多带激励语音编码器仿真实现

汤 敏, 曾毓敏, 谭锡林

(南京师范大学 物理科学与技术学院, 江苏 南京 210097)

[摘要] 介绍多带激励(MBE)模型原理、分析和合成, 该算法突破了二元激励的局限性, 较好的解决了简单二元激励模型导致的合成语音自然度不够以及抗噪声能力差的问题, 是目前低速率语音编码较理想的方案. 此外, 详细阐述了 MBE编码器的仿真实现方法, 并给出了程序流程和仿真结果. 实验结果表明, 此仿真系统合成的语音在频谱和波形上都和原始语音比较接近, 并具有良好的清晰度和可懂度, 任何熟悉的人都可以辨别说话人.

[关键词] 多带激励, 语音编码, 仿真

[中图分类号]TN912 [文献标识码]A, [文章编号]1672-1292-(2005)01-0072-04

Simulation of the Multiband Excitation Encoder

TANG Min, ZENG Yumin, TAN Xilin

(School of Physical Science and Technology, Nanjing Normal University, Jiangsu Nanjing 210097, China)

Abstract The paper presents the principle of MBE (Multiband Excitation Model) and its analytical and synthetical methods. MBE applies a new excitation spectrum, thus providing high quality speech reproduction for both clean and noisy speech. The algorithm is an ideal scheme of current low-speed speech encoding. The paper also illustrates the simulation of MBE Encoder and provides the flow chat of the program and the result of the simulation. The result shows the synthetical speech and the original speech are similar on the time region and the frequency region, and the synthetical speech is of high definition and comprehension, so the speaker can be easily distinguished.

Key words multiband excitation, speech coding, simulation

0 引言

语音编码方法大致分成 3大类[1]: 波形编码、参数编码和混合编码. 波形编码在 64 kbit/s至 16 kbit/s之间音质优良, 当速率进一步降低时, 其性能下降较快. 参数编码的优点是速率低, 如 2.4 kbit/s, 但合成语音质量较差, 而且算法复杂. 混合编码的数码率约在 4~16 kbit/s之间, 音质比较好, 复杂程度介乎与波形编码和参数编码之间.

带宽资源的有限性使得中低速语音编码的研究越来越受到重视, 以往的声码器在低速率时不能得到高质量的语音, 其原因之一就是激励模型过于简单, 阻碍其进一步提高音质[2,3]. 美国 MIT大学林肯实验室 1988年提出的多带激励语音编码方案[4,5], 采用了有效的激励模型, 克服了传统声码器的缺点, 是目前较为理想的编码方案, 在 2.4~4.8 kb/s的速率上能够合成出比传统声码器好得多的语音, 且具有良好的自然度和容忍环境噪声的能力.

1 多带激励语音模型[1,6,7]

利用语音信号的短时平稳特性, 可以对输入语音信号 $s(n)$ 加窗以进行分帧(窗函数表示为 $w(n)$, 一般窗长为 20~40 ms, 以适应语音信号短时平稳性):

$$s_w(n) = w(n)s(n) \quad (1)$$

用 $S_w(\omega)$ 表示 $s_w(n)$ 的傅氏变换, 将 $S_w(\omega)$ 看成是系统函数 $H_w(\omega)$ 和激励信号谱 $E_w(\omega)$ 的乘积, 即:

$$S_w(\omega) = H_w(\omega)E_w(\omega) \quad (2)$$

收稿日期: 2004-09-28
基金项目: 江苏省教育厅自然科学基金资助项目(2002WLXTSJB125).
作者简介: 汤 敏(1979-), 硕士研究生, 主要从事语音信号编码等方面的学习和研究. E-mail: mynameo990@sina.com
通讯联系人: 曾毓敏(1962-), 副教授, 主要从事适时信号处理、语言编码器等方面的教学与研究. E-mail: zengyumin@njnu.edu.cn

而重建语音信号可表示为:

$$S_{wr}(\omega) = H_{wr}(\omega)E_{wr}(\omega) \quad (3)$$

式中, $H_{wr}(\omega)$ 和 $E_{wr}(\omega)$ 分别是合成器的系统频域函数和激励信号的频域函数, 可以从原始语音信号中提取.

多带激励声码器与传统声码器相比, 主要差别在于激励信号 $E_{wr}(\omega)$ 的表示方法不同. 在 MBE模型中, 将整个频带划分为以基音谐频为中心的互不交叠的谐频带, 对每个频带独立地进行清浊音判决, 多带激励模型因此得名. 总的激励信号 $E_{wr}(\omega)$ 由各带激励信号相加构成. 清音带采用白噪声谱作为激励信号; 浊音带采用加窗周期脉冲序列的傅氏变换 (用 $P_w(w)$ 表示) 作为激励信号. 系统函数 $H_{wr}(\omega)$ 的作用是确定各个频带分量的相对幅度和相位, 起到将 $E_{wr}(\omega)$ 映射成为 $S_{wr}(\omega)$ 的作用.

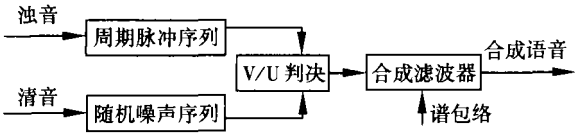

图 1　MBE语音模型

2　多带激励语音分析和合成[1 6 7]

分析过程主要是 3个参数的提取: 基音周期、各谐频带的清浊音判决信息、各谐频对应的谱包络幅度. 为了使合成语音谱在最小均方误差准则条件下逼近原始语音谱, 需同时考虑激励参数和谱包络对合成谱的影响, 因此分两步: 第一步是以合成语音谱和原始语音谱均方误差最小的原则来同时估计基音周期和谱包络参数, 第二步是根据合成谱和原始语音谱之间的匹配程度进行 V/U 判决.

合成的方法有时域合成法和频域合成法, 时域合成法能实现帧与帧之间的基音周期的平滑过渡, 使合成语音更自然, 因此浊音部分用时域合成法; 而由于带通滤波器在频域中易于实现, 而且可以用高效的 FFT 算法, 因此清音用频域合成法. 最后相加, 形成完整的语音.

3　多带激励语音编码实现[1 6 7]

3 1　高、低通滤波[8]

高通滤波器的目的是除去 50Hz工频干扰. 低通滤波器的目的是为了减少高频共振峰和外来高频噪声的影响, 截至频率为 800Hz 分别采用下列 4阶和 5阶的椭圆滤波器滤波:

高通: $H(z) = \dfrac{\sum_{i=0}^{4} A_i z^{-i}}{\sum_{j=0}^{4} B_j z^{-j}}$　低通: $H(z) = \dfrac{\sum_{i=0}^{5} C_i z^{-i}}{\sum_{j=0}^{5} D_j z^{-j}}$

其中: $\{A_i\} = \{0.93563, -3.74083, 5.61033, -3.74083, 0.93563\}$

$\{B_j\} = \{1.0000, -3.9224, 5.7725, -3.7776, 0.92753\}$

$\{C_i\} = \{0.008233, -0.004879, 0.007632, 0.007632, -0.004879, 0.008233\}$

$\{D_j\} = \{1.0000, -3.6868, 5.8926, -5.0085, 2.2518, -0.4271\}$

经过高通滤波后的信号记为 $s_{hp}(n)$, 再经过低通滤波后的信号记为 $s_{lp}(n)$.

3 2　将语音信号 $s_{lp}(n)$ 加窗 $w(n)$ 进行分帧

此处笔者选用 221点的三角窗, 语音信号采样率为 8kHz, 每帧取 20ms, 即 160个样点. 此外, 窗要满足归一化条件, 因此要先对 221点的三角窗进行归一化处理.

3 3　基音提取

基音粗估: 设 P 为可能的基音周期值, 对每一帧语音信号运用下列公式, 对 $P = 20 \sim 147$的所有整数计算 $\xi_B(P)$, 使 $\xi_B(P)$ 最小的 P 值就是基音估计的粗估值:

$$\xi_B = \frac{\sum_{n=1}^{221} w^2(n) s_{lp}^2(n) - P\sum_{k=1}^{\lfloor \frac{221}{P} \rfloor} \Phi(kP)}{\left[1 - P\sum_{n=1}^{221} w^4(n)\right]\left[\sum_{n=1}^{221} w^2(n) s_{lp}^2(n)\right]} \quad (4)$$

其中:

$$\Phi(m) = \sum_{n=-\infty}^{\infty} w^2(n) s(n) w^2(n-m) s(n-m) \quad (5)$$

基音平滑: 以进一步提高粗估准确性, 增强基音周期的连续性, 去除估计错误点. 统计结果表明, 语音信号基音每 ms的变化率最大不超过 1%, 因此此处如果前后帧基音超过 20% 就令当前基音为当前帧基音和前一帧基音的平均值.

基音细搜索: 基音细估的频率范围为 $\frac{2\pi}{148.125} \leqslant \omega_0 \leqslant \frac{2\pi}{18.875}$ 估计的精度为 0.25个样点. 选取十个细搜索候选点, 利用 $\omega = \frac{2\pi}{P}$, 将 $P_i = P - \frac{9}{8}, P - \frac{7}{8}, \ldots, P + \frac{7}{8}, P + \frac{9}{8}$ 转换成频域上的 ω_i. 对每

个 ω_i 进行下列运算,使 ε最小的 ω_i 值即为基音周期的精估值 ω_0:

$$\varepsilon(\omega_i) = \sum_{l=1}^{127} G(l) \mid S_w(l) - S_{wr}(l, \omega_i) \mid^2 \tag{6}$$

$S_w(l)$ 是当前帧语音的频谱,取 256点 DFT;

$S_{wr}(l, \omega_0) = A_m(\omega_0) W\left(?64l - \frac{16\,384}{2\pi}\omega_0 + 0.5\right)$,

$a_m \leqslant l \leqslant b_m$,是合成信号谱;

$$A_m(\omega_0) = \frac{\sum_{l=a_m}^{b_m} S_w(l) W^*\left(?64l - \frac{16\,384}{2\pi} m\omega_0 + 0.5\right)}{\sum_{l=a_m}^{b_m} \left| W\left(?64l - \frac{16\,384}{2\pi} m\omega_0 + 0.5\right)\right|},$$

是第 m 次谐波带的最佳谱包络; a_m、b_m 分别是该频带的下限和上限.

为避免基音细化窗函数谱在频域上产生混叠,应选取旁瓣低的窗函数,但旁瓣降低会导致主瓣加宽.如果主瓣太宽,而基音周期较高时,主瓣也可能产生混叠.考虑到以上因素,本文仿真系统中选用 221点的三角窗做基音细化窗,对其做 16 384点 DFT,得到 $W(l)$.

细搜索主要考虑的是高次谐波带拟合的好坏,因此本文只考虑从 1 562.5~3 800 Hz谐波带内的拟合误差.公式(6)中的求和上下限分别取:

$$50, ?\left(?\frac{0.96\pi}{\omega_0} - 0.5\right)\frac{256}{2\pi}\omega_0」$$

3.4 清浊音判决

通过大量试验可以看到,清音带与浊音带不会频繁交替,而是保持着一定的连续性,这样在编码速率较低时,可以将相邻的几个谐频带划分在一起,共同进行清浊音判决.本文将相邻的 3个谐频带划分在一起,整个频带采用最多分成 12个带的方法进行 V/U判决.

判决阈值采用自适应值,如果拟合误差小于阈值,判为浊音,否则判为清音:

$$\theta(k, \omega_0) = (0.35 + 0.557\omega_0)[1 - 0.475(k-1)\omega_0]F \tag{7}$$

其中: $E_0 = \sum_{l=0}^{??\frac{0.96\pi}{\omega_0}-0.5」\frac{256}{2\pi}\omega_0」} \mid S_w(l) \mid^2$;

$E_{avg}(0) = 0.7E_{avg}(-1) + 0.3E_0$;

$$E_{min}(0) = \begin{cases} 0.5E_{min}(-1) + 0.5E_0 & \text{如果 } E_0 \leqslant E_{min}(-1) \\ 0.975E_{min}(-1) + 0.025E_0 & \text{如果 } E_{min}(-1) \leqslant E_0 \leqslant 2E_{min}(-1); \\ 1.025E_{min}(-1) & \text{其它} \end{cases}$$

$$E_{max}(0) = \begin{cases} 0.5E_{max}(-1) + 0.5E_0 & \text{如果 } E_0 > E_{max}(-1) \\ 0.99E_{max}(-1) + 0.01E_0 & \text{其它} \end{cases}$$

$$F = \begin{cases} 0.5 & \text{如果 } E_{avg}(0) < 200 \\ \dfrac{(E_0 + E_{min}(0))(2E_0 + E_{max}(0))}{(E_0 + 0.0075E_{max}(0))(E_0 + E_{max}(0))} & \text{如果 } E_{avg}(0) \geqslant 200 \text{ and } E_{min}(0) < 0.0075E_{max}(0) \\ 1.0 & \text{其它} \end{cases}$$

3.5 计算各谐频带的谱包络幅度 x_m

浊音带: $x_m = |A_m(\omega_0)|$;

清音带: $x_m = \left[\frac{\sum_{l=a_m}^{b_m} |S_w(m)|^2}{b_m - a_m}\right]^{\frac{1}{2}}$.

3.6 清音部分的合成

产生一个均值为 0,频谱密度值为 1的白噪声序列 $u(n)$,对其加 221点的三角窗,然后进行 256点的 DFT,得到频谱 U_w.令所有判为浊音带的频带包络为 0,而判为清音带的频带包络为: $U_{wc}(l) = \frac{x_m U_w(l)}{\left[\sum_{k=a_m}^{b_m} |U_w(k)|\right]^{\frac{1}{2}}}$.然后对所得信号再进行 256点 IDFT,得到时间序列 $u_{wc}(n)$.为了使合成语音连续,将所得的时间序列与前一帧清音序列作叠接处理,即得当前帧的清音部分 $s_{ru}(n)$.

3.7 浊音部分的合成

同清音部分的合成类似,对各谐波频率处的幅度加以修正.令清音频带的谱包络为 0,浊音频带的谱包络不变,然后进行插值.再用一组余弦波在时域中直接合成得到 $s_{rv}(n)$.

3.8 重建语音的产生

将求出的清音部分和浊音部分相加,即得到最后的合成语音 $s_r(n) = s_{ru}(n) + s_{rv}(n)$.

3.9 算法框图

根据上述步骤,仿真主要分为分析和合成过

程, 其框图分别如下:

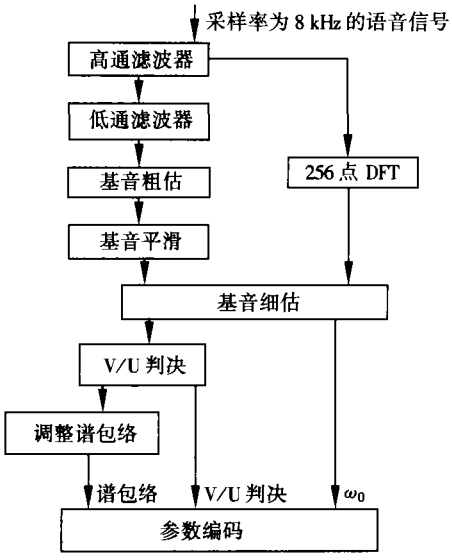

图 2　分析算法框图

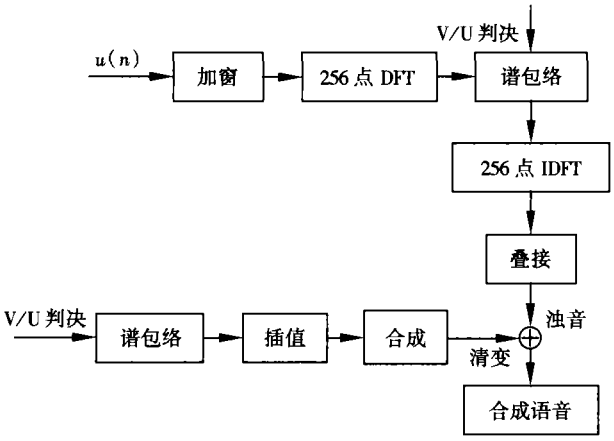

图 3　合成算法框图

3 10　仿真结果

我们取一帧实际语音, 分别作出原始语音频谱 $S_w(\omega)$ 与合成语音频谱 $S_{wr}(\omega)$.

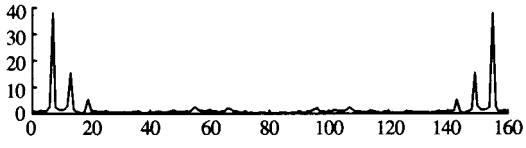

图 4　原始语音谱

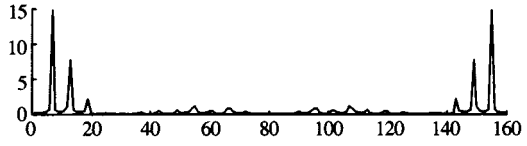

图 5　合成语音谱

下图分别是男女声的原始语音和合成语音的时域信号波形:

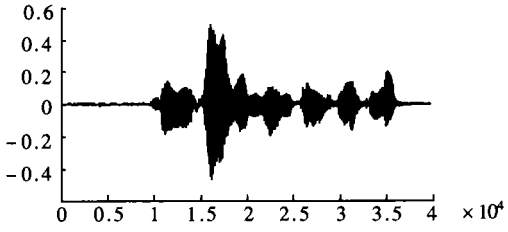

图 6　男声原始语音

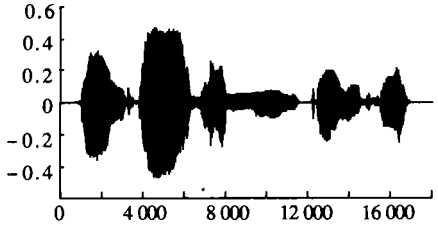

图 7　女声原始语音

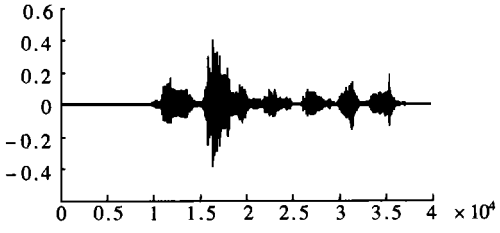

图 8　男声合成语音

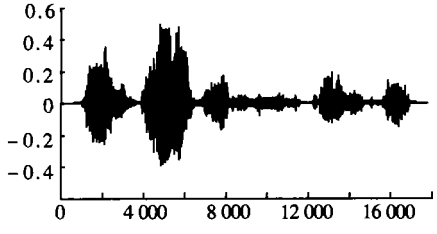

图 9　女声合成语音

无论从频谱还是从波形上看原始语音波形与合成波形都比较接近, 合成语音的主观听觉效果良好, 具有较高的清晰度与可懂度, 任何熟悉的人都可以辨别说话人.

Matlab为平台对多带激励语音编码算法进行了仿真, 仿真结果与原始语音比较接近, 并具有良好的自然度和可懂度. 但多带激励需要的运算量大, 很难做到实时处理, 解决此问题需要进一步的研究.

4　结论

(下转第 94页)

本文深入研究了多带激励语音模型, 并采用

数递推关系式的特征多项式: $C(x) = x^3 - 5x^2 + 8x - 4$ 该特征多项式有 3个根 $x_{1,2} = 2$(二重根), $x_3 = 1$, 则递推关系式的解即 $t(n)$为:

$t(n) = (A_0 + A_1 n) \cdot 2^n + A_2 \cdot 1^n$, 其中 A_0, A_1, A_2为待定常系数, 根据初值求得 $A_0 = 3$, $A_1 = -1$, $A_2 = -3$, 所以该递归算法的时间复杂度 $t(n) = 3 \cdot 2^n - n \cdot 2^n - 3$

1.4 推论 3

设某一递归算法时间复杂度为 $t(n)$, 其递推关系式所对应的特征多项式 $C(x)$有不同的复根, 此时可按照"推论 1"的办法处理. 不过复数有它的特点, 假如特征多项式 $C(x)$的两个共轭复根 x_1, x_2可化成下列形式: $x_1 = \rho(\cos\theta + i\sin\theta)$, $x_2 = \rho(\cos\theta - i\sin\theta)$. 例如将特征多项式 $C(x) = x^2 + x + 1$的两个复根 $x_1 = (-1 + \sqrt{3}i)/2$, $x_2 = (-1 - \sqrt{3}i)/2$可写成:

$$x_1 = \cos\frac{2}{3}\pi + i\sin\frac{2}{3}\pi,$$

$$x_2 = \cos\frac{2}{3}\pi - i\sin\frac{2}{3}\pi \qquad (\rho = 1).$$

此种情况下, 递推关系式的解即递归算法的时间复杂度可根据推论 1有:

$t(n) = k_1 x_1^n + k_2 x_2^n = k_1 \rho^n (\cos n\theta + i\sin n\theta) + k_2 \rho^n (\cos n\theta - i\sin n\theta) = \rho^n (k_1 + k_2)\cos n\theta + i\rho^n (k_1 - k_2)\sin n\theta$

$k_1 + k_2$和 $i(k_1 - k_2)$仍然是待定常数, ρ, θ可由复根求得. 令 $A = k_1 + k_2$, $B = i(k_1 - k_2)$得到:

$t(n) = A\rho^n \cos n\theta + B\rho^n \sin n\theta$

应用实例由于篇幅所限, 作者在此不再赘述.

2 结论

在分析递归算法的时间复杂度 $t(n)$时, 如果 $t(n)$的递推关系式是一个常系数线性递推关系式, 则可以利用母函数与递推关系理论求出其时间复杂度.

[参考文献]

[1] 严蔚敏, 吴伟民. 数据结构[M]. 北京: 清华大学出版社, 1991. 303-305.

[2] 谭浩强. C程序设计[M]. 第 2版. 北京: 清华大学出版社, 2003. 160-163.

[3] 黄国瑜, 叶乃箐. 数据结构(C语言版)[M]. 北京: 清华大学出版社, 2001. 13-18.

[4] 卢开澄. 组合数学[M]. 北京: 清华大学出版社, 2000. 48-79.

[5] 杨骅飞, 王朝瑞. 组合数学及其应用[M]. 北京: 北京理工大学出版社, 1992. 40-48.

[责任编辑: 刘健]

(上接第 75页)

[参考文献]

[1] 王炳锡. 语音编码[M]. 西安: 西安电子科技大学出版社, 2002. 257-274.

[2] Jamrozik M, Gowdy J. Enhanced quality modified multiband excitation model at 2400 bps. Bringing Together Education, Science and Technology[J]. Proceedings of the IEEE, 1996. 223-226.

[3] Jamrozik M, Gowdy J. Modified multiband excitation model at 2400 bps[J]. Proceedings of ICASSP-97, 2: 1603-1606.

[4] Yu W M E, Chan C F. Multiband excitation coding of speech at 2.0 kbps[J]. Proceedings of ISSIPNN-94, 2: 559-562.

[5] Xu Peixia, Chen Zhou, Liu Wenfei. Wavelet analysis based multiband excited vocoder[A]. TENCON'93. Proceedings Computer, Communication, Control and Power Engineering[C]. 1993 IEEE Region 10 Conference on, 1993, (2): 349-352.

[6] Griffith D W, Lim J S. Multiband excitation vocoder[A]. Acoustics, Speech, and Signal Processing[J]. IEEE Transactions on, 1988, 36(8): 1223-1235.

[7] 杨行峻, 迟惠生. 语音信号数字处理[M]. 北京: 电子工业出版社, 1995. 265-283.

[责任编辑: 刘健]