

文章编号:1004-9037(2014)02-0243-05

# 基于替代数据检测的音频信号非线性分析

刘 鑫 鲍长春

(北京工业大学电子信息与控制工程学院,北京,100124)

**摘要:**针对轻微含噪的音频信号,本文提出了一种基于替代数据检测的非线性分析方法。该方法首先假设音频信号来自于线性高斯随机过程,并利用基于随机相位的傅里叶变换法生成多组替代数据,然后分别计算原始数据和替代数据的样本峰度,最终根据假设检验方法判断原始音频中是否包含非线性成分。实验测试中分别对不同信噪比下音调型乐器演奏信号进行了分析。结果表明,相比于传统基于最大 Lyapunov 指数的非线性分析方法,本文所提方法在含噪情况下能更准确地检测到音频信号的非线性成分。

**关键词:**音频信号;非线性分析;替代数据法;假设检验

**中图分类号:** TN912.3

**文献标识码:** A

## Nonlinear Analysis of Audio Signals Using Surrogate Data Test

Liu Xin, Bao Changchun

(School of Electronic Information and Control Engineering, Beijing University of Technology, Beijing, 100124, China)

**Abstract:** A kind of nonlinear analysis method based on the surrogate data test for slightly noisy audio signals is presented. According to the hypothesis that audio signals can be modeled as a linear Gaussian stochastic process, several groups of surrogate data are generated by means of Fourier transform with random phase. The kurtosis measures are respectively calculated for the original audio and surrogate data. Hypothesis test is utilized to detect the nonlinear components in the original audio. In experiments, the audio signals of harmonic instruments were analyzed in the condition of different signal-to-noise ratios. The results show that the proposed method achieves the better performance for detecting the nonlinear components of audio signals than the nonlinear analysis based on the largest Lyapunov exponent in the noisy conditions.

**Key words:** audio signals; nonlinear analysis; surrogate data; hypothesis test

## 引 言

传统方法通常采用线性系统模型(如线性预测模型和傅里叶变换)来分析和描述语音与音频信号。这类线性模型广泛地应用于语音与音频编码、合成和识别等方面,并展现出优异的性能。然而源于对声音生成机制的线性假设,线性系统模型只能对真实声学系统进行一阶近似,并不能有效地描述声音产生过程中形成的湍流以及非线性耦合等现象。为此,有学者针对空气动力学现象进行了理论和实验性研究<sup>[1-2]</sup>,进而验证了声音产生机制中存在非线性过程这一重要假设。

受此启发,音频信号的非线性分析方法得到了广泛的关注。这类方法利用基于延迟重建的相空间重构技术将一维的音频信号转换到多维空间中。根据嵌入定理<sup>[3]</sup>,只要适当地选择嵌入参数,重构的多维相空间就能够与真实声学系统拓扑等价,且保留了其全部的动力学特性。基于此,有学者通过研究音频在相空间中呈现出的几何结构实现对声音特性的分析和分类<sup>[4-7]</sup>。此外,有研究人员通过在相空间中提取的非线性不变量来定量地描述音频信号的非线性特性<sup>[8-9]</sup>。其中,Lyapunov 指数估计了重构相空间中相邻轨迹间收敛或发散的平均指数率,而正的 Lyapunov 指数意味着信号的相轨迹随着时间的演变呈现出发散的特性,即存在非线性

性成分,所以最大 Lyapunov 指数通常用于判断声音信号是否蕴含非线性成分。Kolmogorov 熵对应所有正的 Lyapunov 指数的和,它描述了动力学系统中信息量耗散的平均速率,亦可作为判断非线性成分的有效指标。关联维表示真实声学动力学系统的复杂程度,可用于区分确定性信号和纯随机信号。针对纯净音频信号非线性不变量测度的相关研究成果<sup>[10-11]</sup>表明,语音和音频信号在时域和频域都具有典型的非线性特性,这为非线性特征分析和非线性预测理论应用于语音与音频动力学分析、单音素分类与识别及音频频带扩展等领域奠定了坚实的理论基础。

尽管基于重建相空间和非线性不变量的分析方法有效描述了音频的非线性特性,但是这些方法都对所处理的音频信号做出了特殊的限制:一方面,需要较长的数据才能使得计算得到的嵌入参数和非线性不变量趋近于实际声学系统的真实值;另一方面,在动力学系统理论中并没有明确定义噪声成分的非线性不变量<sup>[4,12]</sup>。实验表明,实际噪声同样可获得正的 Lyapunov 指数和 Kolmogorov 熵以及分数的关联维。因此,含噪条件下音频的非线性特性就很难采用传统的非线性不变量来度量。

针对含噪短时音频数据非线性分析中存在的问题,本文引入了替代数据检测方法来实现对音频信号非线性特性进行分析。替代数据法<sup>[4,13,14]</sup>首先假设音频信号源自平稳的线性高斯随机过程,然后利用傅里叶变换法生成与原始音频信号具有相同线性特性的多组替代数据,并分别计算原始数据和替代数据的样本峰度,最后判断两者的峰度是否存在明显差异。如果不存在明显差异,则接受假设,认为原始数据是由线性系统产生的;否则可以判定音频数据中包含有非线性成分。实验测试过程中,对不同信噪比(Signal to noise ratio, SNR)下音调型器乐信号进行非线性特征分析。结果表明,相比于基于最大 Lyapunov 指数的传统分析方法<sup>[15]</sup>,本文所提方法能够有效地区分噪声信号和实际音频信号。

## 1 基于替代数据法的非线性分析

### 1.1 生成替代数据

在替代数据法中,首先需要做出零假设,即音频信号来自于线性高斯随机过程。令实际音频信号表示为  $x(n)$  ( $n=0,1,2,\dots,N-1$ ),  $N$  为数据长度。根据传统的线性系统模型,该随机过程可采用线性预测模型(自回归模型)<sup>[16]</sup>来表示为

$$x(n) = a_0 + \sum_{i=1}^p a_i x(n-i) + e(n) \quad (1)$$

式中:  $a_i$  ( $i=0,1,2,\dots,p$ ) 为线性预测系数;  $p$  为预测阶数; 而  $e(n)$  为高斯白噪声信号。

根据该模型,最直接的替代数据生成方法需要借助线性预测分析方法估计  $a_i$ , 并利用高斯白噪声来激励线性预测滤波器,从而获得与原始音频具有相同线性特性的替代数据。但是,受到系统误差的影响,实际迭代获得的替代数据可能发散。为此,Theiler 等人提出了基于傅里叶变换的替代数据生成方法<sup>[4,13]</sup>。其主要步骤如下:

#### (1) 傅里叶变换

对于原始音频数据  $x(n)$ , 其傅里叶变换系数可以表示为

$$X(k) = \sum_{n=0}^{N-1} x(n) \exp(-j \frac{2\pi kn}{N})$$

$$k = 0, 1, 2, \dots, N-1 \quad (2)$$

#### (2) 相位随机化

在保持原始音频的幅度谱不变的同时,引入随机相位  $\varphi_k$ , 进而获得随机化的傅里叶变换系数  $X_r(k)$

$$X_r(k) = X(k) \exp(j\varphi_k) \quad (3)$$

式中:  $\varphi_k$  服从均匀分布  $U[0, 2\pi]$ ;  $k=0, 1, \dots, N-1$ 。为了保证生成的替代数据仍然是实数,  $\varphi_k$  还需要满足以下条件

$$\begin{cases} \varphi_0 = \varphi_{\frac{N+1}{2}} = 0 \\ \varphi_k = -\varphi_{N-k} \quad k = 1, 2, \dots, \frac{N-1}{2} \end{cases} \quad (4)$$

#### (3) 傅里叶逆变换

经过傅里叶逆变换,可以获得相位随机化后的替代数据  $x_{SD}(n)$  ( $n=0, 1, 2, \dots, N-1$ ), 有

$$x_{SD}(n) = \frac{1}{N} \sum_{k=0}^{N-1} X_r(k) \exp(j \frac{2\pi kn}{N}) \quad (5)$$

由于  $x_{SD}(n)$  和  $x(n)$  具有相同的幅度谱, 人为生成的  $x_{SD}(n)$  保留了原始音频数据的自相关函数, 而相位随机化过程则基本去除了两者的非线性相关性。因此, 可以认定  $x_{SD}(n)$  和  $x(n)$  是由式(1)中假设的同一线性高斯随机过程生成的。

## 1.2 检验统计量

获得替代数据后, 还需要选择适当的检验统计量来描述原始音频序列和替代数据。由于两者具有一致的均值、方差和自相关函数等特性, 为区分两种数据间的差异需要选择与这些特性无关的参数作为检验统计量。

在传统非线性分析中, 通常采用关联维、Lyapunov 指数和 Kolmogorov 熵等非线性不变量作为检验统计量<sup>[13]</sup>。然而针对短时含噪音频及其替代数据来说, 这些非线性不变量很难准确地描述信

号的特性。为此,本文选择样本峰度  $T_K$  来衡量数据分布曲线的尖锐程度,进而为假设检验提供依据,其定义如下

$$T_K = \frac{\frac{1}{N} \sum_{i=0}^{N-1} (s(n) - m)^4}{\left(\frac{1}{N} \sum_{i=0}^{N-1} (s(n) - m)^2\right)^2} - 3 \quad (6)$$

式中: $s(n)$  ( $n=0, 1, 2, \dots, N-1$ ) 为测试数据,  $N$  为数据长度,  $m$  为  $s(n)$  的均值。

### 1.3 假设检验

受到数据长度的限制,借助随机相位人为构建的替代数据在统计性质上会出现一定的波动,因此,需要生成多组的替代数据,并利用其检测统计量的均值和标准差来和原始音频数据进行差异性比较。令  $m_K$  和  $\sigma_K$  分别表示替代数据峰度的均值和标准差,而  $T_{K0}$  表示原始音频数据的样本峰度。由于零假设中替代数据和原始音频数据源自同一线性随机过程,因此在统计意义上原始音频的检验统计量应与替代数据统计量的均值相等。实际检测中,定义峰度的差异显著度  $D_K$  为

$$D_K = \frac{|T_{K0} - m_K|}{\sigma_K} \quad (7)$$

根据双边  $Z$ -检验方法,在显著性水平为 0.05 的条件下,差异显著度的临界值为 1.96。若差异显著度在  $[0, 1.96]$  区间内,则可以认为原始音频数据和替代数据的检验统计量没有显著差异,并接受零假设,认为两种数据来自同一个线性高斯随机过程。否则,拒绝零假设,即原始音频数据中包含着非线性成分。

## 2 实验结果与分析

为了验证所提方法的有效性,本文分别对不同的音调型器乐信号进行分析。所采用的音频数据来自欧洲广播联盟录制的声音质量评测材料 (Sound quality assessment material, SQAM)<sup>[17]</sup>, 其中包括小提琴、长笛、黑管、萨克斯、小号等音调型器乐演奏的音乐信号。这些信号采用 16 位 PCM 进行数字存储,采样率为 32 kHz,有效带宽为 14 kHz,数据长度为 10 s 左右。非线性分析实验在纯净音频的基础上进一步加入适当的白噪声构成含噪音频,进而实现在不同 SNR 下对音频数据的非线性分析。

### 2.1 纯净信号的非线性分析

为了有效地研究纯净音频信号,在非线性分析

前,须去除所有测试音频中的静音段,并利用 20 ms 窗长进行分帧 (每帧数据包含 640 个采样点)。所有音频数据采用替代数据检测法来对其进行分析。对于不同音调型器乐信号,每帧数据分别采用傅里叶变换法生成 50 组替代数据,并计算原始音频和替代数据样本峰度的差异显著度,如图 1 所示。其中,实线表示差异显著度曲线,而虚线表示差异显著度的临界值 1.96。

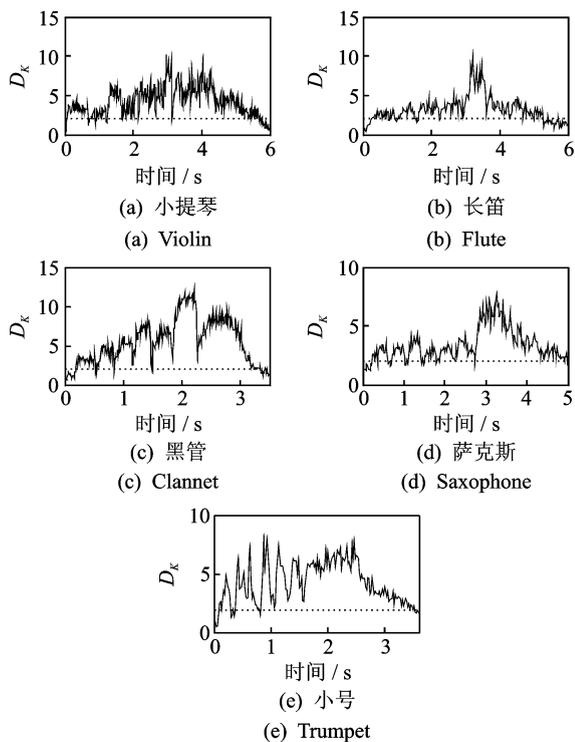


图 1 不同音乐信号差异显著度曲线

Fig. 1 Difference significance curve for different audio signals

由图 1 所示,90% 左右音频帧的差异显著度  $D_K$  大于双边  $Z$ -检验的临界值 1.96,从而可以拒绝零假设。这表明不同器乐音频都具有显著的非线性特性,即采用线性系统模型不能完全描述音频样点间的相关性。

### 2.2 含噪音频的非线性分析

为验证所提方法的鲁棒性,实验测试中还需在纯净音频的基础上加入了不同能量的噪声,实现在不同 SNR 下音频信号的非线性分析。

由于真实采录的噪声是否具有非线性特性还有待进一步研究,因此含噪音频测试中加入的是高斯白噪声信号。测试数据同样来自 SQAM 库中 5 种音调性器乐演奏的纯净信号,共有 15 片段。加噪处理前,所有音频中的静音段须保留下来,并作

为线性随机信号进行对比。同时选择传统的基于最大 Lyapunov 指数非线性分析方法<sup>[15]</sup>作为参考方法。该方法同样对长度为 20 ms 的短时音频数据进行处理,计算得到的最大 Lyapunov 指数用于描述该段信号在相空间中相邻轨迹的收敛或发散特性。若该指数为正,则判定该段信号中包含一定的非线性成分<sup>[6]</sup>。

纯净音乐信号中包含非线性成分帧占总帧数的比例接近 100%,而噪声信号并不包含任何的非线性成分。对于含噪条件,本文分别对不同 SNR 的音频数据进行非线性特性分析,并计算出每一类器乐信号音乐段和静音段中被判定包含非线性成分的音频帧占总帧数的比例,据此对两种方法在含噪条件下分析音频非线性的有效性进行评价,如表 1-3 所示。

表 1 被判定包含非线性成分的音频帧占总帧数的比例 (SNR=35 dB)

Table 1 Proportion of the audio frames determined as non-linear in all the frames (SNR=35 dB)

音频类型		所提方法/%	最大 Lyapunov 指数方法/%
小提琴	音乐段	96.34	98.87
	静音段	5.62	12.32
长笛	音乐段	93.33	92.33
	静音段	4.34	14.37
黑管	音乐段	95.33	95.57
	静音段	5.42	12.34
萨克斯	音乐段	89.53	96.04
	静音段	4.68	12.78
小号	音乐段	92.36	92.21
	静音段	4.24	13.81

表 2 被判定包含非线性成分的音频帧占总帧数的比例 (SNR=25 dB)

Table 2 Proportion of the audio frames determined as non-linear in all the frames (SNR=25 dB)

音频类型		所提方法/%	最大 Lyapunov 指数方法/%
小提琴	音乐段	81.27	81.34
	静音段	6.39	35.23
长笛	音乐段	82.36	85.10
	静音段	5.28	23.43
黑管	音乐段	73.36	84.71
	静音段	5.23	38.54
萨克斯	音乐段	78.53	84.30
	静音段	5.48	31.47
小号	音乐段	82.59	86.21
	静音段	6.65	31.44

表 3 被判定包含非线性成分的音频帧占总帧数的比例 (SNR=15 dB)

Table 3 Proportion of the audio frames determined as non-linear in all the frames (SNR=15 dB)

音频类型		所提方法/%	最大 Lyapunov 指数方法/%
小提琴	音乐段	64.19	71.53
	静音段	6.34	55.23
长笛	音乐段	69.40	79.23
	静音段	6.42	51.87
黑管	音乐段	62.44	84.71
	静音段	4.44	48.35
萨克斯	音乐段	62.10	74.52
	静音段	4.92	56.22
小号	音乐段	64.69	74.12
	静音段	5.76	52.92

从上述结果可以看出,在高 SNR(35 dB)下,两种方法在音乐段中检测出非线性成分的帧比例较高,且静音段中错误检出非线性成分的比例均在 15%以下。随着 SNR 的不断降低,在音乐段中信号的非线性成分逐渐被噪声所掩盖。在 SNR=15 dB 情况下,音乐段中利用替代数据法准确检出非线性特性音频帧的比例下降到 70%以下,而这种方法在静音段中,错误检出非线性成分的比例并没有显著的提高。这说明基于替代数据检测的非线性方法在含噪情况下能更准确地检测到非线性成分。而对于基于最大 Lyapunov 指数的参考算法,尽管在低信噪比下仍能检测出音乐段中非线性成分。但是在静音段中错检出非线性的帧比例比较高。可以看出噪声成分对最大 Lyapunov 指数方法判断非线性特性具有较大的影响。

### 3 结束语

本文提出了一种基于替代数据检测的音频信号非线性分析方法。该算法根据音频的线性假设生成替代数据,并利用假设检验判定音频信号中是否包含非线性成分。实验测试中分别对不同信噪比下音调型器乐演奏信号进行了分析,结果表明,相比于传统基于最大 Lyapunov 指数的非线性分析方法,所提方法在含噪情况下能更准确地检测到非线性成分。

#### 参考文献:

- [1] Barney A, Shadle C, Davies P. Fluid flow in a dynamical mechanical model of the vocal folds and

- tract; part 1 & 2[J]. *J Acoust Soc Amer*, 1999, 105(1): 444-466.
- [2] Maragos P, Dimakis A, Kokkinos I. Some advances in nonlinear speech modeling using modulations, fractals, and chaos[C]//14th International Conference on Digital Signal Processing. Greece: IEEE Press, 2002: 325-332.
- [3] Kantz H, Schreiber T. Nonlinear time series analysis [M]. UK: Cambridge University Press, 1997.
- [4] 郭兴明,李传鹏,卢德林. 定量递归分析和近似熵在心音分类识别中的应用[J]. *数据采集与处理*, 2013, 28(5): 559-564.
- Guo Xingming, Li Chuanpeng, Lu Deling. Application of recurrence quantification analysis and approximate entropy in classification and recognition of heart sound[J]. *Journal of Data Acquisition and Processing*, 2013, 28(5): 559-564.
- [5] Kokkinos I, Maragos P. Nonlinear speech analysis using models for chaotic systems [J]. *IEEE Trans Speech Audio Process*, 2005, 13(6): 1098-1109.
- [6] Zhang Liyan, Bao Changchun, Liu Xin, et al. Audio classification algorithm based on nonlinear characteristics analysis[C]//3rd Asia-Pacific Signal and Information Processing Association Annual Summit and Conference. Xi'an, China: APSIPA, 2011: 214-217.
- [7] Rui Rui, Bao Changchun. Musical instrument classification based on nonlinear recurrence analysis and supervised learning[J]. *Radio Engineering*, 2013, 22(1), 60-67.
- [8] Pitsikalis V, Kokkinos I, Maragos P. Nonlinear analysis of speech signals: generalized dimensions and lyapunov exponents [C]//8th European Conference on Speech Communication and Technology. Geneva, Switzerland: ISCA, 2003: 1-4.
- [9] Prasad S, Srinivasan S, Pannuri M, et al. Nonlinear dynamical invariants for speech recognition[C]//9th International Conference on Spoken Language Processing. Pittsburgh, USA: ISCA, 2006: 2518-2521.
- [10] Zhang Xingtao, Bao Changchun, Liu Xin, et al. Audio bandwidth extension based on maximum lyapunov prediction[C]//3rd Asia-Pacific Signal and Information Processing Association Annual Summit and Conference. Xi'an, China: APSIPA, 2011: 849-852.
- [11] Liu Xin, Bao Changchun, Jia Maoshen, et al. Nonlinear bandwidth extension based on nearest-neighbor matching[C]//2nd Asia-Pacific Signal and Information Processing Association Annual Summit and Conference. Biopolis, Singapore: APSIPA, 2010: 169-172.
- [12] 孟庆芳. 非线性动力系统时间序列分析方法及其应用研究[D]. 济南:山东大学, 2008.
- Meng Qingfang. Nonlinear dynamical time series analysis methods and its application[D]. Jinan: Shandong University, 2008.
- [13] Theiler J, Eubank S, Longtin A, et al. Testing for nonlinearity in time series: the method of surrogate data[J]. *Physica D: Nonlinear Phenomena*, 1992, 58(1): 77-94.
- [14] Small M, Tse C K. Detecting determinism in time series: The method of surrogate data [J]. *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, 2003, 50(5): 663-672.
- [15] Wolf A, Swift J B, Swinney H L, et al. Determining Lyapunov exponents from a time series[J]. *Physica D: Nonlinear Phenomena*, 1985, 16(3): 285-317.
- [16] Markel J E, Gray A H. Linear prediction of speech [M]. New York: Springer-Verlag, 1982.
- [17] European Broadcasting Union. Sound quality assessment material, Recordings for subjective tests-Users' handbook for the EBU-SQAM Compact Disc [S]. EBU-Tech 3253. Geneva, Switzerland: Technical Centre of EBU, 2008.

作者简介:刘鑫(1986-),男,博士生,研究方向:音频信号处理,E-mail:liuxin0930@emails.bjut.edu.cn; 鲍长春(1965-),男,教授,博士生导师,研究方向:语音与音频信号处理。