

一种引入多脉冲提高 AMR_WB 丢包处理质量的前向纠错技术^①

付焕荣, 孙咏, 贾军营

(中国科学院沈阳计算所, 沈阳 110171)

摘要: GSM AMR_WB 是唯一可以作为有线和无线通用的语音编码标准, 提高 AMR_WB 在丢包环境下的语音质量至关重要。本文提出了一种改进的 AMR_WB 丢包处理方法, 发送端利用基于多脉冲的前向纠错技术, 解决丢包时自适应码本带来的错误传播问题。接收端采用基音轮廓修改方法消除语音帧不同步在帧边界带来的恼人噪声。提出的方法在 AMR_WB 12.56kb/s 模式下实现, PESQ 客观评测结果表明, 此方法仅增加了 0.26kb/s 冗余比特率却明显提高了丢包情况下 AMR_WB 12.56kb/s 的语音质量。

关键词: 多脉冲; 前向纠错; 丢包; 基音轮廓修改

A Multipulse-Based Forward Error Correction Technique for Improving the Quality of AMR_WB in Lossy Network

FU Huan-Rong, SUN Yong, JIA Jun-Ying

(Shenyang Institute of Computing Technology, Chinese Academy of Sciences, Shenyang 110004, China)

Abstract: GSM AMR_WB speech codec being used for both Internet and mobile networks, robustness to both frame erasures and random bit errors assumes significance. This paper presents a high-performance concealment algorithm for AMR_WB speech. The proposed algorithm includes using Multipulse-based Forward Error correction technique at encoder to address the error propagation caused by the Adaptive codebook. And at the decoded side, to alleviate audible artifacts at the boundary between the previous and the current frame by modifying the pitch contour of the current frame. The algorithm has been implemented in the AMR_WB 12.56kb/s standard. Subjective tests have demonstrated that this improves perceived quality in the presences of frame erasures.

Key words: multi-pulse; forward error correction; packet loss; modification of the Pitch contour

1 引言

AMR_WB 是由 3GPP/ETSI 在 2001 年制定用于 WCDMA 和 GSM 的宽带语音编解码标准, ITU-T 在 2002 年将其选为 16KHZ 采样率的宽带语音编解码标准 G.722.2。这标志着无线与有线首次采用了同一个语音编码, 意味着 3G 和固定网之间的互通更为容易。

IP 网络先天并不是为实时媒体传输业务设计的, 它具有丢包、延时和抖动等传输特性, 严重影响着 VoIP 的服务质量。AMR_WB 标准^[1]本身具有丢包隐藏机制, 在丢包时产生丢失包的代替帧 (在本文为了描述

方便之后统称其为代替帧), 然而由于 AMR_WB 基于 CELP 模型, 相邻帧的参数之间具有很强的依赖性, 错误的代替帧导致之后的好帧也解码错误, 丢包不仅影响了丢失帧的质量而且影响了随后接收的好帧的质量^[2,3]。

为了减少代替帧对随后接收的好帧的错误传播, Global IP Sound 公司提出了帧间独立的编码方法 ILBC^[4]。然而这种帧间独立的编码为了得到与基于 CELP 的编码同样的质量不得不引入更多的比特率。M.Chibani 等人在 2007 年提出了一种新的方法, 编码

^① 收稿时间:2010-11-21;收到修改稿时间:2011-01-01

时约束自适应码本在周期信号中所占的比重,用固定码本表示部分周期信号。解码丢包后接收到的好帧时,先利用固定码本得到周期语音的一些信息,再使用这些信息同步自适应码本,从而一定程度上减少错误传播^[2]。但这种方法会使语音编码的质量在无丢包率的情况下有所下降。Angel M. Gomez*等人在文献[3]中提出了一种基于多脉冲的 FEC 方法,此方法利用多脉冲表示自适应码本内容,然后将多脉冲的振幅和位置作为冗余信息与当前帧一起传输,解码器处理丢包后接收的第一个好帧时,并不使用不同步的自适应码本产生当前帧的激励信号,而是用冗余的脉冲信息代替自适应码本产生当前帧的激励信号。但是此方法中每帧都传输冗余信息,且会在解码端的代替帧和第一个接收的好帧之间引入恼人噪声^[5]。本文是对方法^[3]的改进和扩展,对语音进行分类,根据当前帧语音类别决定当前帧是否携带冗余信息,只在当前帧为有声语音时才产生冗余信息并传递,从而减少冗余信息所占的平均比特率。在解码端发现丢包后,利用接收的脉冲位置和脉冲振幅解码丢包后接收的好帧,并利用基因轮廓修改方法处理代替帧和接收的第一个好帧之间的边界,进一步提高语音质量。

本文的结构如下:第2部分。简单介绍一下基于 ACELP 的 AMR_WB 编解码的基本思想及传统丢包隐藏算法的缺陷。第3部分介绍本文对文献[3]的改进和扩展。第4部分给出本文所提出的丢包隐藏算法应用到 AMR_WB 12.56kb/s.中的实验结果。最后在第5部分给出结论。

2 AMR_WB 编码原理

AMR_WB 语音编码采用代数码激励线性预测 (Algebraic-Code-Excited Linear-Predictions ,ACELP)技术。ACELP 基于合成-分析模型,选择使加权合成语音与原始语音误差最小的激励作为最终的激励信号。激励 $u(n)$ 被表示为自适应向量 $v(n)$ 和固定向量 $c(n)$ 的加权和

$$u(n) = g_p v(n) + g_c c(n) \quad (1)$$

其中, g_p 和 g_c 分别为自适应码本和固定码本的增益。自适应向量和固定向量分别从自适应码本和固定码本中得到。自适应码本旨在建模激励信号的长期相关性(浊音的声门脉冲和基音周期)。整个自适应码本由当前帧之前的激励信号通过长期预测滤波器(LTP)过

滤得到

$$v(n) = \sum_{k=-(q-1)/2}^{(q+1)/2} p_k u(n-T+k) \quad (2)$$

其中, T 为基音延时, p_k 为长期滤波器参数。固定码本表示的则是激励信号去掉长期周期性后得到的残差信号,它并不依赖于当前帧之前的激励信号^[1]。

当发生丢包后,标准^[6]中的丢包隐藏算法为了降低丢包造成的影响,根据丢包前的好帧产生丢失帧的参数,然后由参数产生丢失帧的代替帧。对于多个丢包的情况,丢包隐藏算法会逐渐削弱代替帧的能量,最后以静音帧为代替帧。丢包隐藏算法在产生代替帧的同时,利用产生的代替帧的激励更新自适应码本中的内容。由于代替帧中的语音与编码端编码的语音不同,产生的自适应码本与实际的自适应码本也不同,这便导致丢包后接收到的一个好帧解码后得到的语音与实际语音也不同,继而接收的第二个、第三个……好帧均发生错误,即自适应码本错误在帧之间传播。严重影响语音质量。

3 基于多脉冲FEC的丢包处理方法

Angel M. Gomez*等人在文献[3]中提出了一种基于多脉冲的 FEC 方法,这种方法每帧均携带冗余信息(激励脉冲位置和脉冲振幅),没有考虑语音分类的情况。本文提出的多脉冲 FEC 丢包处理方法是对文献[3]的改进和扩展,在编码端利用分类信息减少冗余信息所需的平均比特数。在解码端解决由代替帧和接收的第一个好帧之间的不同步引起的帧边界恼人噪声问题。编码端和解码端配合共同提高丢包环境下 AMR_WB 的语音质量。

3.1 编码端

语音信号有清音、浊音和静音之分。浊音是由声带振动产生的语音,在音节上具有高度的周期性。清音是由声带不振动产生的语音,在音节上不具有周期性。静音是不包含活跃语音的声音部分^[7]。由上一节的描述可知,自适应码本用于重建激励信号的周期部分。如果当前帧为浊音,自适应码本对当前帧的激励信号的贡献比较大,反之如果当前帧为清音,自适应码本对当前帧的激励信号的贡献很小。同样,由自适应码本引起的错误传播问题的严重性也因第一个接收的好帧在语音中的位置不同而不同,若第一个接收的好帧为清音,它无明显的周期语音部分,错误的自适

应码本对其影响很小。反之若第一个接收的好的帧为浊音，错误的自适应码本则会导致第一个好帧的解码也出现错误，当丢失的帧的类型为浊音的开始部分时，错误尤为严重。

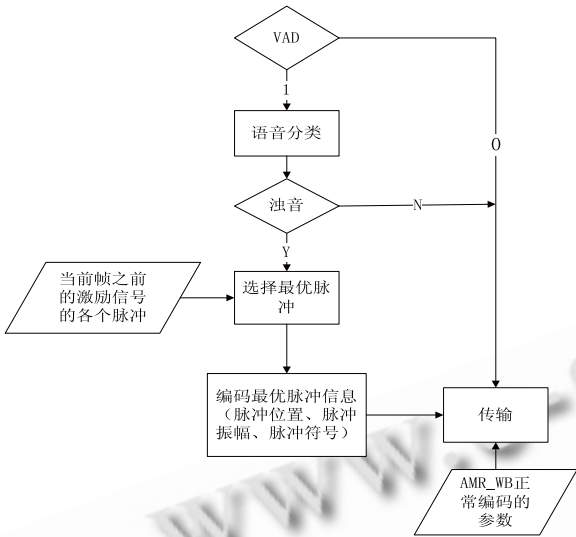


图 1 编码器编码时根据语音的分类信息选择性的携带冗余信息

基于以上的分析，编码端在文献[3]的基础上做了如下改进。在选择能够合成当前帧激励的最优脉冲之前先进行语音帧类型的判断，如图 1 所示，如果 VAD 检测为 1（活跃语音）则进行语音分类，语音分类信息为浊音的情况下再选择最优脉冲，并将前面激励的最优脉冲的位置和振幅与当前帧的其他参数一起传输。如果 VAD 检测为 0（非活跃语音），或 VAD 检测为 1，但语音为清音的情况下，当前帧均不携带冗余信息。

3.2 解码端

在解码端，若无丢包情况发生，则正常解码^[1]。发现丢包后，处理随后收到的第一个好的帧时，若其没有携带冗余信息，则利用丢包隐藏产生的自适应码本解码，若第一个好的帧携带有冗余信息（当前帧之前的激励脉冲），则使用冗余信息中的脉冲代替自适应码本产生当前帧的激励信号的周期部分，如图 2 所示。

按照上面的过程，用当前帧的冗余信息中的脉冲代替丢包隐藏算法更新的 ACB 产生当前帧的激励信号，可以很好的减少自适应码本对当前帧产生的错误，然而这种方法会引起另一个问题——丢包隐藏算法产生的代替帧的激励信号和利用冗余脉冲产生的第一个

好帧的激励信号不同步，如图 3 所示。

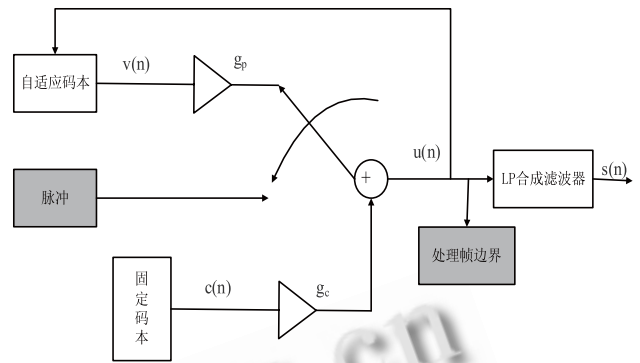


图 2 解码器的结构，灰色模块为新增的模块

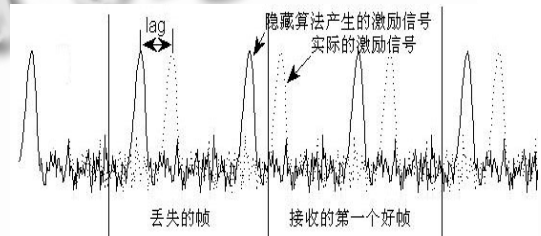


图 3 代替帧与随后接收的第一个好帧之间的不同步

这种不同步会在帧边界处引起恼人的噪音。我们称利用第一个接收的好帧解码得到的语音与代替帧语音之间的相位差为滞后——lag。如果 lag>0,则脉冲的峰值和脉冲的主要的语音特性会在帧边界处重复，如果 lag<0,则会使帧边界处的主要语音特性相互抵消^[5]。语音特性在帧边界重复的情况如图 4 (c) 所示。因此我们在解码器中增加了帧边界处理功能，首先判断第一个好的帧和代替帧之前的相位差，然后根据相位差，使用基音轮廓修改方法同步代替帧和第一个帧，消除帧边界处的恼人噪音。

3.2.1 得到代替帧和第一个好帧之间的滞后

若想利用基音轮廓修改算法消除帧边界噪音，首先要求得代替帧和第一个好帧之间的滞后。求解过程如下所示。如果当前帧含有冗余信息，则解码得到冗余信息：脉冲位置，脉冲振幅，和脉冲符号。(本文提出的算法在文献[3]的基础上将脉冲符号也作为冗余信息传递，因为同步脉冲时不考虑脉冲的方向会导致帧边界处语音特征相互抵消)。为了描述方便，称冗余信息中的脉冲为 PU(0),脉冲的峰值位置为 PS(0)。根据冗余信息中的脉冲 PU(0) 和固定码本产生的激励信号中的脉冲表示为 PU(1)、PU(2) … PU(n),代替帧中的脉

冲表示为 $PU(-n)$ 、 $PU(-n+1) \cdots PU(-1)$ 。如果 $PU(0)$ 的符号为正,则在自适应码本(ACB)中找到最大的正脉冲,如果 $PU(0)$ 为负,则在自适应码本(ACB)中找到最大的负脉冲。ACB 中最后的最大脉冲可以由下式得到

$$PS(-1) = \arg \max(|e_{ACB}(t)|) \quad (3)$$

其中, $-T(-1) < t < -1$, $T(-1)$ 为代替帧的最后一个子帧的自适应码本延时。其他的最大脉冲可以由下式得到

$$PS(-i) = PS(-1) - (i-1)T(-1) \quad (4)$$

$i=2, \dots, N_p-1$, N_p 为代替帧中的脉冲数目。

滞后 lag 则可通过在 ACB 中查找与 $PU(0)$ 最近的脉冲得到,假设 $PU(-k)$ 为离 $PU(0)$ 最近的脉冲,峰值位置为 $PS(-k)$ 滞后 lag 为:

$$lag = PU(0) - PU(-k) \quad (5)$$

如果 $lag=0$,表示代替帧和当前帧之间同步,不需要额外的处理,如果 $lag \neq 0$,则代替帧和当前帧之间不同步,需要应用基音轮廓修改方法消除不同步,从而避免恼人噪音的产生,如 3.2.2 所述。

3.2.2 基音轮廓修改方法

利用 3.2.1 描述的方法求得代替帧和第一个好帧之间的 lag 后,便可根据 lag 值对代替帧和当前帧进行同步,从而消除帧间噪音。基于不在编码器引入任何延时的考虑,在处理丢包后接收的第一个好帧时,代替帧已经输出到播放设备,我们只能移动第一个好帧的相位,使第一个好帧和代替帧同步,并且第一个好帧的相位移动不应影响随后帧的解码。基音轮廓修改方法首先应用在文献[8]中的编码器中,在编码器端通过基音轮廓修改方法,使帧中的各个子帧共用一个预先定义的自适应码本延时,从而降低编码的比特率。基音轮廓修改方法通过移动各个脉冲从而使整个子帧的相位移动,这种方法修改后得到的语音质量和原始的语音质量非常接近。我们可以在解码器端利用基音轮廓修改方法同步代替帧和第一个好帧的激励信号,消除帧边界的恼人噪音。

首先,将当前帧的激励信号 e 分成各个基音脉冲。基音脉冲的峰值位置和峰值(振幅)通过开环基音检测得到,相邻的两个基音脉冲的边界设在两个基音脉冲的中间^[8]。然后将代替帧和第一个好帧之间的滞后 lag (3.2.1 中得到)分配到各个脉冲中,且保证第二个好帧的第一个脉冲位置不变。令 $PS(i)$ 为未同步前第一个好帧中第 i 个脉冲的峰值位置, $\widehat{PS}(i)$ 为第一个好

帧与代替帧同步后第 i 个脉冲的峰值位置。

$$\widehat{PS}(i) = PS(i) - lag - \sum_{j=1}^i \Delta_j, i = 1, \dots, Q_p + 1 \quad (6)$$

其中 Q_p 为第一个好帧中样本总数。 Q_p+1 表示丢包后接收的第二个好帧的第一个脉冲的峰值位置。

按照公式(6)移动后的激励,相当于脉冲 $PU(0)$ 与代替帧中和它本身脉冲峰值位置最近的 $PU(-k)$ 对齐

$$\widehat{PS}(0) = PS(0) - lag = PS(-k) \quad (7)$$

Δ_i 为第 i 个脉冲相对于第 $i-1$ 个脉冲少移动的相位。

$\sum_{j=1}^{Q_p+1} \Delta_j = lag$ 。脉冲 $PU(Q_p+1)$ 的峰值位置保持不变,即丢包后接收的第二个好帧的第一个脉冲的峰值位置保持不变,

$$\widehat{PS}(Q_p + 1) = PS(Q_p + 1) \quad (8)$$

这样。经过基音轮廓修改后的激励信号为:

$$e_{final}(n + \widehat{PS} - \left\lfloor \frac{T(i)}{2} \right\rfloor) = e(n + PS - \left\lfloor \frac{T(i)}{2} \right\rfloor) \quad (9)$$

比如, $Q_p=4, lag=-6$ 然后 $\Delta = \{-2, -1, -1, -1, -1\}$, $\Delta_j \in \Delta$ 。则, p_1, p_2, p 相位分别为 4, 3, 2, 1, 0。如图 4 所示。

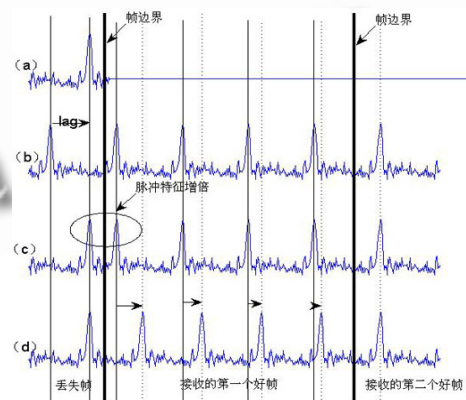


图 4 基音轮廓修改方法消除帧边界恼人噪音的实例 (a)代替帧的激励波形 (b)假若丢包未发生,解码端接收到的语音的波形 (c)使用冗余信息(脉冲)代替自适应码本产生的第一个好帧与代替帧连接形成的激励波形(可看出此波形存在明显的基音特征加倍现象) (d)使用基音轮廓修改后得到的激励信号波形

4 性能分析

本文搭建了一个模拟的网络环境,采用 PESQ 算法对提出的方法进行测试^[9]。实验中使用 TIMIT 语音

资料库作为测试的语音库。由于 PESQ 算法不适于测试很短的句子 (P.826 建议 PESQ 算法用来测试长度在 8-20 s 的语音), 实验中我们把 TIMIT 中的同一个人说的短句拼接起来组成新的长度在 9-14s 的语句, 并选择同样数目的男发音和女发音的语句作为我们最终的测试语音材料。实验中, 在编码器端对需要携带冗余信息的帧数进行统计, 统计结果表明, 以 TIMIT 为语音库, 需要携带冗余信息的帧数占总数的 43%。

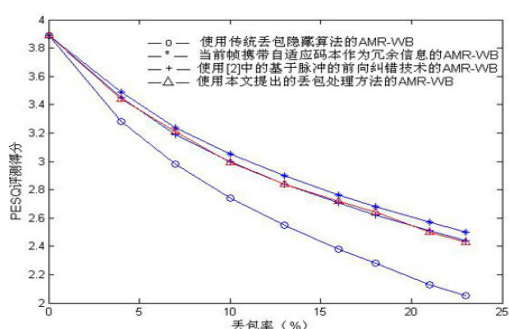


图 5 PESQ 评测结果

图 5 为不同丢包率下使用各种丢包处理方案 (传统丢包隐藏算法-标准 AMR_WB、文献[3]中提到的基于多脉冲的前向纠错技术、本文提出的丢包处理方案和每帧均携带自适应码本作为冗余信息)的 AMR_WB 的 PESQ 得分。在无丢包的情况下, 标准 AMR_WB 的 PESQ 得分为 3.89, 使用其他三种丢包处理方案的 AMR_WB 的 PESQ 得分与标准 AMR_WB 的得分相差无几。由此可见本文提出的丢包处理方法不会降低无丢包情况下编解码 AMR_WB 的语音质量。不管采用四种丢包处理方案中的那一种, 编解码的质量都会随着丢包率的增加而降低。相对来说, 每帧均携带自适应码本作为冗余信息的 AMR_WB, 在发生丢包时, 可用当前帧携带的自适应码本代替丢包隐藏算法更新的错误的自适应码本来产生当前帧的激励, 彻底解决了错误传播问题, 因此丢包情况下语音质量最高, 但是它的代价也最大, 冗余信息所占的比特率决不是 AMR_WB 所能接受的。文献[3]中提出的基于多脉冲的前向纠错技术, 若每帧携带一个脉冲的冗余信息, 丢包情况下的 PESQ 得分虽然没有每帧均携带自适应码本作为冗余信息的丢包处理方案的得分高, 但较标准 AMR_WB, 却要高出 0.2-0.4, 代价是 12.56kb/s 模式下, 增加 0.55kb/s 的冗余比特率。本文提出的丢包

处理方法, 只有浊音帧才会携带冗余信息, 增加 0.26kb/s 的比特率的情况下, 得到与文献[3]中方法非常接近的语音质量。且由于本文采用基因轮廓修改方法消除了帧边界恼人噪音, 丢包情况下语音质量有时还会高于文献[3]中方法的语音质量。

5 结论

针对编解码 ARM_WB 在丢包环境下鲁棒性不高, 丢包后错误传播的问题, 本文提出了一种改进的基于多脉冲的前向纠错技术, 以提高 AMR_WB 丢包情况下的语音质量。此方法包括利用语音分类技术减少基于多脉冲的 FEC 需要传输的冗余比特率, 利用基音轮廓修改技术消除帧间恼人噪声。总之, 本文在增加非常少的比特冗余的情况下, 明显的改善了编解码在丢包情况下的语音质量, 且不会在解码端引入任何的延时。

参考文献

- Wideband coding of speech at around 16 kbit/s using adaptive multi-rate wideband (AMR- WB). Geneva, January ITU-T Recommendation G. 722.2.2002.
- Chibani M, Lefebvre R, Gournay P. Resynchronization of the adaptive codebook in a constrained CELP codec after a frame erasure. Proc. ICASSP, 2006.13-16.
- Gómez AM, Carmona JL, Peinado AM, Sánchez V. A multipulse-based forward error correction technique for robust CELP-coded speech transmission over erasure channels In Proc. IEEE Signal Processing Society, pp.99, Aus.2009.
- Andersen S, Kleijn W, Hagen R, Linden J, Murthi M, Skoglund J. iLBC-a linear predictive coder with robustness to packet losses. Proc. Speech Coding, 2002, IEEE Workshop Proceedings. pp.23-25, October 2002.
- Zopf R, Thyssen J, Chen JH. Time-warping and rePhasing in packet loss concealment. Proc. Interspeech 2007-Eurospeech, pp.1677-1680, Antwerp, Belgium, August 2007.
- Adaptive multi-rate-wideband(AMR-WB) speech codec: error concealment of lost frames. Jun.2007, 3GPP TS 26.91.
- 胡航. 语音信号处理. 哈尔滨: 哈尔滨工业出版社, 2000.
- Tammi M, Jelinek M, Ruopilla V. Signal modification method for variable bit rate wideband speech coding IEEE Trans. Speech Audio Process. 2005, 13(7): 799-810.
- ITU-T P.862, Perceptual evaluation of speech quality (PESQ).