

# 语音的短时关联维分析

胡水清, 徐 歆, 杜功煊

(南京大学声学研究所, 近代声学国家重点实验室 南京 210093)

中图分类号: H017.0423

文献标识码: A

## The short correlation dimension analysis of speech

HU Shui-qing, XU Xin, DU Gong-huan

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

### 1 引言

众所周知, 语音发声系统实质上是一个非常复杂的时变非线性系统, 随着对混沌研究的深入, 逐渐形成了一套非线性动力学分析方法, 主要包括计算维数、Liapunov 指数、Kolmogorov 熵以及短时非线性预测等方法。在确定系统的性态、区分随机的“白噪声”和确定性的“似随机”混沌信号上, 这些非线性动力学方法是传统的线性的信号处理方法所无法替代的。在语音的发声机制上, 非线性动力学研究也取得一系列成果并被广泛关注<sup>[1-5]</sup>。本文讨论了非稳态和静态对于非线性动力学方法的影响, 用短时关联维的方法, 对汉语语音的音素、音节和短语进行分析。结果表明, 短时关联维方法对清音与元音边界划分清晰有效, 对传统的线性语音处理方法能提供一定的有用信息。同时, 本文还用短时非线性预测方法进行语音建模进行了初步的探索。

### 2 理论

由于语音发声系统的内在状态变量是未知的, 我们只能通过语音信号进行相空间重构, 而考虑到语音信号的时变性, 我们对语音信号  $x(t_0), x(t_1), x(t_2), \dots$ , 进行分段处理, 对于第  $k$  段的语音信号, 可重构为:

$$X(t_0 + k\tau) = [x(t_0 + k\tau), x(t_0 + k\tau - \Delta t), \dots, x(t_0 + k\tau - (m-1)\Delta t)]^T$$

$$\vdots$$

$$X(t_N + k\tau) = [x(t_N + k\tau), x(t_N + k\tau - \Delta t), \dots, x(t_N + k\tau - (m-1)\Delta t)]^T$$
(1)

其中  $m$  为嵌入维,  $\tau$  为两端间的移动间隔。在嵌入维  $m > 2D + 1$  情况下 ( $D$  为系统维数), 这种方法

可成功重构动力学系统的原始状态空间, 重构出的相空间反映该段语音发声系统的动力学特性。

关联维可以通过计算关联积分 (Correlation Integral) 求得。通过以上方法重构的相空间, 我们可以得到如下的关联积分表达式:

$$C(N, r, t_0 + k\tau) = \frac{1}{N(N-1)} \sum_{i=1}^N \sum_{j=1}^N \theta(r - \|X_i(t_0 + k\tau) - X_j(t_0 + k\tau)\|), \quad k = 0, 1, \dots, (2)$$

$$\text{其中, } \theta(x) = \begin{cases} 1 & x > 0 \\ 0 & x \leq 0 \end{cases}$$

通过 (2) 式求得的即是第  $k$  段语音段相对应的关联维, 通过计算不同语音段的关联维, 我们就可以得到  $D_2(t_0 + k\tau)$  随  $k\tau$  的时间演化曲线, 即关联维—时间曲线, 它反映了语音发声系统的非线性动力学特性的时间演化。

### 3 数值计算结果

我们选取语音中间较平稳的部分, 用通过相空间重构计算关联维方法对正常语速汉语语音音素进行分析。图 1 为元音 /e/、摩擦音 /c/ 和噪声的关联维曲线。从中可以看到, 元音 /e/ 的关联维随嵌入维的增长是收敛的, 收敛到的关联维数都较低, 一般都在 3 维以下; 相反的, 摩擦音 /c/ 和噪声的关联维都随嵌入维数的增加而增加, 看不出明显的收敛趋势。在不同声调的 5 个汉语单元音 /a/、/e/、/i/、/u/、/ü/ 的 112 个样本和 6 个摩擦音 /c/、/ch/、/s/、/sh/、/z/、/zh/ 的 43 个样本中, 我们发现元音的关联维数较低, 一般在 1—3 维之间, 元音发声系统是一个低维系统, 系统处于极限环或者准周期状态。而摩擦音的关联维数较高, 摩擦音发声系统是一个复杂的高维系统, 符合元音与摩擦音的不同的发声机理。由于元音和摩擦音的关联维有较大区别, 我们可以关联维方法区分元音和摩擦音。

收稿日期: 2000-05-21; 修订日期: 2000-05-15

作者简介: 胡水清 (1975-), 男, 硕士研究生。

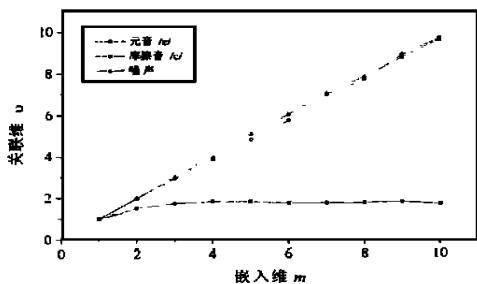


图 1

静态是非线性动力学方法的一个基本要求,即时间序列必须有足够多的周期数,才能保证重构的相空间与原始相空间微分同胚,这就要求用于计算的语音数据足够长;然而,语音发声系统是一个时变的非线性系统,而非线性动力学方法只能对于稳态和相对稳态的系统进行处理,因此语音数据又不能太长,以保证用于计算的语音数据处于准稳态。因此选取语音的一部分,或者全部用于计算是不合适的,因此我们采用短时关联维的办法,计算关联维-时间曲线以反映出语音发声系统随时间变化的动力学特性。通过反复试验,我们发现选择用语音数据段长度为 50ms 能较好的满足静态和稳态条件。

图 2 为语音 /c-i/ 的波形图和关联维-时间曲线。从中可看出,由于摩擦音和元音的不同发声机理,短时关联维方法能够清晰有效的切分摩擦音 /c/ 和元音 /i/。在我们计算的 100 个音节中,短时关联维方法均可以有效的切分元音和清音。

四声是汉语的独特现象,由于不同声调的元音的关联维都是低维的,用短时关联维方法区分元音和清音部首语调的影响。

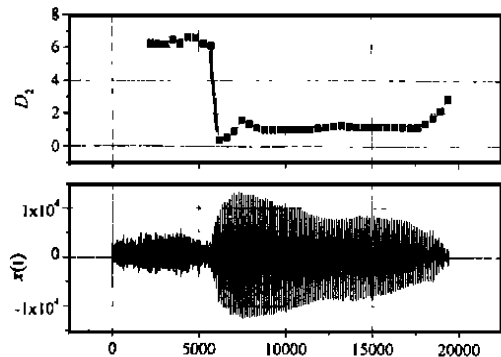


图 2

图 3 自上而下分别为短语 /ming-tai-zhu/ 的波形图、短时过零率、短时能量和短时关联维的结果。从图中可看出,短时关联维的方法可清晰有效地区分语音和清音,并准确的确定元音的开始。短时关联

维的分析有望为区别清浊提供一些有用信息,当然有关这方面的研究尚需进一步展开。但是关联维分析能提供汉语语音的非线性动力学特性的信息,无疑对语音深层次的研究和应用将是十分有意义的。

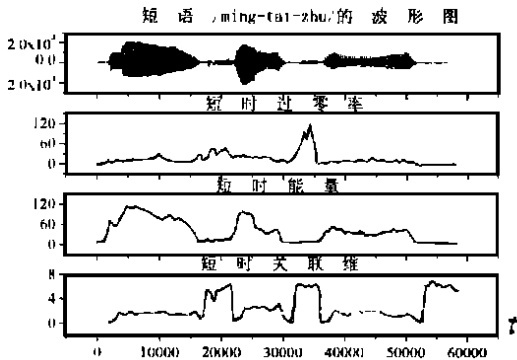


图 3

局域法的短时非线性预测方法,相对于线性预测方法有更好的预测效果,我们初步进行了语音波形预测的实验,实验发现,用非线性预测方法对语音进行预测可以得到较好的预测结果。图 4 为鼻音 /m/ 的预测波形和原始语音波形的对比,从中可以看出,短时非线性预测的方法可以有效的对语音进行预测,在语音建模,语音合成中有一定的前景。

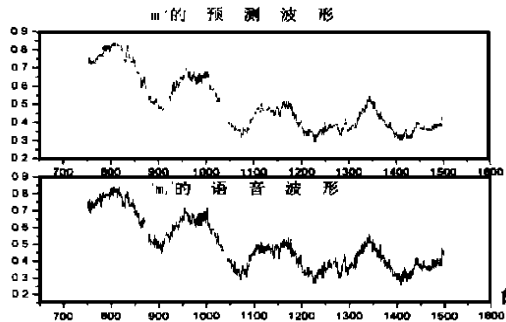


图 4

#### 参考文献

- [1] A. Behrman, Gopal and local dimension of vocal dynamics[J]. J. Acoust. Soc. Am., 1999, 105: 432-443.
- [2] S. S. Narayanan and A. A. Alwen. A nonlinear dynamic system analysis of fricative consonants[J]. J. Acoust. Soc. Am., 1995, 97: 2511.
- [3] M. Bandbrook, S. McLaughlin, and I. Mann. Speech characterization and synthesis by nonlinear methods [J]. IEEE Trans on Speech Audio Process, 1999, 7 (1): 1-17.
- [4] P. Maragos. Fractal dimensions of speech sounds: computation and application to automatic speech recognition [J]. J. Acoust. Soc. Am, 1999, 105(3): 1925-1932.
- [5] S. Q. Hu, Y. Zhang, Y. M. Hua and G. H. Du. Nonlinear dynamic characteristic analysis of speech for Chinese [J]. Chinese Journal of Acoustics(in English), to appear in 2000.