

一种优质、低时延的 8kb/s 语音编码算法*

林嘉宇 易波

(国防科技大学电子技术学院 长沙 410073)

摘要 介绍一种优质、低时延的 8kb/s 语音编码算法,重点描述了算法的构造,分析了算法的性能。本算法可用一片 33MIPS 的通用定点 DSP 芯片实现。对算法所作的编解码信噪比测试和非正式的听音测试表明,该算法具有优良的编解码语音质量。

关键词 语音编码,低时延

分类号 TN912.3

A Toll-quality, Low-delay 8kb/s Speech Coder

Lin Jiayu Yi Bo

(Institute of Electronic Engineering, NUDT, Changsha, 410073)

Abstract This paper presents a 8kb/s speech coder which is based on CELP. The coder can offer toll-quality coded speech and low coding delay. The structure of the algorithm is described and the performance is analyzed. The coder and decoder can be implemented by one general-purposed 33MIPS fixed-point DSP chip. According to the objective SNR test and informal listening test, the coder shows good performance.

Key words speech coder, low-delay

近年来,语音数字化的研究在实用领域中已取得了长足的进步^[1~3]。在进一步研究语音编码体制时,必须顾及编码体制各指标的协调,包括编码速率、恢复语音质量、编码时延、算法复杂度、抗信道误码和环境噪声能力等。

自从 G. 728 于 1992 年问世以后,8kb/s 的语音编码算法成为市场关注的中心之一,其主要可应用于网络通信,多媒体通信,个人移动通信,数字卫星系统,存储/检索,公共电话群路系统^[4]等。CCITT(已并入 ITU)对 8kb/s 的语音编码算法提出过较严格的要求^[5,6]。ITU 于 1996 年 6 月通过了 G. 729 的 8kb/s 标准^[6],其各项性能指标较好。但是,G. 729 的单次编解码延时为 25ms,离 ITU 的目标相差尚远。本文构造的语音编码算法选择 8kb/s 的编码码速率,属 CELP 体制,其重心在于低时延,同时兼顾恢复语音的质量、体制抗信道误码和噪声的能力,以及算法复杂度等,称为 LD-8k 算法。

下文将具体描述 LD-8k 算法的构造,分析其各项技术指标,给出语音编码的性能结果。

1 LD-8k 语音编码算法构造

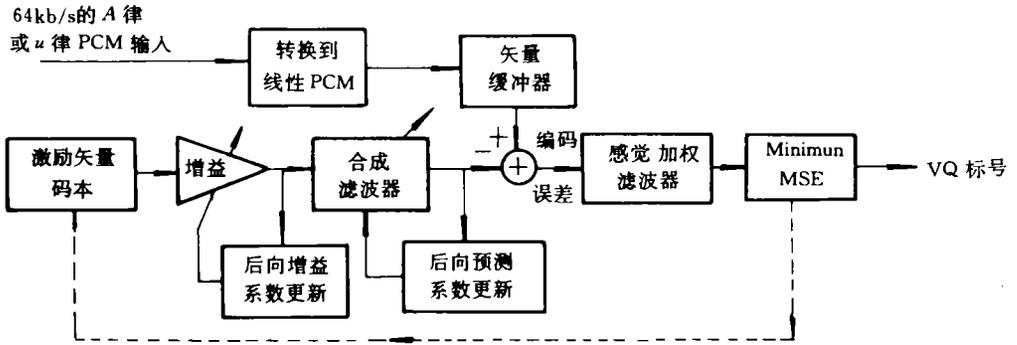
LD-8k 属 CELP 体制。目前 CELP 体制是语音编码体制的主流之一,其应用性设计已趋向成熟^[7]。传统的 CELP 原理可参见文^[7,8]。

1.1 系统结构构造

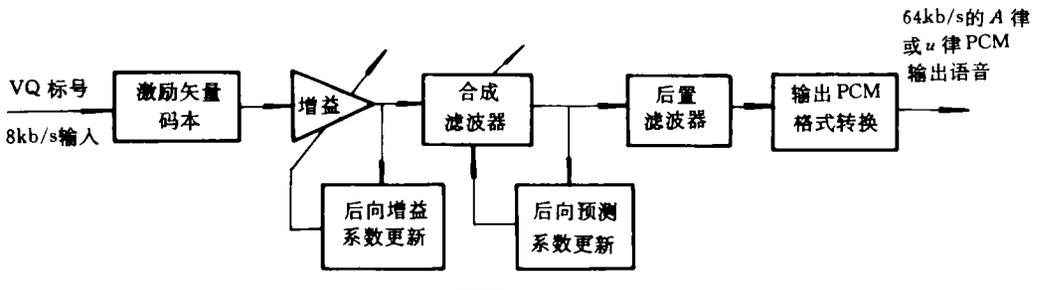
8k-LD 编码器和解码器的原理框图可见图 1。我们构造的 LD-8k 算法的结构类似 G. 728^[9],但针对 8kb/s 的速率和此处的低时延目标要求作了改进。

编码器中,激励码矢量为 8 维,单路编解码时延小于 3ms。码本含 256 个(8bit 编码)纯形状码矢

* 1997 年 9 月 12 日修订
第一作者:林嘉宇,男,1973 年生,博士生



(a) LD-8K 编码器原理图



(b) LD-8K 解码器原理图

图1 LD-8k 编解码器原理图

量。为了降低待量化矢量的动态范围，我们使用后向增益预测器，对输入语音增益进行预测，从而对输入语音进行规一化。根据语音幅度的特性，在对数域中预测增益，可以较灵敏地反映语音幅度的细节变化^[10]。对数域增益预测器为10阶全极点滤波器，它的系数每3个矢量周期（即24个样点，3ms）更新一次。LD-8k算法取消常规CELP体制中的基音预测器，而将全极点合成滤波器的阶数提高为50阶。合成滤波器的系数每3个矢量周期更新一次，以较迅速地反映输入语音的频谱特性；它是通过对预测语音（即解码语音）后向地进行LPC分析得来的。采纳感觉加权滤波器，让语音矢量的量化在感觉加权域中进行。感觉加权滤波器为10阶零极点模型，其系数每3个矢量周期更新一次，这是通过对原始输入语音进行LPC分析得到的。由于解码器形式上并不涉及感觉加权滤波器，所以，算法中唯一的使用前向分析的这一处不会对整个体制的低时延特性产生影响。码本搜索采取全搜索，从候选的256个单一形状码中选中最佳码矢量，该最佳码矢量激励合成滤波器所输出的合成语音，在感觉加权域中，和原始语音比较具有最小的均方误差。另外，LD-8k算法还采用了“零输入响应~零状态响应”方法，来降低码本搜索前对激励码矢量所作的“整形”（表现为卷积）准备工作的运算量^[7]。

为实现编解码的低时延，系统采取“后向自适应”的总体工作方式，这意味着编码器事实上包含解码器。所以，解码器的构成可以参照编码器。为了进一步提高解码语音音质，可以在解码器的输出端级联自适应后置滤波器^[9]。自适应后置滤波器的功用在于，它可以自适应地调节输出语音的频谱响应，将输出语音频谱的谱峰加强，谱谷进一步减弱。

1.2 低时延构造

低时延语音编码技术的发展概况可以参见[11]。依据Chen在文[9]中的讨论，CELP编码器的编解码时延是由语音处理帧的长度决定的。事实上，如果不考虑比特传输延时，单次编解码时延可分解为三部分，即：样点缓存时延，编码算法处理时延，解码算法处理时延。显然，实时处理的总时延不应该超过样点缓存时延的3倍。为实现低时延编码，我们必须采取较短的语音处理帧。同时，我们

的目标是构建 8kb/s 的编码器, 这意味着对原始语音, 平均一样点用一比特编码。为了限制码本搜索的运算量, 选择码本包含 256 个激励矢量, 即 8bit 编码, 这样, 可取语音基本处理帧为含 8 个样点。于是, 样点缓存时延为 1ms (8kHz 采样频), 单次编解码时延不超过 3ms。该结果远优于 ITU 的 10ms 编解码时延目标。

为实现低时延而采取的 8 样点长的基本处理帧, 决定了我们必须采取“后向自适应”的总体工作方式。编码器的各参数, 如预测滤波器的系数, 不再从原始输入语音中求得, 因为不能让编码器等待足够长的时间, 积累足够多的语音数据, 来提取语音参数; 编码器只能使用预测语音, 或即前时的解码语音, 方能不造成缓存延时。“后向自适应”工作方式又进一步决定了系统的其它特征。编码器使用前时的解码语音(预测语音), 就形成了闭环工作状态, 存在系统的反馈。前时的解码语音在解码器中也能获得, 编码器就不必向解码器传输滤波器系数, 解码器可以如同编码器, 后向地求出所需值。因此, 系统只需传输码激励矢量的标号。

1.3 合成滤波器构造

在 LD-8k 算法中, 合成滤波器以“后向自适应”方式工作。我们对前时的合成语音进行 LPC 分析。LPC 分析可以采用自相关法 (Levinson-Durbin 递推)、协方差法 (Choleskey 分解) 和格型法^[12] (Lattice 的 Burg 算法)。综合考虑各算法的计算效率和定点实现时的稳定性保证问题^[13] (数据存贮量的大小倒在其次), 选用自相关法。

使用混合窗法对等待分析的数据加窗^[14], 这是基于加窗效果和实际实现时(定点)计算量之间平衡的结果。混合窗是由对语音新值加窗的非迭代三角函数和对语音旧值加窗的可迭代指数衰减函数组成, 形状可见图 2。可迭代部分可以降低加窗的理论运算量, 非迭代部分易于实时定点 DSP 实现, 且性能效果较好。事实上, 我们也把各自相应的混合窗用于增益预测滤波器和感知加权滤波器的系数的 LPC 分析中。

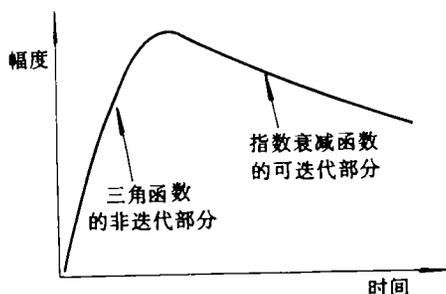


图 2 混合窗形状

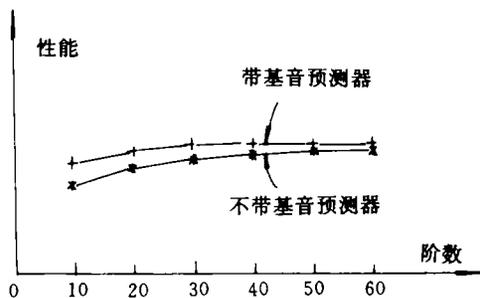


图 3 合成滤波器阶数和体制性能的关系

文 [9] 中指出, 后向基音预测器对信道误码非常敏感, 不宜使用。而体制的低时延特性又要求不使用依前向方式工作的部分, 这样, 只好取消基音预测器; 相应地, 提高合成滤波器的阶数, 以补偿取消基音预测器对女声造成的损伤。理论和仿真都表明, 合成滤波器的阶数取为 50 时, 体制的性能提高就饱和了^[9,12], 参见图 3。因此, 合成滤波器的阶数可取为高阶的 50 阶。高阶的 LPC 分析的稳定性问题应该给予足够的重视, 由于我们选用自相关法求解, 程序可以通过判定自相关法中的反射系数来判断 LPC 分析的稳定性。另外, 高阶 LPC 分析的运算量较大 (运算量大概和阶数的平方成正比), 必须在实现时仔细地安排运算量的整体平衡。

1.4 激励码本构造

语音编码 CELP 体制的激励码本可以表现为随机码矢量、自适应码矢量码本和训练码本等^[7]。由于 LD-8k 算法中取消了基音预测器, 并且训练码本具有较强的广泛性, 我们选择 LD-8k 算法的激励码本为训练码本。

从 LD-8k 算法的原理图描述中可看出, 量化事实上是激励码矢量对预测误差进行波形编码。LD-8k

算法本身是闭环的,因此,码本训练将是闭环的。码本为纯形状码码本,只需对形状码进行设计,但是,由于整个体制含有增量调整控制,这决定了码本设计不能简单地应用常规的 LBG 算法。LBG 算法通常用于设计局部最优化的固定矢量量化码本,其核心包括邻近聚类准则和形心形成准则。但是,加入增益调整之后,LBG 算法的形心形成准则不再是最优的了^[10],必须加以修改。改进的 LBG 算法使用加权形心形成准则,即用增益对原来的形心算式进行加权。

码本设计的具体算法包括:LD-8k 算法的编码器本身,改进的 LBG 算法,Cholesky 分解法,聚类泡腔处理。其中,改进的 LBG 算法中,邻近聚类由编码器的码本搜索完成,形心形成依据加权形心形成准则完成。Cholesky 分解法应用于码本形心方程的求解。我们还对形状码形心进行聚类泡腔处理,使用 ISODATE 聚类中的思想。

码本训练完成后,再对码本的标号进行伪 Gray 编码处理。伪 Gray 编码是一种零冗余的编码方式,它的原理是,让距离较小的码矢量的标号接近。这样,当标号传输出现传输误码时,错误的码矢量和原来的码矢量差别较小。通常采用的伪 Gray 编码算法为 Binary Switching Algorithm。

2. LD-8k 语音编码算法性能分析

2.1 LD-8k 算法的抗误码和噪声性能

LD-8k 算法的抗误码性能是相当好的,主要原因在于采用了如下的几个方面的设计措施。

码本设计:使用伪 Gray 编码方式设计码本标号,单比特错的标号在解码器中所对应的码矢量和原传输的码矢量很近,引起的实际错误不大;取消了基音预测器;用 50 阶的 LPC 预测器代替对信道很敏感的基音预测器,增强了对付信道传输错误的适应性;预测器设计:各预测器设计时,都注意减小预测器中激励响应的有效宽度,以尽可能地降低信道误码的影响。并且,由于 LD-8k 算法属本质为波形编码的 CELP,通常的背景噪声不会对体制有大的影响。仿真实验和 LD-8k 算法的实现也证实了这一点。

2.2 LD-8k 算法实现的语音质量

对 LD-8k 算法进行客观性测试,测试使用全局信噪比 (GlobSNR) 和分段信噪比 (SegSNR)。测试所用语音数据不取自提供训练码本的语音数据的说话者,测试数据长为几秒。附表列出了两个男声、两个女声的测试结果。由于解码语音经后置滤波器处理后,虽然主观听觉感受得到改善,但波形变化了,不再适合用客观性 SNR 测试,故我们的信噪比测试结果并不考虑经后置滤波器的解码语音。图 4 展示了原始语音和未经后置滤波器处理的解码语音。听音结果表明,解码语音的清晰度、自然度都较好;经后置滤波器处理后,语音质量有进一步的提高。

附表 LD-8k 算法的测试结果 (不经后置滤波器)

指标	男声 a	女声 a	男声 b	女声 b
SegSNR (dB)	18.94	19.43	18.38	16.79
GlobSNR (dB)	15.69	14.57	15.49	12.62

2.3 LD-8k 算法传输规则音调信号的能力

用变化的单频正弦波和 Chirp 信号来测试 LD-8k 算法的传输规则音调信号的能力。单频正弦波在 300Hz 到 4000Hz 之间取有代表性的 10 个单频点,每个单频点持续 0.5s。Chirp 信号为在 0.5s 内的从 200Hz 到 4000Hz 的扫频。依照惯例,在测试中取消后置滤波器。测试结果如下。

单频变化信号,其解码信号分段信噪比为 23.05dB,全局信噪比为 15.10dB,和语音信号处同一水平。但可以明显发现,高频信号受损,并且,频间跳转不能迅速反应,存在一定过渡时间。Chirp 信号,其解码信号分段信噪比为 10.69dB,全局信噪比为 9.44dB,较之单频正弦波降低了一个档次。这和单频编码时,频间跳转不能迅速反应是一致的。从测试结果可知,LD-8k 算法可以用来传输波特率不太高的规则音调信号。

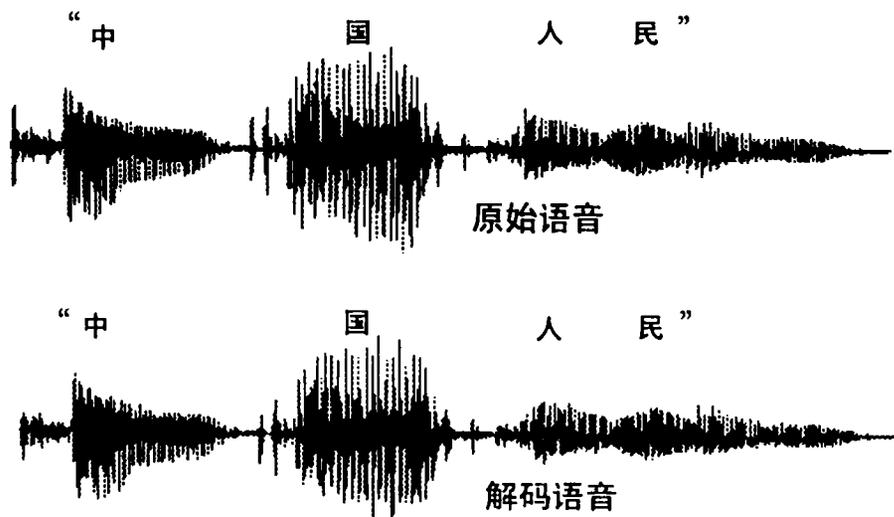


图 4 原始语音和解码语音

3 小结

本文介绍了一种优质、低时延的 8kb/s 的语音编码算法。对算法的设计理论问题进行了讨论，描述了算法的构造，分析了算法的性能。算法实现了单次编解码理论时延达到 3ms。对算法所作的编解码信噪比测试和非正式的听音测试表明，该算法具有优良的编解码语音质量。本体制的算法复杂度在目前通用 DSP 芯片可承受的范围内，算法全双工编解码器可以用一片 33MIPS 的通用 DSP 芯片实现。本算法可以应用于网络通信、卫星通信、多媒体通信、个人移动通信等领域中。

参考文献

- 1 CCITT Recommendation G. 728, 1992
- 2 Gerson I, Jasiuk M. Vector Sum Excited Linear Prediction (VSELP). in Advances in Speech coding, B. S. Atal, etc, Eds. Norwell, MA; Kluwer, 1991
- 3 Campbell J P. etc, The DOD 4. 8kb/s Standard (Proposed Federal Standard 1016). in Advances in Speech coding, B. S. Atal, etc, Eds. Norwell, MA; Kluwer, 1991
- 4 Rabiner L R. Applications of Voice Processing to Telecommunications. Proc. of IEEE. 1994, 82 (2)
- 5 CCITT Temporary Document 77 (XW/2). 1991
- 6 李晓明. 新一代语音编码标准——G. 729. 通信技术, 1996 (4)
- 7 Allen Gersho. Advances in Speech and Audio Compress. Proc. of IEEE. 1994, 82 (6)
- 8 Spanlas A S. Speech Coding: A Tutorial Review. Proc. of IEEE, 1994, 82 (10)
- 9 Chen J H, A Low Delay CELP Coder for the CCITT 16kb/s Speech Coding Standard. IEEE Journ. on SAC, 1992, 10 (5)
- 10 Chen J H, Gersho A. Gain-adaptive Vector Quantization with Application to Speech Coding. IEEE Trans. Commun., 1987 (9)
- 11 Cupetman V, Gersho A. Low delay speech coding. Speech Commun., 1993 (6)
- 12 Cuperman V, Peng R. Lattice Low-Delay Vector Excitation Coding of Speech at 8—16kb/s. IEEE Trans. Commun., 1994, 42 (6)
- 13 Rabiner L R, Schafer R W. Digital Processing of Speech Signals. Prentice Hall, 1978
- 14 Chen J H, A Fixed-Point 16kb/s LD-CELP Algorithm. ICASSP, 1991, pp. 21-24