一种谐波正弦语音模型的最佳相位估计算法

应 娜[1],赵晓晖[2],董 婧[3],方 昕[1]
(1. 杭州电子科技大学通信工程学院,浙江杭州 310018;2. 吉林大学通信工程学院
信息科学实验室,吉林长春 130022;3. 大连理工大学机械工程学院,辽宁大连 116024)

摘 要: 基于谐波正弦语音模型(HSSM),利用最小二乘方法估计语音模型的最佳相位参数,给出了一种估计相位的批处理方法和迭代算法.把利用该算法得到的相位参数用于宽带语音编解码算法进行仿真,其结果与 G.722.2 标准宽带编码算法中的两种编码速率 8.85kbit/s 及 6.60kbit/s 的语音进行了比较,语音波形的比较和主客观测试结果表明该最佳相位估计算法相位参数估计准确有效,可由此建立的语音模型获得较高质量的合成语音.

关键词: 谐波正弦语音模型; 相位估计; 最小二乘; 迭代算法
中图分类号: TN912.3 **文献标识码**: A **文章编号**: 0372-2112 (2009) 04-0860-04

An Algorithm of Optimal Phase Estimation in Harmonic Sinusoidal Speech Model

YING Na[1], ZHAO Xiao-hui[2], DONG Jing[3], FANG Xin[1]
(1. *Communication Engineering College of Hangzhou Dianzi University, Hangzhou, Zhejiang* 310018, *China*;
2. *Laboratory of Information Science, College of Communication Engineering, Jilin University, Changchun, Jilin* 130022, *China*;
3. *School of Mechanical & Engineering, Dalian University of Technology, Dalian, Liaoning* 116024, *China*)

Abstract: Based on harmonic sinusoidal speech model (HSSM) and through Least Square method for the optical estimation of phase parameters, both batch and an iterative phase estimation algorithms are presented. Using these estimated phase parameters in a wideband speech coding algorithm to emulate and compare with 8.85kbit/s and 6.60kbit/s speech coding algorithm in G.722.2 standard, the simulation results of waveform and subjective/objective test demonstrate that the proposed phase estimation algorithm is accurate and effective. And higher quality synthesized speech can be obtained with these estimated phase parameters in HSSM.

Key words: harmonic sinusoidal speech model; phase estimation; least square; iterative algorithm

1 引言

相位信息在高质量语音合成中非常重要,如何建立更好的语音相位模型,使重建语音质量更高、更自然是目前语音信号处理研究的热点之一.通常利用 ARMA 模型或者结合使用 ARMR 模型与量化算法经常用来表示和建立最小相位模型,但由于语音信号并不是完全的最小相位模型,一些编码算法摈弃了相位参数,在合成端利用二、三、五次相位内插方法合成相位[1,2].还有声门起始脉冲估计算法的提出,也为正弦语音模型的相位估计和合成做出了贡献[3].在基于正弦语音模型的编码算法中,McAulay 和 Quatieri 提出混合相位模型中浊音部分的相位参数取自 LPC 合成谱,而清音信号的相位随机[1,3].Ahmadi 和 Spanias 提出在进行 LPC 分析后,利用相位补偿和纠正进行更准确的相位分析.Cheetham 和 Sun 等人提出了用二阶全通滤波器进行相位补偿的方法[4].Zhao 和 Gordon 等人则提出用其它代价函数来代替相位函数[5].还有其它的基于相位模型的语音编码方法也对相位模型的建立进行了研究和讨论[6~8].

基于谐波正弦语音的模型(HSSM)在很多语音编码算法中获得重要应用,它认为正弦语音模型的尖锋处频率是基音的整数倍.由于语音的谐波特性并不完全按照基音倍数存在,存在偏差,为了使该模型更有效地符合原始语音信号,有必要引入相位模型,并对相位进行准确的估计.

本文根据谐波正弦语音模型非严格周期性的特点,依据最小二乘准则,提出了一种新的谐波语音相位模型的最佳相位估计算法,并将此模型应用于宽带语音编解码中,以验证该模型的准确性.

谐波正弦语音模型如式(1)

收稿日期:2007-02-01;修回日期:2008-12-15
基金项目:浙江省自然科学基金(No. KYZ040306017)

© 1994-2010 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

$$\hat{s}(m) = \sum_{l=1}^{L} B_l \exp\{ j[l\omega_0 (m - m_0) + \Phi_s(l\omega_0)]\} \quad (1)$$

其中, m 是帧序列数, B_l 是谐波尖峰的幅度, ω_0 是基音, $\Phi_s(\omega)$ 是系统相位, m_0 是声门脉冲的起始时间,表示每帧语音里最靠近帧中心基音脉冲出现的时间. 但是考虑语音变化及外界环境干扰等因素,对声门起始时间的估计(相位估计)计算量大、准确度低. 就此问题,本文研究一种新的谐波正弦语音模型的相位估计方法.

2　相位模型和最佳相位的估计

2.1　相位模型及其最佳估计

假设根据 SEEVOC 算法寻找到第 k 帧、第 l 次尖峰时[9],其幅度为 B_l^k,频率固定为 $\omega_l^k = l\omega_0$,则相位 $\phi_l^k \approx - l\omega_0 m_0 + \Phi_s(l\omega_0)$. 当幅度和频率固定时,用最佳相位 ϕ_{ls}^k 表示声门起始时间和系统相位的组合,则第 k 帧语音 $s^k(m)$ 表示为

$$s^k(m) = \sum_{l=1}^{L} B_l^k \mathrm{Re}\{ \exp[j(\omega_l^k m + \phi_{ls}^k)]\},\ m = 1, 2, \cdots, M \quad (2)$$

其中, $\mathrm{Re}(\cdot)$ 函数表示取实数部分, $\mathrm{Img}\{\cdot\}$ 函数则表示取虚数部分. 用矩阵表示第 k 帧的各个参数,则可得幅度矩阵 B 和频率指数函数矩阵 M 分别可表为

$$\mathrm{B} = \begin{bmatrix} B_1 & & 0 \\ & O & \\ 0 & & B_L \end{bmatrix}$$

$$\mathrm{M} = \begin{bmatrix} \exp(j\omega_0) & \exp(j\omega_0 \cdot 2) & \cdots & \exp(j\omega_0 \cdot M) \\ \exp(j2\omega_0) & \exp(j\omega_0 \cdot 2) & \cdots & \exp(j2\omega_0 \cdot M) \\ \cdots & \cdots & \cdots & \cdots \\ \exp(jL\omega_0) & \exp(jL\omega_0 \cdot 2) & \cdots & \exp(jL\omega_0 \cdot M) \end{bmatrix}$$

其中, L 表示每帧尖峰总数. 设 ϕ_l 和其指数函数 $\exp(j\phi_l)$ 的矢量形式为 $\phi_{ls} = [\phi_1, \cdots, \phi_L]$, $\psi_e = [\exp(j\phi_1), \cdots, \exp(j\phi_L)]$,式(2)可以表达成

$$\tilde{\mathrm{S}} = \mathrm{Re}(\psi_e \mathrm{BM}) \quad (3)$$

其中, $\tilde{\mathrm{S}}$ 表示语音信号矢量. 已知信号矢量 S,幅度矩阵 B 和频率矩阵 M,其中 S 和 B 为实数矩阵, M 为非相关矩阵,均不可能为零矩阵. 于是可以根据最小二乘方法求得最佳相位矩阵,设代价函数

$$J = [\mathrm{Re}(\psi_e \mathrm{BM}) - \mathrm{S}]\mathrm{Re}(\psi_e \mathrm{BM}) - \mathrm{S}]^{\mathrm{T}} \quad (4)$$

根据矩阵求导公式,有

$$\frac{\mathrm{d}J}{\mathrm{d}\phi_{ls}} = 2 \cdot \frac{\mathrm{d}[\mathrm{Re}(\psi_e \mathrm{BM}) - \mathrm{S}]}{\mathrm{d}\phi_{ls}} [\mathrm{Re}(\psi_e \mathrm{BM}) - \mathrm{S}]^{\mathrm{T}} \quad (5)$$

$$\frac{\mathrm{d}[\mathrm{Re}(\psi_e \mathrm{BM}) - \mathrm{S}]}{\mathrm{d}\phi_{ls}} = \frac{\mathrm{d}[\mathrm{Re}(\psi_e \mathrm{BM})]}{\mathrm{d}\phi_{ls}} = \mathrm{Re}\left[\frac{\mathrm{d}(\psi_e \mathrm{BM})]}{\mathrm{d}\phi_{ls}} \right] = \mathrm{Re}[j\psi_e \mathrm{BM})] \quad (6)$$

已知 $\mathrm{Re}[j\psi_e \mathrm{BM}]$ 不为零矩阵,若令 $\frac{\mathrm{d}J}{\mathrm{d}\phi_{ls}} = 0$,则必有

$$[\mathrm{Re}(j\psi_e \mathrm{BM}) - \mathrm{S}]^{\mathrm{T}} = 0 \quad (7)$$

把频率和相位指数矩阵分别分解为实数部分和虚数部分

$$\mathrm{M} = \mathrm{M}_r + j\mathrm{M}_i \quad (8)$$

其中:

$$\mathrm{M}_r = \mathrm{Re}(\mathrm{M}) = \begin{bmatrix} \cos(\omega_0) & \cos(\omega_0 \cdot 2) & \cdots \cos(\omega_0 \cdot M) \\ \cos(2\omega_0) & \cos(2\omega_0 \cdot 2) & \cdots \cos(2\omega_0 \cdot M) \\ \cdots & & \\ \cos(L\omega_0) & \cos(L\omega_0 \cdot 2) & \cdots \cos(L\omega_0 \cdot M) \end{bmatrix}$$

$$\mathrm{M}_i = \mathrm{IMG}(\mathrm{M}) = \begin{bmatrix} \sin(\omega_0) & \sin(\omega_0 \cdot 2) & \cdots \sin(\omega_0 \cdot M) \\ \sin(2\omega_0) & \sin(2\omega_0 \cdot 2) & \cdots \sin(2\omega_0 \cdot M) \\ \cdots & & \\ \sin(L\omega_0) & \sin(L\omega_0 \cdot 2) & \cdots \sin(L\omega_0 \cdot M) \end{bmatrix}$$

$$\psi_e = \psi_r + j\psi_i \quad (9)$$

其中, $\psi_r = [\cos(\phi_1), \cdots, \cos(\phi_L)]$, $\psi_i = [\sin(\phi_1), \cdots, \sin(\phi_L)]$. 把式(8)和式(9)代入式(7),有

$$\mathrm{Re}[j\psi_r + j\psi_i) \mathrm{B}(\mathrm{M}_r + j\mathrm{M}_i)] - \mathrm{S} = 0 \quad (10)$$

则根据上式分解,可得

$$\psi_r \mathrm{BM}_r - \psi_i \mathrm{BM}_i = 0 \quad (11)$$

$$\psi_r \mathrm{BM}_i + \psi_i \mathrm{BM}_r + \mathrm{S} = 0 \quad (12)$$

根据式(11)和式(12)及矩阵伪逆公式,解出

$$\psi_r = - \mathrm{SZB}^{-1} \quad (13)$$

$$\psi_i = \psi_r \mathrm{BM}_r \mathrm{PB}^{-1} = - \mathrm{SZM}_r \mathrm{PB}^{-1} \quad (14)$$

其中:

$$\mathrm{Z} = \mathrm{Y}_M^{\mathrm{H}} \left(\mathrm{Y}_M \mathrm{Y}_M^{\mathrm{H}} \right)^{-1} \quad (15)$$

$$\mathrm{Y}_M = \mathrm{M}_i + \mathrm{M}_r \mathrm{PM}_r \quad (16)$$

$$\mathrm{P} = \mathrm{M}_i^{\mathrm{H}} \left(\mathrm{M}_i \mathrm{M}_i^{\mathrm{H}} \right)^{-1} \quad (17)$$

再由 ψ_r 和 ψ_i 求出相位 ϕ_{ls}. 根据反三角函数及正弦、余弦角在各象限的正负值,假设角 ϕ_l 的正弦值为 a_i,余弦值为 b_i,则有

$$\phi_l = \begin{cases} \arcsin(\phi_l), & b_i \geq 0 \\ \arcsin(\phi_l) + 0.5\pi, & a_i \geq 0 \text{ 且 } b_i \leq 0,\ i = 1, \cdots, L \\ \arcsin(\phi_l) - 0.5\pi, & a_i \leq 0 \text{ 且 } b_i \leq 0 \end{cases} \quad (18)$$

由此,得到最佳相位的最小二乘估计. 在估计算法中,矩阵逆的求解计算比较复杂,而且要占计算机较多的内存单元和计算时间. 因此,有必要在此结果的基础上推导出相位模型估计的迭代算法.

2.2　相位模型估计的迭代算法

假设式(17)的第 k 次观测表示为

$$\mathrm{P}_k = \mathrm{M}_{ik}^{\mathrm{H}} \left(\mathrm{M}_{ik} \mathrm{M}_{ik}^{\mathrm{H}} \right)^{-1} \quad (19)$$

$k+1$ 次新的观测表示为

© 1994-2010 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

$$P_{k+1} = M^{H}_{i(k+1)}\left[M_{i(k+1)} M^{H}_{i(k+1)}\right)^{-1} \quad (20)$$

为方便起见,定义

$$p_k = \left(M_{ik}M^{H}_{ik}\right)^{-1},\ p_k^{-1} = M_{ik}M^{H}_{ik} \quad (21)$$

则,式(19)变为

$$P_k = M^{H}_{ik}\, p_k \quad (22)$$

式(20)变为

$$P_{(k+1)} = M^{H}_{i(k+1)}\, p_{k+1} \quad (23)$$

对矩阵 $P_{(k+1)}$ 进行分块处理,有

$$\begin{aligned} p_{k+1} &= \left(M_{i(k+1)} M^{H}_{i(k+1)}\right)^{-1} \\ &= \left\{ \left[M_{ik} \quad M_{i(k+1)}\right] \left[M_{ik} \quad M_{i(k+1)}\right]^{H}\right\}^{-1} \\ &= \left(p_k^{-1} + M_{i(k+1)} M^{H}_{i(k+1)}\right)^{-1} \end{aligned} \quad (24)$$

矩阵 $\left(p^{-1} + H^{H}R^{-1}H\right)$ 的逆

$$\left(p^{-1} + H^{H}R^{-1}H\right)^{-1} = p - pH^{H}\left(HpH^{H} + R\right)^{-1} Hp$$

利用上式,可得

$$\left(p^{-1} + HR^{-1}H^{H}\right)^{-1} = p - pH\left(H^{H}pH + R\right)^{-1} H^{H}p \quad (25)$$

所以将式(25)代入式(24),令 $A = M^{H}_{i(k+1)}\, p_k M_{i(k+1)} + I$,将其代入式(23),则得到

$$P_{(k+1)} = M^{H}_{i(k+1)}\left\{ p_k - p_k M_{i(k+1)} A^{-1} M^{H}_{i(k+1)}\, p_k\right\} \quad (26)$$

其中,$p_1 = \left(M_{i1} M^{H}_{i1}\right)^{-1}$. 如果 M 的维数为 $L \times K$,则 p_k 是一个 $L \times L$ 矩阵(已知观测次数 k 足够高,远远大于 L),当利用式(22)和式(23)求得估计时,要计算一个 $L \times L$ 矩阵的逆矩阵,用 LU 分解方法求则计算量为 $o(1/3L^2 + L^3)$;但是如果用式(26),因为 $M^{H}_{i(k+1)}$ 是 $1 \times L$ 的向量,则 $(M^{H}_{i(k+1)}\, p_k M_{i(k+1)} + I)$ 是一个数量,求逆是简单除法.

同理,式(15)可变为

$$Z_{k+1} = Y^{H}_{M(k+1)} \left\{ q_k - q_k Y_{M(k+1)}\left[Y^{H}_{M(k+1)}\, q_k Y_{M(k+1)} + I\right]^{-1} Y^{H}_{M(k+1)}\, q_k\right\} \quad (27)$$

其中

$$q_{k+1} = q_k - q_k Y_{M(k+1)}\left[Y^{H}_{M(k+1)}\, q_k Y_{M(k+1)} + I\right]^{-1} Y^{H}_{M(k+1)}\, q_k \quad (28)$$

$$q_1 = \left(Y_{M1} Y^{H}_{M1}\right)^{-1} \quad (29)$$

因为 B 是已知的对角矩阵,所以

$$Y_{Mk} = M_{ik} + M_{rk} P_k M_{rk} \quad (30)$$

由此,最佳相位模型的估计可以由式(13)、式(14)、式(18)和式(26)~式(30)迭代计算得到.

3 仿真分析

3.1 波形比较

通过基于该语音模型的低速语音编码算法来验证本文提出方法的有效性和正确性. 由上节的最佳相位估计算法得到最佳相位,对语音进行分析和重建. 在对语音进行预处理、低通滤波、分帧、清浊音分类以及基音提取,并提取基音与倍频基音处的幅度值后,根据式(15)和式(16)提取相位参数;在解码端,根据清浊音的不同,浊音部分用式(2)表示,清音部分用高斯白噪声表示,并叠加合成. 为方便起见把该编码算法简称为 HPM 编码算法,其仿真结果如图 1 所示. 可以看出,重建语音与原始语音信号相似;同原始语音相比,合成语音略有一些毛刺. 为看清语音波形的细节,再分别在原始语音信号与重建语音信号中取小段语音进行比较,如图 2 所示,可以看到波形非常近似,甚至在细节上也表现地非常一致. 这说明,利用本文提出的相位估计算法得到的语音模型能得到良好的语音重建波形.

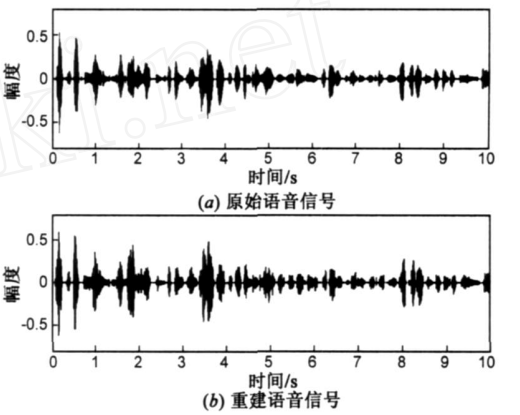


图1 原始语音与重建语音信号图

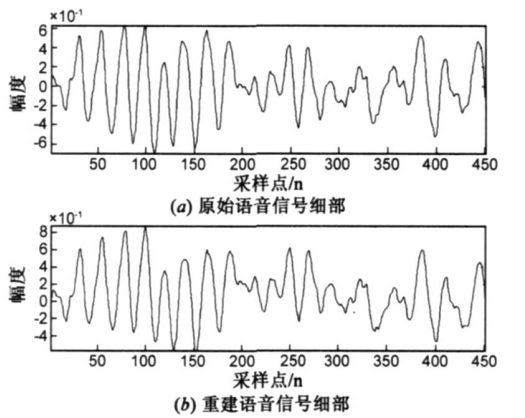


图2 原始语音与重建语音截取图

3.2 语音质量的客观质量评价及比较

根据语音质量的评价指标,对利用最佳相位估计构造的语音模型得出的 HPM 语音编码算法进行性能评测,以反映本文提出的相位估计算法的性能. 首先对原始语音用多种语音编码器进行编解码,然后根据各种客观评价方法对其进行比较. 将 HPM 算法与基于正弦语音模型的编码方案相比[2,7],具有更好的吻合度和清晰度. HPM 算法的传输比特率是宽带条件下作者设计的 8.56kbit/s 的编码方案,因此采用标准宽带语音编码

© 1994-2010 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

算法 G.722.2 中 6.60kbit/s(记为 G.722(1))和 8.85kbit/s(记为 G.722.2(2))的语音进行比较.得到以下结论:HPM 的各种语音评价测度数据在不同的测试标准下都接近 G.722.2(1)、(2)的算法.这可以说明 HPM 算法具有较好的客观评价结果,具有较高的语音重建质量,从而间接地验证了相位估计的准确性和由此建立起来的语音模型的精确性.

3.3　主观评价

采用 38 人测听实际语音信号,按 MOS 标准给出分值.语音信号包括中英文女声、中英文男声、以及女孩、男孩的语音信号.带噪声语音的 MOS 测试中(如表 1 所示),HPM 算法的重建语音 MOS 分数对于原始语音信号并没有下降太多,但是不如 G.722.2(2)的算法质量,在信噪比降低的情况下,与 G.722.2(1)的算法质量比较接近.在各信噪比的情况下,重建语音质量稳定.通过这些不同角度的仿真语音可以看出,由本文提出的语音模型相位的估计方法具有精度高、稳定性好和易实时实现等优点.它可以为基于语音正弦模型的语音编码算法提供可靠的相位估计.

表 1　带噪声语音的 MOS 测试

信噪比 / 编码算法	20dB (各种声音)	15dB (各种声音)	10dB (各种声音)	5dB (各种声音)
原始语音	3.97	3.50	3.60	3.40
HPM(8.56kbit/s)	3.60	3.58	3.50	3.48
G.722.2(2)	4.19	3.82	3.93	3.92
G.722.2(1)	3.67	3.55	3.55	3.45

4　结论

根据语音特性的非严格周期性特点,基于谐波正弦语音模型,依据最小二乘准则,给出了一种新的谐波语音相位的相位最佳估计方法.在批处理算法的基础上得到了迭代算法.通过将此算法应用到宽带语音编码算法中进行了大量的计算机仿真及客观和主观的测试对比,结果表明,该最佳相位估计算法具有语音重建波形细节上非常接近原始语音;语音主客观评价几乎接近 G.722.2 的 8.85kbit/s 及 6.60kbit/s 的语音编码标准.这说明了该算法对谐波正弦语音模型的相位估计准确.

参考文献:

[1] R J McAulay, T F Quatieri. Low-rate Speech Coding Based on the Sinusoidal Model[M]. New York, Marcel Dekker, 1992.

[2] S Ahmadi, A S Spanias. A new phase model for sinusoidal transform coding of speech[J]. IEEE Transactions on Speech and Audio Processing, 1998, 6(5): 495 - 501.

[3] R J McAulay, T F Quatieri. Sinusoidal Coding[M]. Kleijn W B. New York, Elsevier Science B V, 1995.

[4] B M G Cheetham, X Q Sun, P A Barrett. Phase modeling for low bit-rate speech coding[A]. IEE Seminar on Speech Coding for Algorithms for Radio Channels[C]. London, UK, April 2000. 1 - 5.

[5] Q Zhao, G L Stuber. Convergence behavior of iterative phase synchronization and detection[J]. Selected Areas in Communications of IEEE, 2005, 23(5): 932 - 943.

[6] M R Morelande, A A Zoubir. Model selection of random amplitude polynomial phase signals[J]. IEEE Transactions on Signal Processing, Mar 2002, 50(3): 578 - 589.

[7] A S Karunajeewa, U R Abeyratne, C Hukins. Mixed-phase modeling of snore sounds within a nonlinear framework for component identification[A]. Nonlinear Signal and Image Processing[C], May 2005, Sapporo, Japan. 2005. 42 - 42.

[8] Lu Kun, Wang Jiong, Liu Xing-zhao. A piecewise parametric method based on polynomial phase model to compensate ionospheric phase contamination [A]. ICASSP 2003 [C], Hong Kong, China, Vol. 2, April 2003. 6 - 10.

[9] D B Paul. The spectral envelope estimation vocoder[J]. IEEE Transaction on Acoustic, Speech and Signal and Proc, 1981. 786 - 94.

作者简介:



应　娜　女,1978 年 12 月出生于浙江仙居.通信与信息系统博士,杭州电子科技大学通信工程学院讲师,主要研究方向:信号处理及其在通信网络中的应用,语音信号处理等.
E-mail: yingna@hdu.edu.cn.



赵晓晖　男,满族,1957 年 11 月出生于吉林长春.吉林大学通信工程学院教授,博士生指导教师.主要研究方向为通信中的信号处理理论与技术.近年来在国内外学术刊物和会议上发表学术论文 80 多篇.

© 1994-2010 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net