第 28 卷第 3 期
2003 年 5 月

声 学 学 报
ACTA ACUSTICA

Vol. 28, No. 3
May, 2003

汉语语音的非线性动力学特征及其降噪应用*

郑能恒 王新龙 马伟菁 眭 群

(南京大学声学研究所 近代声学国家重点实验室 南京 210093)

2001 年 9 月 6 日收到

2002 年 2 月 8 日定稿

摘要 分析了汉语语音的相关维、最小嵌入维数以及重构相图。分析结果表明汉语语音具有混沌特征。根据这些非线性特征可以有效区分汉语中的浊音、清音和随机噪声，从而可以用于语音降噪。介绍了本地投影法混沌语音降噪的原理与算法，并利用该算法对一些典型的元音和辅音进行降噪，获得了较好的降噪效果。

PACS 数: 43.25, 43.50

Nonlinear characteristics of Chinese speech and its application in noise reduction

ZHENG Nengheng WANG Xinlong MA Weijing SUI Qun

(*State Key Lab of Modern Acoustics and Institute of Acoustics, Nanjing University* Nanjing 210093)

Received Sept. 6, 2001

Revised Feb. 8, 2002

Abstract Some invariants of nonlinear dynamics of Chinese speech are analyzed. The results reveal the chaotic characters in Chinese speech and the difference among voiced, unvoiced sound and random noise, which verify the feasibility of denoising Chinese speech using noise reduction algorithms for chaotic data. The results of denoising experiments show good effects on voiced and unvoiced sounds.

引言

语音作为人类交流的主要信息载体，一直是当代信息科学研究的重要内容。目前的语音研究有一个基本的前提假设，即语音是一个短时平稳的物理过程。例如，在语音产生的时域模型中，浊音和清音模型就分别被简化为周期脉冲发生器和随机噪声发生器[1]。这种假设在大多数场合是可行的，但也不可避免的带来某些局限性，限制了语音研究的进一步发展和重大突破。例如，在语音通信中，降噪处理是必不可少的环节。目前已有的语音降噪方法有：(1) 基于傅里叶变换的信号频谱滤波，(2) 基于语音生成模型的自适应滤波，如迭代维纳滤波和卡尔曼滤波，(3) 基于短时谱幅度估计的滤波，如谱相减法及其改进算法，等等[1]。这些方法已经应用于许多降噪场合，但又各有其局限性。频谱滤波要求信号和噪声在频域上可分，因此，对大部分信号来说，用低通或带通滤波器进行滤波时无法除去带内噪声。自适应滤波需要参考通道，在很多场合，例如电话通信中是不实用的。尤其值得指出的是，在上述方法中，都把清音的产生模型简化为随机信号发生器，因此难以有效地消除其中的噪声。

事实上，正如我们即将看到的，语音本质上是一个非平稳的非线性 (混沌) 过程。为了深入探讨语音的物理本质，本文首先对汉语语音的基本音素进行非线性动力学特征 (包括相关维、最小嵌入维以及相空间重构图等) 分析。结果发现，语音实际上是一个混沌信号：元音和部分浊辅音具有明显的低维混沌特征，而清辅音则是一个高维混沌信号，不能简单地认为是随机信号。利用这些非线性特征，可以有效区分汉语语音信号和噪声信号，从而可以借助于混沌

* 国家杰出青年科学基金和国家自然科学基金资助项目

降噪方法进行语音降噪。接着介绍了一种常用的混沌降噪算法。该算法在对元音和辅音的降噪实验中都获得了令人满意的效果。

1 汉语语音的非线性动力学分析

自然界的一切现象本质上是非线性的。线性化只不过是人们在对非线性的本质尚未完全把握、缺乏非线性分析手段时对其作的近似描述。在线性范畴内，人们往往把一些复杂的现象当作随机信号来处理。上述语音信号的时域模型就是一个典型特例。随着非线性动力学理论和计算机硬件技术的发展，研究非线性现象的方法手段不断得到丰富和完善，人们可以更深入地研究分析产生非线性现象的动力学机理，例如语音的发声机理。上世纪 90 年代初，I.Steinecke[2] 、H.Herzel[3] 、D.A.Berry[4] 等人从语音 (主要是浊音) 发音时的声道模型着手，研究声带振动的力学机理，在声带振动轨迹的相图中发现了分岔和混沌现象，并利用相图、频谱等分析手段指出浊音信号具有低维混沌的特征。胡水清等人[5] 也曾研究过汉语语音的相关维，但在文献 5 中，清音还是被当作随机信号。本文将通过相关维、最小嵌入维数以及重构相轨迹等常用的描述混沌信号的特征参量来研究汉语语音，从中可以发现汉语语音的浊音和辅音中都存在不同程度的混沌特征。

首先简要回顾一下非线性动力学的一个重要概念 —— 相空间重构。为了从一个混沌时间序列获得其动力学系统模型的信息， Packard 等人提出了延时坐标重建相空间方案[6]: 对于时间序列 $x(t)$, 构造延时坐标矢量为： $\boldsymbol{X}(t)=\{x(t),x(t+d\Delta t),\cdots,x(t+(m-1)d\Delta t)\}$, 这里 Δt 为采样时间间隔， d 为延迟数据个数， $d\Delta t$ 即延迟时间， m 为嵌入维。 Takens 证明[7], 当嵌入维 $m>2D+1$(D 为吸引子的分数维) 时，重建相空间保持了原吸引子的拓扑特性和几何不变性。这里， $m>2D+1$ 并非是一个必要条件，实际上所需的最小 m 往往小于 $2D+1$ 。例如 Lorenz 系统，根据 Takens 定理， m 至少为 5, 但实际上 $m=3$ 就足够。事实上，最小嵌入维表征了一个动力学系统的复杂性程度。所以，去除维数冗余，找出最小嵌入维在实际应用中显得十分重要。一个常用的方法是根据相关维 (correlation dimension) 来求最小嵌入维：令 m 从 1 逐渐递增构造延时坐标向量 $\boldsymbol{X}(t,m)$, 求相关积分：

$$c(\varepsilon,m)=\frac{1}{N^2}\sum_{i,j=1}^{N}\vartheta\left[\varepsilon-|\boldsymbol{X}(i,m)-\boldsymbol{X}(j,m)|\right], \quad (1)$$

其中：

$$\vartheta(x)=\begin{cases}1, & \text{当 } x>0 \text{ 时},\\ 0, & \text{当 } x\le 0 \text{ 时},\end{cases}$$

取其双对数关系 $\log c(\varepsilon,m)\sim\log\varepsilon$ 中的直线段的斜率作为相关维 $D_2(m)$ 。对于混沌序列， $D_2(m)$ 随 m 递增并在最小嵌入维 M 处达到饱和值 $D_2(M)$, $D_2(M)$ 就是该序列的相关维。而对于随机序列，则相关维不存在饱和现象： $D_2(m)$ 随 m 无限递增[8,9] 。至于延迟时间 $d\Delta t$ 的最优选择应该能把吸引子充分展开在重构的相空间，常用方法有自相关函数法，互信息 (mutual information) 法等[8,9] 。

本文所分析的是汉语中一些基本音素：元音 [a], [o], [e], [i], [u], [ü] 和辅音 [c], [f], [h], [k], [s], [t], [l], [m], [n], [r] 。用 Creative SoundBlaster 声卡采集女性声音，单声道， 24 k 采样频率， 16 位量化。其中，元音长度约为 1.5 s, 36000 个采样数据，辅音长度约为 0.8 s, 20000 个采样数据。

首先求元音 [a] 和辅音 [c] 的相关积分和相关维。图 1 是相关积分的双对数坐标曲线。图 2 是由图 1 求得的相关维随嵌入维变化曲线。作为比较，还计算了正弦序列以及高斯白噪声序列 (数据长度均为 20000) 的相关积分 (图 1(a) 和 1(b)) 和相关维。周期正弦序列的 D_2 很快就达到饱和 (图 2(a)), $m=2$; 白噪声序列正好相反， D_2 随着 m 的增大而无限增大 (图 2(b)) 。语音信号则介于二者之间：元音 [a] 的 D_2 先是随着 m 的增加而增加，当 $m=4$ 时，达到饱和值 $D_2\approx 2.1$(图 2(c)), 辅音 [c] 则要在 $m\approx 28$ 时 D_2 才达到饱和值 (图 2(d)) 。还计算了汉语 6 个元音 [a], [o], [e], [i], [u], [ü] 4 个声调的相关维 (表 1) 和最小嵌入维 (表 2), 以及部分辅音的相关维和最小嵌入维 (表 3) 。元音具有较小的相关维和最小嵌入维，辅音中的浊音 [l], [r] 和元音类似，鼻音 [m], [n] 的相关维较小，但最小嵌入维较高，而清音 (摩擦音和塞音) 虽然具有很高的相关维和最小嵌入维 (20~30 维), 但和随机噪声仍有本质的区别，它仍然是一个维数有限的混沌信号。

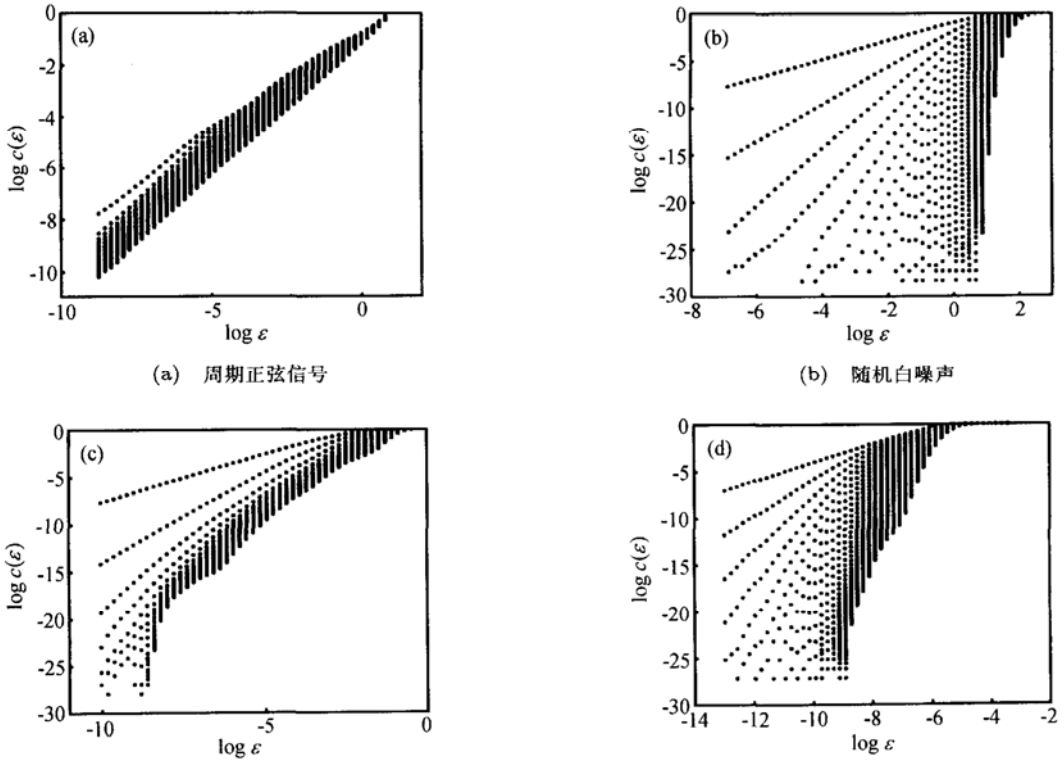
利用求互信息法[9] 容易求得元音 [a] 的最优延迟时间为 $8\Delta t$, 辅音 [c] 的最优延迟时间为 Δt 。现在根据时间序列分别构造延时坐标矢量：

$$\boldsymbol{X}_a(t)=\{x(t),x(t+8\Delta t),x(t+16\Delta t),x(t+24\Delta t)\},$$
$$\boldsymbol{X}_c(t)=\{x(t),x(t+\Delta t),x(t+2\Delta t),\cdots x(t+27\Delta t)\},$$

它们在二维相平面和三维相空间上的相图如图 3 所示。由于 [a] 是低维混沌信号，其重构相图在二维相平面和三维相空间上可以较好地展现其吸引子，而 [c] 的吸引子维数很高，在低维相空间上看起来和噪

声没有很大的区别。

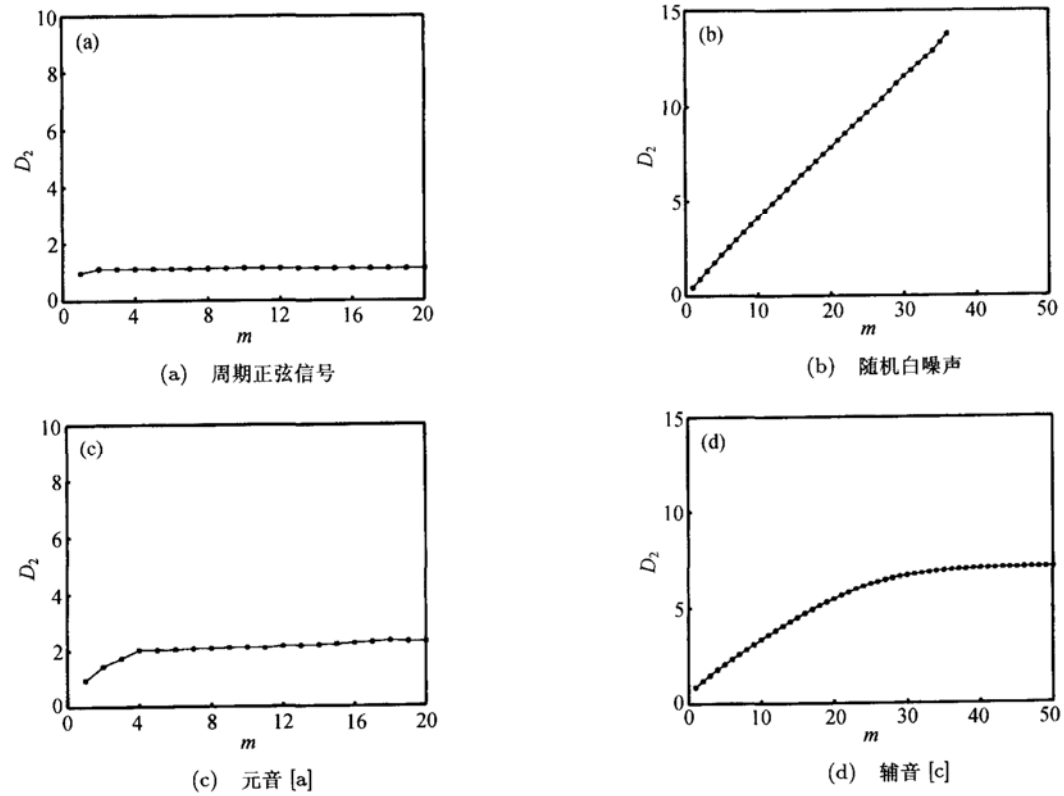
由以上对汉语语音的非线性动力学分析可知，汉语语音中确实存在混沌特征。因此，可以利用混沌降噪的原理进行降噪。



(a)　周期正弦信号　　(b)　随机白噪声

(c)　元音 [a]　　(d)　辅音 [c]

图 1　相关积分的双对数曲线



(a)　周期正弦信号　　(b)　随机白噪声

(c)　元音 [a]　　(d)　辅音 [c]

图 2　相关维 ～ 最小嵌入维曲线

表 1 6 个基本元音及其四调的相关维

	阴平	阳平	上声	去声
[a]	2.19±0.16	2.40±0.08	2.76±0.22	3.02±0.06
[o]	2.31±0.12	2.19±0.17	2.65±0.22	2.71±0.09
[e]	1.84±0.05	2.50±0.15	2.60±0.29	2.20±0.07
[i]	2.15±0.12	2.42±0.06	2.15±0.07	2.45±0.04
[u]	2.25±0.04	2.47±0.11	2.36±0.08	2.42±0.18
[ü]	2.19±0.19	2.48±0.20	2.40±0.14	2.62±0.13

表 2 6 个基本元音及其四调的最小嵌入维

	阴平	阳平	上声	去声
[a]	4	5	5	5
[o]	4	3	3	5
[e]	3	3	4	3
[i]	5	4	4	5
[u]	3	4	3	3
[ü]	4	4	3	3

表 3 部分辅音的相关维和最小嵌入维

	[l]	[r]	[m]	[n]	[k]
D_2	2.50±0.11	2.58±0.17	2.31±0.21	2.24±0.16	4.38±0.09
m	6	6	12	13	20
	[t]	[c]	[f]	[h]	[s]
D_2	3.31±0.06	7.21±0.16	4.28±0.32	4.19±0.20	4.35±0.16
m	22	28	20	28	25

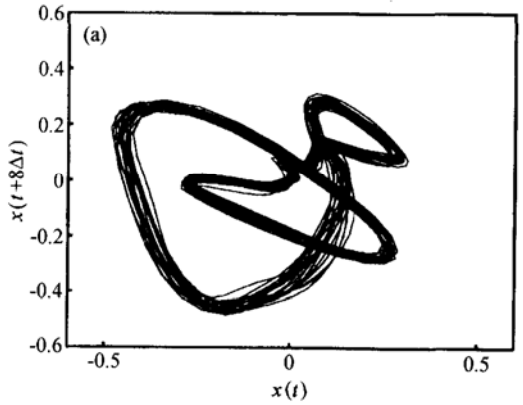


图 3(a) [a] 的二维重构相图

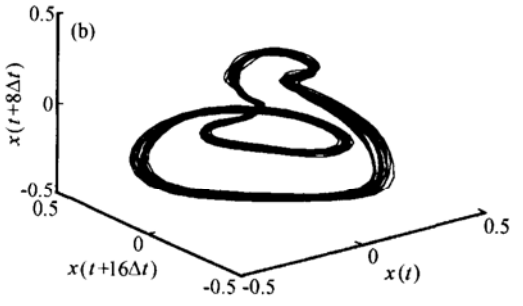


图 3(b) [a] 的三维重构相图

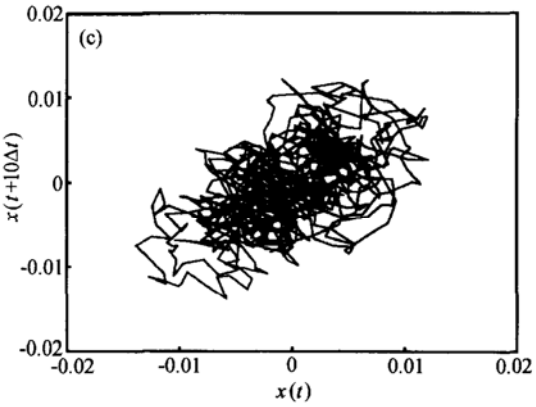


图 3(c) [c] 的二维重构相图

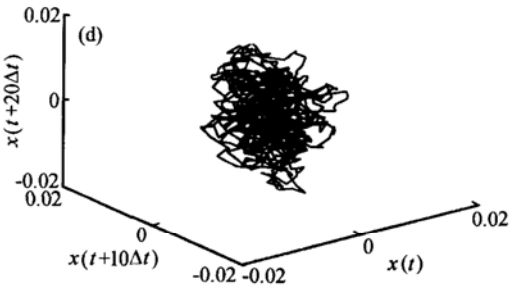


图 3(d) [c] 的三维重构相图

2 混沌语音降噪

2.1 混沌语音降噪的理论与算法

关于混沌语音降噪的研究最早见于 R.Hegger 等人的文章[10]。但到目前为止，国际上的研究都仅限于对低维浊音的降噪。汉语语音的发音与西方语言有所不同，尤其是汉语中的四声调是西方语言所没有的，因此有必要研究混沌语音降噪对汉语的降噪效果。首先介绍一种常用的混沌降噪算法 —— 本地投影法[11]。

对一个 m 维非线性动力学系统 $F: \Gamma \to \Gamma \subset R^m$，其时间序列 $x(n)$ 在 m 维嵌入空间的流形可由如下方程表示：

$$\boldsymbol{X}(n+1) = F(\boldsymbol{X}(n)), \tag{2}$$

其中的函数形式 $F(\cdot)$ 未知。为了求 F，可以在以很小的邻域 $H(n,\varepsilon)$ (ε 是邻域半径) 内将 F 线性化

$$F(\boldsymbol{X}(n)) = A(n)\boldsymbol{X}(n) + B(n), \tag{3}$$

确定 $A(n)$, $B(n)$ 可以用奇异值分解法[12]。通过奇异值分解可以求得 $\boldsymbol{X}$ 的协方差矩阵 $\boldsymbol{C}_x$ 的特征值矩阵 $\boldsymbol{\Sigma}$，它表征了系统能量的 m 个特征方向 (即由 $A(n)$, $B(n)$ 确定的空间方向)，以及各个方向上所包含能量的大小。

对于受噪声 $\xi(n)$ 干扰的混沌序列 $y(n)=x(n)+\xi(n)$, 嵌入向量 $\boldsymbol{Y}(n)$ 将不在由 (2) 给出的流形上，如图 4 所示。因此降噪也就是要使得降噪后 $\boldsymbol{Y}'(n)$ 尽可能地落在该流形上。在小噪声情况下，C_y 的最大的 m 个特征值表征了由 $A(n)$, $B(n)$ 确定的流形的空间方向，其余 $r-m$ 个特征值则表征了噪声的空间方向。可以简单地令这 $r-m$ 个特征值等于 0, 再做上述奇异值分解的逆变换，即可求得降噪后的 $\boldsymbol{Y}'(n)$。

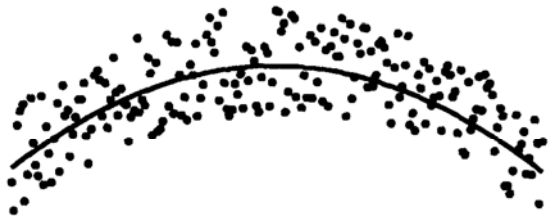


图 4　纯净信号的流形和含噪信号的空间分布

整个降噪算法步骤为：

(1) 对时间序列 $y(n)=x(n)+\xi(n)$, 构造 r 维延时嵌入向量：

$$\boldsymbol{Y}(n)=\{y(n),y(n+d\Delta t),y(n+2d\Delta t),\cdots,y(n+(r-1)d\Delta t)\}.$$

(2) 在 r 维嵌入空间内找到 $Y(n)$ 的 N 个邻点：

$$\boldsymbol{Y}(k)\in H(n),\quad k=1,2,\cdots,N.$$

(3) 计算 $\boldsymbol{Y}(k)\in H(n)$ 的协方差矩阵 $\boldsymbol{C}_y$ 并求其特征值。

(4) 将 $\boldsymbol{Y}(n)$ 投影到由 m 个特征值确定的空间上，可以求得 $\boldsymbol{Y}'(n)$。

(5) 对所有 $\boldsymbol{Y}'(n)$ 中具有相同序号的 $y'(n)$ 做平均，可以得到降噪后的序列 $\{y'(1),y'(2),\cdots y'(n)\}$。

(6) 对 $\{y'(1),y'(2),\cdots y'(n)\}$ 重复一上降噪步骤，直到满足要求为止。

2.2　降噪实验结果

根据上述算法，对部分典型元音和辅音进行降噪实验。这里语音的采样方法与 1 中相同，考虑到在正常语速下单音节的持续时间较短，采样时间为元音 600 ms, 辅音 100 ms。以元音 [a](最小嵌入维 4, 延迟时间 $8\Delta t$) 和辅音 [t](最小嵌入维 22, 延迟时间 Δt) 为例，分别加上 0, 5, 10, 15, 20 dB 的高斯白噪声，用上述混沌语音降噪算法进行降噪。作为比较，用谱减法对同样的含噪语音进行降噪实验。降噪前后的信噪比、Itakura 距离和主观评价 (MOS 分) 如表 4 和表 5 所示。这里信噪比和 Itakura 距离的计算采用 Hansen 的算法[13]。为保证结果的可比性，对所有降噪后的语音均先进行幅值调整，使其最大幅值与纯净语音相等，再进行相应的计算。采用非正式的主观评价，试听者 12 人，在较为安静的实验室内进行。

表 4　元音 [a] 的降噪实验结果

含噪语音信噪比 /dB	降噪后信噪比 /dB 混沌降噪 / 谱减法	Itakura 距离 降噪前 / 混沌降噪 / 谱减法	MOS 分 降噪前 / 混沌降噪 / 谱减法
20	23.41/23.15	5.24/3.86/3.88	3.05/4.14/4.23
15	19.74/20	6.09/4.22/3.91	2.3/3.67/3.5
10	16.33/16.7	6.91/4.92/4.30	1.83/3.67/3.75
5	13.74/12.18	7.66/6.0/6.23	1.5/3.17/2.25
0	9.10/5.95	8.32/6.70/7.36	1.17/2.67/2

表 5　辅音 [t] 的降噪实验结果

含噪语音信噪比 /dB	降噪后信噪比 /dB 混沌降噪 / 谱减法	Itakura 距离 降噪前 / 混沌降噪 / 谱减法	MOS 分 降噪前 / 混沌降噪 / 谱减法
20	22.26/22.64	2.62/1.53/1.38	3/3.5/3.5
15	18.64/17.44	3.31/1.88/1.81	2.67/3.6/3.4
10	15.00/13.75	4.13/2.04/2.47	2.5/3.25/3.08
5	10.82/9.10	4.99/2.47/3.42	2.4/3.0/2.72
0	6.72/3.73	5.81/4.31/3.86	2.3/2.92/2

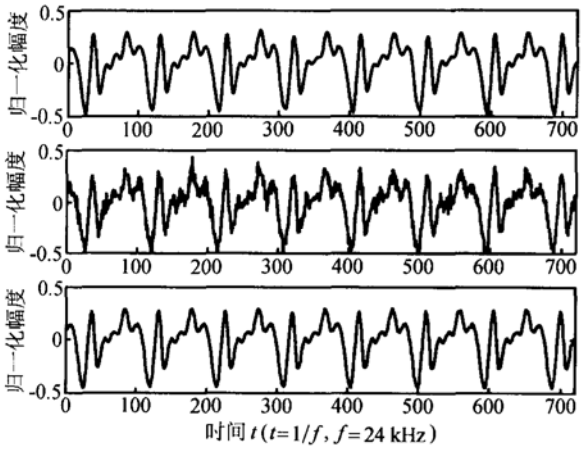


图 5 30 ms 元音 [a] 的波形从上到下依次为纯净语音、信噪比为 10 dB 的语音、混沌降噪后的语音

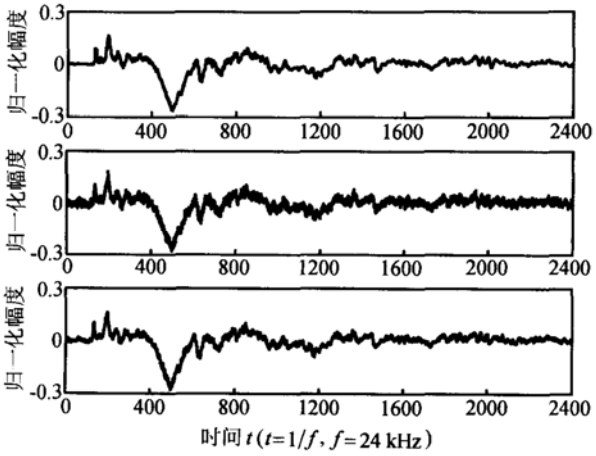


图 6 100 ms 辅音 [t] 的波形从上到下依次为纯净语音、信噪比为 10 dB 的语音、混沌降噪后的语音

比较表 4 和表 5, 发现:

(1) 对于元音, 当信噪比高于 10 dB 时, 两种方法降噪效果差不多, 而信噪比低于 10 dB 时, 混沌降噪效果优于谱减法。

(2) 对于辅音, 当信噪比低于 20 dB 时, 混沌降噪效果优于谱减法。

(3) 混沌降噪算法对元音的降噪效果优于辅音。

谱减法的基本原理是语音和噪声的能量相减, 因此当噪声能量很高时 (0 dB) 或者语音的频谱分布类似于噪声时 (辅音), 其降噪效果较差。而混沌降噪算法是根据二者在相空间分布的不同特点 (维数不一样) 来降噪的。根据本文上述, 元音、辅音和噪声在相空间的分布有明显的差别, 因此该降噪算法能获得较好的降噪效果。至于实验结果 (3), 据 2.1 所述, 首先要把含噪语音重构到r维相空间, 然后投影到 m 维子空间。 m 是纯净语音吸引子的维数, 一般是固定的。但是 r 的选择则要仔细考虑, 理论上最好要等于噪声的吸引子维数 (无限大), 这显然不现实, r 太大将极大地增加计算的复杂性。在可行的范围内, r 越大越好, 实验表明当 $r/m < 2$ 时, 降噪效果将急剧下降。实验中, 元音 [a] 的 $m=4$、$r=21$; 辅音 [t] 的 $m=20$、$r=41$, 由于 [t] 的 $r/m \approx 2$ 较小导致降噪效果下降。另一方面, 算法中很重要的一点是邻域的确定。邻域半径 ε 既要足够小, 以满足流形的线性化近似要求, 同时又要足够大, 以满足统计特性要求。对于辅音来说, 由于其持续时间很短, 数据太少, 难以同时满足这两个条件, 这也影响了降噪效果。与谱减法相比, 混沌降噪算法的缺点是其运算量较大。

图 5 和图 6 分别给出了 [a] 和 [t] 的纯净语音、信噪比为 10 dB 的带噪语音和混沌降噪后语音的某一段时域波形图。很明显, [a] 降噪后波形有明显的改善, [t] 降噪后的波形改善虽然不如 [a], 但还是可以看到大部分噪声被消除了。

3 结论

本文分析了汉语基本音素的相关维、最小嵌入维和重构相图。分析结果表明汉语语音具有混沌信号的特征。元音和浊辅音可以看作是低维混沌信号, 其重构相图在二维相平面和三维相空间上可以较好地展现吸引子。对于清音, 虽然在三维相空间上其重构相图看起来和噪声没有很大区别, 但从计算结果可知, 其相关维和最小嵌入维都是有限的, 在 20~30 维的相空间上可展现其吸引子, 而理论上噪声的相关维和最小嵌入维都是无限大的, 在可以计算的范围内也没有看到饱和的趋势。因此利用语音的混沌特征参数可以有效区分浊音、清音和随机噪声, 从而可以利用混沌降噪方法进行降噪。对部分汉语元音和辅音分别用混沌语音降噪方法和谱减法进行对比实验。结果表明, 混沌降噪方法对于辅音和低信噪比元音的降噪效果都要好于谱减法。该算法的不足之处是运算量较大。但随着计算机硬件技术的发展, 对语音非线性特征的研究必将促进语音处理技术的进一步发展。

参 考 文 献

1 杨行峻, 迟惠生. 语音信号数字处理. 北京: 电子工业出版社, 1995

2 Steinecke I, Herzel H. Bifurcations in an asymmetric vocal-fold model. *J. Acous. Soc. Am.*, 1995; **97**(3): 1874—1884

3 Herzel H. Bifurcations and chaos in voiced signals. *Appl. Mech. Rev.*, 1993; **46**(2): 399—413

4 Berry D A, Titze I R, Herzel H, Krischer K. Interpretation of biomechanical of normal and chaotic vocal fold oscillations with empirical eigenfunctions. *J. Acous. Soc. Am.*, 1994; **95**(6): 3595—3604

5 胡水清, 徐 敢, 杜功焕. 语音的短时相关维分析. 声学技术, 2000; **19**(3): 150—151

6 Packard N H, Crutchfield J P, Farmer J D, Shaw R S. Geometry from a time series. *Phys. Rev. Lett.*, 1980; **45**(9): 712—716

7 Takens F. Detecting strange attractors in turbulence. *Dynamical Systems and Turbulence.*, 1981; **898**(1): 365—381

8 Eckmann J P, Ruelle D. Ergodic theory of chaos and strange attractors. *Rev. Mod. Phys.*, 1985; **57**(3): 617—656

9 Abarbanel H D I, Brown R, Sidorowich J J. The analysis of observed chaotic data in physical systems. *Rev. Mod. Phys.*, 1993; **65**(4): 1331—1392

10 Hegger, Kantz H, Matassini L. Denoising human speech signals using chaoslike features. *Phys. Rev. Lett.*, 2000; **84**(14): 3197—3200

11 Kostelich J, Schreiber T. Noise reduction in chaotic time-series data: A survey of commom methods. *Phys. Rev. E.*, 1993; **48**(3): 1752—1763

12 Moor B D, Dooren P V. Generalization of the singular value and QR decompositions. *SIAM J. Matrix Anal. Appl.*, 1993; **13**(3): 993—1014

13 Hansen J H L, Pellom B. An effective quality evaluation protocol for speech enhamcement algorithms. ICSLP-98: Inter. Conf. on Spoken Language Processing, 1998(7): 2819—2822