基于定量递归分析的清浊音判决

闫润强　　朱贻盛
(上海交通大学生命科学学院　上海　200030)

摘　要：在语音信号处理中，清浊音判决的准确与否直接关系到后续语音处理的质量。该文通过分析不同的语音音素动力学物理模型在其递归图上的表现，统计定量递归分析中确定性和归一化最长对角线这两种特征参数，得到清浊音的显著差异。设定灵活合理的阈值判决语音信号的清浊音，得到了良好的试验结果。和其他传统判决方法比较，误判率有明显降低，为语音特征提取和识别研究提供了新的途径。

关键词：清浊音判决；语音动力学；递归图；定量递归分析；特征提取
中图分类号：TN912.3　　**文献标识码**：A　　**文章编号**：1009-5896(2007)07-1703-03

Voiced/unvoiced Decision Based on Recurrence Quantification Analysis

Yan Run-qiang　　Zhu Yi-sheng
(*College of Life Science & Biotechnology*, *Shanghai Jiao Tong University*, *Shanghai* 200030, *China*)

Abstract: Voiced/unvoiced decision is an important component in speech signal processing. In this paper, different topological structures in Recurrence Plots (RPs) are described for the different physical models of speech production. By statistically analyzing the determinism and the normalized maximal length of diagonal structures acquired from Recurrence Quantification Analysis (RQA), a flexible and efficient decision framework is proposed. Comparing with some traditional methods, the proposed algorithm has lower wrong decision rate. The method provides a new way for feature extraction and speech recognition.

Key words: Voiced/unvoiced decision; Speech dynamics; Recurrence plot; Recurrence quantification analysis; Feature extraction

1　引言

在语音信号处理中，清浊音判决的准确与否直接关系到后续语音处理的质量。目前的研究成果表明语音的产生过程具有多样化的非线性动力学特征[1]，其时域的不规则和频域的宽频谱特性，使得传统的AMDF (平均幅度差函数法)和AUTOC(自相关法)难以获得精确的清浊音判决结果。如何利用语音的非线性特性提取音素的特征参数，并有效地进行清浊音判决是语音识别的重要课题。

非线性语音信号研究中，音素作为最小的发声单元，产生音素的非线形动力学物理模型有 3 种[2]，(1)振荡模型(浊音)、(2)湍流源模型(清音)、(3)混合模型。根据混沌理论，不同的动力学系统在其相空间上表现为不同的吸引子结构。为了描述这种结构，Eckmann等[3]引入了一种从二维图形上观察非线性时间序列内部动力学机理递归的分析方法：递归图法(Recurrence Plot, RP)。为了量化递归图中表现出来的系统递归现象，Zbilut和Webber[4]提出了定量递归分析(Recurrence Quantification Analysis, RQA)。

本文分析了不同的语音音素非线性动力学物理模型在其递归图上的表现，统计定量递归分析中确定性和归一化最长对角线这两种特征参数，得到清浊音的显著差异。设定灵活合理的阈值进行语音信号的清浊音判决，对比传统方法，给出结论。

2005-12-22 收到，2006-08-29 改回
国家 973 计划(2005CB724303)资助课题

2　递归图和定量递归分析

2.1 递归图

对于混沌系统，混沌吸引子在它的相空间中体积是有限的。因此构成吸引子的非稳态轨道，在有限的吸引子空间中不断的近似逼近又分叉远离，轨道状态的这种递归现象成为混沌系统的基本特征之一。递归图法[3]就是通过从系统中提取的时间序列重现系统动力学递归行为的方法。根据Takens[5]的相空间重构理论，选择合适的嵌入维数 m 和延迟时间 τ，可以将一维的非线性时间序列 $\{x(i),\ i=1,2,\cdots\}$ 重构出向量 $\boldsymbol{X}_i=[x(i),x(i+\tau),\cdots,x(i+(m-1)\tau)]$。这些有时间标记的向量序列 $\{\boldsymbol{X}_i,\ i=1,2,\cdots,N\}$ 构成了系统的 m 维相空间轨道。用这些相空间上的点作为行和列构成 $N\times N$ 的矩阵递归图，图中的每个节点由对应的行、列向量点之间的距离来描述：

$$R_{i,j}=\Theta\left(\varepsilon-\left\|\boldsymbol{X}_i-\boldsymbol{X}_j\right\|\right),\ \ i,j=1,2,\cdots,N \tag{1}$$

公式(1)中 ε 为预先确定的阈值常数，表示临界距离。符号 $\|*\|$ 表示取向量的 Euclidean 范数。$\Theta(x)$ 是 Heaviside 函数。当 $R_{i,j}$ 的值为 1 时，在递归图中 (i,j) 位置上表示为一个黑点；当值为 0 时，(i,j) 位置上表示为一个白点。因此，递归图将一个 m 维相空间的轨道关系映射到了一个二维图上。

汉语普通话语音从发音上可以分为 31 种基本音素单元{a, b, c, ch, d, e, er, f, g, h, i, i1, i2, j, k, l, m, n, ng, o, p, q, r, s, sh, t, u, ü, x, z, zh}。在分析复杂多样的语音动力学信号的时候，由于没有动力学方程等先验知识，本文分别采用邻接误差法和平均互信息法[6]统计得到重构相空间嵌入维数和延迟时间。

图 1 给出了语音信号最小嵌入维数的选择，横轴表示嵌入维数。A1 画出 6 种音素{b, f, l, n, o, ü}的邻接误差曲线，纵轴表示误差比。A2 为 31 种基本音素单元最小嵌入维数统计表，纵轴表示音素出现个数， $m=5$ 时出现个数最多(14 种音素)。

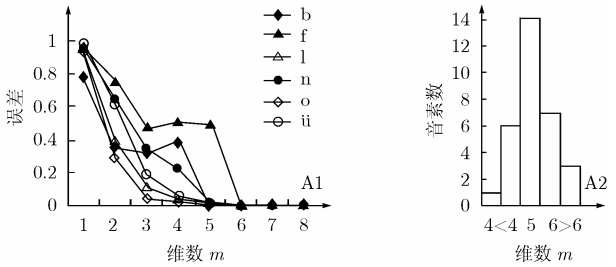

图 1 语音信号最小嵌入维数的选择

图 2 给出了语音信号最小延迟时间的选择，横轴表示延迟时间。A1 画出相应音素(如图 1)的平均互信息曲线，纵轴为互信息值。A2 为最小延迟时间统计表，纵轴表示音素出现个数， $\tau=3$ 时出现个数最多(10 种音素)。

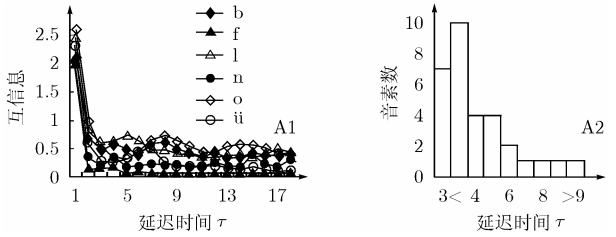

图 2 语音信号最小延迟时间的选择

由图 1 和图 2 可以得到 $m=5$ 和 $\tau=3$ 为汉语普通话语音信号重构相空间参数的最佳选择。同时，临界距离 ε 的选择对于递归图也非常重要， ε 选取过大和过小都会造成对非线性系统错误的递归行为解释。通过实验对比， $\varepsilon=\sigma$ 作为语音信号的临界距离比较合适，其中 σ 为时间序列方差。

递归图的研究包括宏观模式和微观模式两个方面，宏观模式是从全局拓扑结构上观察递归图特征，分为均态、周期、漂移和突变 4 种[3]。微观模式主要研究的是递归图中基本图形(线段和点)的分布以及图形代表的动力学意义，这些图形主要包括平行于主对角的线段、竖直(水平)线段和孤立点[7]。从系统动力学行为角度上讲，对角线的出现表明该时间段的轨道持续逼近现象，竖直(水平)线段表明轨道运行的缓慢变化，孤立点表明空间点的瞬间逼近后迅速分离。

汉语普通话元音单元都是浊音，辅音单元除舌尖后擦音 r, 鼻音 n, m, ng 和边音 l 这 5 个浊音外，其他都是清音。图 3 给出了 6 种具有代表意义的汉语单音素语音波形及其递归图，图 3 中的 A1 和 A2 分别表示元音 e 的语音波形及其递

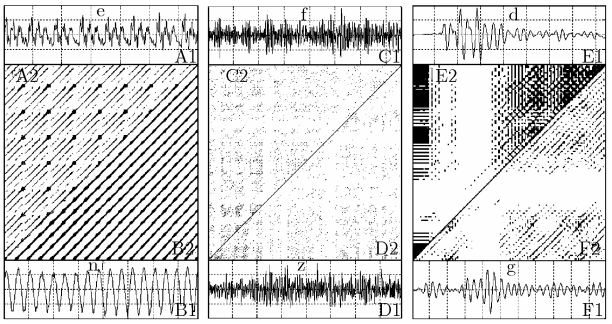

图 3 6 种音素的语音波形及其递归图

归图；B1 和 B2 表示鼻音 n；C1 和 C2 表示擦音 f；D1 和 D2 表示擦音 z；E1 和 E2 表示爆破音 d；F1 和 F2 表示爆破音 g。由于递归图关于主对角线的对称性，本文只画出了其中的一半(左上三角或右下三角)。音素波形时间序列采样频率 16000Hz，幅值范围[−0.5, 0.5] ，递归图参数 $m=5$ ， $\tau=3$ 和 $\varepsilon=\sigma$ 。

气流通过声带，引起声带的震动产生了浊音，声带中间的空隙(声门)的周期性张合决定了浊音的基频周期的变化。从图 3 中的 A2 和 B2 可以看到规则的等间距平行于主对角线的长线段，这种周期模式的递归图揭示了浊音的振荡模型特征。和浊音不同，清音在产生过程中声带没有震动，气流在发生器官管腔的约束下产生湍流现象。在均态模式的递归图 C2 和 D2 中，散点均匀的分布，很少出现较长的线段结构(本文中线段长度定义为递归图中连续黑点的个数)。爆破音由于闭塞段的存在和压力的突然释放使得这些发音具有很大的不稳定性，其递归图中往往出现大区域的白块，如 E2 和 F2 所示。

2.2 定量递归分析

递归图给出了动力学系统相空间轨道运行关系的二维映射图，为了量化递归图中表现出来的系统递归现象，Zbilut 和Webber[4]提出了定量递归分析。在他们的论文中给出了 5 种量化参数：递归度、确定性、最长对角线、熵和趋势。不同的量化参数描述了系统不同的动力学行为。其中，确定性描述了轨道周期递归的程度，值越大表明确定性越强，相反则表明随机性越强；最长对角线(主对角线除外)和邻接轨道的分离速率有关，因此从另一个角度反映了最大Lyapunov 指数的大小。

递归度计算公式如下：

$$RR=\frac{1}{N^2}\sum_{i,j=1}^{N}R_{i,j} \tag{2}$$

确定性计算公式如下：

$$\mathrm{DET}=\frac{\sum_{l=l_{\min}}^{N-1}lP(l)}{\sum_{i,j=1}^{N}R_{i,j}} \tag{3}$$

其中 $P(l)$ 为对角线长度为 l 的出现概率， $l_{\min}$ 为对角线长度的统计初值， $2\le l_{\min}\le N-1$ 。

针对分析信号的长度不同，归一化的最长对角线计算公式如下：

$$\mathrm{MAX}=\frac{L_{\max}}{N-1} \tag{4}$$

表 1 给出了图 3 中 6 音素的递归度、确定性和归一化最长对角线。可以看出浊音 e 和 n 的确定性和归一化最长对角线都远大于擦音 f 和 z。由于爆破音 d 和 g 的发音不稳定特征，其在确定性参数上也表现出较大值，然而其递归图上的大区域白块使得归一化最长对角线比浊音小的多。

表 1　6 种音素(同图 3)的递归度、确定性和归一化最长对角线

音素	递归度	确定性	归一化最长对角线
e	0.031	0.844	0.747
n	0.134	0.987	0.999
f	0.031	0.145	0.021
z	0.012	0.161	0.019
d	0.233	0.623	0.225
g	0.064	0.446	0.138

3　清浊音判决的实现和性能分析

由图 3 和表 1 可以看出振荡模型的浊音表现出很强的确定性，而湍流源模型的清音则表现出类随机的特性。作为衡量动力学系统的相空间演化轨道变化快慢程度的最大Lyapunov指数，统计上表明清音远大于浊音，这与语音的发声机理相吻合[8]。因此，本文利用定量递归分析中的确定性和归一化最长对角线作为判决语音信号清浊音的特征参数。

为了得到合适的语音信号清浊音判决阈值，在安静环境下(信噪比为 30dB)采集多名成年男女发音语音，采样频率16000Hz，量化精度 16bit，样本的清浊音分类通过人工对波形的观察和对声音的感知获得。以 20 ms 时间窗切分清浊音段，分别得到 2000 帧和 1000 帧作为阈值训练集和判决测试集。训练集和测试集中，清音和浊音各占 50%，包含 31 种基本音素单元，且分布平衡。图 4 给出了训练集中清浊音语音帧确定性和归一化最长对角线参数统计直方图。

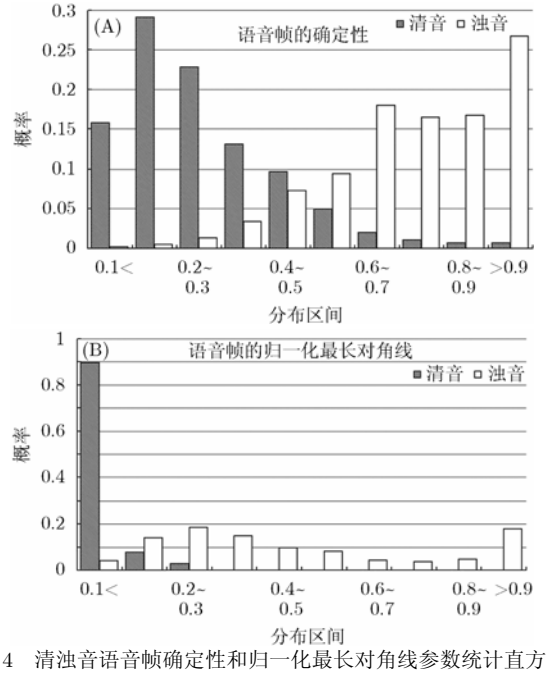

图 4　清浊音语音帧确定性和归一化最长对角线参数统计直方图

由于训练集中孤立的语音帧从连续语音中切分得到，因此必然存在浊音音素间不稳定的过渡段帧(清音不存在这种现象)，图 4 表明连续语音信号的非线性动力学特性改变对确定性的影响较小，而使得浊音的归一化最长对角线分布区域发散。同时，图 4 揭示全部清音的归一化最长对角线参数小于 0.3。

清浊音判决错误包括浊音误判为清音(V/UV)和清音误判为浊音(UV/V)两种情况。根据统计图 4，若采用单一阈值 MAX=0.3，则得到 V/UV 个数为 368，UV/V 个数为 0，此时清浊音判决总误判率为 18.40%。而若采用单一阈值 MAX=0.2，V/UV 个数为 183，UV/V 个数为 27，此时清浊音判决总误判率为 10.50%。可以看出采用单一的归一化最长对角线阈值判决语音清浊音误判率较高，还不能达到要求。

图 5 给出采用确定性和归一化最长对角线联合特征参数判别孤立帧语音信号清浊音流程，其中 S 表示语音帧，Thmax、Thdet 分别为归一化最长对角线阈值和确定性阈值，V 表示浊音，UV 表示清音。

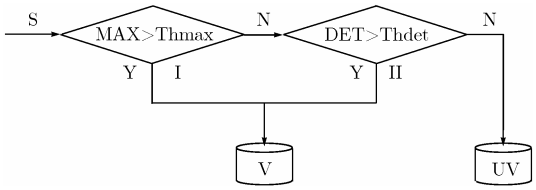

图 5　采用确定性和归一化最长对角线联合特征参数判别语音帧清浊音流程图

针对训练集，方法 1：取 Thmax=0.3，Thdet=0.6。第Ⅰ阶段结束后有 368 帧浊音、1000 帧清音进入第Ⅱ阶段判断过程，第Ⅱ阶段结束后 V/UV 个数为 86，UV/V 个数为 47，此时清浊音判决总误判率为 6.65%。方法 2：取 Thmax=0.2，Thdet=0.6。第Ⅰ阶段结束后有 183 帧浊音、973 帧清音需要判断，第Ⅱ阶段结束后 V/UV 个数为 44，UV/V 个数为 73，此时清浊音判决总误判率为 5.85%。可见联合确定性阈值相比单一阈值判决，误判率有了很大程度的降低。同时方法 1 和方法 2 分别侧重清、浊音判决要求，使得本算法具有很大的灵活性。

由于递归图和定量递归分析技术基于分析语音信号的非线性动力学特性，因此在满足语音帧统计量充分的条件下，具有话语人无关、语种无关特征。对于测试集，采用同样的确定性和归一化最长对角线阈值，流程如图 5。孤立帧语音信号清浊音判决统计结果见表 2，可见得到了相当高的判别准确率，测试集相比训练集，误判率变化不大。

表 2　基于定量递归分析的语音帧清浊音判决结果

样本	判决法	浊音误判率 (%)	清音误判率 (%)	总误判率 (%)
训练集	方法 1	4.30	2.35	6.65
	方法 2	2.20	3.65	5.85
测试集	方法 1	4.35	2.55	6.90
	方法 2	2.30	3.65	5.95

AUTOC和AMDF是传统的进行清浊音判决的方法，表

3 列出了使用这两种方法进行清浊音判决的统计结果[9]。将这些数据与本文介绍的方法所得结果进行比较可以发现，误判率有明显的降低。

表 3 两种传统清浊音判决算法的性能比较

判决法	浊音误判率 (%)	清音误判率 (%)	总误判率 (%)
AUTOC	2.7	6.9	9.6
AMDF	5.7	8.4	14.1

4 结束语

语音信号具有多样化的非线形动力学特性，本文分析了不同的汉语语音音素动力学物理模型在其递归图上的表现，如周期模式的浊音振荡模型、均匀模式的清音湍流源模型和爆破音的突变模式。通过统计定量递归分析中确定性和归一化最长对角线这两种特征参数，得到清浊音的显著差异。给出两种不同的分别侧重于清音、浊音的判决框架，都得到了良好的试验判决结果。和传统的自相关法、平均幅度差函数法比较，误判率有明显的降低，为语音特征提取和识别研究提供了新的分析手段。

本文方法的不足之处在于计算量较大，主要是语音帧构造递归图过程中递归点的计算，对于 M 点的一帧语音，重构相空间点的个数为 N ($N = M - \tau \times (m-1)$)，由于递归图关于其主对角线的对称性，构造递归图计算复杂度约 $o\left((N \times m)^2/2\right)$，确定性和归一化最长对角线计算量相对较小约 $o(N)$，因此本文中算法总的计算复杂度约为 $o((N \times m)^2/2) + o(N)$，比自相关法、平均幅度差函数法的复杂度 $o(N^2)$ 要高。如何降低算法的计算复杂度、提高处理的速度和降低清浊音的误判率是本课题组进一步研究的重点。

参 考 文 献

[1] Faundez-zanuy M, Kubin G, and Kleijin W B, *et al.*. Nonlinear speech processing: Overview and applications. *Control and Intelligent Systems*, 2002, 30(1): 1–10.

[2] Kleijn W B and Paliwal K K. Speech coding and synthesis. Elsevier: Amsterdam, 1995: 560–562.

[3] Eckmann J P, Kamphorst S O, and Ruelle D. Recurrence plots of dynamical systems. *Europhysics Letter*, 1987, 4(1): 973–977.

[4] Zbilut J P, Webber J, and Charles L. Embeddings and delays as derived from quantification of recurrence plots. *Physics Letters A*, 1992, 171(1): 199–203.

[5] TAKENS F. Detecting strange attractors in turbulence. *Lecture Notes in Mathematics*, 1981, 898(1): 366–381.

[6] Hegger R, Kantz H, and Schreiber T. Practical implementation of nonlinear time series methods: The TISEAN package. *Chaos*, 1999, 9(1): 413–435.

[7] Gao J B and Cai H Q. On the structures and quantification of recurrence plots. *Physics Letters A*, 2000, 270(1): 75–87.

[8] 韦岗，陆以勤，欧阳景正. 混沌、分形理论与语音信号处理. 电子学报，1996, 24(1): 2–8.
Wei Gang, Lu Yi-qin, and Quyang J Z. Chaos and fractal theories for speech signal processing. *Acta Electronica Sinica*, 1996, 24(1): 2–8.

[9] Rabiner L R, Cheng M J, and Rosenberg A E. A comparative performance study of several pitch decision algorithm. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 1976, 24(5): 399–418.

闫润强： 男，1973 年生，博士生，研究方向为非线性语音信号处理.

朱贻盛： 男，1945 年生，教授，博士生导师，主要研究方向为生物医学信息检测与处理方法、人工视网膜的电子电路实现技术.