物理学报 Acta Physica Sinica
Chinese Physical Society
Institute of Physics, CAS

语音信号序列的Volterra预测模型

张玉梅　胡小俊　吴晓军　白树林　路纲

Volterra prediction model for speech signal series

Zhang Yu-Mei　Hu Xiao-Jun　Wu Xiao-Jun　Bai Shu-Lin　Lu Gang

引用信息 Citation: Acta Physica Sinica, 64, 200507 (2015)　DOI: 10.7498/aps.64.200507
在线阅读 View online: http://dx.doi.org/10.7498/aps.64.200507
当期内容 View table of contents: http://wulixb.iphy.ac.cn/CN/Y2015/V64/I20

语音信号序列的Volterra预测模型*

张玉梅[1)2)3)] 胡小俊[2)] 吴晓军[1)2)3)†] 白树林[4)] 路纲[1)2)]

1) (陕西师范大学, 现代教学技术教育部重点实验室, 西安 710062)

2) (陕西师范大学计算机科学学院, 西安 710062)

3) (西北工业大学自动化学院, 西安 710072)

4) (西北工业大学电子信息学院, 西安 710072)

(2014 年 11 月 12 日收到; 2015 年 6 月 2 日收到修改稿)

对给定的英语音素、单词和语句进行了采集并完成预处理. 分别应用互信息法和 Cao 氏法确定了实际采集的语音信号序列的延迟时间和嵌入维数, 以完成语音序列的相空间重构. 通过计算实际采集的语音信号序列的最大 Lyapunov 指数, 完成了语音信号的混沌特性识别, 判定其具有混沌特性. 引入 Volterra 级数, 提出了一种具有显式结构的语音信号非线性预测模型. 为克服最小均方误差算法在 Volterra 模型系数更新时固有的缺点, 在最小二乘法基础上, 应用基于后验误差假设的可变收敛因子技术, 构建了一种基于 Davidon-Fletcher-Powell 算法的二阶 Volterra 模型 (DFPSOVF), 并将其应用于具有混沌特性的语音信号序列预测. 仿真结果表明: DFPSOVF 非线性预测模型对于单帧和多帧语音信号均具有更好的预测精度, 优于线性预测模型, 并且能够很好地反映语音序列变化的趋势和规律, 完全可以满足语音预测的要求; 可以根据语音信号序列的嵌入维数选取预测模型的记忆长度. 所提出模型可以为语音信号重构和压缩编码开辟一条新途径, 以改善语音信号处理方法的复杂度和处理效果.

关键词: 语音信号, 混沌, Volterra 预测模型, Davidon-Fletcher-Powell 算法

PACS: 05.45.Tp **DOI:** 10.7498/aps.64.200507

1 引 言

语音信号处理是信号处理领域重要的研究课题, 是数字通信、语音识别、语音存储与保密通信的基础. 随着对语音信号的深入研究以及非线性理论的发展, 已经表明语音信号的产生含有大量非线性过程. 而且严格的声学及空气动力学理论已经证明: 语音信号不是一个确定性过程, 也不是纯随机过程, 而是一个复杂的非线性过程 [1]. 线性预测模型已经不能完全满足现代语音信号处理的需求. 非线性理论与技术的发展为人们研究语音信号提供了新的思路, 它使人们的研究从线性领域转向非线性领域, 并开始尝试使用非线性方法来分析和处理语音信号, 非线性研究已经成为语音信号处理领域的热点 [2,3]. 近年来, 伴随着小波理论 [4]、人工神经网络 [5]、混沌理论 [6−8] 等非线性技术的不断发展与完善, 其在语音信号处理研究领域中的应用也取得了较为丰富的研究成果. 人工神经网络 [5] 是语音信号处理领域最常用的非线性方法, 但神经网络模型存在不能获得显式模型结构的缺点, 并且针对不同的样本需要重新构建神经网络模型, 限制了其应用. 文献 [8] 证明, 重构后的语音信号更利于对其特征进行深入和准确的分析. 因此, 引入混沌理论及其时间序列分析理论等新的方法研究语音信号中的混沌特征并构建非线性语音信号处理模型具有非常重要的理论和应用价值. 综上所述, 目前关于语音信号非线性建模的大量研究虽然能够得到比

* 国家自然科学基金 (批准号: 11502133, 11172342, 11372167, 61202153)、陕西省重点科技创新团队项目 (批准号: 2014KTC-18)、西安市科技计划 (批准号: CXY1437(1)) 和榆林市科技计划 (批准号: 2014cxy-09, sf13-43, 2012cxy3-6) 资助的课题.

† 通信作者. E-mail: wythe@snnu.edu.cn

传统线性模型更为优秀的特性却未能像传统线性模型一样给出具有显式结构的语音信号处理模型.

本文通过对线性预测 (linear prediction, LP) 方法的分析与借鉴, 基于语音信号采样点之间的相关性, 引入Volterra级数对语音信号建立非线性预测模型. 由于Volterra模型同时考虑了线性因素和非线性因素的影响, 因此其在非线性时间序列预测、回声消除及系统辨识中得到了广泛使用[9−15]. 文献[9—11]研究了混沌时间序列的多种非线性Volterra预测模型. 文献[12—15]研究了LMS (least mean square), NLMS (normalized least mean square) 及RLS (recursive least square) 等多种算法在Volterra 模型系统辨识中的应用. 尽管LMS算法已经广泛应用在Volterra模型系数更新中, 但其收敛速度对于输入信号频谱的依赖是算法的固有缺点[12−15]. 同时, 在将LMS算法应用于Volterra模型时也存在如何合理选择收敛因子的问题.

基于以上分析, 本文对实际采集的语音信号序列, 在分别应用互信息法和Cao氏法确定其延迟时间和嵌入维数的基础上, 完成了对语音信号序列的相空间重构. 然后, 对语音信号序列的最大Lyapunov指数混沌特征量进行计算, 以判定语音信号序列的混沌特性. 在对Volterra模型应用LMS 算法进行系数更新的基础上, 构建了一种基于Davidon-Fletcher-Powell算法的具有可变收敛因子技术的二阶Volterra自适应预测模型 (DFP-based second of volterra filter, DFPSOVF)[16]. 并将该模型应用于单帧和多帧语音信号预测, 以获得对语音信号预测的新方法、新思路, 为语音信号重构和压缩编码开辟一条新途径.

2 语音信号数据采集与预处理

2.1 语音信号数据采集

本文研究以实验为基础, 需要通过采集不同录音者对给定英语音素、单词和语句的发音建立语料库, 录音的采样率为8000/s. 录制过程中严格控制环境因素, 使采集到的语音信号具有更高的信噪比. 音素是从音质角度及自然属性划分出来的最基本的、最小的并且是不可分解的组成单位, 也是分析语音信号的基础. 所以本文研究主要是基于语料库中的音素对语音信号进行分析, 并通过单词、语句验证其一般性. 实验中的语音信号数据分别为音素样本 [ai], [ɔ:], [a:] 和 [əu], 语音信号序列波形图如图1所示.

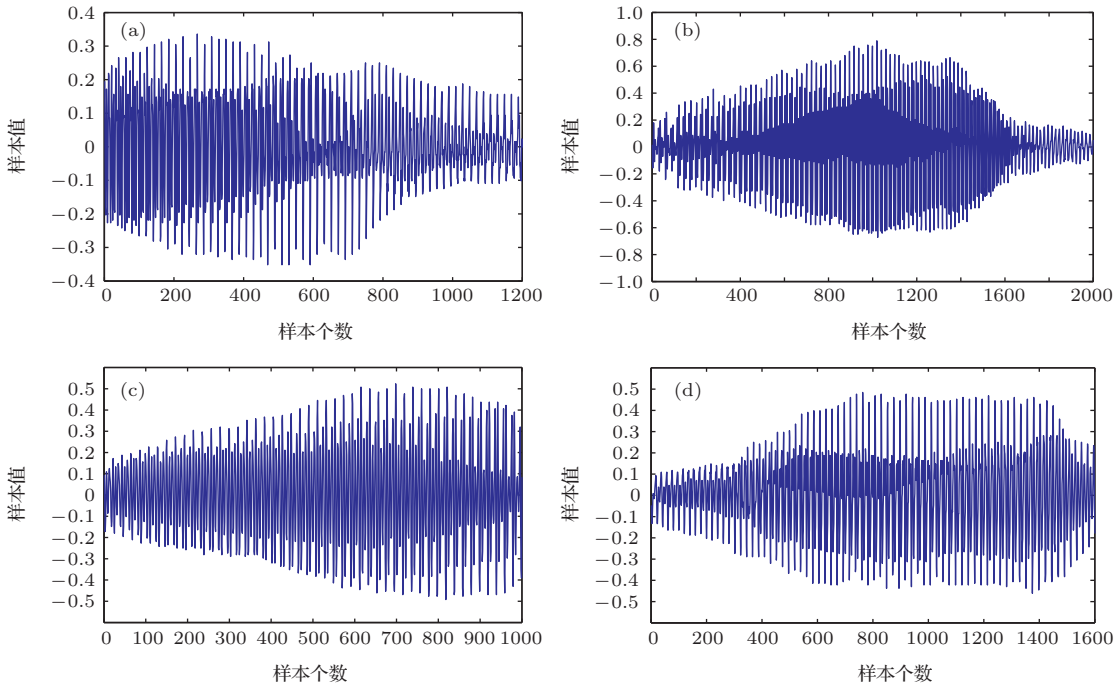

图1 (网刊彩色) 语音信号音素样本波形图 (a) 样本1-音素 [ai]; (b) 样本2-音素 [ɔ:]; (c) 样本3-音素 [a:]; (d) 样本4-音素 [əu]

Fig. 1. (color online) Waveform charts of speech signal phoneme samples: (a) Sample 1-phoneme [ai]; (b) sample 2-phoneme [ɔ:]; (c) sample 3-phoneme [a:]; (d) sample 4-phoneme [əu].

2.2 语音信号数据预处理

在对语音信号样本进行建模之前, 需要进行预处理, 包括预加重、加窗和分帧. 本文预加重在A/D 变换之后进行, 使用具有6 dB/倍频程的提升高频特性的一阶数字滤波器实现, 如(1)式:

$$H(z) = 1 - \mu z^{-1}, \tag{1}$$

其中, μ为接近于1的值, 本文取0.9378.

语音信号的短时平稳性表明, 虽然语音信号具有时变特性, 但是在一个短时间范围内(10—30 ms) 其特性变化却有限. 语音信号分析一般建立在短时的基础上, 即对语音信号分帧处理. 本文中的语音信号非线性模型建立在短时分析的基础上, 需要对数据进行加窗和分帧. 常用的窗函数有矩形窗、汉明窗和汉宁窗. 本文选择汉明窗, 如(2)式:

$$h(n) = \begin{cases} 0.54 - 0.46\cos[2\pi n/N], \\ \quad 0 \leqslant n \leqslant N-1, \\ 0, \qquad \text{其他}, \end{cases} \tag{2}$$

其中, N为窗中的语音数据量. 语音信号分帧预处理时, 本文取帧长为30 ms, 即在8000/s采样率的条件下, 每帧数据含有240个样点值.

3 语音信号序列相空间重构

设预处理得到的语音时间序列为

$$\{x(n)\}_{n=1}^{N}, \quad x(n) = x(t_0 + n\Delta t), \tag{3}$$

其中, t_0为初始时间, Δt为采样时间间隔. 选择适当的嵌入维数m和延迟时间τ, 重构一个m维的状态向量[16]:

$$\begin{aligned} \boldsymbol{x}(n) &= \left[x(n), x(n-\tau), \cdots, x(n-(m-1)\tau)\right]^{\mathrm{T}}, \\ n &= N_0, N_0+1, \cdots, N, \end{aligned} \tag{4}$$

其中, $N_0 = (m-1)\tau + 1$, $\boldsymbol{x}(n)$为相点, m维序列$\{\boldsymbol{x}(n)\}_{n=N_0}^{N}$构成一个相型, 它表示语音系统在某一瞬间的状态, 状态空间$\boldsymbol{x}(n) \to \boldsymbol{x}(n+1)$的演化反映了语音系统的演化, 即可以由历史数据进行预测.

3.1 互信息法选取延迟时间

设观测序列为$\{\boldsymbol{x}(i)\}_{i=1}^{N}$, 则$i$和$i+\tau$时刻观测量之间的互信息函数[17]可表示为

$$\begin{aligned} I(\tau) = &\sum_{i=1}^{N} P[x(i), x(i+\tau)] \\ &\times \log_2\left[\frac{P[x(i), x(i+\tau)]}{P[x(i)]P[x(i+\tau)]}\right], \end{aligned} \tag{5}$$

其中, $P[x(i)]$为点$x(i)$的概率密度, $P[x(i), x(i+\tau)]$为点$x(i)$和$x(i+\tau)$的联合概率. 本文选取$I(\tau)$第一个局部最小时的τ为延迟时间[16,17], 此时冗余最小且独立性最大.

对图1所示语音信号序列, 利用互信息法计算的延迟时间τ的结果分别见图2和表1. 图2给出了四个语音样本平均互信息与延迟时间之间的关系曲线, 表1给出了样本1-音素 [ai] 和样本2-音素 [ɔ:] 的延迟时间τ从1增加到30的平均互信息. 由于τ均为2时取得第一个局部最小值, 因此, 四个样本最佳延迟时间均取2.

表1 平均互信息与延迟时间对应关系表

Table 1. The relationship of average mutual information and delay time.

样本1-音素 [ai]					
延迟时间	平均互信息	延迟时间	平均互信息	延迟时间	平均互信息
1	1.5639	11	1.5061	21	1.6722
2	1.5217	12	1.5256	22	1.5207
3	1.5883	13	1.4780	23	1.5668
4	1.5346	14	1.5205	24	1.5538
5	1.5678	15	1.5634	25	1.5553
6	1.5223	16	1.5363	26	1.5419
7	1.4647	17	1.5323	27	1.5068
8	1.5188	18	1.5613	28	1.4822
9	1.4898	19	1.5548	29	1.5357
10	1.5099	20	1.9764	30	1.5242
样本2-音素 [ɔ:]					
延迟时间	平均互信息	延迟时间	平均互信息	延迟时间	平均互信息
1	1.6937	11	1.5760	21	1.6171
2	1.5683	12	1.5928	22	1.6168
3	1.5933	13	1.5687	23	1.6230
4	1.7251	14	1.5595	24	1.6029
5	1.6000	15	1.5698	25	1.6162
6	1.5674	16	1.5961	26	1.5985
7	1.5748	17	1.6504	27	1.6038
8	1.6185	18	1.6394	28	1.5908
9	1.5610	19	1.6164	29	1.5921
10	1.5566	20	1.6054	30	1.5760

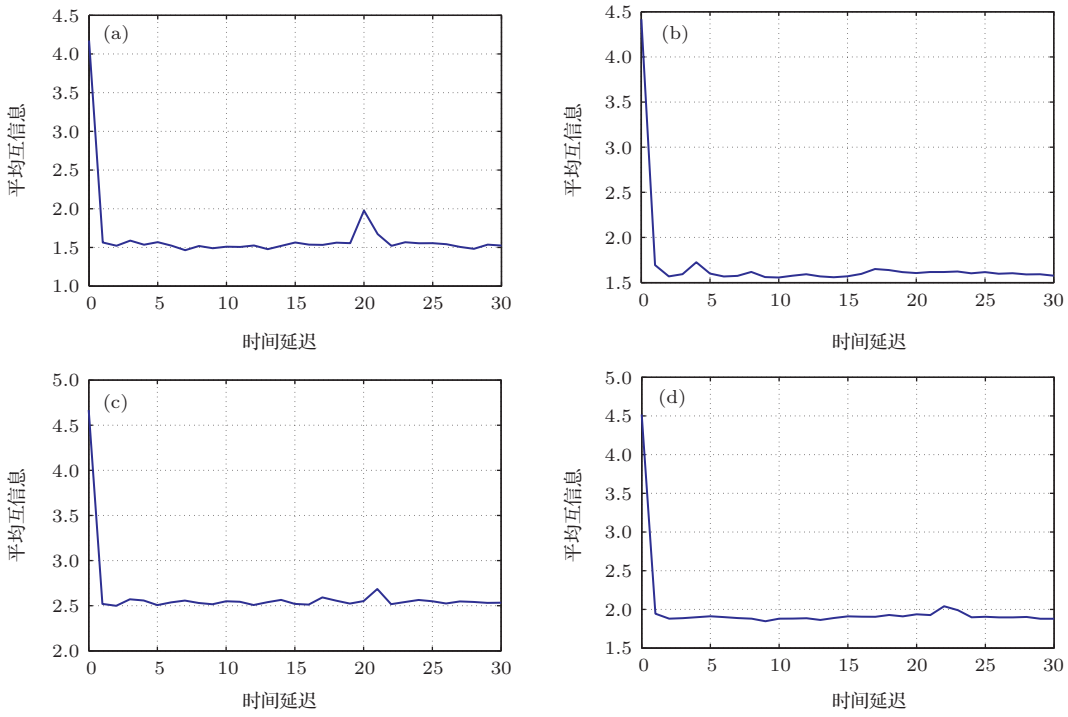

图 2 平均互信息与延迟时间的关系 (a) 样本 1- 音素 [ai]; (b) 样本 2-音素 [ɔ:]; (c) 样本 3-音素 [a:]; (d) 样本 4-音素 [əu]

Fig. 2. The relationship of average mutual information and delay time: (a) Sample 1-phoneme [ai]; (b) sample 2-phoneme [ɔ:]; (c) sample 3-phoneme [a:]; (d) sample 4-phoneme [əu].

3.2 Cao 氏法选取嵌入维数

Cao 提出的 Cao 氏法 [18] 主要用来求取时间序列的嵌入维数, 定义

$$E(m) = \frac{1}{N - m\tau} \sum_{i=1}^{N-m\tau} a_2(i, m), \tag{6}$$

$$E_1(m) = \frac{E(m+1)}{E(m)}, \tag{7}$$

其中,

$$a_2(i, d) = \frac{|\boldsymbol{X}_{d+1}(i) - \boldsymbol{X}_{d+1}^{NN}(i)|}{\|\boldsymbol{X}_d(i) - \boldsymbol{X}_d^{NN}(i)\|},$$

$\boldsymbol{X}_d(i)$ 和 $\boldsymbol{X}_d^{NN}(i)$ 为 d 维相空间的第 i 个向量和它的最临近点, $\boldsymbol{X}_{d+1}(i)$ 和 $\boldsymbol{X}_{d+1}^{NN}(i)$ 是 $d+1$ 维相空间的第 i 个向量和它的最临近点.

对于一个确定的时间序列, 利用 Cao 氏法计算嵌入维数时, 当 m 大于某个值 m_0 时, $E_1(m)$ 将不再发生变化, 则说明该时间序列的嵌入维数是存在的. 若 $E_1(m)$ 随着 m 的增大不断变化, 则说明这个时间序列是随机信号. 但在实际计算中, $E_1(m)$ 具体是已经稳定不变还是在缓慢变化不容易判断, 因此需要补充一个判断准则:

表 2 嵌入维数-E_1 & E_2 对应关系表

Table 2. The relationship of embedding dimension-E_1 & E_2.

样本 1-音素 [ai]						样本 2-音素 [ɔ:]					
嵌入维数	E_1	E_2	嵌入维数	E_1	E_2	嵌入维数	E_1	E_2	嵌入维数	E_1	E_2
1	0.4947	0.5296	11	0.9959	0.9895	1	0.4285	0.5462	11	0.9935	1.0012
2	0.3402	0.4723	12	0.9978	1.0076	2	0.2984	0.5836	12	0.9948	1.0076
3	0.4819	0.5853	13	0.9998	1.0058	3	0.4402	0.4899	13	0.9926	1.0025
4	0.7352	0.7638	14	1.0006	1.0300	4	0.7978	0.9027	14	0.9914	1.0112
5	0.8907	0.9065	15	0.9977	1.0021	5	0.9066	0.9847	15	0.9932	1.0153
6	0.9269	0.9155	16	0.9975	1.0201	6	0.9573	1.0447	16	0.9962	1.0125
7	0.9484	0.9356	17	0.9988	1.0008	7	0.9782	1.0400	17	0.9935	1.0316
8	0.9690	0.9492	18	0.9985	0.9975	8	0.9881	1.0375	18	1.0006	1.0384
9	0.9764	0.9840	19	0.9999	1.0109	9	0.9827	0.9970	19	0.9945	1.0296
10	1.0001	1.0195				10	0.9897	1.0161			

$$E^*(m)=\frac{1}{N-m\tau}\sum_{i=1}^{N-m\tau}\left|x(i+m\tau)-x^{NN}(i+m\tau)\right|, \tag{8}$$

$$E_2(m)=E^*(m+1)/E^*(m). \tag{9}$$

对于随机时间序列, 由于数据之间没有相关性, 因此不能预测, $E_2(m)$ 将始终为1; 对于确定性的时间序列, 由于数据之间存在相关性, 且其和嵌入维数的大小有关, 因此存在一些 m 值使得 $E_2(m)$ 不等于1.

利用Cao氏法求图1所示语音样本嵌入维数 m 的计算结果分别如图3和表2所示. 图3给出了语音序列嵌入维数与 $E_1(m)$ 和 $E_2(m)$ 的关系曲线, 表2给出了样本1-音素 [ai] 和样本2-音素 [ɔ:] 的嵌入维数 m 从1增加到19的对应值. 得到语音序列样本1, 2, 3, 4的嵌入维数在 m 分别取10, 11, 10, 13时, $E_1(m)$ 已经不再发生变化, 而 $E_2(m)$ 不等于1, 由此可得出四个样本的最小嵌入维数分别为10, 11, 10, 13.

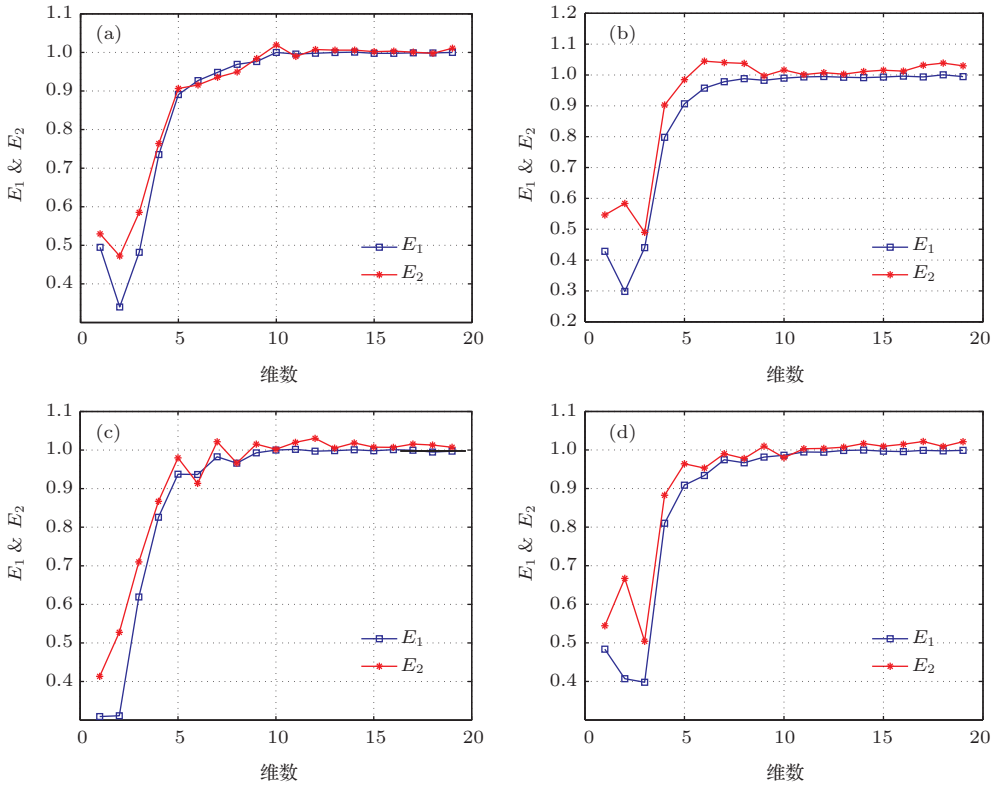

图3 (网刊彩色) 嵌入维数-E_1 & E_2 关系 (a) 样本1-音素 [ai]; (b) 样本2-音素 [ɔ:]; (c) 样本3-音素 [a:]; (d) 样本4-音素 [əu]

Fig. 3. (color online) The relationship of embedding dimension-E_1 & E_2: (a) Sample 1-phoneme [ai]; (b) sample 2-phoneme [ɔ:]; (c) sample 3-phoneme [a:]; (d) sample 4-phoneme [əu].

3.3 相空间重构

将3.1和3.2计算得到的延迟时间和嵌入维数代入(4)式, 即实现了对如图1所示的语音信号的相空间重构, 即

$$\begin{matrix} x(N_0) & x(N_0+1) & \cdots & x(N) \\ x(N_0-\tau) & x(N_0+1-\tau) & \cdots & x(N-\tau) \\ x(N_0-2\tau) & x(N_0+1-2\tau) & \cdots & x(N-2\tau) \\ \vdots & \vdots & \vdots & \vdots \\ x(N_0-(m-1)\tau) & x(N_0+1-(m-1)\tau) & \cdots & x(N-(m-1)\tau) \\ \boldsymbol{x}(N_0) & \boldsymbol{x}(N_0+1) & \cdots & \boldsymbol{x}(N) \end{matrix}$$

其中, $x(\cdot)$ 为序列中某一时刻的采样值, $\boldsymbol{x}(n)$ 为相空间重构得到的相点, 对于本文研究的语音序列, 延迟时间 τ 取 2, 嵌入维数 m 分别取 10, 11, 10, 13.

4 语音信号序列混沌特性识别

本文从语音信号序列是否对于初始条件敏感来判断其是否具有混沌特性. 下面采用小数据量法计算最大 Lyapunov 指数, 对语音信号序列的混沌特性进行识别.

采用小数据量法 [19] 计算语音信号序列的最大 Lyapunov 指数, 演化步数 i 分别取 100, 200, 400, 600, 800 和 1000, 计算结果如表 3 所列. 图 4 和图 5 分别给出样本 3-音素 [a:] 和样本 4-音素 [əu] 当 i 为 800 的计算结果. 表 3 得到所研究的语音信号序列的最大 Lyapunov 指数均大于零, 表明语音信号序列具有混沌特性.

表 3 样本 1—4 音素序列的最大 Lyapunov 指数

Table 3. The largest Lyapunov exponent of samples 1 to 4-phoneme series.

样本 1-音素 [ai] 序列		样本 2-音素 [ɔ:] 序列		样本 3-音素 [a:] 序列		样本 4-音素 [əu] 序列	
演化步数 i	λ	演化步数 i	λ	演化步数 i	λ	演化步数 i	λ
100	0.2831	100	3.1065	100	1.1997	100	1.0667
200	0.2932	200	3.3271	200	1.2548	200	0.9541
400	0.3558	400	3.1172	400	1.4182	400	1.0460
600	0.3215	600	3.1283	600	1.4286	600	0.9681
800	0.3405	800	3.3096	800	1.3046	800	1.0023
1000	0.3405	1000	3.2045	1000	1.3255	1000	0.8958

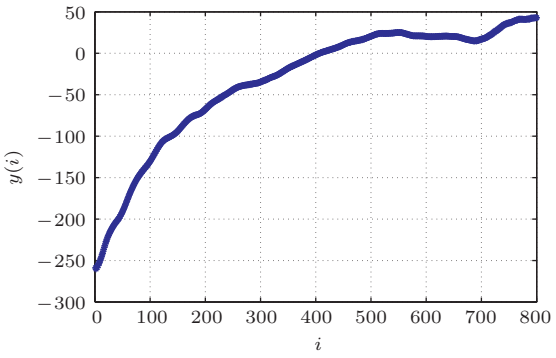

图 4 样本 3-音素 [a:] 序列的最大 Lyapunov 指数

Fig. 4. The largest Lyapunov exponent of sample 3-phoneme [a:] series.

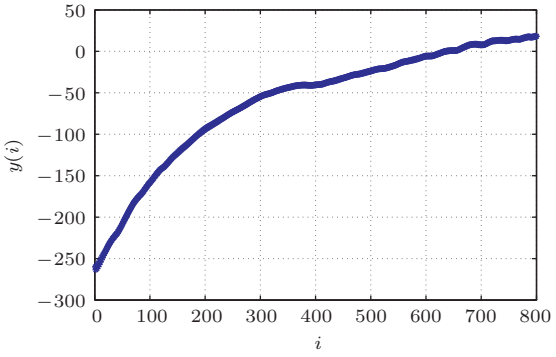

图 5 样本 4-音素 [əu] 序列的最大 Lyapunov 指数

Fig. 5. The largest Lyapunov exponent of sample 4-phoneme [əu] series.

5 DFPSOVF 模型

二阶截断 Volterra 模型 [16] 定义如下:

$$\begin{aligned}\hat{y}(n) = h_0 &+ \sum_{i=0}^{m-1} h_1(i; n-1)x(n-i) \\ &+ \sum_{i=0}^{m-1}\sum_{j=0}^{m-1} h_2(i,j; n-1)x(n-i) \\ &\times x(n-j), \end{aligned} \tag{10}$$

其中, $x(n)$ 和 $\hat{y}(n)$ 分别表示 n 时刻的输入信号和自适应输出, h_0 表示常数项, m 为模型记忆长度, $h_1(i; n-1)$ 和 $h_2(i,j; n-1)$ 分别为 n 时刻需要更新的线性项系数和平方项系数. 记

$$\begin{aligned}\boldsymbol{H}(n-1) = [&h_1(0; n-1), h_1(1; n-1), \cdots, \\ &h_1(m-1; n-1), h_2(0,0; n-1), \\ &h_2(0,1; n-1), \cdots, \\ &h_2(m-1, m-1; n-1)]^{\mathrm{T}}, \end{aligned} \tag{11}$$

$$\begin{aligned}\boldsymbol{X}(n) = [&x(n), x(n-1), \cdots, x(n-(m-1)), \\ &x^2(n), x(n)x(n-1), \cdots, \\ &x^2(n-(m-1))]^{\mathrm{T}}, \end{aligned} \tag{12}$$

其中, $(\cdot)^{\mathrm{T}}$ 表示向量转置. 因此, (10) 式可表示为如下向量形式:

$$\hat{y}(n) = \boldsymbol{H}^{\mathrm{T}}(n-1)\boldsymbol{X}(n). \tag{13}$$

令 $e(n)$ 为 n 时刻的先验误差信号, $y(n)$ 为 n 时刻的期望输出, 则 n 时刻的平方误差为

$$e^2(n) = (y(n) - \hat{y}(n))^2. \tag{14}$$

对于 (13) 式, 应用LMS算法的系数向量 $\boldsymbol{H}(n-1)$ 的迭代公式为

$$\boldsymbol{H}(n) = \boldsymbol{H}(n-1) + 2\mu e(n)\boldsymbol{X}(n), \tag{15}$$

其中, μ 为收敛因子, 其控制LMS算法的收敛速度与稳定性. 为消除输入信号频谱对LMS-Volterra模型系数更新收敛速度影响, 考虑使用拟Newton法[20]. 将拟Newton法应用至 (15) 式, 得到

$$\boldsymbol{H}(n) = \boldsymbol{H}(n-1) + 2e(n)\hat{\boldsymbol{R}}^{-1}(n-1)\boldsymbol{X}(n), \tag{16}$$

其中, $\hat{\boldsymbol{R}}^{-1}(n-1)$ 表示输入向量 $\boldsymbol{X}(n-1)$ 的自相关逆矩阵, 即

$$\hat{\boldsymbol{R}}^{-1}(n-1) = \left(\boldsymbol{X}(n-1)\boldsymbol{X}^{\mathrm{T}}(n-1)\right)^{-1}. \tag{17}$$

同时, 在 (16) 式中引入可变收敛因子. 将 (16) 式改写为

$$\boldsymbol{H}(n) = \boldsymbol{H}(n-1) + 2\mu(n)e(n)\hat{\boldsymbol{R}}^{-1}(n-1)\boldsymbol{X}(n). \tag{18}$$

因此, 得到DFPSOVF的系数更新过程如下.

初始化 分别给定记忆长度 m、序列长度 N、给定输入信号 $x(n)$ 和期望值 $y(n)$, 令 $\boldsymbol{H}(0) = \boldsymbol{H}(1) = \cdots = \boldsymbol{H}(m-1) = \boldsymbol{0}$, $\hat{\boldsymbol{R}}^{-1}(0) = \hat{\boldsymbol{R}}^{-1}(1) = \cdots = \hat{\boldsymbol{R}}^{-1}(m-1) = \boldsymbol{I}$, 其中 $\boldsymbol{0}$ 和 $\boldsymbol{I}$ 分别为 $(m^2+3m)/2$ 维零向量和 $\left(\dfrac{m^2+3m}{2}\right) \times \left(\dfrac{m^2+3m}{2}\right)$ 维单位矩阵.

计算 令 $n = m$,

步骤 1 构造向量

$$\begin{aligned}\boldsymbol{X}(n) = [&x(n), x(n-1), \cdots, x(n-(m-1)),\\ &x^2(n), x(n)x(n-1), \cdots,\\ &x^2(n-(m-1))]^{\mathrm{T}};\end{aligned}$$

步骤 2 $\hat{y}(n) = \boldsymbol{H}^{\mathrm{T}}(n-1)\boldsymbol{X}(n)$;

步骤 3 $e(n) = y(n) - \hat{y}(n)$;

步骤 4 $\mu(n) = \left[2\boldsymbol{X}^{\mathrm{T}}(n)\hat{\boldsymbol{R}}^{-1}(n-1)\boldsymbol{X}(n)\right]^{-1}$;

步骤 5

$$\boldsymbol{H}(n) = \boldsymbol{H}(n-1) + 2\mu(n)e(n)\hat{\boldsymbol{R}}^{-1}(n-1)\boldsymbol{X}(n);$$

步骤 6

$$\begin{aligned}\hat{\boldsymbol{R}}^{-1}(n) = &\hat{\boldsymbol{R}}^{-1}(n-1)\\ &+ \frac{\hat{\boldsymbol{R}}^{-1}(n-1)\boldsymbol{X}(n)\boldsymbol{X}^{\mathrm{T}}(n)\hat{\boldsymbol{R}}^{-1}(n-1)}{\boldsymbol{X}^{\mathrm{T}}(n)\hat{\boldsymbol{R}}^{-1}(n-1)\boldsymbol{X}(n)}\\ &\times (\mu(n) - 1);\end{aligned}$$

步骤 7 $n = m+1$. 如果 $n \leqslant N$, 转到步骤1, 否则结束.

6 模型评价

实验样本取自采集的语料库, 在对样本进行预处理时, 取帧长为30 ms, 即每帧语音数据含有240个样点值. 本实验选取四个语音音素样本 (标记为样本1, 2, 3, 4)、一个语音单词样本 (标记为样本5) 和两个语音语句样本 (标记为样本6和7). 帧数较少的音素样本用于确定DFPSOVF模型记忆长度, 帧数较多的单词、语句样本用于测试DFPSOVF模型的泛化能力.

实验首先比较不同记忆长度时得到的样本均方根误差值, 确定非线性模型记忆长度. 再利用LP模型和DFPSOVF模型对单帧语音数据进行预测, 通过预测的数据波形以及均方根误差的比较, 对本文提出的非线性预测模型的有效性进行验证. 为了更好地验证模型有效性及泛化能力, 利用LP模型和DFPSOVF模型分别对多帧数据进行了预测效果的对比.

6.1 语音信号序列的DFPSOVF预测

使用如图1所示的语音音素样本作为预测的原始数据, 选取每个样本的其中一帧数据 (240个采样点) 作为DFPSOVF预测模型预测的样本. 将语音信号序列数据分为两部分: 第一部分 (前80个采样点) 为训练样本; 第二部分 (后160个采样点) 作为测试样本. 为评价预测结果的优劣, 以预测均方根误差 (root mean square error, RMSE) 作为评价标准.

应用DFPSOVF模型对上述单帧样本进行预测. 为研究记忆长度对DFPSOVF模型的影响, 且考虑到四个样本序列的嵌入维数分别为10, 11, 10, 13, 分别取记忆长度3, 5, 7, 9, 11, 13, 15, 17和19等涵盖嵌入维数且均匀分布的9个值的预测结果进

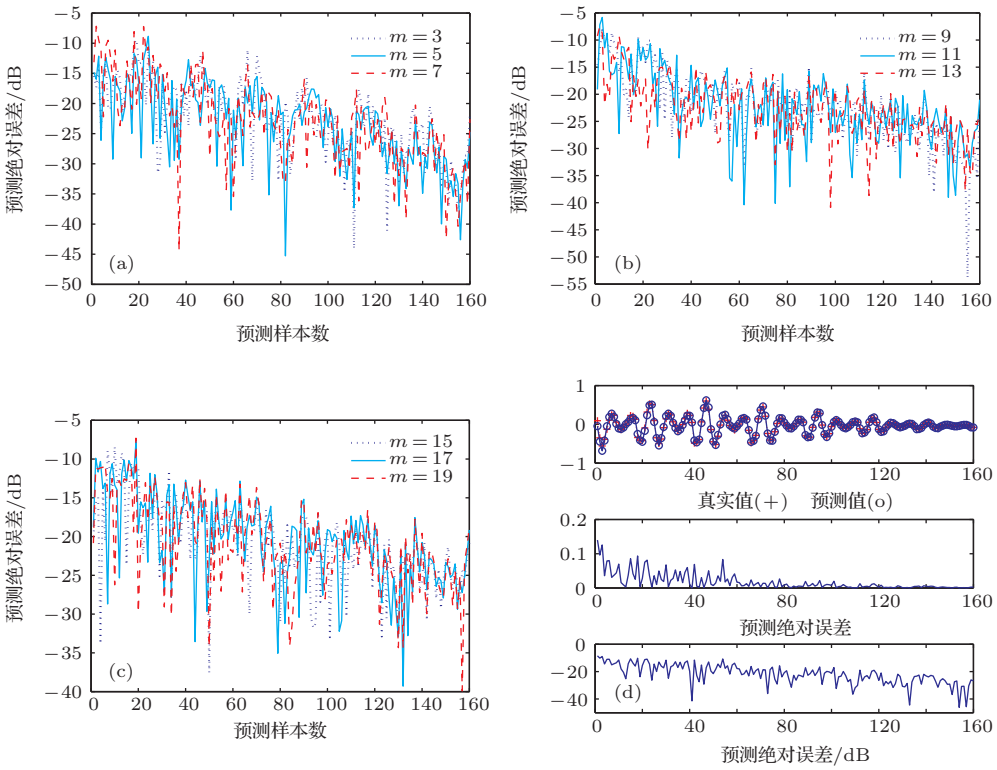

图 6 (网刊彩色) m 分别取不同值时, 样本 2-音素 [ɔ:] 序列的 DFPSOVF 预测误差 (a) m 分别取 3, 5, 7; (b) m 分别取 9, 11, 13; (c) m 分别取 15, 17, 19; (d) m 取 11

Fig. 6. (color online) DFPSOVF prediction error of sample 2-phoneme [ɔ:] series when m is taken different values: (a) m is taken 3, 5, 7, respectively; (b) m is taken 9, 11, 13, respectively; (c) m is taken 15, 17, 19, respectively; (d) m is 11.

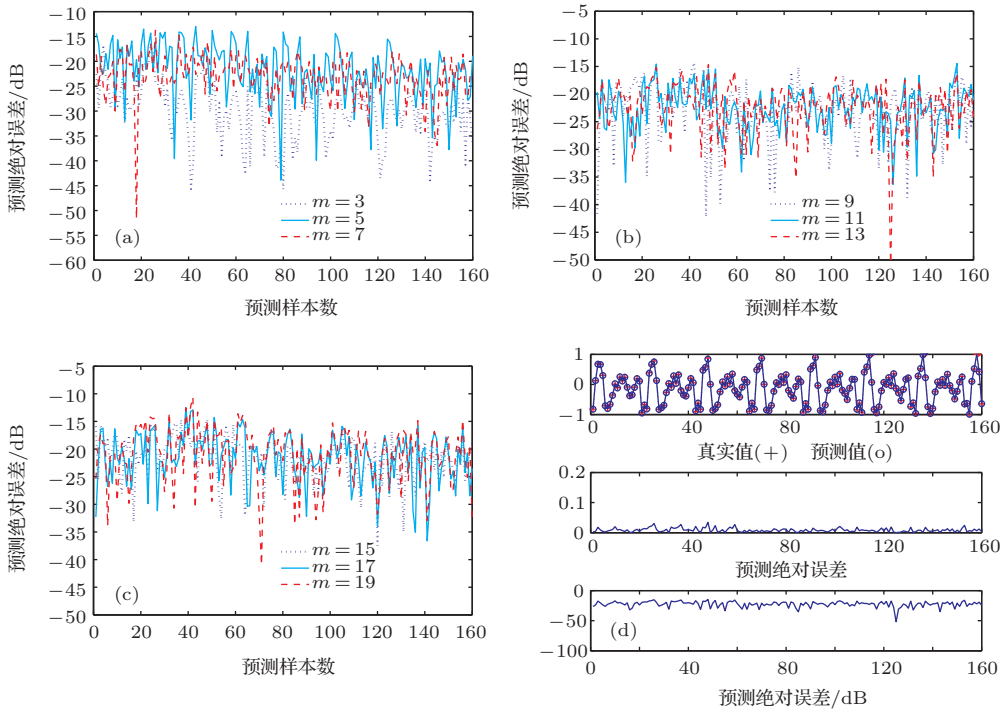

图 7 (网刊彩色) m 分别取不同值时样本 4-音素 [əu] 序列的 DFPSOVF 预测误差 (a) m 分别取 3, 5, 7; (b) m 分别取 9, 11, 13; (c) m 分别取 15, 17, 19; (d) m 取 13

Fig. 7. (color online) DFPSOVF prediction error of sample 4-phoneme [əu] series when m is taken different values: (a) m is taken 3, 5, 7, respectively; (b) m is taken 9, 11, 13, respectively; (c) m is taken 15, 17, 19, respectively; (d) m is 13.

行比较. 实际语音样本与预测结果之间的绝对误差曲线如图6 和图7 所示, 预测的均方根误差见表4

表4 记忆长度不同时, 样本2-音素 [ɔ:] 序列和样本4-音素 [əu] 序列的预测RMSE

Table 4. Prediction RMSE of sample 2-phoneme [ɔ:] series and sample 4-phoneme [əu] series when different memory length is taken.

样本2-音素 [ɔ:] 序列		样本4-音素 [əu] 序列	
记忆长度	RMSE	记忆长度	RMSE
3	2.251×10^{-2}	3	1.372×10^{-2}
5	2.082×10^{-2}	5	1.526×10^{-2}
7	3.509×10^{-2}	7	1.900×10^{-2}
9	3.123×10^{-2}	9	2.237×10^{-2}
11	3.539×10^{-2}	11	2.469×10^{-2}
13	2.733×10^{-2}	13	2.594×10^{-2}
15	2.870×10^{-2}	15	3.050×10^{-2}
17	3.098×10^{-2}	17	3.022×10^{-2}
19	2.967×10^{-2}	19	2.756×10^{-2}

(这里仅给出样本2-音素 [ɔ:] 和样本4-音素 [əu] 的实验结果). 仿真结果可见, 记忆长度对预测精度影响有限. 因此, 依据上述相空间重构的嵌入维数来选取DFPSOVF模型的记忆长度. 依据各个样本的嵌入维数选择DFPSOVF模型的记忆长度为10, 11, 10, 13, 对选定样本帧进行预测, 真实值与预测值的比较和预测的绝对误差分别如图6 (d) 和图7 (d) 所示. 上述仿真表明, DFPSOVF模型能够很好地反映语音序列变化的趋势和规律, 预测精度较高, 完全可以满足语音预测的要求.

6.2 单帧语音信号预测

分别利用LP线性模型和DFPSOVF非线性模型对单帧语音数据进行预测. 通过对比预测数据波形以及均方根误差, 对本文提出的DFPSOVF非线

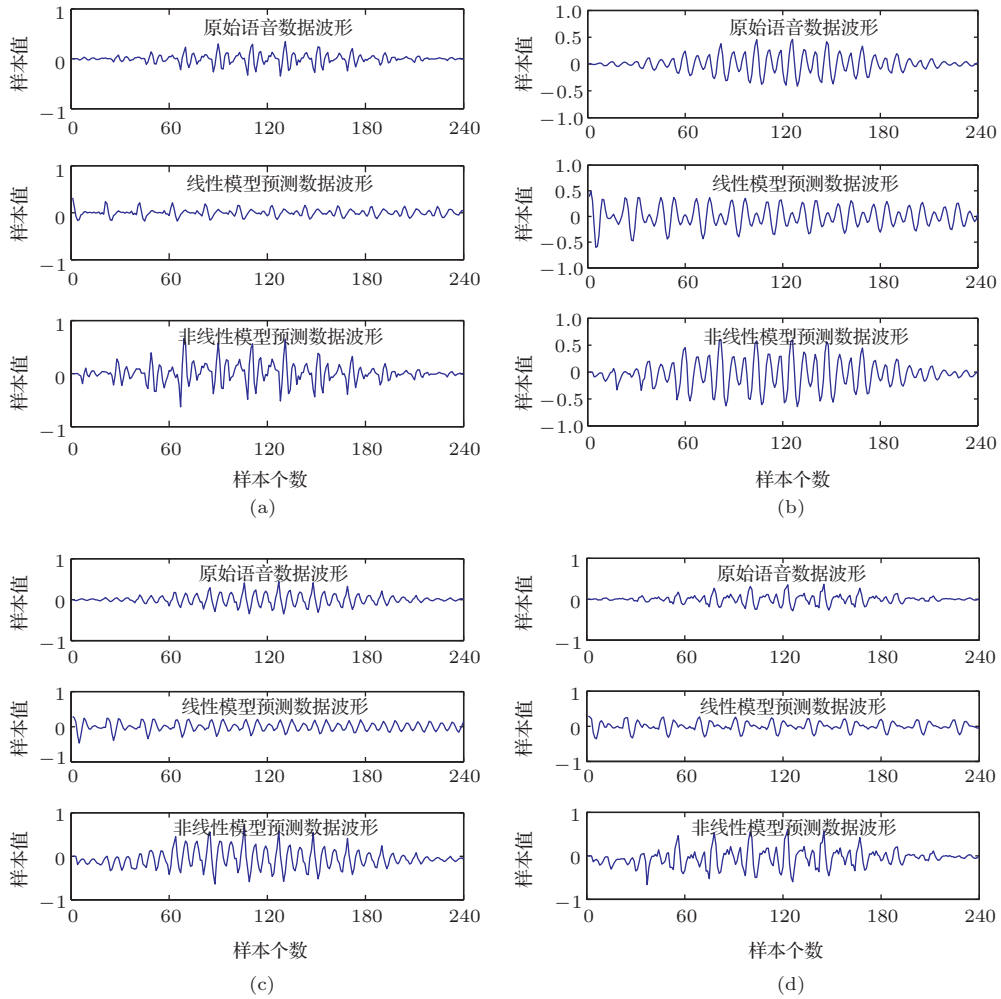

图8 LP线性模型和DFPSOVF非线性模型对单帧语音样本1—4预测值对比 (a) 样本1-音素 [ai]; (b) 样本2-音素 [ɔ:]; (c) 样本3-音素 [a:]; (d) 样本4-音素 [əu]

Fig. 8. Comparisons of prediction value of single-frame speech samples from 1 to 4 when using linear model and DFPSOVF nonlinear model, respectively: (a) Sample 1-phoneme [ai]; (b) sample 2-phoneme [ɔ:]; (c) sample 3-phoneme [a:]; (d) sample 4-phoneme [əu].

性预测模型的有效性进行初步验证. 线性预测模型中参数的求解采用经典 Levinson-Durbin 算法, 模型阶数 P 为 12, 预测时采用脉冲激励作为初始输入, 而对非线性模型参数的优化采用 DFP 算法. 预测数据波形对比如图 8 所示.

由以上四个样本的预测结果可以看出, 线性模型的预测效果较差, 而非线性模型的预测值能够较直观地反映原始信号值, 且得到的信号值与原始值更为接近. 此外, 线性预测模型得到的单帧数据 RMSE 均明显大于非线性预测模型, 如表 5 所列, RMSE1 和 RMSE2 分别为线性模型均方根误差和非线性模型均方根误差, 表明 DFPSOVF 非线性模型对于单帧语音信号数据有更好的预测效果.

表 5 LP 线性模型和 DFPSOVF 非线性模型对单帧语音样本 1—4 均方根误差比较

Table 5. Comparisons of RMSE of single-frame speech samples from 1 to 4 when using linear model and DFPSOVF nonlinear model, respectively.

语音样本	RMSE1	RMSE2	语音样本	RMSE1	RMSE2
样本 1	0.129265	0.0337189	样本 3	0.166817	0.0289157
样本 2	0.217397	0.0249918	样本 4	0.171582	0.0222915

6.3 多帧语音信号预测

本实验对提出的非线性预测模型在更多语音信号中的有效性进行验证. 取自语料库的实验数据来自不同的采样者, 分别标记为样本 5 (China)、样本 6 (I'm Chinese)、样本 7 (I love my country), 含有 29, 57 和 98 帧不同的语音数据. 实验利用两种模型进行预测并对比其波形图和均方根误差值. 两个样本的实验结果分别如图 9—图 11 所示.

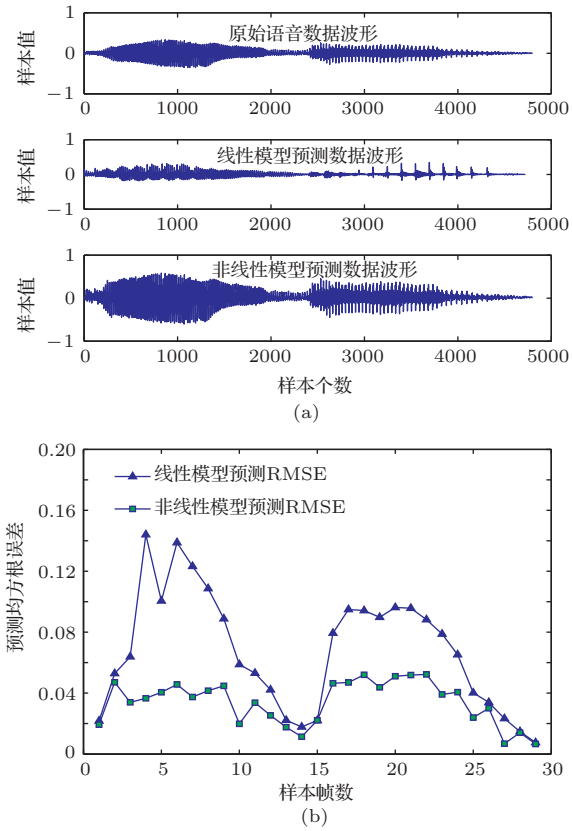

图 9 (网刊彩色) LP 线性模型和 DFPSOVF 非线性模型对多帧语音样本 5(China) 的预测对比 (a) 波形对比; (b) 均方根误差对比

Fig. 9. (color online) Prediction comparisons of multi-frame speech sample 5 (China) when using LP linear model and DFPSOVF nonlinear model, respectively: (a) Waveform comparison; (b) RMSE comparison.

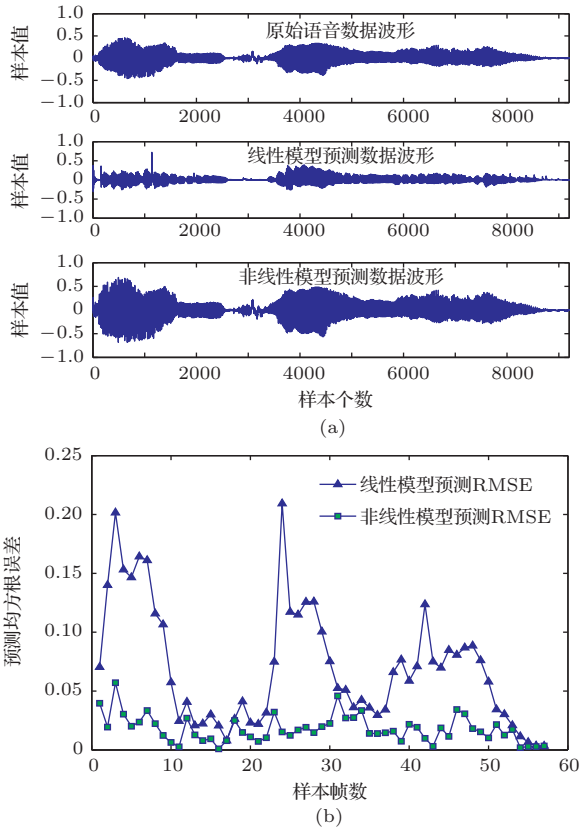

图 10 (网刊彩色) LP 线性模型和 DFPSOVF 非线性模型对多帧语音样本 6 (I'm Chinese) 的预测对比 (a) 波形对比; (b) 均方根误差对比

Fig. 10. (color online) Prediction comparisons of multi-frame speech sample 6 (I'm China) when using LP linear model and DFPSOVF nonlinear model, respectively: (a) Waveform comparison; (b) RMSE comparison.

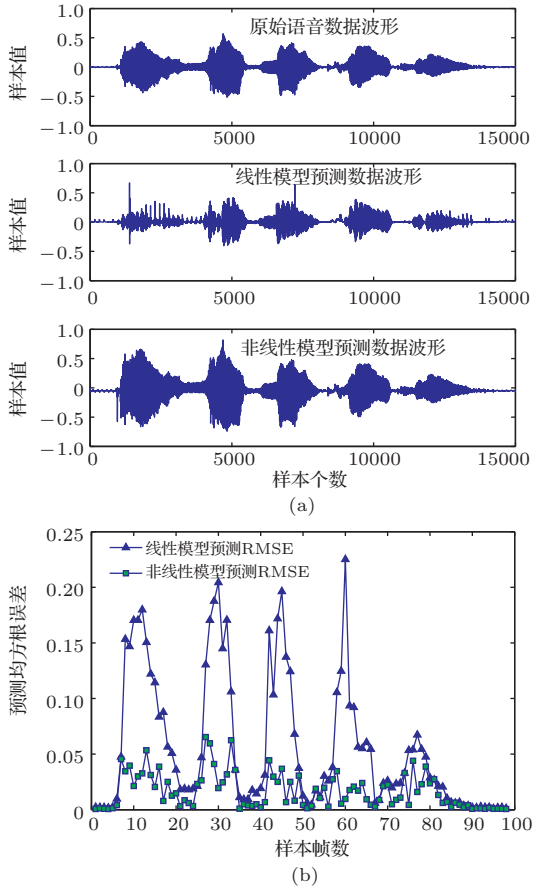

图 11 (网刊彩色) LP 线性模型和 DFPSOVF 非线性模型对多帧语音样本 7 (I love my country) 的预测对比 (a) 波形对比; (b) 均方根误差对比

Fig. 11. (color online) Prediction comparisons of multi-frame speech sample 7 (I love my country) when using LP linear model and DFPSOVF nonlinear model, respectively: (a) Waveform comparison; (b) RMSE comparison.

由图 9、图 10 和图 11 可以看出, 两种模型的预测值都能够较直观地反映原始信号值. 但在中间波动较大的部分, 非线性预测模型得到的信号值与原始值更为接近. 此外, 对于三个样本中的每一帧数据, 采用非线性预测模型都具有更小的均方根误差值, 说明本文提出的 DFPSOVF 非线性预测模型对于多帧语音信号同样具有更好的预测效果. 另外, 通过模型分析可以看出, 本文提出的非线性预测模型既考虑到线性预测部分又兼顾非线性部分, 因而从理论上具有比单纯的线性预测模型更为精确的描述能力.

7 结 论

语音信号是非线性的时间序列, 本文对语音信号序列进行了相空间重构, 并对其混沌特性进行了识别, 判定了语音信号序列的混沌特性. 针对将 LMS 算法应用于 Volterra 模型时存在的对输入信号特性的依赖问题, 研究 DFPSOVF 模型. 最后, 将 DFPSOVF 模型应用于实测单帧和多帧语音信号序列的预测. 对本文所做的工作归纳如下.

1) 分别采用互信息法和 Cao 氏法计算了语音信号序列的延迟时间和嵌入维数, 完成了语音信号序列的相空间重构. 利用小数据量法对语音信号序列的最大 Lyapunov 指数进行计算, 结果表明其最大 Lyapunov 指数大于零, 表明语音信号序列具有混沌特性.

2) 针对具有混沌特性的语音信号序列, 构建了一种具有可变收敛因子特性的 DFPSOVF 模型. 应用 DFPSOVF 模型对语音信号序列进行单帧和多帧预测. 仿真结果表明, DFPSOVF 模型对具有混沌特性的语音信号序列预测的有效性和准确性, 并克服了 LMS 算法存在的参数选择困难. 因此, 本文提供的预测模型为语音信号的准确预测提供了一种新方法和思路.

3) 研究了 DFPSOVF 模型的记忆长度对语音信号预测精度的影响. 结果表明 DFPSOVF 模型记忆长度对语音信号预测精度影响不明显, 可以依据语音信号序列的嵌入维数选择 DFPSOVF 模型的记忆长度.

4) 仿真结果表明本文提出的非线性预测模型具有比传统线性预测模型更好的预测能力, 说明本文提出的非线性模型可在一定程度上替代线性预测模型, 为语音信号预测与非线性分析提供了更为有效的模型结构.

参考文献

[1] Maragos P 1991 *ICASSP* 417
[2] Wu X J, Yang Z Z 2013 *Appl. Soft Comput.* **13** 3314
[3] Max A L 2011 *Advances in Nonlinear Speech Processing* **7015** 9
[4] Cheng X F, Zhang Z 2013 *Acta Phys. Sin.* **62** 168701 (in Chinese) [成谢锋, 张正 2013 物理学报 **62** 168701]
[5] Chen D Y, Liu Y, Ma X Y 2012 *Acta Phys. Sin.* **61** 100501 (in Chinese) [陈帝伊, 柳烨, 马孝义 2012 物理学报 **61** 100501]
[6] Iasonas K, Petros M 2005 *IEEE Trans. Speech Audio Process.* **13** 1098
[7] Sun J F, Zheng N H, Wang X L 2007 *Singal Process.* **87** 2431
[8] Maciej O 2002 *IEEE Int. Symp.Circuits Syst.* 564
[9] Xiao X C, Li H C, Zhang J S 2005 *Chin. Phys.* **14** 2181

[10] Zhang J S, Li H C, Xiao X C 2005 *Chin. Phys.* **14** 49

[11] Thyssen J, Nielsen H, Hansen S D 1994 *ICASSP* 185

[12] Sigrist Z, Grivel E, Alcoverro B 2012 *Signal Process.* **92** 1010

[13] Mathews V J 1991 *IEEE Signal Process. Mag.* **8** 10

[14] Wei R X, Han C Z, Zhang Z L 2005 *Acta Electron. Sin.* **33** 656 (in Chinese) [魏瑞轩, 韩崇昭, 张宗麟 2005 电子学报 **33** 656]

[15] Guerin A, Faucon G, Le Bouquin-Jeannes R 2003 *IEEE Trans. Speech Audio Proc.* **11** 672

[16] Zhang Y M, Wu X J, Bai S L 2013 *Acta Phys. Sin.* **62** 190509 (in Chinese) [张玉梅, 吴晓军, 白树林 2013 物理学报 **62** 190509]

[17] Abarbanel H D I, Masuda N, Rabinovich M I,Tumer E 2001 *Phys. Lett. A* **281** 368

[18] Cao L Y 1997 *Physica D* **110** 43

[19] Rosenstein M T, Collins J J, de Luca C J 1993 *Physica D* **65** 117

[20] de Campos M L R, Antoniou A 1997 *IEEE Trans. Circ. Syst.* **44** 924

Volterra prediction model for speech signal series*

Zhang Yu-Mei[1)2)3)] Hu Xiao-Jun[2)] Wu Xiao-Jun[1)2)3)†] Bai Shu-Lin[4)] Lu Gang[1)2)]

1) (*Key Laboratory of Modern Teaching Technology, Ministry of Education, Shaanxi Normal University, Xi'an 710062, China*)

2) (*School of Computer Science, Shaanxi Normal University, Xi'an 710062, China*)

3) (*School of Automatic Control, Northwestern Polytechnical University, Xi'an 710072, China*)

4) (*School of Electronics and Information, Northwestern Polytechnical University, Xi'an 710072, China*)

(Received 12 November 2014; revised manuscript received 2 June 2015)

Abstract

The given English phonemes, words and sentences are sampled and preprocessed. For these real measured speech signal series, time delay and embedding dimension are determined by using mutual information method and Cao's method, respectively, so as to perform phase space reconstruction of the speech signal series. By using small data set method, the largest Lyapunov exponent of the speech signal series is calculated and the fact that its value is greater than zero presents chaotic characteristics of the speech signal series. This, in fact, performs the chaotic characteristic identification of the speech signal series. By introducing second-order Volterra series, in this paper we put forward a type of nonlinear prediction model with an explicit structure. To overcome some intrinsic shortcomings caused by improper parameter selection when using the least mean square (LMS) algorithm to update Volterra model efficiency, by using a variable convergence factor technology based on a posteriori error assumption on the basis of LMS algorithm, a novel Davidon-Fletcher-Powell-based second of Volterra filter (DFPSOVF) is constructed and is performed to predict speech signal series of the given English phonemes, words and sentences with chaotic characteristics. Simulation results under MATLAB 7.0 environment show that the proposed nonlinear model DFPSOVF can guarantee its stability and convergence and there are no divergence problems in using LMS algorithm; for single-frame and multi-frame of the measured speech signals, when root mean square error (RMSE) is used as an evaluation criterion the prediction accuracy of the proposed nonlinear prediction model DFPSOVF in this paper is better than that of the linear prediction (LP) that is traditionally employed. The primary results of single-frame and multi-frame predictions are given. So, the proposed DFPSOVF model can substitute linear prediction model on certain conditions. Meanwhile, it can better reflect trends and regularity of the speech signal series and fully meet requirements for speech signal prediction. The memory length of the proposed prediction model may be selected by the embedding dimension of the speech signal series. The proposed model can present a nonlinear analysis and more valuable model structure for speech signal series, and opens up a new way to speech signal reconstruction and compression coding so as to improve complexity and process effect of speech signal processing method.

Keywords: speech signal, chaos, Volterra prediction model, Davidon-Fletcher-Powell algorithm

PACS: 05.45.Tp **DOI:** 10.7498/aps.64.200507

* Project supported by the National Natural Science Foundation of China (Grant Nos. 11502133, 11172342, 11372167, 61202153), the Key Science and Technology Innovation Team in Shaanxi Province, China (Grant No. 2014KTC-18), the Science and Technology Plan of Xi'an City, China (Grant No. CXY1437(1)), and the Science and Technology Plan of Yulin City, China (Grant Nos. 2014cxy-09, sf13-43, 2012cxy3-6).

† Corresponding author. E-mail: wythe@snnu.edu.cn