

一种基于 VoIP 的改进 WSOLA 丢包隐藏算法

杜水明 张 聪 王 赞 梁瑞凡

(武汉工业学院数学与计算机学院 湖北 武汉 430000)

摘 要 丢包现象严重影响 VoIP 的通话语音质量。WSOLA(Waveform Similarity Based Overlap-Add)算法是一种基于接收端的丢包隐藏方法,可以较好地提高语音质量。在介绍 WSOLA 算法原理的基础上,针对该算法中计算互相关系数所需计算量较大,会增加过多计算延时的问题,提出一种互相关系数计算的改进方法。最后通过仿真对重建语音信号质量进行了对比。

关键词 VoIP 丢包隐藏 WSOLA 互相关系数

中图分类号 TP39 文献标识码 A DOI:10.3969/j.issn.1000-386x.2014.10.064

AN IMPROVED WSOLA PACKET LOSS CONCEALMENT ALGORITHM BASED ON VOIP

Du Shuiming Zhang Cong Wang Zan Liang Ruifan

(School of Mathematic and Computer Science, Wuhan Polytechnic University, Wuhan 430000, Hubei, China)

Abstract The packet loss phenomenon has seriously affected the calling voice quality of VoIP. WSOLA algorithm is a receiving end-based packet loss concealment method, it can better improve the voice quality. This paper introduces the principle of WSOLA algorithm, and based on that it proposes an improved method for cross correlation coefficient calculation to solve the problem that the calculation load of cross correlation coefficient in WSOLA is heavy which leads to too much computation delay. In end of the paper the quality of reconstructed voice signal is compared through simulation.

Keywords Voice over Internet Protocol (VoIP) Packet loss concealment Waveform similarity based overlap-add (WSOLA) Cross correlation coefficient

0 引 言

在目前运用较广的网络电话 VoIP 当中,语音数据的分组(打包)传输在遇到网络阻塞或延时比较长时会出现语音包丢失现象,严重影响了通话的质量。针对 VoIP 网络丢包现象国内外主要采用丢包补偿策略来提高语音质量,而丢包补偿方法主要有两种形式:丢包恢复和丢包隐藏^[1]。丢包恢复重点在于从接收端收到的数据中,得到与原始输入语音信号没有差别的数据,通常是直接利用所传递的信息冗余来抗误码和干扰,达到降低接收端的实际丢包率的效果;丢包隐藏的重点则在于利用人耳的掩蔽特性,在接收端所收到语音信息中,选取一段或几段按照一定的逻辑仿造出已经丢失了的语音包^[2]。由于人耳在听声音的时候会在潜意识中对不完整和不正确的波形进行修补,因此使用丢包隐藏技术所恢复出来的波形,通常不会让人明显地感觉到错误,可以在很大程度上提升接收端人耳的听觉感受^[3]。

在目前的一些丢包隐藏算法中,波形相似叠加 WSOLA 算法^[4]恢复的语音质量可以取得比较令人满意的效果。但是,WSOLA 会引入一定的计算延时,尤其在计算其互相关系数时所

需计算量较大,增加计算延时过多,进而影响语音通话质量的提高。

1 VoIP 的基本原理

VoIP 是采用全球互联的 IP 网络环境来实现语音通信的,语音信号在 IP 网络当中的传输过程是:在发送端,输入语音信号为模拟语音信号,经过采样、量化、编码等步骤转换成数字信号。根据通信环境采用适当的压缩编码技术对该数字信号进行压缩,以应对网络传输的需求。压缩后的语音数据根据实时传输协议 RTP、用户数据报协议 UDP 和网络协议(IP)分组打包^[5],原始输入语音信号就形成了 IP 包的形式,最后就是通过 IP 网络将形成的数据包传输出去;在接收端,将接收到的数据包组合,经解码器解压缩处理后恢复成模拟语音信号,经声卡播放,实现 VoIP 通话。图 1 显示了 VoIP 实现语音通信的基本过程。

收稿日期:2013-03-11。国家自然科学基金项目(61272278);湖北省自然科学基金项目(2012FFB04801);武汉工业学院 2012 年研究生创新基金项目(2012cx007)。杜水明,硕士生,主研领域:语音信号处理。张聪,教授。王赞,硕士生。梁瑞凡,硕士生。

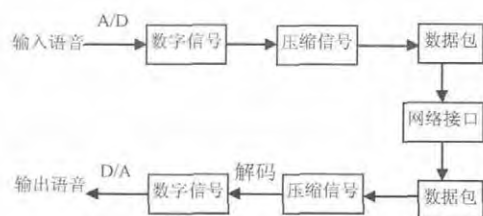


图 1 VOIP 语音信号传输过程

IP 网络提供的是尽力的、不可靠的数据传输服务,网络的负载和拓扑可能会动态变化,因此网络可能会出现拥塞,导致数据包不能有序并且全部到达接收端^[6]。从而在 VoIP 的实际通话过程中,一旦遇到网络阻塞,数据分组就会滞留在 IP 网络中的各个网络节点,而且 IP 网络为了解决网络拥塞,就会选择丢失分组。所以,当一个网络节点的分组数量超过其最大缓冲能力时就会有部分数据丢失—VoIP 网络丢包。在传输中某些语音包丢失,就会造成与此相对应的声音波形不能重建出来,那么在接收端重建的声音波形中就会有許多断裂处,使得最终听到的声音音质下降。

2 改进 WSOLA 算法

2.1 WSOLA 算法思想

丢包隐藏方法中的时域扩展技术是采用丢包处两端的语音波形通过一定的算法来填补丢失的语音信息^[7]。由于时域扩展方法采用了特定的算法来实现,因此不可避免地引入了一定的计算延时。WSOLA 也是一种在 VoIP 网络接收端应用的时域扩展算法,它将丢包处前面一段语音拉长来覆盖丢失了的语音。采用 WSOLA 这种时域处理的办法就会引入一定的计算时延,但它可以较好地保持原始输入语音信号的基音周期,使重建后的语音信号尽可能小的损失音质。因此,采用 WSOLA 丢包隐藏算法可以较好地解决由于丢包而带来的通话质量下降问题。

WSOLA 算法的基本思想如图 2 所示:选取丢失包前面几个正确接收到的数据包作为输入信号,从输入信号中分解出几个 L 样点长的语音段,再把分解得到的几个语音段按 Δy 的步长进行部分重叠相加。并且,在相加之前需要给每个语音段乘上一个和其长度相同的汉宁窗,以使重组后的语音段依旧能够保持正常的语音幅值。语音段的分解步长不固定,而叠加步长固定。同时为使重组语音的基频不变,WSOLA 通过计算互相关系数查找最相似波形来使两段语音在最佳位置重合相加^[8]。组合后的语音段相对于原语音断长度将会增加或减少,这由扩展常数决定。

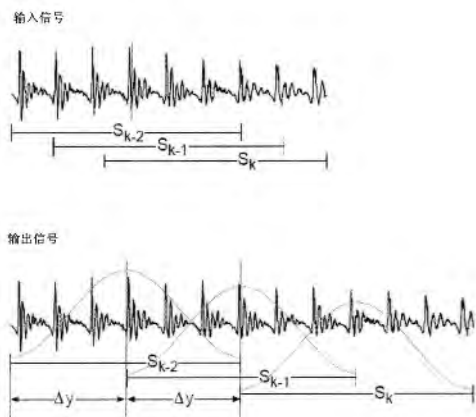


图 2 WSOLA 丢包隐藏原理

2.2 WSOLA 算法过程

对于 WSOLA 算法,其核心部分为计算互相关系数,找到最相似波形的起始位置,对语音段进行分解。对于语音段的分解如图 3 所示, S_{k-1} 语音段为已知参考语音段, S_{ki} 为备选语音段,它的个数取决于搜索区域的长度,WSOLA 算法要找到最相似波形也就是从所有的备选语音段中找出最合适的作为最终进行重叠相加的语音段 S_k 。

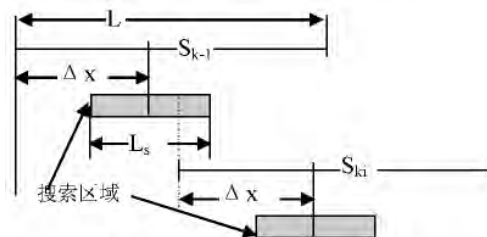


图 3 确定语音段分解位置

给定扩展常数 α , 则:

$$\Delta x = \Delta y / \alpha \quad (1)$$

其中, Δx 是语音段 S_{ki} 的起始位置搜索区域的中点与上一个语音段 S_{k-1} 的起始位置之间的距离, $\Delta y = L/2$, L 为分解语音段长度。 L_s 为搜索区域的长度,将搜索区域内的点作为起点可以得到 L_s 个备选语音段 S_{ki} ,计算每一个 S_{ki} 前半部分与 S_{k-1} 后半部分的互相关系数:

$$C_i = \sum_{j=0}^{\Delta y-1} S_{ki}(j) \times S_{k-1}(\Delta y + j) \quad (2)$$

其中, $i = 1, 2, \dots, L_s$ 。该公式表示 S_{ki} 前半部分与 S_{k-1} 后半部分对应位置的值相乘再求和。搜索区域当中的每一个位置对应一个互相关系数,以互相关系数取得最大值时的 S_{ki} 作为最终 S_k 语音段。再将已知的 S_k 作为参考语音段重复以上过程,分解完输入信号。

将 WSOLA 应用于丢包隐藏时,仅采用丢失包前面接收到的两个完好包的语音数据(如图 4 所示)来重建丢失包的语音,这样有助于减少计算延时。同时为得到语音段 S_1 以及其后的语音段,引入一个辅助语音段 S_0 ,可根据经验(式(3))确定所对应的 S_0 搜索区域的位置:

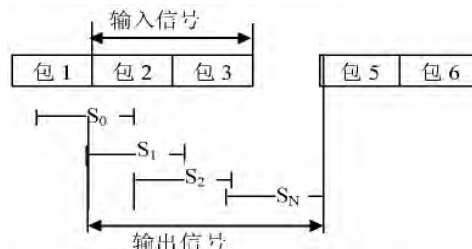


图 4 WSOLA 进行丢包隐藏

$$r_0 = L_s - 80 \left(\frac{L_{out}}{L_{in}} \right) \quad (3)$$

其中 L_s 是搜索区域的长度, L_{in} 为输入信号长度, L_{out} 为输出信号长度, r_0 为与 S_0 语音段所对应的搜索区域的初始位置。在寻找 S_0 的初始位置时,采用计算互相关系数的方法,用在搜索区域中得到的每个语音段与原始采用的语音中的对应位置分别相乘得出。

一般来说,为了使分解出的最后一个语音段的长度可以达到 L 个样点,将固定最后一个语音段的搜索区域在最少离输入信号最后一个样点还剩 L 个样点的位置。计算总语音段的段数

N 为:

$$N = \lfloor \frac{L_{out}}{100} - 1 \rfloor \tag{4}$$

为了使丢包处被输出的语音段覆盖, N 和 L 应满足:

$$\frac{L}{2}(N + 1) > L_{out} \tag{5}$$

由式(4)和式(5)可确定分解语音段的长度:

$$L = 2 \times \lceil \frac{L_{out}}{N + 1} \rceil \tag{6}$$

2.3 改进的互相关系数计算方法

原始 WSOLA 当中给出的互相关系数计算式(式(2))在寻找最相似波形时所需的计算量较大,尤其是当搜寻区域很大(即基频很低)或语音段长度过长时,会引入更多的计算延时^[9]进而降低了通话语音质量。因此,文章提出了一种新的相似波形寻找方法和互相关系数计算方法:根据原始互相关系数计算式中的 S_{k-1} 语音段后半部分最大值位置和备选语音段 S_{ki} 前半部分最大值位置之间的距离来计算互相关系数,两个最大值位置之间的距离越小, S_{ki} 语音段前半部分波形与 S_{k-1} 语音段后半部分波形越相似,互相关系数则越小。互相关系数取得最小值时的备选语音段即为要寻找的最相似波形。由此可得到改进后的互相关系数计算公式:

$$C_i = |L_k - L_{k-1}| \tag{7}$$

其中, $i = 1, 2, \dots, L_s$ 。式中 L_{k-1} 为参考语音段 S_{k-1} 中后半段最大值的位置到 S_{k-1} 中间位置之间的长度, L_k 为备选语音段 S_{ki} 前半段最大值的位置到 S_{ki} 初始位置之间的长度(如图 5 所示)。式(7)表示参考语音段 S_{k-1} 中后半段最大值的位置与备选语音段 S_{ki} 前半段最大值的位置之间的距离。在寻找相似波形时,将式(7)取得最小值时的 S_{ki} 作为最终的 S_k 语音段。改进的 WSOLA 算法过程与原始 WSOLA 算法过程相同,主要不同为互相关系数计算方法。如图 5 所示,改进的 WSOLA 算法在计算互相关系数时只需计算出 S_{k-1} 语音段后半部分最大值位置和备选语音段 S_{ki} 前半部分最大值位置之间的距离,而原始 WSOLA 算法则需要将 S_{k-1} 后半部分与 S_{ki} 前半部分对应位置的值相乘再求和。

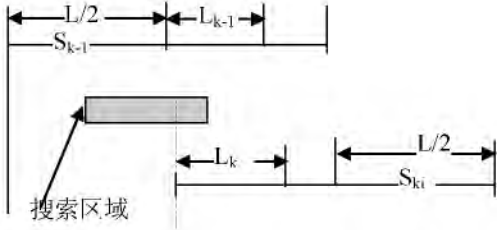


图 5 改进互相关系数细节

2.4 仿真结果

采用 Matlab 对 WSOLA 算法和改进 WSOLA 算法进行仿真,重建出丢包处的数据,以验证对 WSOLA 改进的可行性,同时对两种方法恢复语音的质量进行评价以确定改进后的 WSOLA 算法的优劣。在本次仿真试验中,取 $N = 3$ $L_{in} = 320$ $L_{out} = 480$ $L = 240$ 。同时选取一个声音文件中的一段数据(760 个采样点)为研究对象。假设单个包所含的样点数为 160,搜索区域长度为 40,仿真时将单个包的采样点值置零来模拟丢包。图 6 为模拟丢包时的声音数据,图 7 为采用 WSOLA 算法进行丢包隐藏时所重建的声音数据,图 8 为采用改进的 WSOLA 算法所得结果。

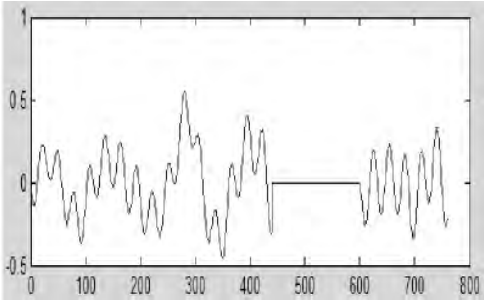


图 6 丢包数据

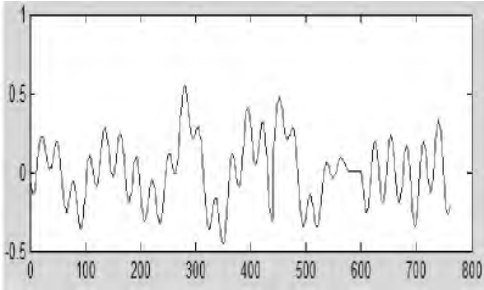


图 7 WSOLA 丢包隐藏数据

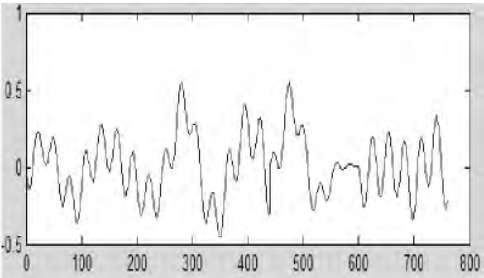


图 8 改进的 WSOLA 丢包隐藏数据

通过比较图 7、图 8 的波形可知,采用改进的互相关系数计算方法,可以取得与 WSOLA 算法相近的结果,表明这种改进方法是可行的。

采用 ITU-T P. 862 建议书提供的客观 MOS 值评价方法 PESQ^[10]对重建语音进行评价,来确定改进后的 WSOLA 算法恢复语音质量。测试语音数据采样频率 8 KHz,量化精度 16 bit,总共有 24 077 个采样点。在模拟一定丢包率的情况下,通过 MATLAB 仿真,可以得到 WSOLA 算法和改进的 WSOLA 算法恢复的语音信号,然后以原始语音为基准,得到 PESQ_MOS 值如表 1 所示。

表 1 PESQ 评价结果

丢包率	PESQ_MOS	
	WSOLA	改进 WSOLA
2%	3.947	3.927
6%	3.856	3.848
8%	3.469	3.485

分析表 1 的数据可知,在相同丢包率的情况下 WSOLA 算法和改进的 WSOLA 算法所得到的 PESQ_MOS 值差别不大,可认为所得到的语音质量相近。

比较 WSOLA 算法与改进的 WSOLA 算法的计算量,主要是在计算互相关系数时有所不同,在其它的方面,如加窗、重叠叠加等,计算量一样。对于改进的 WSOLA 算法只需要进行简单
(下转第 272 页)

通过对比实验可以看出:使用 PCA 方法来对样本进行特征降维比使用 IG 方法的效果要好,其分类结果 F-Measure 从 0.789 上升到 0.819、AUC 值从 0.807 上升到 0.831,究其原因,是因为 WEBSpam-UK2007 样本集中的特征数量较多,但特征之间相关度较高,具有冗余,使用 PCA 方法降维后生成的是一份正交的、无冗余的特征集合,而 IG 方法降维后是选出贡献度较大的特征,但并不能保证这些特征之间是正交、无冗余的。

在分类器的训练中加入 AdaBoost 训练方法显著提高了分类器的性能,但是不足之处在于分类器对正常样本的误判率也相应升高了 4 个百分点,初步判定问题出在 AdaBoost 训练方式下各个分类器的权重分配,后续会对这一问题进行深入研究,通过和随机森林算法的对比,证明了本文提出的组合算法有更高的垃圾网页检测率。

4 结 语

本文针对垃圾网页的检测,设计了一种集成主成分分析与 SVM 的检测方案,并给出了一种在训练样本极度倾斜的情况下通过预处理获得均衡样本的方案,实验证明该方案构建的均衡样本对训练一个有效的分类器起了巨大的作用,同时在训练时使用主成分分析法对样本特征进行降维约减,并引入 AdaBoost 与 SVM 的组合算法,这一模式进一步提高了分类的准确率,相信这一方案能够有效应对那些因为异常样本难以收集、好坏样本数量极不平衡、样本特征维数过高以及样本特征过度冗余而导致分类正确率低的问题。从实验中可以看到,本方案虽然具有最高的垃圾网页检测率,但对正常网页的误检率也较高,未来的工作将致力于优化检测算法,降低对正常网页的误检率。

参 考 文 献

- [1] Craig S, Hannes M, Monika H, et al. Analysis of a very large web search Engine query log[C]//SIGIR Forum, 1999.
- [2] Castillo C, Davison B D. Adversarial Web Search[C]//Foundations and trends in Information Retrieval 2011: 377-488.
- [3] Web Spam Challenge [EB/OL]. [2012]. <http://webspam.lip6.fr/>, 2007.
- [4] Cheng Z, Gao B, Sun C, et al. Let Web Spammers Expose Themselves [C]//Proceedings of the fourth ACM international conference on Web search and data mining. WSDM '11, ACM, New York, NY, 2011: 525-534.
- [5] Becchetti L, Castillo C, Donato D, et al. Link-based characterization and detection of Web spam [C]//Proceedings of the 2nd International Workshop on Adversarial Information Retrieval on the Web (AIR-Web), Seattle, USA, 2006.
- [6] 贾志洋, 李伟伟, 张海燕. 基于内容的搜索引擎垃圾网页检测[J]. 计算机应用与软件, 2009, 26(11): 165-167.
- [7] 魏小娟, 李翠平, 陈红. Co-Training—内容和链接的 Web Spam 检测方法[J]. 计算机科学与探索, 2010, 4(10): 899-908.
- [8] Castillo C, Donato D, Gionis A, et al. Know your neighbors: Web spam detection using the Web topology [C]//Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval 2007.
- [9] Yuchun Tang, Sven Krasser, Yuanchen He, et al. Support Vector Ma-

chines and Random Forests Modeling for Spam Senders Behavior Analysis[C]//IEEE GLOBECOM 2008.

- [10] Witten I H, Frank E. Data mining—practical machine learning tools and techniques with java implementation [C]. Morgan Kaufmann, 2000.
- [11] Carlos Castillo, Jumar Chellapilla, Ludovic Denoyer. AIRWeb 2008 Fourth International Workshop on Adversarial Information Retrieval on the Web[R/OL]. Inc. Yahoo, Microsoft, 2008.

(上接第 268 页)

寻找最大值运算,在计算每一个 C_i 时只需要一次减法及一次求取绝对值运算,而 WSOLA 算法需要 $L/2$ 次乘法运算以及 $L/2 - 1$ 次的加法运算,并且总的运算量随着 L_s 的增大而大幅增加。由此可知改进的 WSOLA 算法的计算量相比较 WSOLA 算法大大减少。

3 结 语

WSOLA 算法应用于 VoIP 网络的丢包隐藏时,能够在尽可能小的损失音质的情况下,把通话信号的持续时间拉长,但会引入计算延时。本文提出了 WSOLA 丢包隐藏算法中的互相关系数计算的一种改进方法。通过仿真测试以及恢复语音评价表明,该方法在保证语音质量的情况下,大大减少了计算量,避免了高延时的引入,从而有助于提高 VoIP 的通话质量。

参 考 文 献

- [1] Perkins C, Hodson O, Hardman V. A Survey of packet loss recovery techniques for streaming audio [J]. IEEE Network, 1998, 12(5): 40-48.
- [2] 李默嘉. 基于无线网络的 VoIP 中 WSOLA 算法的改进研究[D]. 北京:北京邮电大学, 2009.
- [3] Wang J, Jerry D G. Parameter interpolation to enhance the frame erasure robustness of CELP coders in packet networks [C]//2001 IEEE International Conference on Acoustics, Speech, and Signal Processing, Salt Lake City, UT, 2001: 745-748.
- [4] Verhelst W, Roelands M. An overlap-add technique based on waveform similarity (WSOLA) for high quality time-scale modification of speech [C]//Proceedings of International Conference on Acoustics, Speech, and Signal Processing. [S. l.]: IEEE Press, 1993: 554-557.
- [5] 李如玮, 鲍长春. VoIP 丢包处理技术的研究进展[J]. 通信学报, 2007, 28(6): 103-110.
- [6] 祝轶群, 王宏伟, 张华, 等. VoIP 中丢包恢复的方法[J]. 计算机工程, 2008, 34(23): 123-125.
- [7] 王培明, 施寅. VoIP 中丢包隐藏技术研究[J]. 计算机技术与发展, 2006, 16(7): 26-28.
- [8] Sanneck H, Stenger A, Younes K B, et al. A new technique for audio packet loss concealment [C]//Proceedings of IEEE Global Internet (Jon Crowcroft and Henning Schulzrinne, eds.). London: IEEE Press, 1996: 48-52.
- [9] 黄华华, 邱小军. 基于 PAOLA 的丢包补偿算法[J]. 电声技术, 2007, 31(4): 53-55.
- [10] ITU-T Recommendation P. 862. Perceptual Evaluation of Speech Quality (PESQ), an Object Method for End-to-end Speech Quality Assessment of Narrow-band Telephone Networks and Speech Codes[S]. 2001.