

基于线性预测系数自适应前后向量化的 可变速率 MELP 语音编码¹

铁满霞*** 王都生* 樊昌信*

*(西安电子科技大学信息科学研究所 西安 710071)

** (北京大学视觉与听觉信息处理国家重点实验室 北京 100871)

摘要 该文基于 LPC 的自适应前后向量化技术, 提出了一种可变速率的混合激励线性预测 MELP 语音编码算法。该算法中, 采用当前语音帧(前向 LPC)或前面某帧已合成语音帧(后向 LPC)进行线性预测, 当采用后向 LPC 时, 只需传输时间序列编码, 故减少了 LPC 系数的平均编码比特。计算机模拟表明, 该算法与标准 MELP 算法合成的语音质量相当, 但显著减少了 LPC 的传输带宽, 从而明显降低了 MELP 平均编码速率。

关键词 混合激励线性预测编码 MELP, 线性预测, 自适应前后向量化, 可变速率

中图分类号 TN912.3

1 引言

在美国联邦政府选择替代 LPC10(e) 的 2400bit/s 语音编码新标准过程中, Alan.V.McCree 等人提出了一种新的混合激励线性预测 (MELP, Mixed Excitation Linear Prediction) 声码器模型^[1], 该模型是在传统线性预测编码 (LPC) 声码器的基础上附加了混合脉冲与噪声激励、非周期脉冲、自适应谱增强、脉冲散布以及傅氏幅度模型 5 个改进特征, 这些附加特征的引入使得 MELP 编码算法能够模拟自然语音的更多特征, 在低比特率上产生高质量的语音。MELP 与先进多带激励 (AMBE, Advanced Multi-Band Excitation)、正弦变化编码 (STC, Sinusoidal Transform Coding)、波形插值 (WI, Waveform Interpolation) 等其它候选编码方案比较, 总评价最优, 因此 2400bit/s 的 MELP 语音编码器被美国国防部数字声音处理协会 (DDVPC) 选为美国联邦政府的新标准, MELP 算法也成为目前低速率语音编码算法中最有潜力的方法之一。

速率为 2400bit/s 的 MELP 编码器 A/D, D/A 采样

率为 8kHz, 帧长度为 22.5ms, 帧编码长度为 54bit, 编码比特分配如表 1 所示。在此类 LPC 声码器中, 为避免由于 LPC 系数的直接量化导致语音质量的下降, 通常将 LPC 系数转化成线谱频率 (LSF) 参数再进行量化。在 MELP 编码算法中, 10 个 LSF 参数多级 (四级) 矢量量化 (MSVQ, Multi-Stage Vector Quantization) 成 25bit/frame, 帧长度为 22.5ms, 即预测系数每隔 22.5ms 更新一次, 则传输 LSF 参数的编码信息需 1111bit/s, 而 MELP 编码器总的编码速率为 2400bit/s, 因此用于传输 LSF 参数的带宽超过了 46%。如何减少 LPC 系数的编码比特, 降低 MELP 算法的编码速率正是本文的研究目标。

MELP 算法中预测系数基于当前语音帧确定, 即采用所谓的前向线性预测方法确定^[2]。众所周知, 语音信号通常随时间缓慢变化, 是非平稳的, 当前语音帧与其前邻近语音帧的统计特性可能经常相似, 从而预测系数也非常相近^[3]。低延迟码激励线性预测 (CELP, Code Excited

表 1 MELP 编码比特分配

编码参数	浊音帧	清音帧
线谱频率	25	25
傅氏幅度	8	
增益	8	8
基音周期	7	7
带通清浊音 判决 U/V	4	
非周期标志	1	
差错保护		13
同步位	1	1
总比特数	54	54

¹ 1999-12-28 收到, 2000-05-08 定稿

北京大学视觉与听觉信息处理国家重点实验室开放课题基金资助

Linear Predictive) 编码 (G.728)^[4] 正是利用这种特性, 采用后向线性预测方案, 即预测系数通过前面合成语音样点来确定, 虽然性能劣于前向线性预测, 但它不需要 LPC 参数的传输, 从而使编码速率大大下降。

借鉴前向线性预测和后向线性预测算法的优点, 本文提出了以下的自适应前后向编码方案: 如果前面邻近的某合成语音信号帧的 LPC 系数在某种程度上已足够接近由当前语音帧计算出来的未经量化的 LPC 系数, 将采用后向 LPC 方案, 即基于前面的那帧合成语音帧的 LPC 系数用于对当前语音帧的编码, 否则采用前向线性预测。

本文采用对数谱失真 (LSD) 测度来评估预测系数间的相似程度。根据当前帧与前面已合成帧的 LSD 测度, 决定采用前向或后向线性预测。若采用前向线性预测, 一组新的 LSF 的量化编码仍需传送, 但若采用后向线性预测, 则只需传送时间序列编码, 从而形成一种可变速率 MELP 编码方案, 此时传输 LPC 系数所需的带宽显著减小, 平均编码速率也明显降低。

2 线性预测系数的自适应前后向量化方案

将输入语音信号分成长度为 L 、样点互不重叠的帧。对每一帧, 用 Levinson-Durbin 算法计算其 LPC 系数。就预测残差信号的能量最小化, LPC 系数即 a_1, a_2, \dots, a_p 对当前语音帧而言是最佳的。下面介绍自适应前后向量化方案。

首先定义一个 LPC 自适应前后向码本。每个码本有 S 个码矢量, 每个码矢量维数为 p , 其中 p 为线性预测的阶数。第 i 个码矢量, 即 $\hat{a}_1^{(i)}, \hat{a}_2^{(i)}, \dots, \hat{a}_p^{(i)}$ ($i = 0, 1, \dots, S-1$) 是根据前面邻近第 i 个合成语音帧确定的 LPC 系数, 也就是第 i 个语音帧编码输出的 LSF 对应的 LPC 系数。

用 LSD 测度来评估 LPC 系数的相似程度, 自适应前后向码本中第 i 个码矢量的 LSD 算法如下:

$$\text{LSD}^{(i)} = \frac{10}{\ln 10} \sqrt{\frac{1}{2\pi} \int_{-\pi}^{\pi} |V^{(i)}(\omega)|^2 d\omega}, \quad i = 0, 1, \dots, S-1 \quad (1)$$

其中 $V^{(i)}(\omega)$ 为

$$V^{(i)}(\omega) = \ln \frac{1}{|A(\omega)|^2} - \ln \frac{1}{|A^{(i)}(\omega)|^2}, \quad i = 0, 1, \dots, S-1 \quad (2)$$

其中 $A(z) = 1 + \sum_{l=1}^p a_l z^{-l}$, $A^{(i)}(z) = 1 + \sum_{l=1}^p \hat{a}_l^{(i)} z^{-l}$, $z = e^{-j\omega}$ 。

可见每个代表码矢量都需计算对应的 LSD。从自适应 LPC 码本中选择具有最小谱失真即 $\text{LSD}^{(\text{index})}$, 其中 $\text{index} = \arg \min_i \text{LSD}^{(i)}$ 的码矢量。如果 $\text{LSD}^{(\text{index})} > T$, 其中 T 为预先定义的门限, 则采用前向线性预测。当前帧 LPC 系数 a_1, a_2, \dots, a_p 用于语音编码, 将其转换成 LSF, 经四级矢量量化后传给合成端; 否则即 $\text{LSD}^{(\text{index})} \leq T$, 则采用后向线性预测, 相应的 LPC 系数即 $\hat{a}_1^{(\text{index})}, \hat{a}_2^{(\text{index})}, \dots, \hat{a}_p^{(\text{index})}$ 被用来进行语音编码, 在此只需将自适应码本的索引值 (index) 即时间序列编码传到合成端。另外还需一个额外的比特标志来通知合成端, 编码端采用的是前向还是后向预测。

在此之前, 对于每个新的语音帧, 自适应码本中的码矢量需进行更新, 即需将最邻近合成语音帧的 LPC 系数替换掉最旧的码矢量加入到自适应码本中。

下面归纳该 LPC 系数的前后向量化方法。

在编码端, 按如下的 5 个步骤进行:

(1) 计算当前语音帧的 LPC 系数, 即 a_1, a_2, \dots, a_p 。

(2) 将最邻近的合成语音帧的 LPC 系数替换掉最旧的码矢量来更正自适应码本, 即更新为 $\hat{a}_1^{(0)}, \hat{a}_2^{(0)}, \dots, \hat{a}_p^{(0)}$ 。

(3) 计算自适应码本中每个码矢量对应的 LSD 测度。

(4) 从自适应码本中选出具有最小谱失真即 $LSD^{(index)}$, 其中 $index = \operatorname{argmin}_i LSD^{(i)}$ 的码矢量 $\hat{a}_1^{(index)}, \hat{a}_2^{(index)}, \dots, \hat{a}_p^{(index)}$.

(5) 如果 $LSD^{(index)} \leq T$ 预先设定的门限值, 则 LPC 系数采用在第 4 步中确定的 $\hat{a}_1^{(index)}, \hat{a}_2^{(index)}, \dots, \hat{a}_p^{(index)}$, $index$ 值被编码并传送到合成端, 并将标志比特设为 0, 以通知合成端, 编码端采用的是后向线性预测。否则即 $LSD^{(index)} > T$, 则 LPC 系数采用当前语音帧即在第 1 步中确定的系数, 用它对当前语音帧编码, 经矢量量化后送到合成端。将标志比特设定为 1, 以通知合成端, 编码端采用的是前向线性预测。

在合成端, 按如下的 3 个步骤进行:

(1) 如果接收到的标志比特为 0, 则表示编码端采用的是后向线性预测, 根据接收到的时间序列编码值, 采用前邻近对应合成语音帧的 LPC 系数, 跳至 (3);

(2) 如果比特标志为 1, 表明编码端采用前向线性预测编码, 则采用接收到的 LPC 系数;

(3) 用 (1) 或 (2) 确定的 LPC 系数对当前语音帧进行合成, 并用此 LPC 系数更新合成端的自适应码本。

3 性能评估

采用 LPC 自适应前后向量化的 MELP 编码器, 帧长度 L 为 180 个样点, 线性预测的阶数 p 为 10, 当利用后向线性预测时, LSF 编码比特、帧最短编码长度及相应最低编码速率如表 2 所示。该算法的模拟是在 PC 机上利用 Borland C++ 语言实现的, 模拟时的语音输入文件是由声卡采集的 PCM 格式文件, 采样率为 8kHz, 时间大约 4min, 包含 10000 多帧, 每个样点用 16bit 线性码表示, 模拟的输出语音文件也是 PCM 格式文件。

表 2 自适应 LPC 码本 S 取不同值时, 采用后向线性预测 MELP 编码情况

自适应码本 长度 S	线谱频率编码		帧最短编码 长度 (bit)	最低编码速 率 (bit/s)
	时间序列 编码 (bit)	标志比 特 (bit)		
2	1	1	31	1378
4	2		32	1422
8	3		33	1467
16	4		34	1511
32	5		35	1556
64	6		36	1600
128	7		37	1644

在算法中通过改变自适应 LPC 码本的长度 S 与门限值 T 来计算 LSD、平均编码比特速率以及后向线性预测的采用率等指标客观评估系统的性能。选择不同的 T 值、不同的自适应 LPC 码本的长度 S , MELP 编码器平均编码速率、平均 LSD、平均后向 LPC 采用率的情况如表 3 至表 9 所示。

表 3 $T = 3.0\text{dB}$ 时的计算结果

自适应 码本 S	平均编 码速率	平均对数 谱失真	平均后向 LPC 采用率
2	2281	0.284588	0.116697
4	2267	0.337721	0.136445
8	2244	0.411670	0.166966
16	2240	0.446213	0.179533
32	2229	0.505227	0.202873
64	2190	0.652466	0.262118
128	2152	0.810686	0.328546

表 4 $T = 3.5\text{dB}$ 时的计算结果

自适应 码本 S	平均编 码速率	平均对数 谱失真	平均后向 LPC 采用率
2	2205	0.525126	0.190305
4	2172	0.653741	0.233393
8	2144	0.762867	0.274686
16	2135	0.830640	0.298025
32	2121	0.921364	0.330341
64	2088	1.070205	0.389587
128	2034	1.321949	0.484740

表 5 $T = 4.0\text{dB}$ 时的计算结果

自适应码本 S	平均编码速率	平均对数谱失真	平均后向 LPC 采用率
2	2108	0.882005	0.285458
4	2052	1.112514	0.355476
8	2013	1.286209	0.414722
16	1983	1.467762	0.468582
32	1965	1.614199	0.516260
64	1955	1.694257	0.556558
128	1931	1.836626	0.621105

表 6 $T = 4.5\text{dB}$ 时的计算结果

自适应码本 S	平均编码速率	平均对数谱失真	平均后向 LPC 采用率
2	2004	1.317327	0.387792
4	1959	1.515502	0.450628
8	1929	1.666612	0.504499
16	1907	1.833993	0.554758
32	1891	1.990528	0.603232
64	1884	2.063108	0.644524
128	1849	2.294416	0.728905

表 7 $T = 5.0\text{dB}$ 时的计算结果

自适应码本 S	平均编码速率	平均对数谱失真	平均后向 LPC 采用率
2	1897	1.810964	0.491921
4	1859	2.001436	0.552962
8	1827	2.186924	0.614004
16	1810	2.353946	0.664273
32	1801	2.491407	0.709156
64	1804	2.546785	0.745063
128	1795	2.633452	0.800718

表 8 $T = 5.5\text{dB}$ 时的计算结果

自适应码本 S	平均编码速率	平均对数谱失真	平均后向 LPC 采用率
2	1815	2.232565	0.572711
4	1782	2.415892	0.631957
8	1750	2.617617	0.696589
16	1733	2.804100	0.750449
32	1728	2.941192	0.795332
64	1744	2.939381	0.820467
128	1757	2.984878	0.850987

表 9 $T = 6.0\text{dB}$ 时的计算结果

自适应码本 S	平均编码速率	平均对数谱失真	平均后向 LPC 采用率
2	1732	2.697457	0.653501
4	1703	2.848047	0.707361
8	1685	3.020238	0.766607
16	1687	3.103606	0.802513
32	1687	3.189653	0.843806
64	1709	3.216803	0.863555
128	1727	3.221981	0.890485

对于相同的 T 值, 自适应 LPC 码本的长度 S 从 2 变到 128。由表 3 至表 9 可见, 随着自适应 LPC 码本长度的增加, 后向 LPC 的采用频率相对于前向 LPC 均有所增加, 但却导致了平均谱失真的增加。当 T 值选的较小时, 即其值处于 3.0dB 至 4.5dB 范围内, 由表 3 至表 6 可见, 随着自适应 LPC 码本长度的增加, 用于传输 LPC 系数的带宽即编码器的平均编码速率呈下降趋势, 但当 T 值选得较大时, 即其值处于 5.0dB 至 6.0dB 范围内, 由表 7 至表 9 可见, 用于传输 LPC 系数

的带宽即编码器的平均编码速率随着自适应 LPC 码本长度的增加, 并未呈绝对下降趋势。这一点可解释为大多情况下, 后向预测时基于最临近的合成语音帧的 LPC 系数被选中的概率远高于远离当前帧的已合成语音帧的 LPC 系数被选中的概率, 因而当 T 值选得较大时, 自适应码本越大, 后向 LPC 采用率的提高不足以弥补大码本对编码速率造成的影响, 就会出现这一现象。因此可见自适应码本并不绝对是越大越好, 它只需包含相对小数量的码矢量。

对同一自适应 LPC 码本, 即 S 保持不变, T 值以 0.5dB 的步长从 3dB 增至 6dB。由表 3 至表 9 可见, 当门限 T 值增大时, 允许更大的谱失真, 后向 LPC 采用率逐步上升, 传输 LPC 系数所需的带宽逐步减小, 即编码速率逐步下降。容易推出, 当 T 趋于无穷大时, 几乎全部采用后向线性预测, 但当 T 趋于 0 时, 全部采用前向线性预测。

主观性能的评估是通过数个听众对 8 句男女语音非正式 A-B 听力测试进行的, A 为标准 2400bit/s MELP 合成语音, B 为采用 LPC 自适应前后向量化 MELP 合成语音。自适应码本 S 小于 8 时, 门限 T 在 3.0dB 至 6.0dB 任意选取, 听众们认为两者语音质量相当; 当自适应码本 S 大于或等于 8 时, 门限 T 不超出 4.5dB, 语音质量也无下降, 但若门限超出 4.5dB, 语音质量略有下降, 不很明显, 由表可见, 此时后向 LPC 的采用率已达到并超过 50%。因此, 选

择合适的自适应码本 S 与判决门限 T , 采用上述自适应前后向量化得到的合成语音与 MELP 标准算法的合成语音完全可以媲美, 但同时却达到了显著降低比特速率的目的。

4 结束语

本文介绍的自适应前后向量化方案考虑了语音信号的统计特性, 使传输 LSF 系数所需的带宽显著减少。该方案用在 MELP 编码算法中, 经大量的计算机实验表明合成的语音质量无明显下降, 而平均编码比特速率却大大降低。

参 考 文 献

- [1] Federal Information Processing Standards Publication (MELP), Jan. 7, 1998, Draft.
- [2] A. Gersho, Advanced in speech and audio compression, Proc. IEEE, 1994, 82(6), 900-918.
- [3] Tozsef Vass, Adaptive forward-backward quantizer for low bit rate high-quality coding, IEEE Trans. on Speech and Audio Processing, 1997, 5(6), 552-557.
- [4] J. chen *et al.*, A low-delay CELP coder for the CCITT 16kb/s speech coding standard, IEEE J. on Select. Areas Commun., 1992, 10(6), 830-849.

VARIABLE-RATE MELP SPEECH CODER BASED ON ADAPTIVE FORWARD-BACKWARD QUANTIZATION OF LINEAR PREDICTIVE COEFFICIENTS

Tie Manxia* ** Wang Dusheng* Fan Changxin*

**(Information Science Institute, Xidian University, Xi'an 710071, China)*

*** (Video and Audio Info. Proces. Lab., Beijing University, Beijing 100871, China)*

Abstract In this paper, based on LPC adaptive forward-backward quantization, a novel variable-rate MELP speech coder is proposed, in which linear prediction is done by using either the current (forward LPC) or previously decoded (backward LPC) speech frames. The backward LPC scheme shall be applied, i.e., the LPC coefficients based on the previously decoded optimal speech frame are used to encode the current frame and only the time sequence code shall be transmitted to the decoder, so, average LPC bit number becomes smaller. Computer simulation shows significant average overall bit rate reduction is achieved without compromising the decoded speech quality.

Key words Mixed-Excited Linear Prediction (MELP), Linear prediction, Adaptive forward-backward quantization, Variable rate

铁满霞: 女, 1968 年生, 硕士, 副教授, 研究方向为语音编码、图像处理等。

王都生: 男, 1962 年生, 博士, 副教授, 研究方向为语音编码。

樊昌信: 男, 1931 年生, 教授, 博士生导师, 研究方向为语音编码、图像处理及卫星通信等。