时还要判定帧 $N+1$ 自己是否有突变(基音或则能量),来决定是不是需要附加冗余信息. 在文献[5]中,作者根据两帧语音的信噪比值作为是否加载冗余信息的标准,这种方案主要基于分组长度为 10ms 的编码方法,不容易实现,并且计算复杂度较高. AMR 编码算法,帧长较大(20ms),在一帧内语音发生突变(基音和能量)的概率大大增加,所以应该同时考虑帧内变化和帧间变化,防止错判.

(2) 选择器设计

为了保证编码的实时性,采用的判决方法要有小的计算复杂度和高的判决准确度,同时判决参数应该易于提取. 选择器设计如图 4 所示.

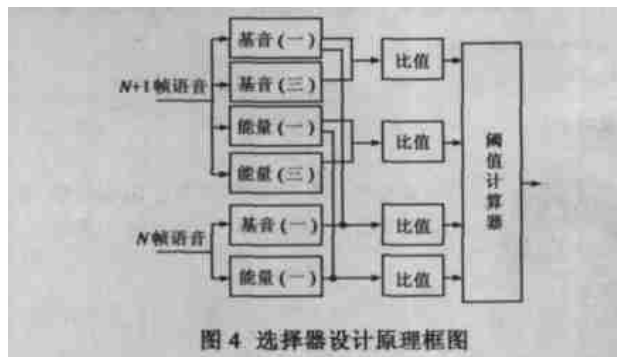


图 4 选择器设计原理框图

在编码过程中提取第 N 帧中第一子帧的基音值,计算第一子帧的能量值;提取第 $N+1$ 帧中第一子帧、第三子帧的基音值,计算第一子帧、第三子帧的能量值. 能量计算公式如下:

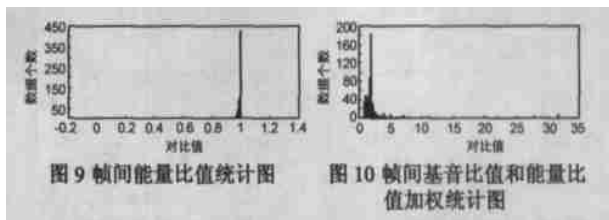
$$E = \sum_{n=0}^{39} s^2(n) \quad (2)$$

然后,求得第 $N+1$ 帧中一、三子帧基音的比值 P_1 、一、三子帧能量比值 E_1 ;求得 $N+1$ 帧和 N 帧中第一子帧基音的比值 P_2 , $N+1$ 帧和 N 帧中第一子帧能量比值 E_2 . P_1 、 P_2 、 E_1 、 E_2 输入阈值计算器,输出值作为冗余信息判决值.

(3) 判断阈值

上面给出了判决器的原理图,可以看出阈值对判定准确度有很大的影响,图 5~10 给出了语音长度为 20s 的一组实验数据统计图,图中横轴为对比值,纵轴为数据个数. 在图 5、图 6、图 8、图 9 中, P_1 、 E_1 、 P_2 、 E_2 值应该在 1 附近. 因为基音的影

响要大于能量的影响,本文用加权后的值 w_1 、 w_2 作为选择器判决的阈值. 其中: $w_1 = p_1 + 0.8 \times E_1$, $w_2 = p_2 + 0.8 \times E_2$. 帧内加权值统计值如图 7, 帧间的加权值如图 10. 表 3 是一段 12 秒语音的统计数据.



选择器在判断中,如果同时满足 w_1 [1.12, 2.5] 和 w_2 [1.12, 2.5], 选择器输出 1, 表示本帧不用附加冗余信息, 如果本帧丢失, 完全可以通过前一帧波形替代来补偿; 如果不能同时满足, 选择器输出 0, 表示本帧需要附加冗余信息, 如果本帧丢失, 用冗余信息来进行补偿.

从表 3 中可以看出

表 3 12 秒语音统计值表

数据范围	概率
$0.75 \leq p_1 \leq 1.35$	41.2 %
$0.75 \leq p_2 \leq 1.35$	32.8 %
$0.9 \leq E_1 \leq 1.2$	85 %
$0.9 \leq E_2 \leq 1.2$	91.2 %
$1.12 \leq w_1 \leq 2.5$	58.8 %
$1.12 \leq w_2 \leq 2.5$	55.3 %
$1.12 \leq w_1 \leq 2.5$ and $1.12 \leq w_2 \leq 2.5$	25.5 %
$0.8 \leq w_1 \leq 3.2$ and $0.8 \leq w_2 \leq 3.2$	40.7 %

4 性能比较

实验采用一段长度为 12 秒的男声语音, 主编码器采用 AMR 中的 MR122 模式, 次编码器分别为 AMR 中 MR475 模式, 选择器采用上文介绍的方法, w_1 阈值均为 [1.12, 2.5]. 在不

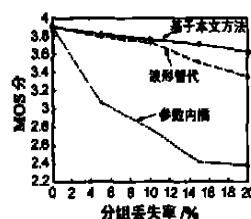


图 11 丢失补偿方法比较

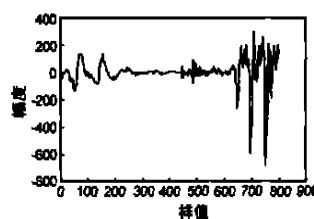


图 12 无丢失分组的解码语音信号

同的分组丢失率下, 主观语音质量如图 11 所示. 图中“基于本文方法”是应用本文介绍的主次型编解码方法, 主编码器为 MR122 模式, 次编码器为 MR475 模式; 波形替代法和参数内插法是前面介绍的常用分组丢失补偿方法. 从图中可以看出, 采用主次型编解码器后, 分组丢失恢复质量明显高于波形替

代方法和参数内插法,特别是当分组丢失率比较大时,优势更为明显.基于本文的方法,即使分组丢失率达到 15%,也不会对主观听觉质量造成太大的影响.上面 MOS 分采用 PESQ 方法测得^[6].

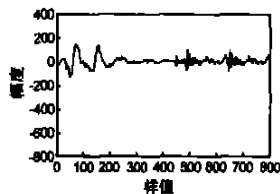


图 13 采用波形替代补偿方法

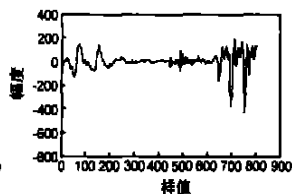


图 14 采用本文方法

图 13 和图 14 是采用波形替代法和主次型编解码法进行丢失补偿的一个例子.丢失分组通过判决器输出 0 值,也就是有基音或者能量突变.从图中可以看出,本文给出的方法在突变帧的补偿上有很大的优越性,关键是能够准确的判断出突变帧语音.

5 结论

本文介绍的主次型编解码方法是一种冗余信息分组丢失补偿方法.它有以下几个特点:(1)冗余编码($N+1$ 帧)利用前一帧本地解码语音信息,这样不用传送编码器状态信息,节约带宽.(2)应用了一种新的选择器能够更好的对每一帧语音状态进行判决,精确的判决能够更好的提高恢复语音质量.(3)能够根据对语音质量不同的要求和网络状态的好坏调整选择器阈值.根据实验可以得出:主次型编解码方法计算复杂度较小,判决参数易于获取,在分组丢失补偿中能够取得理想的结果,因此具有实际的应用性.

参考文献:

- [1] ETSI Draft En301704, Digital 的 Cellular Telecommunication System (Phase 2+); Adaptive Multi-Rate (AMR) Speech Transcoding[S].
- [2] Rec G.722.2, "Wideband coding of speech at around 16 kbit/s using

Adaptive Multi-rate Wideband (AMR-WB)"[S]. Tech Rep, ITU-T, Jan. 2002.

- [3] Colin Perkins, Oron Hodson. A survey of packet-loss recovery techniques for streaming audio[J]. IEEE/ACM Transactions on Networking, 1998, 11(5): 40-48.
- [4] T Morinaga, K Mano, et al. Robust speech coding under packet loss condition using recovery sub-codec for Broad-band IP network[A]. Proc IEEE ICASSP '2002[C]. Orlando, Florida: IEEE, 2002, III - 2713 ~ III - 2716.
- [5] T Morinaga, K Mano, T Kaneko. The forward-backward recovery sub-codec (FB-RSC) method: a robust form of packet-loss concealment for use in broadband IP networks, Speech Coding, 2002[A]. IEEE Workshop Proceedings[C]. Japan: IEEE, 2002, 62-64.
- [6] ITU-T Recommendation P. 862, Methods for objective and subjective assessment of quality[S]. 1996.

作者简介:



贾龙涛 男, 1978 年出生于山东曲阜, 北京工业大学硕士, 研究方向为 VoIP 语音通信.



鲍长春 男, 1965 年出生于内蒙古赤峰市, 博士, 教授、博士生导师, 中国电子学会理事, 信号处理学会委员, 通信学报与信号处理学报编委, 主要研究领域为数字信号处理、语音编码、数字通信和 DSP 应用等. E-mail: chchbao@bjut.edu.cn.