

doi:10.11835/j.issn.1000-582X.2018.01.012

基于 Linux 平台上线性预测语音编码器性能分析

田 媛

(西安航空学院 电子工程学院, 西安 710003)

摘要:线性预测编码器是一类非常重要的语音编码器。文中主要描述了几种线性预测语音编码算法的软件及实现,即码激励线性预测编码(CELP)、低延迟 CELP(LD-CELP)和混合励磁线性预测(MELP),编码速率分别为 4.8、16、2.4 kb/s。几种语音编码器的 C 语言程序已在 Linux 平台上可编译和执行,并对结果进行了以 MOS 为标准的主观测试。波形分析主要使用了 Praat 和 Adobe Audition 软件。结果表明,MELP 和 CELP 的质量相当,而 LD-CELP 编码器的质量则要高得多,虽然是以牺牲较高的比特率为代价。

关键词:线性预测编码;语音编码器;Linux 平台

中图分类号:TN912.31

文献标志码:A

文章编号:1000-582X(2018)01-108-07

The performance comparison of linear predictive vocoders on Linux platform

TIAN Yuan

(Department of Electronic Engineering, Xi'an Aeronautical University, Xi'an 710003, P.R.China)

Abstract: Linear predictive coders are a kind of significant vocoders. The speech coders using three linear predictive vocoding algorithms are described, i.e. CELP (code excited linear prediction), LD-CELP (low delay-CELP) and MELP (mixed-excitation linear predictive) working at 4.8 kb/s, 16 kb/s and 2.4 kb/s, respectively. The C programs for the three vocoders are compiled and executed on the Linux platform and the results are subjectively tested with taking MOS as criterion. Waveform analysis mainly uses Praat and Adobe Audition software. The results show that the performances of MELP and CELP are of comparable quality, while the performance of LD-CELP is much higher, albeit at the expense of higher bit rates.

Keywords: linear prediction coding; vocoder; Linux platform

语音编码是语音信号的编码,比特率比原始的数字语言的传输速率低。人类的听觉系统只能捕捉到语音信号的某些方面。因此,感知语音信号的相关信息可以提取其特征,从而在更低的带宽下产生一个等效的波形。线性预测技术于 1947 年由维纳提出,并于 1967 年由板仓等人应用于语音的分析与合成方面,该技术已广泛应用于语音信号的压缩。

目前,线性预测编码器多应用码激励线性预测(CELP)和低延迟的码激励线性预测(LD-CELP)方法。它们分别使用向前和向后的线性预测方法,即查表获得编码器与语音剩余信号对应^[1-2]。在混合激励线性预测(MELP)^[3]中,引入了另一类语音——过渡语音,混合激励即将周期和噪声激励混合,是 MELP 模型的另一个显著特征。

收稿日期:2017-09-23

作者简介:田媛(1989—),女,硕士,主要从事片上机系统设计及微型网络通信,(E-mail)sweetytian1019@foxmail.com。

以上几种方法为目前主流的处理方法,但是对于处理结果、选择条件等并没有完整的相关数据。文中主要针对比特率和感知质量方面对 3 种线性预测模型进行了比较。在比较声音编码器的过程中,对声音质量的主观测试是评估语音编码器的一个重要步骤。只有在人类对声音质量满意的情况下,一个语音编码器才会被市场接受。在考虑这个因素的情况下,使用平均意见评分(MOS)这一最常见的评估方法对结果进行了主观评价。

1 语音编码算法

线性预测模型以 CELP、ld-CELP 和 MELP 算法为核心,如图 1 所示。在 LPC 模型中,声音信号被分为语音和非语音信号。当声带产生振动,并在 LPC 源滤波器模型中被表示为周期性的激励时,就产生了声音信号。非语音信号由声道的气流产生,而不涉及声带,这些信号通常被表示为白高斯噪声。



图 1 以 CELP、LD-CELP 和 MELP 算法为核心的 LPC 模型

Fig.1 LPC model as the core of CELP, LD-CELP and MELP algorithms

LPC 编码器由带有自适应延迟和增益因子的线性预测器组成^[4]。由于语音是由语音和非声音信号的组合产生的,所以,它被作为重要的感知语音观察。

1.1 码激励线性预测算法

预测剩余信号可用于提高编码器输出的感知质量。一般来说,在有码矢表的码本中获得与剩余信号最接近的匹配。通过码矢指数的选择,可以得到误差度量的最小值。通过码本来获得用于语音信号建模代码的方法叫做语音编码,发送端与接收端使用同样的码本。但是,在接收端,其指数用于获得码矢和合成滤波器。

CELP 算法主要利用了分析-合成(Analysis-by-Synthesis, AbS)方法,在此方法中,发送端分析信号产生线性预测系数,然后,利用这些系数来合成发送端内部的语音信号。当生成了 1 个错误信号时,可以从码本中选择 1 组码矢以减少感知加权平均平方误差。

在 CELP 发送端中,首先,将输入语音分解为大约 30 ms 的帧,而后进行短期线性预测,即对每个输入帧的共振峰(光谱包络线峰值)进行估计。CELP 算法的感知加权滤波器的传递函数为

$$W(z) = \frac{1 - Q(z)}{1 - Q\left(\frac{z}{\gamma}\right)}, \quad (1)$$

$$Q(z) = \sum_{i=1}^M q_i z^{-i},$$

$$Q\left(\frac{z}{\gamma}\right) = \sum_{i=1}^M \gamma^i q_i z^{-i}, 0 < \gamma < 1, \quad (2)$$

其中: M 为 LPC 预测序列长度; q_i 为量化 LPC 系数。

在找到短期 LPC 系数后,每一帧都被分成 4 个子帧,即每个 7.5 ms。每一子帧均作为长期预测滤波器的输入,用以估计语音信号的音高和强度。在完成解码后,后滤波阶段可以被选择性的添加,以提高输出信号的质量。CELP 算法中,位的分配^[5]如表 1 所示。

表 1 CELP 的位分配

Table 1 The CELP bit allocation

参数	序号/帧	位数/帧
线性预测系数	10	34
音高周期	4	28
自适应码本增益	4	20
随机码本指数	4	36
随机码本增益	4	20
合成	1	1
误码纠正	4	4
未来拓展	1	1
共计位数/30 ms 帧	—	144

在 CELP 算法中 30 ms 的帧划分相当于在 8 000 Hz 采样频率下的 240 个采样。因此, CELP 的比特率是 $14/30 \text{ ms}=4.8 \text{ kb/s}$ 。

1.2 低延迟码激励线性预测算法

CELP 和 LD-CELP 算法只在在线性预测中适应的类型上存在差异——分别是向前和向后。低延迟是通过使用后向自适应预测器和短激励向量^[2]来实现的。在传输过程中, 只有激励码本的指数被传输, 而所有其他参数都是通过对先前量子化语音的向后适应进行更新的。LD-CELP 算法使用了一个调整系统函数来作为过滤器权重。

$$W(z) = \frac{1 - Q\left(\frac{z}{\gamma_1}\right)}{1 - Q\left(\frac{z}{\gamma_2}\right)}, \quad 0 < \gamma_1 < \gamma_2 \leq 1, \quad (3)$$

其中, $Q(z)$ 与 $Q(z/r)$ 的表达式与公式(2)相同, γ_1, γ_2 均为向后调整系数。在优化的过程中, 可以通过不断调整 γ_1 和 γ_2 的数值, 获得最优的编码语音质量, LD-CELP 的位分配如下表 2 所示。

表 2 LD-CELP 位分配

Table 2 The LD-CELP bit allocation

参数	位序列
激励指数	7
激励增益	2
码本增益	1
共计位数/2.5 ms 矢量	10
共计位数/10 ms 帧	40

当采样频率为 8 000 Hz 时, 一个矢量对应 0.625 ms, LD-CELP 的比特率是 $10/0.625 \text{ ms}=16 \text{ kb/s}$ 。

1.3 混合激励线性预测算法

在 MELP 编码器中, 有 3 类语音信号——语音信号、非语音信号和转换语音。第 3 种通常是在语音发生转换时出现, 也就是说, 产生了非周期的但不是完全随机的激励。另一个特点是, 激励脉冲同样也可以从输入信号中提取出来。脉冲整形滤波器和噪音滤波器用于过滤脉冲序列和白噪音激励。MELP 算法位分配如表 3 所示^[6]。

表 3 MELP 位分配
Table 3 The MELP bit allocation

参数	语音	非语音
LSFs 参数	25	25
傅里叶幅值级数	8	—
增益(2 每帧)	8	8
音高	7	7
带通语音	4	—
非周期标志	1	—
错误保护	—	13
同步位	1	1
共计位数/22.5 ms 帧	54	54

MELP 的 22.5 ms 帧对应采样频率为 8 000 Hz 时的 180 个样本。因此, MELP 的比特率是 $54/22.5 \text{ ms} = 2.4 \text{ kb/s}$ 。

2 研究方法

语音编码器的 C 语言程序利用 GCC(GNU Compiler Collection)编译,并使用 Linux 平台上的 GNU Make 工具构建。对于波形转换,使用了声音交换(SoX)软件。SoX 是一个用于语音文件处理的开源工具。在每个语音编码器中使用的分析和合成有特定命令如下:

2.1 CELP 命令

a)分析:

```
./celp -i inputfile.wav -o outputfile
```

b)合成:

```
./celp -c outputfile.chan -o outputsynth
```

c)拷贝 spd 文件到 raw 文件:

```
cp outputsynth.spd outputsynth.raw
```

d)转换为 wav 文件:

```
sox -r 8000 -b 16 -c 1 -e signed-integer outputsynth.raw outputsynth.wav
```

e)播放文件:

```
padsp play outputsynth.wav
```

2.2 LD-CELP 命令

a)分析:

```
./ccelp inputfile.wav encoderout.out
```

b)合成:

```
./dcelp encoderout.out outputsynth.raw
```

在 LD-CELP 和 MELP 中,将 raw 格式转换为 wav 文件格式和播放输出文件的方式与 CELP 的方式相同,在此不再赘述。

2.3 波形分析

文中波形分析主要是用 Praat 来进行的。Praat 是一种专业用于语音分析的工具,它采用标准的语音样本格式.wav 进行波形分析。语音文件的时域波形,音高和强度波形均由此工具绘制。声谱分析是在 Adobe Audition 软件中完成。

2.4 语音质量测试:MOS 评价

文中对语音质量的评估是用平均意见评分(MOS)测试来完成的。由于时间的限制,共对这些编码器进

行了 10 次 MOS 评估。以若干英语词组记录的语音样本被作为信号源输入,在一个安静的环境中,3 个样本被记录下来,而 2 个语音样本则在音量很大的背景音乐中被记录。过程中,使用 Logitech h110 立体声耳机进行录音和回放。评估者接受了对 MOS 测试的初步培训,并记录了他们的分数。MOS 测试评分体系如表 4 所示。

表 4 MOS 评分体系
Table 4 MOS scores and interpretations

MOS	语音质量
5	优秀
4	良好
3	正常
2	较差
1	差

在 MOS 评分体系中,4 或 5 分表明了高质量的语音,而 1 或 2 分则表示较为差的语音质量。

3 仿真结果与讨论

3.1 波形分析

使用 Praat 软件的波形分析结果如图 2~图 4 所示。图 2 展示了原始语音、CELP、LD-CELP 和 MELP 编码器输出的语音信号时域波形图。图 3 描述了原始和合成的语音的声谱图。图 4 则表现了不同编码器在音高和强度上的比较。

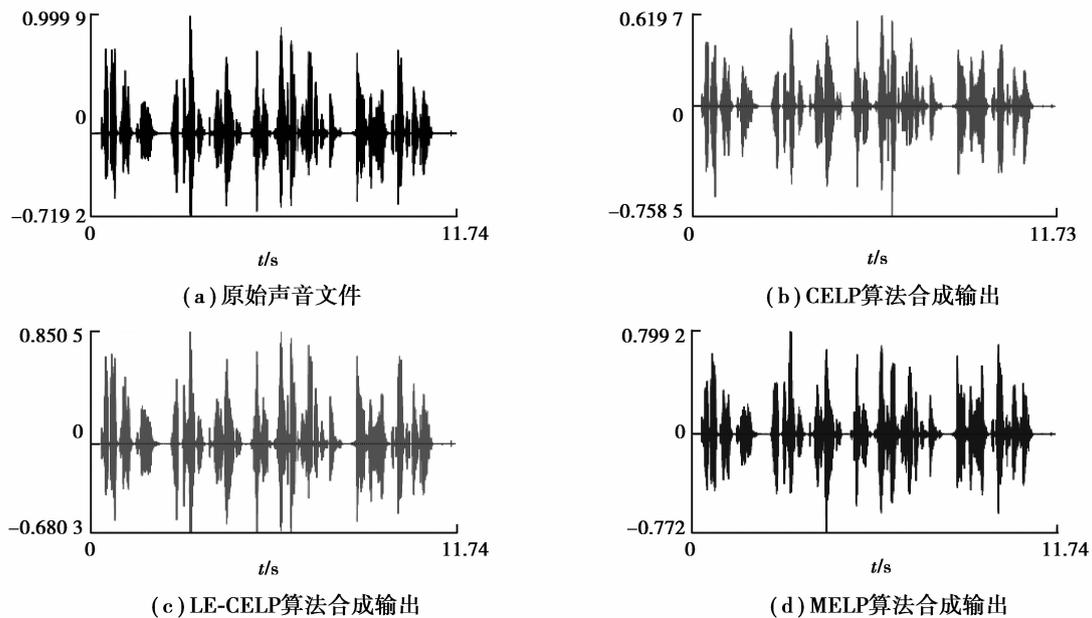


图 2 时域信号波形

Fig.2 Time domain representation

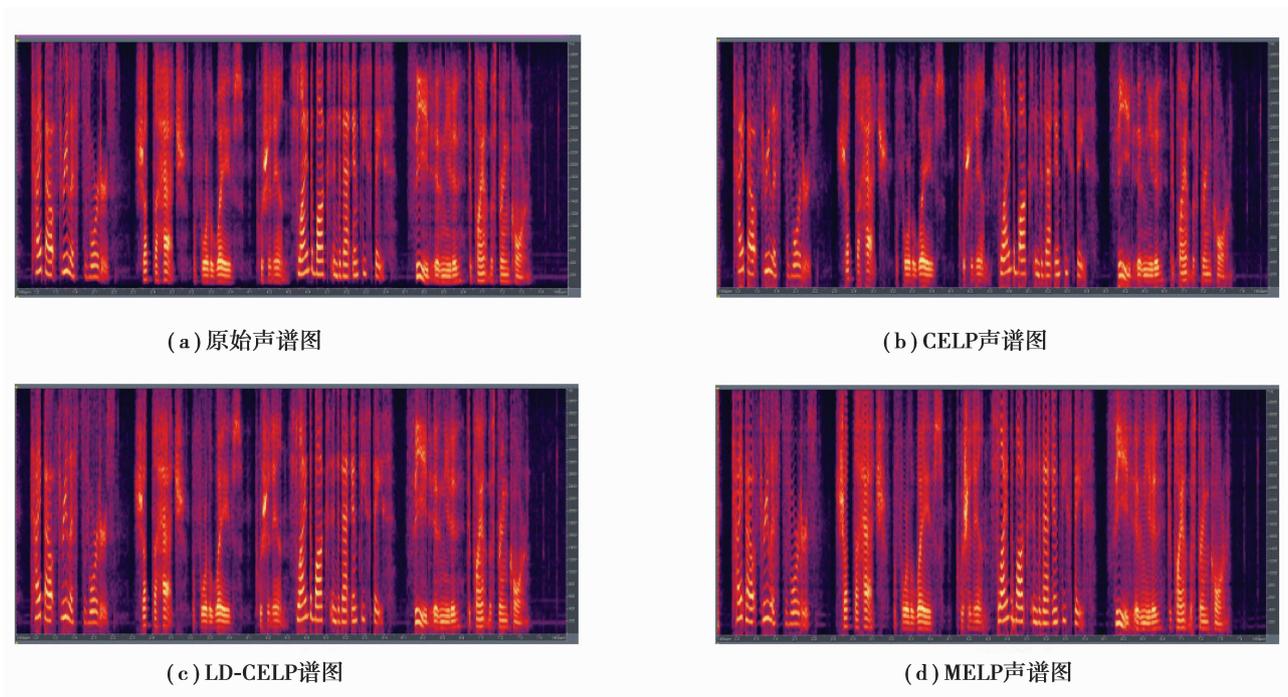


图 3 语音信号声谱图

Fig.3 Spectrograms of speech signals

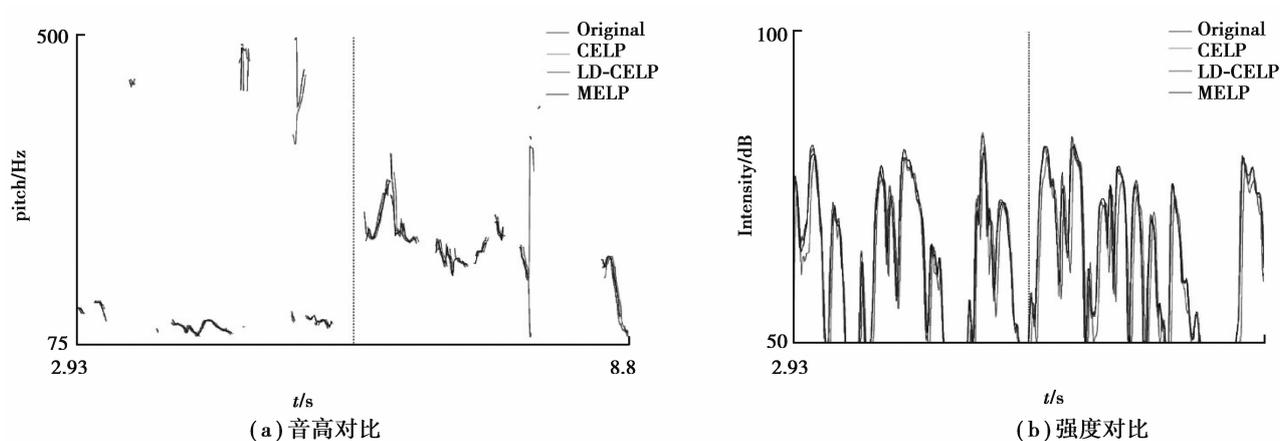


图 4 音高和强度波形

Fig.4 Pitch and intensity waveforms

从时域波形可以看出,大部分原始波的整体形状被保留了下来。然而,在某些部分,峰值被剪掉了,这导致了语音输出的清晰度下降。声谱图显示了语音波形在时间函数上的频率内容。音高和强度对比图也显示了在输出端 3 种编码方法有微小的变化。因此,对感知质量的可靠估计只能通过使用人类的听力进行主观测试来完成。

3.2 主观 MOS 评价

所有记录的输入语音文件都以 8 kHz 的速率进行采样。表 5 中显示了语音输入文件的详细信息。可以看到,输入语音文件平均大小为 393.6 kB,平均长度为 24.59 s。与每个输入语音文件对应的 MOS 分数和每个解码器获得的平均 MOS 分数如表 6 所示。

表 5 输入语音文件详情
Table 5 Input speech file details

文件名	文件大小/kB	时间/s
Male_quiet.wav	170	10.65
Female_quiet.wav	176	10.98
Male_ferr_conver.wav	319	19.91
Male_noisy.wav	447	27.93
Female_noisy.wav	856	53.49

表 6 测试者对编码器 MOS 评分
Table 6 MOS score for vocoders

文件名	CELP	LD-CELP	MELP
Male_quiet.wav	3.10	4.06	2.82
Female_quiet.wav	3.24	4.02	2.76
Male_ferr_conver.wav	3.10	3.76	3.12
Male_noisy.wav	3.00	4.58	3.24
Female_noisy.wav	2.52	4.26	1.72
平均 MOS	2.992	4.136	2.732

从表 6 的 MOS 评分可以看出,LD-CELP 在三者中具有最高的语音质量(听觉感知质量),而 CELP 与 MELP 则评价较低。与此同时,CELP 的质量得分略高于 MELP。

4 结 论

通过仿真与测评的比较表明,编码方案的选择主要取决于编码器的应用条件。在需要低延迟和高质量的应用中,如双向通信系统,16kb/s 的 LD-CELP 算法是理想的候选方法。在低比特率更重要的情况下,MELP 是最好的候选方法,可以在低于 2.4 kb/s 的比特率工作时,仍输出不过于失真(听觉感知质量 2~3)的语音。当需要低比特率和高质量时,4.8 kb/s 的 CELP 编码器相比之下是最合适的编码器。在研究中,由于时间限制,对 MOS 测试的评估者数量限制为 10。在未来的研究中,为了获得更准确的结果,需要增加评估者的数量或加入客观评价的体系。

参考文献:

- [1] Sooraj S, Anselam A S, Pillai S S. Performance analysis of CELP codec for Gaussian and fixed codebooks[J]. International Conference on Communication Systems and Networks, 2016: 211-215.
- [2] Juinhwey C, Cox R V, Lin Y C, et al. A low-delay CELP coder for the CCITT 16 kb/s speech coding standard[J]. IEEE Journal on Selected Areas in Communications, 1992,10(5): 830-849.
- [3] 黄李海,胡晓毅,解永军,等.基于 MELP 的水下实时语音通信机的研究与实现[J].电子技术应用,2013,39(3):17-19.
HUANG Lihai, HU Xiaoyi, XIE Yongjun, et al. Research and implementation of underwater acoustic real-time voice communication prototype based on MELP[J]. Application of Electronic Technique, 2013, 39(3): 17-19. (in Chinese)
- [4] Makhoul J. Linear prediction: A tutorial review[J].Proceedings of the IEEE, 1975,63:561-580.
- [5] 李强,舒勤军,谢虹恩,等.一种基于混合 MELP/CELP 的 4 kbit/s 声码器[J].重庆邮电大学学报(自然科学版),2017, 29(2):143-148.
LI Qiang, SHU Qinjun, XIE Hongen, et al. A 4 kbit/s vocoder based on hybrid MELP/CELP[J]. Journal of Chongqing University of Posts and Telecommunicatons (Natural Science Edition), 2017, 29(2): 143-148. (in Chinese)
- [6] Lynn M, Supplee R P, Cohn, et al. Mc-Cree, MELP: the new Federal Standard at 2 400 bps[J]. IEEE International Conference on Acoustics, Speech, and Signal Processing, 1997(2): 1591-1594.
- [7] Mallat Y, Ayadi M, Ayari A, et al. QoS/QoE-CAODV: Routing protocol for cognitive radio Ad-Hoc network[C]// International Conference on Advanced Information Networking and Applications Workshops Crans-Montand, Switzerland: IEEE, 2016:748-753.