

# 隐相空间的 DUPSO-RPSOVF 语音预测模型研究

吴 霞<sup>1,2</sup>, 吴晓军<sup>1,2</sup>, 史素真<sup>3</sup>, 张其进<sup>2</sup>, 张玉梅<sup>1,2</sup>

(1. 陕西师范大学现代教学技术教育部重点实验室, 陕西西安 710062;  
2. 陕西师范大学计算机科学学院, 陕西西安 710062;  
3. 河南机电职业学院信息工程学院 河南郑州 451191)

**摘 要:** 提出了一种基于二阶 Volterra 级数的语音信号非线性预测模型. 为克服传统的最小均方 (Least Mean Square, LMS) 算法在模型核系数更新时的固有缺点, 引入耗散均匀搜索粒子群优化算法 (Dissipative Uniform Particle Swarm Optimization, DUPSO) 求解核系数, 并构建了 DUPSO-SOVF 预测模型; 为避免传统方法中相空间的重构过程, 构建了隐相空间 DUPSO-SOVF 预测模型, 在求解模型核系数时动态地求解出最优嵌入维数和延迟时间; 为降低模型复杂度, 在误差允许范围内进行模型关键项的提取, 从而减少了核系数个数, 构建了少参数的 DUPSO-RPSOVF (Reduced Parameter SOVF, RPSOVF) 预测模型. 将英语音素、单词和短语作为实验样本数据进行仿真, 结果表明: 隐相空间 DUPSO-SOVF 模型能够准确的计算出相空间重构参数, DUPSO-SOVF 和 DUPSO-RPSOVF 两种预测模型对单帧和多帧语音信号均具有较高的预测精度, 优于 PSO-SOVF 和 LMS-SOVF 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, 可以满足语音序列预测的要求.

**关键词:** 语音信号; 预测; Volterra 模型; DUPSO 算法; 隐相空间重构

**中图分类号:** TP713 **文献标识码:** A **文章编号:** 0372-2112 (2019)09-1875-08

**电子学报 URL:** <http://www.ejournal.org.cn> **DOI:** 10.3969/j.issn.0372-2112.2019.09.009

## Research on DUPSO-RPSOVF Speech Prediction Model with Hidden Phase Space

WU Xia<sup>1,2</sup>, WU Xiao-jun<sup>1,2</sup>, SHI Su-zhen<sup>3</sup>, ZHANG Qi-jin<sup>2</sup>, ZHANG Yu-mei<sup>1,2</sup>

(1. Key Laboratory of Modern Teaching Technology, Ministry of Education, Shaanxi Normal University, Xi'an, Shaanxi 710062, China;  
2. School of Computer Science, Shaanxi Normal University, Xi'an, Shaanxi 710062, China;  
3. Information Engineering Department, Henan Mechanical and Electrical Vocational College, Zhengzhou, Henan 451191, China)

**Abstract:** A type of nonlinear prediction model for speech signals based on second-order Volterra series is put forward. In order to overcome some intrinsic shortcomings caused by using the classic least mean square (LMS) algorithm to update Volterra model kernel coefficients, a dissipative uniform particle swarm optimization (DUPSO) algorithm is applied to obtain the kernel coefficients and then a DUPSO-SOVF prediction model can be constructed. A DUPSO-SOVF prediction model with hidden phase space is constructed by dynamically obtaining parameters of embedding dimension and time delay in the process of solving model kernel coefficients rather than using traditional phase space reconstruction process. On the purpose to reduce model complexity, the key model kernels are extracted within the margin of the allowable error and the model kernels are then reduced, and the reduced parameter DUPSO-SOVF (RPSOVF) prediction model is proposed. Simulation results for samples of English phonemes, words and phrases show that, the DUPSO-SOVF model with hidden phase space can accurately calculate parameters of embedding dimension and delay time of phase space reconstruction; both of the DUPSO-SOVF model and the DUPSO-RPSOVF model exhibit higher prediction accuracy on single frame and multi-frame speech signal than PSO-SOVF and LMS-SOVF models. Also, the two proposed models can better reflect trends and regularities of the speech signal series and meet requirements for speech signal prediction.

**Key words:** speech signal; prediction; Volterra model; DUPSO algorithm; hidden phase space reconstruction

收稿日期: 2018-09-27; 修回日期: 2018-12-27; 责任编辑: 梅志强

基金项目: 国家重点研发计划项目 (No. 2017YFB1402102); 国家自然科学基金 (No. 11872036, No. 11502133, No. 11772178); 陕西省重点研发计划项目 (No. 2019GY-217, No. 2019ZDLSF07-01)

## 1 引言

随着语音信号的深入研究,语音信号已被证实具有较强的时变特性,并包含大量的非线性特征<sup>[1]</sup>.近年来,计算机技术及信号处理领域取得了飞速发展,人们对语音质量的要求也进一步提高,传统的线性预测模型已不能很好地满足现代语音信号处理的需求,因此如何构建语音信号的非线性模型、提出更高精度的预测方法成为语音信号处理领域的一个重要研究热点<sup>[2]</sup>.在非线性和相关问题研究中,Volterra 模型是应用最广泛而有效的方法之一<sup>[3]</sup>,它对系统的线性特征和非线性特征均具备良好的描述能力,从而在非线性的时间序列预测、伪迹消除等研究中取得了丰富成果<sup>[4-8]</sup>.随着语音信号非线性建模的进一步发展,Volterra 模型已被应用于语音信号序列的研究中,如文献[7]通过引入 Volterra 模型很好地提高了语音信号的预测精度并获得了显式结构的非线性语音预测模型.利用 Volterra 模型研究语音信号不但比传统线性模型具有更优异的预测性能,而且比神经网络<sup>[9]</sup>等其他语音信号的非线性模型更容易获得模型的显式结构.

然而,随着预测精度的增加,Volterra 模型会存在维数灾难问题,要解决该问题则需要一个高效的优化算法.现有研究中,最小均方(Least Mean Square, LMS)算法被广泛应用于 Volterra 模型的系数更新,但是该算法涉及矩阵的求逆运算.当自相关矩阵出现奇异时,会存在误差求解难题<sup>[10]</sup>,此外其收敛速度对于输入信号具有依赖性也是该算法的一个固有缺点.而在非线性优化问题的求解过程中,粒子群优化算法(Particle Swarm Optimization, PSO)通常被用来解决复杂系统的参数优化问题,它具有参数较少和容易实现的特点,可用于解决大量的非线性、不可微和多峰值等复杂问题<sup>[11]</sup>,但是 PSO 算法存在易陷入局部最优解的算法缺陷.相较而言,基于均匀搜索思想改进后的耗散均匀搜索粒子群算法(Dissipative Uniform PSO, DUPSO)在提高单峰值以及非均匀多峰值函数的精度等方面具有更明显的优势<sup>[12]</sup>.因此,为解决 Volterra 模型参数优化过程中的维数灾难问题,并获取模型的全局最优解,提高预测精度,本文将利用 DUPSO 算法实现二阶 Volterra 模型(Second of Volterra Filter, SOVF)核系数的求解.

此外,相关研究表明,对重构相空间后的语音信号进行研究,更有利于深入分析其非线性动力学特征并能更准确地对其进行预测.张家树等人<sup>[13]</sup>在相空间重构思想的基础上提出了一种混沌时间序列 Volterra 非线性预测模型,很好地实现了多种不同低维混沌序列的预测研究.文献[14]基于相空间重构理论将自相关法和 Cao 方法相结合得到重构参数向量,然后构建以

Hermite 正交基函数为奇次函数的 Hermite 神经网络预测模型,将模型参数优化问题转化为多维空间上的函数优化问题.这类基于相空间重构理论的方法,其关键步骤是嵌入维数和延迟时间的确定.基于 Takens 理论选取最佳嵌入维数与延迟时间,已成为混沌时间序列相空间重构的重要方法之一<sup>[15]</sup>,但是应用中的语音信号并不满足 Takens 理论中要求无限长及无噪声的条件,因此在相空间重构前,必须精确给出最佳嵌入维数与延迟时间,否则会对重构的相空间质量造成极大影响.因此,如何获得相空间的最佳嵌入维数和延迟时间成为相空间重构过程中的一个重要研究问题.

基于以上分析,本文将引入 DUPSO 优化算法,把相空间参数求解过程隐含于 SOVF 模型核系数优化过程中并进行同步求解,从而建立了隐相空间 DUPSO-SOVF 语音预测模型;为了降低模型复杂度,解决维数灾难问题,在误差允许范围内对隐相空间 DUPSO-SOVF 模型进行关键项提取,建立了少参数的隐相空间 DUPSO-RPSOVF 语音预测模型;最后通过仿真将本文所构建的两个模型分别应用于英语音素、单词和短语预测.

## 2 语音信号数据采集与预处理

### 2.1 语音信号数据采集

本文给定一组英语音素、单词和短语,并针对不同被试者采集其对应的发音信号并建立语料库.为了得到更高信噪比的语音信号,在录制过程中需严格控制环境因素,减少干扰噪声,此次采集实验给定的语音音素样本分别为[dr],[dz],[m],[ð],[ŋ],[u]等,录音的采样率是 8000/s,数据位数为 8 位,给出部分语音信号序列波形图如图 1 所示.

### 2.2 语音信号数据预处理

语音信号的短时平稳性表明,在一个短时间范围内(10ms~30ms)其时变特性有限,所以语音信号分析大多建立在短时基础上,且对语音信号样本进行建模前,需要进行预加重、加窗和分帧等预处理.在本文中,采用具有 6dB/倍频程的提升高频特性的一阶数字滤波器对 A/D 变换之后的语音信号进行处理,实现预加重过程;然后选取汉明窗完成预处理中的加窗步骤;最后在语音信号分帧预处理中,令帧长为 30ms,即在 8000/s 采样率条件下,每帧数据含有 240 个样点值.

### 2.3 语音信号相空间预测模型

设  $x(t), t=1, 2, \dots, N$  为语音信号时间序列,其在  $m$  维相空间中的状态转移形式为

$$\mathbf{Y}(t+1) = f(\mathbf{Y}(t)) \quad (1)$$

其中,  $\mathbf{Y}(t)$  为相空间中的点,且

$$\mathbf{Y}(t) = (x(t), \dots, x(t+(m-1)\tau)) \quad (2)$$

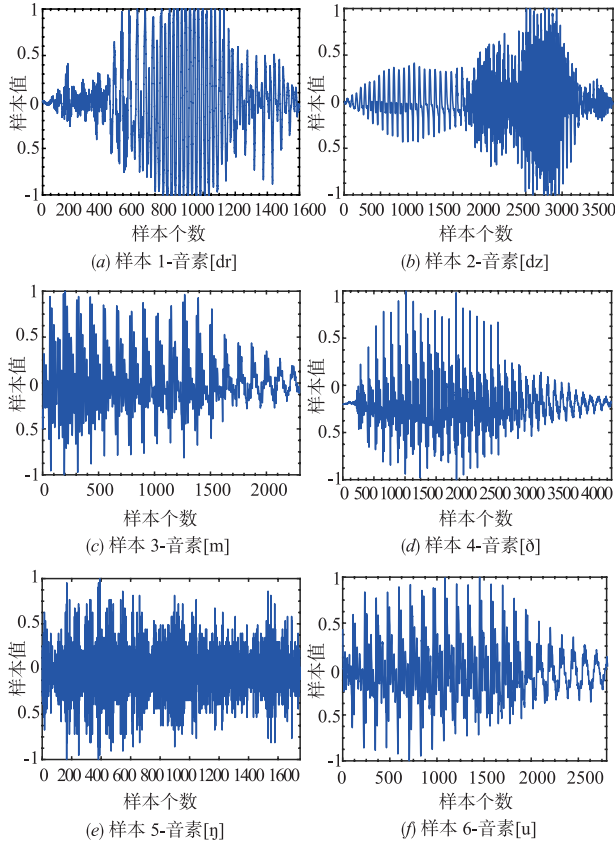


图1 语音信号音素样本波形图

其中,  $\tau$  为延迟时间,  $m$  为嵌入维数, 展开式(1)得

$$\begin{aligned} & (x(t+1), \dots, x(t+1+(m-1)\tau)) \\ & = f(x(t), \dots, x(t+(m-1)\tau)) \end{aligned} \quad (3)$$

对混沌时间序列进行预测, 等价于求解式(3)的映射函数  $f$ , 使之能以重构后向量的前一个或前几个向量作为输入, 预测之后的向量. 传统语音信号预测模型在重构相空间时, 先利用 Cao 方法和互信息法分别求取语音信号的  $m$  和  $\tau$ , 然后令第  $n+\tau$  个向量的前  $m-1$  个分量为第  $n$  个向量的后  $m-1$  个分量, 如表 1 所示. 已知时间序列  $\{x_1, x_2, \dots, x_{15}\}$ , 假设其  $m$  为 4,  $\tau$  为 3, 对其进行相空间重构可得到向量  $\mathbf{X}_1 \cdots \mathbf{X}_6$ , 其中第四个向量  $\mathbf{X}_4$  的前 3 个分量为第一个向量  $\mathbf{X}_1$  的后 3 个分量, 第五个向量  $\mathbf{X}_5$  的前三个分量为第二个向量  $\mathbf{X}_2$  的后三个分量.

表 1 相空间重构

向量序号	向量的分量			
$\mathbf{X}_1$	$x_1$	$x_4$	$x_7$	$x_{10}$
$\mathbf{X}_2$	$x_2$	$x_5$	$x_8$	$x_{11}$
$\mathbf{X}_3$	$x_3$	$x_6$	$x_9$	$x_{12}$
$\mathbf{X}_4$	$x_4$	$x_7$	$x_{10}$	$x_{13}$
$\mathbf{X}_5$	$x_5$	$x_8$	$x_{11}$	$x_{14}$
$\mathbf{X}_6$	$x_6$	$x_9$	$x_{12}$	$x_{15}$

### 3 隐相空间 DUPSO-SOVF 预测模型

#### 3.1 SOVF 模型

SOVF 模型是二阶截断 Volterra 模型, 定义如下:

$$\begin{aligned} \hat{y}(n) = & h_0 + \sum_{i=0}^{s-1} h_1(i, n-1)x(n-i) \\ & + \sum_{i=0}^{s-1} \sum_{j=0}^{s-1} h_2(i, j; n-1)x(n-i)x(n-j) \end{aligned} \quad (4)$$

其中,  $x(n)$  和  $\hat{y}(n)$  分别表示  $n$  时刻的输入信号和输出信号,  $s$  为模型记忆长度<sup>[7]</sup>,  $h_0$ 、 $h_1(i; n-1)$  和  $h_2(i, j; n-1)$  分别为常数项、线性项系数和平方项系数, 其中线性项系数和平方项系数又称为核系数. 记:

$$\begin{aligned} \mathbf{H}(n-1) = & [h_1(0; n-1), h_1(1; n-1), \dots, h_1(s-1; n-1), \\ & h_2(0, 0; n-1), h_1(0, 1; n-1), \dots, \\ & h_1(s-1, s-1; n-1)]^T \end{aligned} \quad (5)$$

$$\begin{aligned} \mathbf{X}(n) = & [x(n), x(n-1), \dots, x(n-(s-1)), x^2(n), \\ & x(n)x(n-1), \dots, x^2(n-(s-1))]^T \end{aligned} \quad (6)$$

其中,  $[\cdot]^T$  表示向量转置. 因此, 式(4)可表示为如下向量形式:

$$\hat{y}(n) = \mathbf{H}^T(n-1)\mathbf{X}(n) \quad (7)$$

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:

$$e(n) = (y(n) - \hat{y}(n))^2 \quad (8)$$

引入 LMS 算法对式(8)进行求解, 其系数向量  $\mathbf{H}(n)$  的迭代公式为:

$$\mathbf{H}(n) = \mathbf{H}(n-1) + 2\mu e(n)\mathbf{X}(n) \quad (9)$$

其中,  $\mu$  为收敛因子, 控制着 LMS 算法的收敛速度与稳定性.

#### 3.2 隐相空间 DUPSO-SOVF 语音预测模型

基于式(4), 将相空间重构过程隐含于 DUPSO-SOVF 模型求解过程之中得到:

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

该模型的最优解记为  $Q(m, \tau, \mathbf{H}^*)$ , 其中  $\mathbf{H}^*$  为模型核系数最优解. 求解模型最优解时, 先随机初始化  $m$  和  $\tau$ , 并令有  $N_i$  个粒子, 且第  $i$  个粒子的位置用  $m$  维向量  $\mathbf{H}_i$  表示, 则粒子群的位置信息可以由一个  $N_i \cdot m$  的矩阵  $\mathbf{H}_{N_i \cdot m}$  表示. 将  $m$ 、 $\tau$  和参数矢量  $\mathbf{H}_{N_i \cdot m}$  作为 DUPSO 优化算法中的粒子位置进行迭代求解. 模型在求解过程中可动态地完成隐相空间重构. 在搜索第  $i$  个粒子的最优位置时, 适应度值的计算函数为:

$$F(\mathbf{H}_i) = \frac{1}{L} \sum_{t_i=1}^L (y(t_i) - \mathbf{H}_i^T \mathbf{X}(t_i))^2 \quad (11)$$

其中,  $i = 1, 2, \dots, N$ ,  $L$  为试验数据的长度. 则求解式(10)寻找最优粒子位置使得式(11)目标函数的适应度

值最小. 具体算法过程如下:

**步骤 1** 首先随机初始化  $m$  和  $\tau$  以及每个粒子的初始状态, 包括初始位置  $\mathbf{H}_i(0)$  和初始速度  $\mathbf{V}_i(0)$ . 然后求得使适应度函数值最小的粒子解集, 并令第  $i$  个粒子的当前最佳位置为  $\mathbf{P}_{\text{best}_i}(0) = \mathbf{H}_i(0)$ . 利用式(11)分别求出当前状态下的各个粒子的适应值, 并把使目标函数取值最小的粒子位置记为全局最优位置  $\mathbf{G}_{\text{best}}(0)$ .

**步骤 2** 在更新粒子的速度、位置以及模型的最优  $m$  和  $\tau$  时, 分别采用式(12)和式(13).

$$\mathbf{V}_{\text{id}}(t+1) = \omega \mathbf{V}_{\text{id}}(t) + c_1 \cdot r_1(t) \cdot (\mathbf{P}_{\text{best}_{\text{id}}}(t) - \mathbf{H}_{\text{id}}(t)) + c_2(1 - r_1(t)) \cdot (\mathbf{G}_{\text{best}_{\text{id}}}(t) - \mathbf{H}_{\text{id}}(t)) \quad (12)$$

$$\mathbf{H}_{\text{id}}(t+1) = \mathbf{H}_{\text{id}}(t) + \mathbf{V}_{\text{id}}(t+1) \quad (13)$$

其中,  $c_1$  为学习系数,  $r_1(t)$  为  $0 \sim 1$  之间均匀分布的随机数,  $\omega$  为惯性系数.

**步骤 3** 根据粒子感知信息更新粒子当前的局部最优位置, 并计算此时粒子的适应度值, 当  $F(\mathbf{H}_i(t)) < F(\mathbf{P}_{\text{best}_i}(t))$ , 更新  $\mathbf{P}_{\text{best}_i}(t) = \mathbf{H}_i(t)$ , 否则  $\mathbf{P}_{\text{best}_i}(t)$  不变; 然后由所有粒子的社会感知信息比较  $F(\mathbf{P}_{\text{best}_i}(t))$ , 由式(14)可对全局最优位置进行更新.

$$\mathbf{G}_{\text{best}}(t) = \underset{1 \leq i \leq N}{\operatorname{argmin}} [F(\mathbf{P}_{\text{best}_i}(t))] \quad (14)$$

**步骤 4** 当迭代步数  $t$  达到设定的迭代次数时, 终止当前程序并记录数据. 此时得到的  $\mathbf{G}_{\text{best}}(t)$  即为全局最优位置, 并输出模型的最优解为  $Q(m, \tau, \mathbf{H}^*)$ , 否则返回步骤 2 并继续迭代.

### 3.3 隐相空间 DUPSO-RPSOVF 语音预测模型

基于隐相空间的 DUPSO-SOVF 模型构建, 在误差  $\sigma$  的允许范围内, 选出模型关键项, 从而对 SOVF 模型进行结构优化, 提高 DUPSO 优化算法的计算效率. 步骤如下:

**步骤 1** 定量计算相空间重构后的输入信号  $\mathbf{X}(n)$  的每一列对于输出信号  $y(n)$  的影响, 其关系函数如式(15)

$$y = \mathbf{X}_i \cdot \mathbf{H}, i = 1, 2, 3, \dots, (m^2 + 3m)/2 \quad (15)$$

其中,  $\mathbf{X}_i$  表示只保留  $\mathbf{X}(n)$  中第  $i$  列且其余列为零的向量, 采用 DUPSO 算法求解模型的最优解.

**步骤 2** 计算出  $\mathbf{X}(n)$  每一项的均方根误差 (Root Mean Square Error, RMSE), 并进行排序, 比较  $\mathbf{X}(n)$  的每一项对输出  $y(n)$  的作用.  $\mathbf{X}(n)$  每一项的 RMSE 值可反应其对预测值的影响程度, 即 RMSE 越小说明当前对应的列对预测值越关键, 对输出的影响越大, 反之, RMSE 越大表明其对应的列对输出影响越小.

**步骤 3** 初始化矩阵  $\mathbf{V} = [0 : 0]$ , 选取 SOVF 模型的一阶和二阶的关键项系数并将其分别做为矩阵  $\mathbf{V}$  的两个独立部分, 对  $\mathbf{X}(n)$  中的列进行逐个搜索, 判断是否移除第  $i$  列从而优化矩阵  $\mathbf{V}$  中的列. 计算移除第  $i$  列时的 RMSE 值, 当移除第  $i$  列使得  $\text{RMSE} > \sigma$ , 则保留第  $i$  列, 否则将其移出. 当遍历完  $\mathbf{V}$  中的每一列则停止判断, 得到最

优输入矩阵  $\mathbf{V}'$ . 利用 DUPSO 优化算法获得  $\mathbf{V}$  对应的核系数矩阵  $\mathbf{H}'$ , RPSOVF 模型可表示为式(16).

$$y = \mathbf{V}' \cdot \mathbf{H}' \quad (16)$$

## 4 模型评价

### 4.1 相空间重构参数

求解隐相空间 DUPSO-SOVF 模型获得最优  $m$  和  $\tau$ , 与传统 Cao 方法求取  $m$  和互信息法求取  $\tau$  进行对比, 结果如表 2 所示. 由表 2 可以看出, 隐相空间 DUPSO-SOVF 模型在求取  $m$  和  $\tau$  时与 Cao 方法<sup>[14]</sup> 和互信息法<sup>[15]</sup> 的求取结果基本一致, 这表明本文提出的隐相空间 DUPSO-SOVF 模型在求解语音信号的  $m$  和  $\tau$  时, 具有良好的准确性. 同时所构建的模型减少了传统方法中基于相空间重构进行预测时需要预先确定  $m$  和  $\tau$  的步骤.

表 2 隐相空间 DUPSO-SOVF 模型与 Cao 和互信息法比较

试验样本	隐相空间 DUPSO-SOVF 模型求解 $m$	Cao 方法 求解 $m$	隐相空间 DUPSO-SOVF 模型求解 $\tau$	互信息法 求解 $\tau$
[b]	13	10	12	12
[d]	9	9	7	7
[d <sub>5</sub> ]	13	13	15	14
[dz]	11	11	8	5
[dr]	5	5	15	15
[m]	9	9	9	10
[n]	11	9	9	10
[s]	9	10	15	15
[g]	11	11	12	11
[u]	9	9	15	15
[ð]	9	14	15	14
good	10	11	11	12
jack	13	13	2	3
this	14	14	8	8
video	11	11	10	9

### 4.2 LMS-SOVF、PSO-SOVF 和隐相空间 DUPSO-SOVF 模型的比较

本实验对 LMS-SOVF 模型<sup>[10]</sup>、PSO-SOVF 模型<sup>[16]</sup> 和隐相空间 DUPSO-SOVF 模型进行对比, 以语音音素 [dz] 和 [u] 作为样本数据, 令学习因子为 1.5, 惯性系数为 0.78, 迭代次数为 100 次. 仿真结果如表 3 和图 2 至图 4 所示. 从表 3 可以看出, 在相同的记忆长度下, 当以 RMSE 作为对比标准时, 隐相空间 DUPSO-SOVF 模型对音素 [dz]、[u] 的预测比其它两种方法更精确. 由图 2、图 3 和图 4 的波形图可以看出, 隐相空间 DUPSO-SOVF 模型获得的预测波形最接近原始语音波形, 表明隐相空间 DUPSO-SOVF 模型比 PSO-SOVF 模型和 LMS-SOVF 模型具有更佳的预测效果.

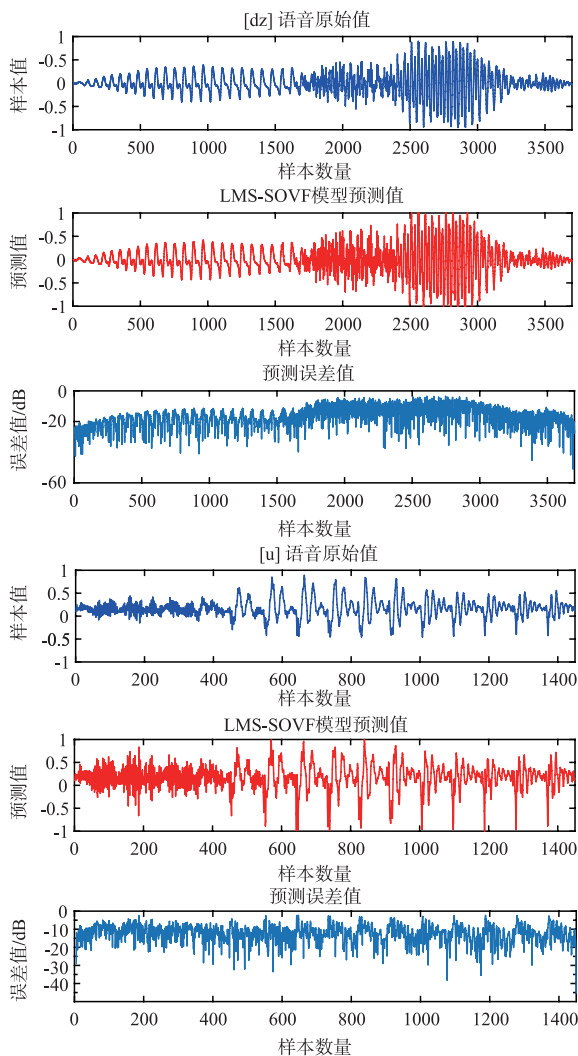


图2 LMS-SOVF模型对音素[dz]、[u]的预测

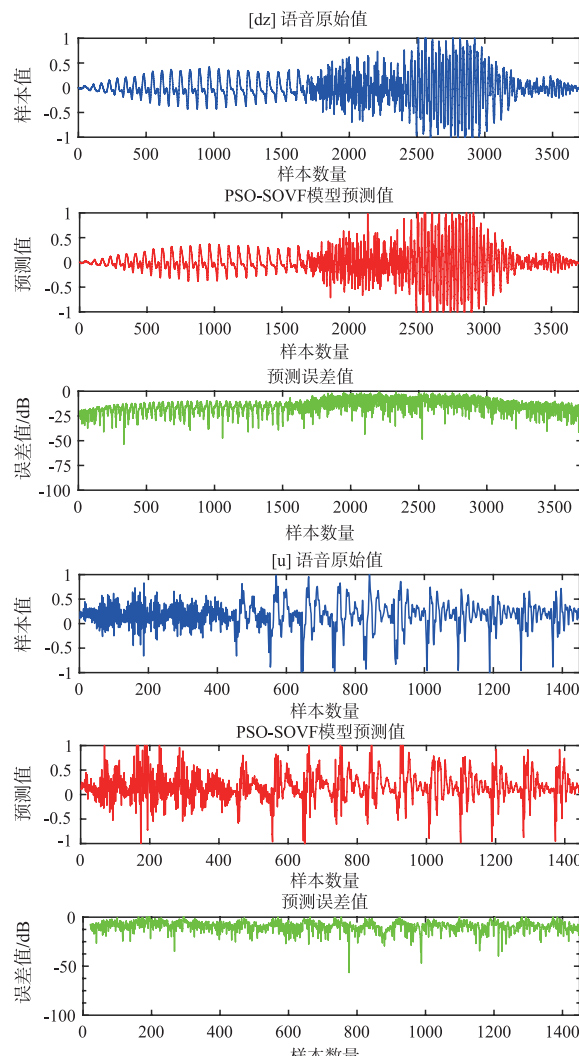


图3 PSO-SOVF模型对音素[dz]、[u]的预测

表 3 LMS-SOVF、PSO-SOVF 和隐相空间 DUPSO-SOVF 模型对音素 [dz] 和 [u] 预测的 RMSE 对比

语音音素	LMS-SOVF	PSO-SOVF	隐相空间 DUPSO-SOVF
[dz]	0.2634	0.2547	0.2058
[u]	0.2186	0.1874	0.1547

表 4 隐相空间 DUPSO-SOVF 与 PSO-SOVF 模型对音素 [dr]、[dz]、[m]、[u] 预测的 RMSE 及时耗对比

语音音素		[dr]	[dz]	[m]	[u]
PSO-SOVF	RMSE	0.5876	0.2547	0.5361	0.2058
	Time(s)	285.78	755.12	501.28	487.85
隐相空间 DUPSO-SOVF	RMSE	0.4479	0.2014	0.4521	0.1547
	Time(s)	201.47	632.98	385.62	321.35

以时耗及 RMSE 作为评判指标,比较隐相空间 DUPSO-SOVF 模型和 PSO-SOVF 模型对音素 [dr]、[dz]、[m]、[u] 的预测结果,仿真结果如表 4 所示。从表 4 可以看出,在预测过程中,基于隐相空间 DUPSO-SOVF 模型比 PSO-SOVF 模型的 RMSE 值小,且在相同

的运行环境下,隐相空间 DUPSO-SOVF 模型所需要的运行时间较短,即隐相空间 DUPSO-SOVF 模型比 PSO-SOVF 模型预测性能更高且时耗更小。

### 4.3 关键项提取与隐相空间 DUPSO-RPSOVF 模型预测

以音素 [dr] 为实验样本,基于关键项选择步骤,仿真结果如表 5 所示。从表 5 可以看出,隐相空间 DUPSO 优化算法在求解 SOVF 模型核系数时,会出现数值较小的值,这说明该参数对模型的影响较小,不属于关键项。求取音素 [dr] 的一阶核系数和二阶核系数,如图 5 所示。由图 5 可知,模型一阶核系数中  $H(1)$  属于关键项,二阶核系数中  $H(0,0)$ 、 $H(1,1)$ 、 $H(1,3)$  属于关键项。将模型核系数从一阶到二阶按表 5 顺序排列,模型的关键项分别标记为 2,6,12,14。按照相同的方法分别对音素 [δ]、[m]、[u] 进行关键项提取,其结果如表 6 所示。由表 6 看出隐相空间 DUPSO-RPSOVF 模型能准确的提取实验样本的关键项,从而有效地提高了模型的预测

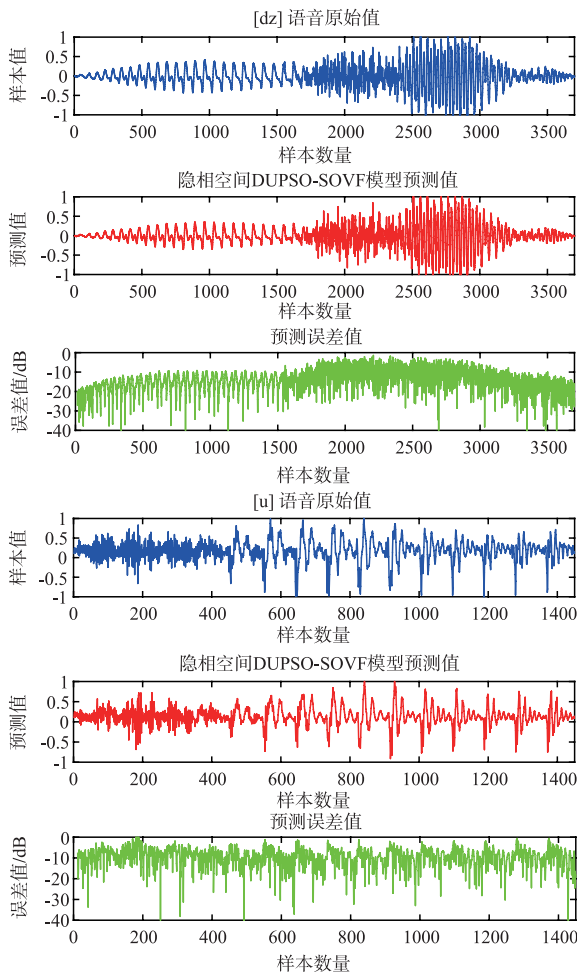


图4 隐相空间DUPSO-SOVF模型对音素[dz]、[u]的预测

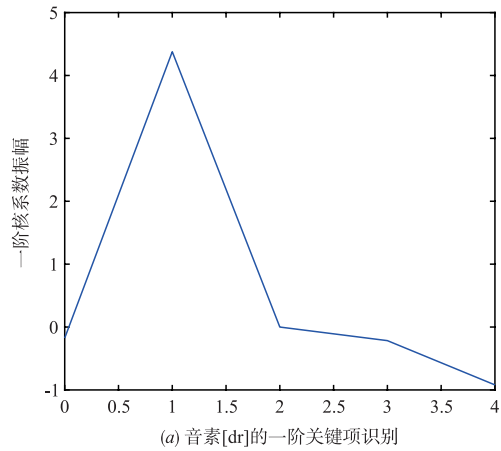
效率. 基于提取的关键项构建隐相空间 DUPSO-RPSOVF 模型, 对音素 [dr] 和 [ŋ] 进行预测, 结果如图 6 所示. 由图 6 可看出, 通过关键项提取所构建的 DUPSO-RPSOVF 模型在进行语音因素预测时, 预测值与样本值具有较高的相似度且误差较小, 这表明 DUPSO-RPSOVF 模型具有更高效的预测性能.

表 5 隐相空间 DUPSO-RPSOVF 模型对 [dr] 预测的核系数值

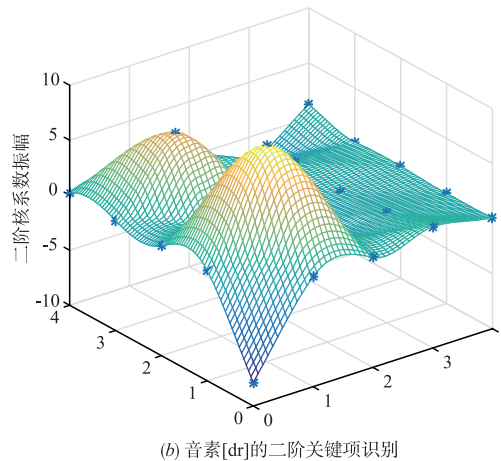
H(0)	H(1)	H(2)	H(3)	H(4)	H(0,0)	H(0,1)	H(0,2)
-0.175	4.378	0	-0.216	-0.920	-7.852	0	0
H(1,0)	H(1,1)	H(1,2)	H(1,3)	H(1,4)	H(2,1)	H(2,2)	
-0.001	9.634	0.003	6.152	0.285	0	-0.169	-0.584
H(3,0)	H(3,1)	H(3,2)	H(3,3)	H(3,4)	H(4,0)	H(4,1)	H(4,1)
1.027	0	-0.425	0	-2.146	0.104	0	0.187

表 6 隐相空间 DUPSO-RPSOVF 模型对音素 [dr]、[ð]、[m]、[u] 预测的关键项提取

音素	[dr]	[ð]	[m]	[u]
RMSE	0.1856	0.1008	0.1289	0.0915
关键项	2,6,12,14	23,40,77,78	10,45,89	16,25,56,75



(a) 音素[dr]的一阶关键项识别



(b) 音素[dr]的二阶关键项识别

图5 隐相空间DUPSO-RPSOVF模型对音素[dr]的一阶和二阶关键项识别

隐相空间 DUPSO-RPSOVF 模型在进行音素 [dr]、[ð]、[m]、[u] 的预测时, 其 RMSE 值比隐相空间 DUPSO-SOVF 模型有了减小, 且隐相空间 DUPSO-RPSOVF 模型能有效利用模型中对语音信号影响较大的关键项, 从而优化预测模型, 获得更高效的预测效果.

#### 4.4 隐相空间 DUPSO-RPSOVF 模型预测单词和短语

本实验对提出的 DUPSO-RPSOVF 模型在更多语音信号中的有效性进行验证, 对多帧单词样本 hand 和短语样本 yesterday once more 进行预测实验, 并对比其波形图和均方根误差值, 仿真结果如图 7 所示. 由图 7 可以看出, DUPSO-RPSOVF 模型的预测值能够较直观地反映原始信号值. 此外, 图 6 的预测样本数为 1600, 图 7 的预测样本数分别为 25000 和 50000, 且 RMSE 分别为 0.0350 和 0.0284. 由此可知, 预测样本数的增大并没有对隐相空间 DUPSO-RPSOVF 模型的预测精度造成影响, 所以该模型具有良好的稳定性.

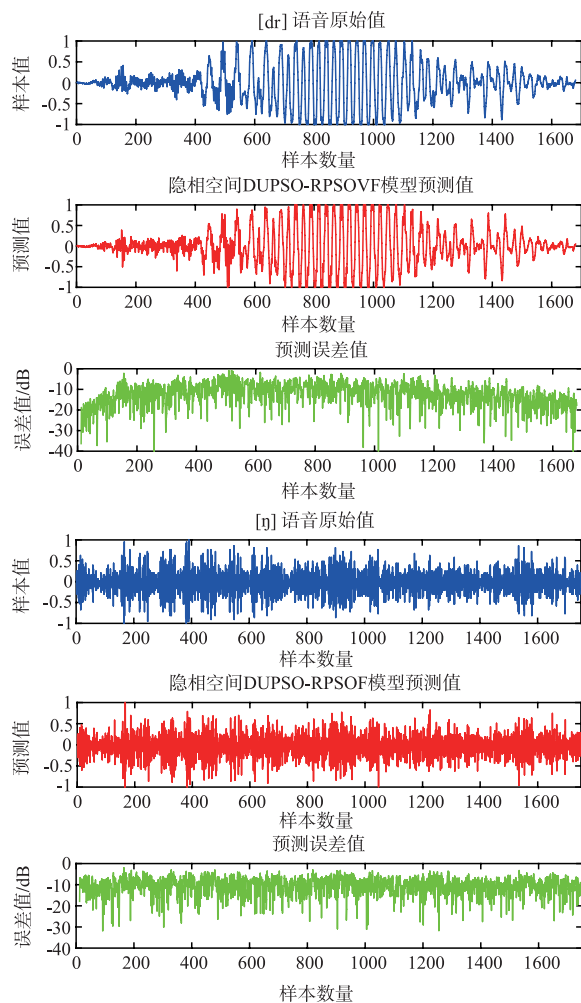


图6 隐相空间DUPSO-RPSOVF模型对音素[dr]和[ŋ]的预测

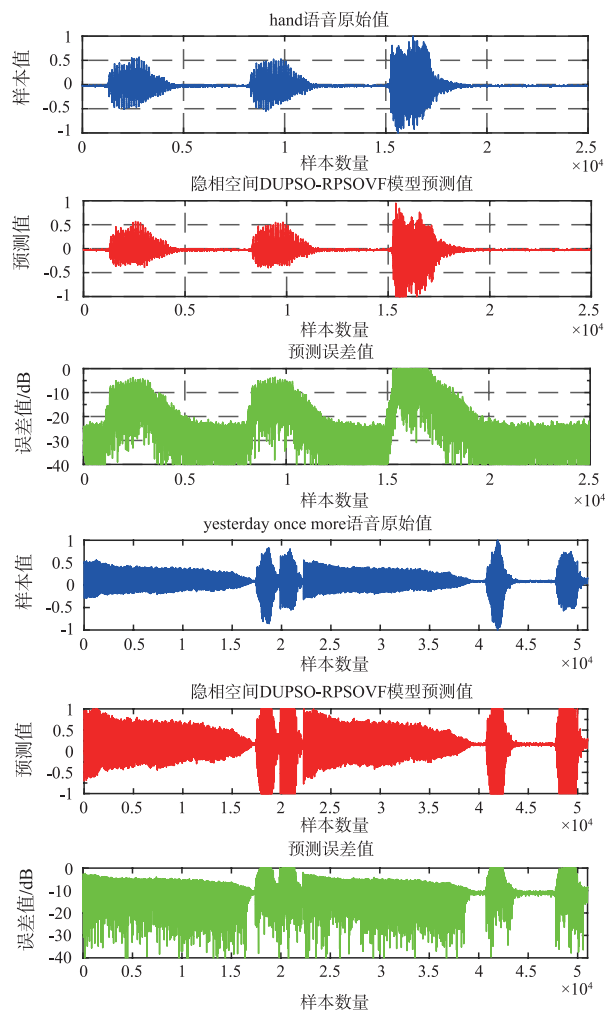


图7 隐相空间DUPSO-RPSOVF模型对单词hand和短语yesterday once more的预测

## 5 结论

本文基于 Volterra 模型提出了一种具有显式结构的隐相空间语音信号非线性预测模型. 通过引入 DUPSO 优化算法求解 SOVF 模型核系数, 并将相空间重构过程隐含于语音时间序列模型的求解过程之中, 得到了隐相空间 DUPSO-SOVF 模型, 避免了传统方法中需要提前进行相空间重构的步骤. 为了降低模型复杂度, 在误差允许范围内对 DUPSO-SOVF 模型进行关键项的提取, 减少了核系数数量, 构建了少参数的隐相空间 DUPSO-RPSOVF 模型. 最后将英语音素、单词和短语分别作为实验样本数据, 通过仿真得到如下结论: 隐相空间 DUPSO-SOVF 模型能够准确的计算出相空间重构核心参数, 隐相空间 DUPSO-SOVF 和 DUPSO-RPSOVF 语音预测模型均对单帧和多帧语音信号具有优良的预测精度, 且优于 PSO-SOVF 和 LMS-SOVF 语音预测模型, 能够满足语音信号预测的要求.

## 参考文献

- [1] Sadeghi M, Marvi H, Ali M. The effect of different acoustic noise on speech signal formant frequency location[J]. International Journal of Speech Technology, 2018, 23(1): 741 - 752.
- [2] Su L. Prediction of multivariate chaotic time series with local polynomial fitting[J]. Computers & Mathematics with Applications, 2010, 59(2): 737 - 744.
- [3] Cheng C M, Peng Z K, Zhang W M, et al. Volterra-series-based nonlinear system modeling and its engineering applications: A state-of-the-art review[J]. Mechanical Systems & Signal Processing, 2017, 87: 340 - 364.
- [4] 张玉梅, 吴晓军, 白树林, 基于 DFP 的二阶 Volterra 滤波器及其在混沌序列预测中的应用[J]. 中国科学: 物理学力学天文学, 2013, 43(4): 530 - 537.  
Zhang Y M, Wu X J, Bai S L. A DFP-method-based second-order Volterra filter and its application to chaotic time

- series prediction [J]. Scientia Sinica, 2013, 43 (4) : 530 – 537. (in Chinese)
- [5] Tan L, Jiang J. Adaptive Volterra filters for active control of nonlinear noise processes [J]. IEEE Trans Signal Process, 2019, 49(8) : 1667 – 1676.
- [6] 张玉梅, 吴晓军. 交通流量序列混沌特性分析及 DFPSOVF 预测模型 [J]. 物理学报, 2013, 62 (19) : 190509-1 – 190509-9.  
Zhang Y M, Wu X J, Bai S L. Chaotic characteristic analysis for traffic flow series and DFPSOVF prediction model [J]. Acta Physica Sinica, 2013, 62 (19) : 190509-1 – 190509-9. (in Chinese)
- [7] 张玉梅, 胡小俊, 吴晓军, 等. 语音信号序列的 Volterra 预测模型 [J]. 物理学报, 2015, 64 (20) : 200507-1 – 200507-13.  
Zhang Y M, Hu X J, Wu X J, et al. Volterra prediction model for speech signal series [J]. Acta Physica Sinica, 2015, 64(20) : 200507-1 – 200507-13. (in Chinese)
- [8] Sekar R C G, Murugesan K. Numerical solutions of nonlinear system of higher order Volterra integro-differential equations using generalized STWS technique [J]. Differential Equations & Dynamical Systems, 2017(1) : 1 – 13.
- [9] Sainath T N, Weiss R J, Wilson K W, et al. Multichannel signal processing with deep neural networks for automatic speech recognition [J]. IEEE/ACM Transactions on Audio Speech & Language Processing, 2017, 25(5) : 965 – 979.
- [10] Hassani M, Karami M R. Noise estimation in electroencephalogram signal by using Volterra series coefficients [J]. Journal of medical signals and sensors, 2015, (5)3: 192 – 200.
- [11] Wu Y, Miao Q, Ma W, et al. PSOSAC: Particle swarm optimization sample consensus algorithm for remote sensing image registration [J]. IEEE Geoscience & Remote Sensing Letters, 2018, 15(2) : 242 – 246.
- [12] 吴晓军, 杨战中, 赵明. 均价搜索粒子群算法 [J]. 电子学报, 2011, 39(6) : 1261 – 1266.  
Wu X J, Yang ZZ, Zhao M. A uniform searching particle swarm optimization algorithm [J]. Acta Electronica Sinica, 2011, 39(6) : 1261 – 1266. (in Chinese)
- [13] 张家树, 肖先赐. 混沌时间序列的 Volterra 自适应预测 [J]. 物理学报, 2000, 49(3) : 403 – 408.  
Zhang J S, Xiao X C. Predicting low-dimensional chaotic time series using Volterra adaptive filters [J]. Acta Physica Sinica, 2000, 49(3) : 407 – 408. (in Chinese)
- [14] 李瑞国, 张宏立, 范文慧, 等. 基于改进教学优化算法的 Hermite 正交基神经网络混沌时间序列预测 [J]. 物理学报, 2015, 64(20) : 200506-1 – 200506-13.  
Li R G, Zhang H L, Fan W H, et al. Hermite orthogonal basis neural network based on improved teaching-learning-based optimization algorithm for chaotic time series prediction [J]. Acta Physica Sinica, 2015, 64(20) : 200506-1 – 200506-13. (in Chinese)
- [15] Stark J, Broomhead D S, Davies M E, et al. Takens embedding theorems for forced and stochastic systems [J]. Nonlinear Analysis Theory Methods & Applications, 1997, 30(9) : 5303 – 5314.
- [16] Wei D C, Ching L C, Shih S P, et al. Parameter estimation algorithms for Volterra digital systems [J]. International Journal of Future Computer and Communication, 2017, 6(3) : 115 – 118.

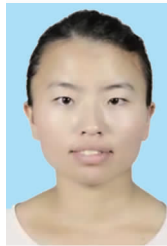
## 作者简介



吴霞 女, 1993 年 9 月出生于贵州安顺, 陕西师范大学博士研究生, 研究方向为非线性动力学, 混沌系统和信号处理等。  
E-mail: wuxia\_9309@snnu.edu.cn



吴晓军 男, 1970 年 12 月出生于陕西凤翔, 陕西师范大学教授, 研究方向为人工智能、大数据、复杂系统理论与应用等。  
E-mail: xjwu@snnu.edu.cn



史素真 女, 1990 年出生于河南濮阳, 河南机电职业学院助教, 研究方向为非线性理论及信号系统。  
E-mail: ssz\_0312@126.com



张其进 男, 1992 年 1 月出生于山东莱芜, 陕西师范大学硕士研究生, 研究方向混沌理论、语音信号处理等。  
E-mail: 513164238@qq.com



张玉梅 (通讯作者) 女, 1977 年 10 月出生于陕西绥德, 陕西师范大学副教授, 研究方向为非线性理论、复杂系统、信号处理等。  
E-mail: zym0910@snnu.edu.cn