

VoIP 中一种信包丢失隐藏算法

谢青松, 魏 维, 罗 凯, 闫玲博

(西安通信学院一系, 西安 710106)

摘 要: 为减轻基于 IP 语音(VoIP)网络中因信包丢失而造成的语音失真, 提出一种基于双边线性预测和基音调整的包丢失隐藏算法。该方法利用丢失信包的前一信包或邻接信包(在后一信包可获得的情况下)预测丢失的信包。线性加权经过基音调整后的双边线性预测样点以获得最终的重建信号。重建信号在相位连续性上表现更加合理。经过 ITU-T P.862 协议推荐的 PESQ 算法测试证明, 该算法重建语音信号的质量有了较为明显的改善。

关键词: 基于 IP 语音; 信包丢失隐藏; 双边线性预测; 基音调整

Packet Loss Concealment Algorithm in VoIP

XIE Qing-song, WEI Wei, LUO Kai, YAN Ling-bo

(First Department, Xi'an Communication Institute, Xi'an 710106)

【Abstract】 This paper presents a two-sided linear prediction and pitch-adjusted based packet loss concealment algorithm to alleviate the speech distortion caused by lost packets in VoIP. This algorithm predicts lost packets from speech segments via either preceding, or both preceding and subsequent packets (if subsequent packet is available). Predicted samples, whose pitches are modified, are smoothed by linear weighting to obtain the final reconstructed signal which appears more reasonable in phase continuity. The assessment of resultant speech quality via objective measurement with PESQ algorithm from ITU-T P.862 shows that the algorithm significantly improves the performance.

【Key words】 VoIP; Packet Loss Concealment(PLC); two-sided linear prediction; pitch-adjusted

1 概述

信包丢失是基于 IP 语音(VoIP)中语音感知质量下降的主要原因。当信包不能按时到达接收端用于解码时, 语音的质量就会降低。为了减轻信包丢失对语音感知质量的影响, 很多算法被提了出来, 这些方法分为基于发端和基于收端 2 类^[1]。基于发端的信包丢失隐藏(Packet Loss Concealment, PLC)方法由收、发两端共同参与完成; 基于收端各类 PLC 算法根据收端正常接收到的信包、丢失信包编号以及预先知道的编码方式尽可能地恢复出原来的语音。基于收端的技术不需要发端数据的参与, 不会增加网络的带宽和时延。本文主要考虑基于接收端的信包丢失隐藏方法。基于收端常用的 PLC 方法有静音替代、前一信包重复、模板匹配、基音波形复制^[2]和线性预测^[3]。这些方法都没有考虑丢失信包的后续信包。在“尽力而为”型服务的 IP 网络中, 由于接收缓冲器的存在, 丢失信包后面的一个信包(即未来信包, 此后均用后一信包表示)经常可以获得。适当利用丢失信包的后续样点重建丢失信包, 在 VoIP 系统中是完全可以做到的。

本文采用双边线性预测技术, 用丢失信包的前一信包或前一信包及后一信包重建丢失信包。在此基础上, 当只有一个信包丢失时, 对丢失信包两边的语音进行清浊音判断。若两边都是浊音, 则同时考虑基音周期的变化, 对前、后向预测信号进行基音调整及双边加权获得重建信号。经过测试证明, 重建信号的质量有了较明显的改进。

2 双边线性预测

本文的基于双边线性预测和基音调整的 PLC 算法只在接收端实施, 算法的基本操作是从丢失信包的邻接信包中估计出丢失语音分段的线性预测系数 $a(k)$ 和激励信号 $e(n)$, 再通

过线性预测综合和基音周期调整重建丢失信包。

2.1 系统结构

图 1 给出了本算法的系统结构, 分为 5 部分: 前向线性预测, 后向线性预测, 基音调整, 幅度调整和线性加权。当丢失信包的后一信包接收到时, 实施双边线性预测; 反之, 只实施前向线性预测。预测信号产生后, 经基音调整、幅度调整和线性加权得到丢失信包的重建信号并输出。

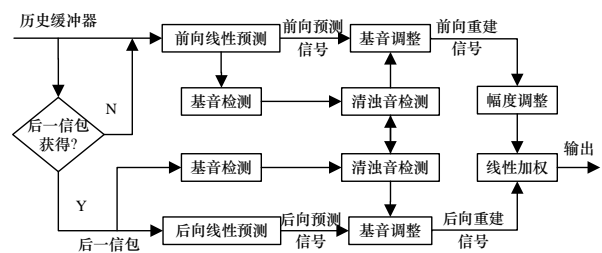


图 1 基于双边线性预测和基音调整的信包丢失隐藏系统结构

2.2 前向线性预测

前向线性预测是指以丢失信包前的信号为参考样点, 通过线性预测分析综合方法合成丢失信包的前向预测信号 $\hat{x}^f(n), 0 \leq n < N + 2D$ (N 为信包长度, D 为输出信号的延时), 具体过程见文献[3]。

2.3 后向线性预测

后向线性预测与前向线性预测类似, 不同之处主要是在基音检测和激励信号产生这 2 个模块。

作者简介: 谢青松(1982—), 男, 硕士, 主研方向: 语音信号处理; 魏 维, 教授; 罗 凯, 硕士; 闫玲博, 硕士研究生

收稿日期: 2008-08-13 **E-mail:** xqsmonkey@163.com

2.3.1 基音检测

在后向线性预测时,采用基于归一化自相关函数的方法估算基音周期。

$$P_b = \arg \max_{i=P_{\min}, \dots, P_{\max}} \left(\frac{\sum_{n=0}^{N-i-1} x(n)x(n+i)}{\sqrt{\sum_{n=0}^{N-i-1} x(n+i)x(n+i)}} \right) \quad (1)$$

其中, $P_{\min}=20, P_{\max}=120$ 分别为基音周期的最小值和最大值; $x(n), 0 \leq n < N$ 为丢失信包的最后一信包。

2.3.2 清浊音判断

在丢失信包的最后一信包可以获得的情况下,对前一信包和后一信包进行清浊音判断,判断方法如下:

在计算基音周期的同时,借助于语音信号的自相关函数求出基音周期的置信度 CP 。 CP 定义为 $x(n)$ 归一化自相关函数中异于最大峰值 1 的第 1 大峰值。对于浊音, P_b 是在 2.5 ms~15 ms 之间的一个数,并有较高置信度;对于清音,它是在这个范围之外的一个数,或者是在这个范围之内,但是置信度很低。如果 CP 比阈值 0.3 大,同时该信包的平均功率(均方根)也在阈值 0.5 之上,该信包被认为是浊音段,否则是清音段。

2.3.3 激励信号的产生

在后向线性预测综合时,通过周期性重复最后一信包最近一个基音周期的残差信号产生丢失信包的激励信号。

$$\hat{e}^b(n) = \begin{cases} e^b(N-L-P_b+n) & 0 \leq n < P_b \\ \hat{e}^b(n-P_b) & P_b \leq n < N \end{cases} \quad (2)$$

2.3.4 合成后向预测信号

激励信号 $\hat{e}^b(n)$ 与描述声道特性的 L 个线性预测系数 $a^b(k)$ 线性预测综合合成后向预测信号 $\hat{x}^b(n)$ 。

$$\hat{x}^b(n) = \sum_{k=1}^L a^b(k) \hat{x}^b(n-k) + \hat{e}^b(n) \quad 0 \leq n < N \quad (3)$$

在进行后向线性预测综合时,用最后一信包的前 L 个样点作为参考信号:

$$\hat{x}^b(n) = x^b(N+n) \quad -L \leq n \leq -1 \quad (4)$$

基音调整、幅度调整和线性加权的策略与信包丢失的类型有关,将在信包重建过程详细叙述。

3 信包重建过程

3.1 信包丢失前的正常接收帧

在收到语音帧 $x_j(n), 0 \leq n < N$ 时,为了隐藏将来丢失的信包要做 2 个准备:拷贝输出信号最后 B 个样点到历史缓冲器 $h_j(n), 0 \leq n < B$; 输出信号延时 D 个样点。在第 1 个信包丢失期间,历史缓冲器用于计算线性预测系数、产生残差信号、估计基音周期、生成激励信号和合成预测信号;算法延时用于重叠相加操作、平滑真实信号到重建信号的过渡。

为了进行实时处理,历史缓冲器更新如下:

$$\begin{aligned} h_j(n) &= h_{j-1}(n+N) & 0 \leq n < B-N \\ h_j(B-N+n) &= x_j(n) & 0 \leq n < N \end{aligned} \quad (5)$$

此后,接收信号经过 D 个样点延时后输出:

$$y_j(n) = \begin{cases} h_{j-1}(B-D+n) & 0 \leq n < D \\ x_j(n-D) & D \leq n < N \end{cases} \quad (6)$$

3.2 第 1 个丢失的信包

该信包的隐藏方法与最后一信包是否正确接收到有关,具体隐藏方法如下:

3.2.1 后一信包丢失

在该情况下采用前向预测信号作为丢失信包的重建信号

$$\hat{x}(n) = \hat{x}^f(n+D) \quad 0 \leq n < N \quad (7)$$

3.2.2 后一信包收到

当单个信包丢失,其邻接信包都收到时,实施前、后向线性预测,得到丢失信包的前、后向预测信号 $\hat{x}^f(n)$ 和 $\hat{x}^b(n)$,重建步骤如下:

(1)平滑过渡。为了平滑真实信号到重建信号的过渡,减轻丢失信包边界处的听觉失真,用历史缓冲器最后 D 个样点重叠相加前向预测信号前 D 个样点的结果以替换历史缓冲器最后 D 个样点。

$$\hat{h}(n) = \frac{D-n}{D+1} h(B-D+n) + \frac{n+1}{D+1} \hat{x}^f(n) \quad 0 \leq n < D \quad (8)$$

(2)双边线性加权。前、后向预测信号在各自的起始段,在一二个基音周期内都很好逼近了原始语音样点。重建信号可通过线性加权这 2 个预测信号获得。

$$\hat{x}(n) = \frac{N-n}{N+D+1} \hat{x}^f(n+D) + \frac{n+1}{N+D+1} \hat{x}^b(n) \quad 0 \leq n < N \quad (9)$$

(3)更新历史缓冲器。产生第 1 个丢失信包的重建信号 $\hat{x}(n)$ 后,用该信号更新历史缓冲器。

$$\begin{aligned} h_{j+1}(n) &= h_j(n+N) & 0 \leq n < B-D-N \\ h_{j+1}(B-D-N+n) &= \hat{h}(n) & 0 \leq n < D \\ h_{j+1}(B-N+n) &= \hat{x}(n) & 0 \leq n < N \end{aligned} \quad (10)$$

3.2.3 基音调整

由于语音信号在丢失信包前、后可能发生变化,因此在丢失一帧,且前一信包和最后一信包均为浊音时,考虑基音周期的变化重建丢失信包。根据语音信号的短时平稳特性,假设语音信号的基音周期在信包丢失内线性变化,当前一信包和最后一信包基音周期差值的绝对值小于阈值 T 时,对前、后向预测信号进行基音调整,调整方法如下:

(1)确定基音分段的调整个数。设前一信包和最后一信包的基音周期分别为 P_f, P_b , 前、后向预测信号中包含基音分段的个数分别为 N_f, N_b (不足一个时算一个), 则

$$N_f = \lceil \frac{N}{P_f} \rceil, N_b = \lceil \frac{N}{P_b} \rceil \quad (11)$$

取 N_f, N_b 中的较大者作为前、后向基音分段的调整个数 N_p 。

(2)确定每个基音分段的长度。以 $P_f < P_b$ 为例,确定每个基音分段的长度如下:

假设基音周期在丢失信包内线性变化,则基音周期的变化速率为 $P_b - P_f / N_p$ 。

在前向预测信号中,第 i 个基音分段的长度为

$$PF(i) = P_f + \frac{P_b - P_f}{N_p} \times i \quad 1 \leq i \leq N_p \quad (12)$$

在后向预测信号中,从后向前数第 j 个基音分段的长度为

$$PB(j) = P_b - \frac{P_b - P_f}{N_p} \times j \quad 1 \leq j \leq N_p \quad (13)$$

当 $P_f > P_b$ 时,每个基音分段的长度确定方法与此类似。

(3)基音分段调整。当变调后的基音周期 P 与原基音周期 P_0 之比为有理数时,可以通过在原信号样点之间线性内插,再等间隔抽取的方法实现。对前、后向预测信号进行基音分段调整后,信号的长度可能会大于帧长,超过帧长部分的信

号直接删去。

图2和图3给出了在浊音段单帧丢失且其邻接信包的基音周期发生变化时,基于双边线性预测的重建信号和在此基础上考虑邻接信包基音周期变化的重建信号。可以看出,考虑基音周期变化重建的信号与原信号非常接近,在峰值处吻合得较好。

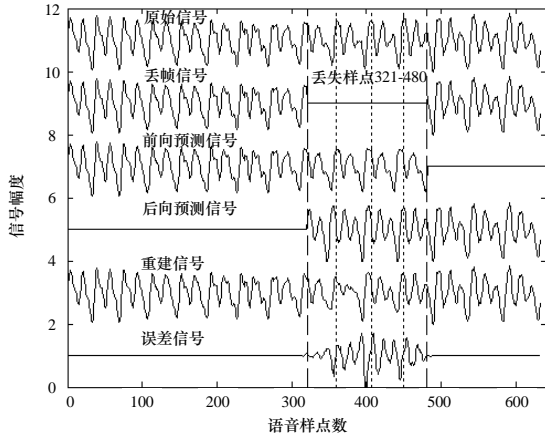


图2 基于双边线性预测的重建过程

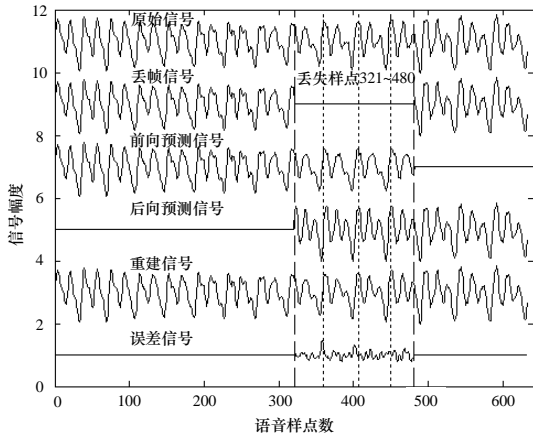


图3 基于双边线性预测和基音调整的重建过程

3.3 后续丢失的信包

(1) 后一信包丢失

在后一信包丢失的情况下,只采用经过幅度调整的前向线性预测信号作为丢失信包的重建信号。LP系数和基音周期与前一信包的值相同。

(2) 后一信包接收到

在收到后一信包时,线性加权经过幅度调整的前向线性预测信号和后向线性预测信号作为丢失信包的重建信号。

3.4 信包丢失后的第1个正常接收帧

为了平滑重建信号到真实信号的过渡,用目前信包前 D 个样点重叠相加前一信包前向重建信号最后 D 个样点的结果替代该信包的前 D 个样点。

$$x_j(n) = \frac{D-n}{N+D+1} \hat{x}_{j-1}^f(N+D+n) + \frac{n+1}{N+D+1} x_j(n) \quad 0 \leq n < D \quad (14)$$

3.5 幅度调整

保持第1个丢失信包对应前向线性预测信号的幅度不变;从第2个丢失信包开始,将前向线性预测信号的幅度每20ms线性衰减20%。

4 测试结果

衡量不同PLC算法性能最根本的原则就是比较重建语音

信号的质量。本文采用ITU-T推荐P.862中的PESQ^[4]算法作为重建语音质量的客观评价。它能够比较待测试语音信号与指定参考信号之间的听觉距离,提供一个客观语音质量打分PESQ MOS。其取值在-0.5~4.5之间,与主观语音质量打分MOS的关联程度非常高(0.935)。

测试语音取自ITU-T P.23^[5]数据库,两男两女讲英语,每人有46个长8s的句子,采样频率为16kHz,16比特编码保存。在测试句子时先读出原始记录,低采样到8kHz,然后按照ITU-T推荐^[6]进行PCM编解码,再对PCM解码信号进行语音激活检测。在语音活动区域按照指定的信包丢失率随机选择一些信包设置为丢失,将其中样点用零替换。

测试参数设置如下:历史缓冲器长度 $B=240$,信包长度 $N=160$,线性预测阶数 $L=10$,输出延时 $D=8$,基音调整的阈值 $T=15$,信包丢失率为2%~10%。

表1按照5种信包丢失率,对5种PLC算法(前一信包重复、模板匹配、ITU-T推荐G.711附录I^[2]、文献[3]的算法和本文算法)重建语音信号的质量进行了比较。

表1 重建语音信号质量(PESQ MOS)

信包丢失率/(%)	前一信包重复	模板匹配	文献[2]算法	文献[3]算法	建议方法
2	3.64	3.66	3.75	3.77	3.99
4	3.31	3.35	3.53	3.56	3.85
6	3.12	3.16	3.36	3.39	3.74
8	2.97	3.02	3.21	3.26	3.65
10	2.84	2.88	3.09	3.15	3.57
平均值	3.18	3.20	3.39	3.43	3.76

可以看出,本文算法重建语音的质量在所有信包丢失率下均优于其他算法。与文献[3]相比,本文算法在不同信包丢失率下的PESQ MOS得分平均增加0.37。

5 结束语

本文提出了一种基于双边线性预测和基音调整的信包丢失隐藏算法。该方法对丢失语音段两侧的邻接信包进行清浊音判断,当单个信包丢失,其邻接信包都是浊音段时,考虑基音周期的变化重建丢失信号。标准数据库中大量语音信号的测试证明,重建语音信号的质量得到了有效的提高。

参考文献

- [1] Perkins C, Hodson O, Hardman V. A Survey of Packet Loss Recovery Techniques for Streaming Audio[J]. IEEE Networks, 1998, 12(5): 40-48.
- [2] ITU-T Recommendation G.711 Appendix I. A High Quality Low-complexity Algorithm for Packet Loss Concealment with 711[EB/OL]. (1999-09-09). <http://www.itu.int/rec/T-REC-G.711-99909-1!AppI/en>.
- [3] Gunduzhan E, Momtahan K. A Linear Prediction Based Packet Loss Concealment Algorithm for PCM Coded Speech[J]. IEEE Trans. on Speech and Audio Processing, 2001, 9(8): 778-785.
- [4] ITU-T Recommendation P.862. Perceptual Evaluation of Speech Quality(PESQ), an Objective Method for End-to-end Speech Quality Assessment of Narrow-band Telephone Networks and Speech Codecs[S]. 2001.
- [5] ITU-T Recommendation P Supplement 23 ITU-T Coded-Speech Database[S]. 1998.
- [6] ITU-T Recommendation G.711. Pulse Code Modulation(PCM) of Voice Frequencies[EB/OL]. (1988-11-11). <http://www.itu.int/rec/T-REC-G.711-198811-1/en>.

编辑 张正兴