

## 非言語音声のもつ非線形的特徴と感性情報

著者	今井 順一, 田中 秀典, 長島 知正
雑誌名	サテライト・ベンチャー・ビジネス・ラボラトリー 年報
巻	6
ページ	39-42
発行年	2004
URL	<a href="http://hdl.handle.net/10258/324">http://hdl.handle.net/10258/324</a>

# 非言語音声のもつ非線形的特徴と感性情報

今井順一<sup>1)</sup>、田中秀典<sup>2)</sup>、長島知正<sup>3)</sup>

1) 室蘭工業大学 サテライト・ベンチャー・ビジネス・ラボラトリー

2) 室蘭工業大学大学院 工学研究科生産情報システム工学専攻

3) 室蘭工業大学 工学部情報工学科

## 1. はじめに

音声から話者に関する多くの情報を抽出する技術は、コンピュータを通じた「感性コミュニケーション」の実現のために非常に重要なものとなってきている。特に、話者の感情や心理状態などは円滑なコミュニケーションのためには欠かせない情報である。言語音声だけでなく、泣き声や叫び声、笑い声などの非言語音声もまた感情に関する多くの情報を含んでおり、周波数解析の手法を用いて乳幼児の泣き声を解析するなどの研究例が報告されている。

一方、近年になって、決定論的非線形力学系の視点からの音声解析が注目を集めている。これらの研究により、母音の時系列データが非線形性を有していること、さらにその非線形性が音声の「自然さ」に対して重要な役割を果たしていることが明らかとなった。これらの非線形的な特徴は話者の感性情報を抽出するための重要な手掛かりとなると期待されている。

本稿では、非言語音声から非線形的特徴量に基づいて感性情報を抽出するための第一歩として、サロゲートデータ法 [1] により泣き声、叫び声、笑い声の時系列データに対する非線形性の存在の検定を行う。

## 2. 方法

### 2.1 音声データ

非言語音声（乳幼児や子供が発声した泣き声、叫び声および笑い声；各 4 種類ずつ）から切り出した時系列データを解析する。これらはいずれも音声サンプル集 CD “The Nightingale Voice Box: Just Kids and Babies” (Nightingale Music Productions, 2001) に収録されているデータである。各音声データについての詳細な説明を表 1 に示す。

各データのサンプリング周波数は 44.1 kHz である。また、時系列データ長  $N$  は、泣き声および叫び声については 21,870 点（およそ 0.50 秒に相当）、笑い声については 10,368 点（およそ 0.24 秒に相当）とした。

### 2.2 サロゲートデータ

非言語音声データに関して、次の帰無仮説  $H_0$  を立てる。

$H_0$ : 「観測された時系列信号は、線形確率過程から作り出されたが、観測する際に静的な単調非線形変換を施されたことにより得られたデータである。」

この帰無仮説の下、Iterative Amplitude Adjusted Fourier Transform (IAAFT) アルゴリズム [2] によってサロゲートデータを作成する。IAAFT によって作成されたサロゲートデータはオリジナルデータと同一の頻度分布をもち、またオリジナルデータとほぼ等しい自己相関関数をもつ。本稿では、後述するように各音声データに対して

表 1. 音声データ

#	説明
泣き声	1 1 歳の赤ちゃんが怒って泣く声
	2 1 歳の赤ちゃんが哀れっぽく長い間泣く声
	3 男の子が長くむせび泣く声
	4 女の子がむせび泣く声
叫び声	1 子供が何度も金切り声で叫ぶ声
	2 男の子が長い間叫ぶ声
	3 子供が恐怖のため金切り声で叫ぶ声
	4 子供が恐怖のため金切り声で数度叫ぶ声
笑い声	1 10 歳の女の子が楽しそうに笑う声
	2 2 歳の女の子が機嫌良さそうに甲高く笑う声
	3 10 歳の男の子が意地悪げに思い切り笑う声
	4 1 歳の赤ちゃんが機嫌良さそうに笑う声

39 個のサロゲートデータを用いる。サロゲートデータの作成には IAAFT アルゴリズムが実装されているフリーのソフトウェア・パッケージ “TISEAN” [3-4] を利用した。

### 2.3 検定のための非線形統計量

本稿では、非線形性を検定するための統計量として最大リアプノフ指数および非線形予測誤差を用いる。これらの統計量の推定には、サロゲートデータの作成と同様 “TISEAN” を利用した。

#### (i) 最大リアプノフ指数

最大リアプノフ指数 (Largest Lyapunov Exponent; LLE) は決定論的力学系の軌道不安定性を定量化する指標である。

リアプノフ指数を推定するためには、まずアトラクタの再構成 [5] が必要となる。1 次元の時系列（すなわちオリジナルデータもしくはサロゲートデータ） $x(t)$  ( $t = 1, 2, \dots, N$ ) を以下のように  $m$  次元ベクトルの列  $\mathbf{x}(t)$  ( $t = 1, 2, \dots, N - (m - 1)\tau$ ) に変換する。

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

ここで  $\tau$  は時間遅れパラメータである。 $m$  は埋め込み次元とよばれる。

リアプノフ指数は、点  $\mathbf{x}(t)$  と再構成アトラクタ上におけるその近傍点との距離の時間変化を調べることで推定される。リアプノフ指数の推定には Sano & Sawada [6] のアルゴリズムを用いた。LLE はリアプノフ指数のうち最大のものである。

本稿では、時間遅れ  $\tau$  は各音声に対して高次自己相関関数から定まる適切な値に設定する [7]。また埋め込み次元  $m$  はパラメータとする。

(ii) 非線形予測誤差

非線形予測手法の一つである局所線形近似法 [8] による非線形予測誤差 (Nonlinear Prediction Error; NPE) を 2 つめの統計量とする。この手法では、 $x(n+p)$  と  $x(n)$  との関係が次の 1 次式で近似される。

$$x(n+p) \cong a_n \cdot x(n) + b_n \quad (2)$$

式 (2) における係数  $a_n$  および  $b_n$  は  $x(n)$  の近傍に関する最小 2 乗法により決定する。

$p$  ステップ先の予測  $\hat{x}(t+p) (= a_t \cdot x(t) + b_t)$  の精度は次の相対予測誤差 (非線形予測誤差) NPE で評価される。

$$NPE = \frac{\sqrt{\sum_{t=1}^P (\hat{x}(t) - x(t))^2}}{\sqrt{\sum_{t=1}^P (x(t) - \bar{x})^2}} \quad (3)$$

ここで  $\bar{x} = \frac{1}{P} \sum_{t=1}^P x(t)$ ,  $P$  は予測時系列のデータ長を表す。この予測誤差 NPE を検定のための統計量として用いる。

本稿では、埋め込み次元  $m$  または予測ステップ数  $p$  のいずれかをパラメータとする。  $m$  を可変とするときは  $p = 1$  とし、逆に  $p$  を可変とするときは  $m = 5$  に設定する。

### 2.4 非線形性の検定

帰無仮説  $H_0$  はモンテカルロ法を用いて次のように検定される。

$t_0$  をオリジナルデータに対する統計量 (LLE または NPE) とする。また  $t_k$  ( $k = 1, 2, \dots, B$ ) をサロゲートデータに対する統計量とし、これらは降順に並んでいるものとする。ここで  $B$  はサロゲートデータの数を表す。両側検定では、 $t_0 > t_{(B+1)\alpha/2}$  または  $t_0 < t_{(B+1)-(B+1)\alpha/2}$  のとき  $H_0$  は有意水準  $\alpha$  で棄却される。

本稿では、 $B = 39$ ,  $\alpha = 0.05$  とする。したがって、 $t_0$  がすべての  $t_k$  よりも大きい、あるいはすべての  $t_k$  よりも小さいとき、 $H_0$  は有意水準 0.05 で棄却される。

## 3. 結果

### 3.1 泣き声

典型的な場合の例として、泣き声 #1 について非線形統計量を求めた結果を図 1 に示す。

$m = 3$  の再構成アトラクタを図 1 (a) に示す。図から、この音声時系列に何らかの非線形構造が存在することが示唆される。

図 1 (b) は推定した LLE を、(c) および (d) はそれぞれ埋め込み次元  $m$  および予測ステップ数  $p$  に対する NPE を表す。これらの図から、オリジナルデータに対する統計量はサロゲートデータに対するそれよりも小さいことがわかる。したがって、2.2 節で立てた帰無仮説  $H_0$  は有意水準  $\alpha = 0.05$  で棄却される。今回の実験では、他のすべての泣き声についても同様に  $H_0$  は有意水準 0.05 で棄却された。

各データに対する LLE ( $m = 5$ ) および NPE ( $m = 5, p = 1$ ) の値を表 2 に示す。すべてのデータにおいて、LLE は絶対値は小さいものの正の値をとっている。これらの結果から、泣き声の音声は非線形性を有していると考えられることができる。

表 2. 泣き声に対する非線形統計量の推定値

#	LLE ( $m = 5$ )	NPE ( $m = 5, p = 1$ )
1	$4.759 \times 10^{-2}$	$7.433 \times 10^{-2}$
2	$5.067 \times 10^{-2}$	$6.285 \times 10^{-2}$
3	$3.425 \times 10^{-2}$	$7.028 \times 10^{-2}$
4	$2.463 \times 10^{-3}$	$1.226 \times 10^{-1}$

### 3.2 叫び声

典型的な場合の例として、叫び声 #1 について求めた非線形統計量を図 2 に示す。図 2 (a) に示した再構成アトラクタ ( $m = 3$ ) は同心円のような形状をしているように見える。図 2 (b), (c) および (d) を見ると、オリジナルデータに対するグラフの線は対応するサロゲートデータの線によってほとんど隠れてしまっている。したがって、帰無仮説  $H_0$  は有意水準 0.05 で棄却することはできない。その他の叫び声についても同様の結果であり、叫び声 #4 についてのみ  $H_0$  は有意水準 0.05 で棄却された。

参考のため、各データに対する LLE ( $m = 5$ ) および NPE ( $m = 5, p = 1$ ) の値を表 3 に示す。叫び声 #3 に対する LLE は負の値を取っている。

表 3. 叫び声に対する非線形統計量の推定値

#	LLE ( $m = 5$ )	NPE ( $m = 5, p = 1$ )
1	$8.181 \times 10^{-3}$	$6.345 \times 10^{-2}$
2	$3.174 \times 10^{-2}$	$1.137 \times 10^{-1}$
3	$-3.092 \times 10^{-4}$	$1.236 \times 10^{-2}$
4	$2.052 \times 10^{-3}$	$1.119 \times 10^{-1}$

### 3.3 笑い声

笑い声 #3 に対する結果を図 3 に示す。図 3 において、オリジナルデータに対するほとんどすべての統計量は対応するサロゲートデータのそれよりも小さい。したがって、それらの統計量に対しては、帰無仮説  $H_0$  は有意水準 0.05 で棄却される。しかし  $m = 2$  および 3 における LLE については、それよりも高い埋め込み次元の場合とは異なり、 $H_0$  を棄却することはできない。これは、 $m = 2$  あるいは 3 はこのデータを埋め込むには低すぎる次元であったためと考えられる。その他の笑い声については  $H_0$  は棄却される。

各データに対する LLE ( $m = 5$ ) および NPE ( $m = 5, p = 1$ ) の値を表 4 に示す。泣き声の場合と同様、すべての笑い声データについて LLE は正の値をとる。

表 4. 笑い声に対する非線形統計量の推定値

#	LLE ( $m = 5$ )	NPE ( $m = 5, p = 1$ )
1	$4.672 \times 10^{-2}$	$3.852 \times 10^{-2}$
2	$2.149 \times 10^{-2}$	$1.815 \times 10^{-1}$
3	$3.706 \times 10^{-2}$	$4.934 \times 10^{-2}$
4	$2.309 \times 10^{-2}$	$1.149 \times 10^{-1}$

## 4. 考察

上述の通り、泣き声と笑い声のすべてのデータに対して帰無仮説  $H_0$  は棄却され、LLE は正の値をとる。これらの結果から、泣き声と笑い声の音声時系列には非線形性が存在することが示唆される。一方、叫び声の多くのデータについては  $H_0$  を棄却することができなかった。

したがって、これらの結果からは叫び声における非線形性の存在を主張することはできない。これに関してはさらなる解析が必要である。

こうした非線形的な特徴の果たす役割として、(1) 音声の「自然さ」、(2) 音声の個人差、(3) 音声に含まれる感性情報の表現、の3つの可能性が考えられる。特に、非線形特徴量をもとに非言語音声から感性情報を抽出することは非常に有望な応用の一つである。非言語音声はしばしば言語音声よりも多くの感性情報を含む。言語能力が十分に発達していない乳幼児ならなおさらである。例えば同じ乳児が発する似たような泣き声が、しばしば「お腹が空いた」「痛い」「眠い」などの異なる意思の表明となっていることがある。従来の周波数解析と決定論的非線形力学系の技術とを相補的に用いて解析することにより、これらの音声から感性情報を効果的に抽出することが可能になると期待される。

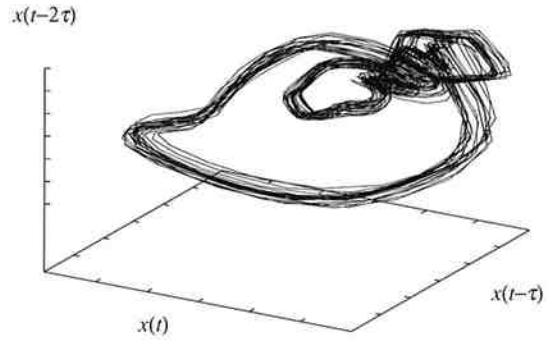
### 5. おわりに

本稿では、それぞれ4種類ずつの泣き声、叫び声、笑い声について最大リアプノフ指数および非線形予測誤差を推定し、サロゲートデータ法によってその非線形性の検定を行った。その結果として、泣き声および笑い声の音声中に非線形性が存在することを統計的に確認した。

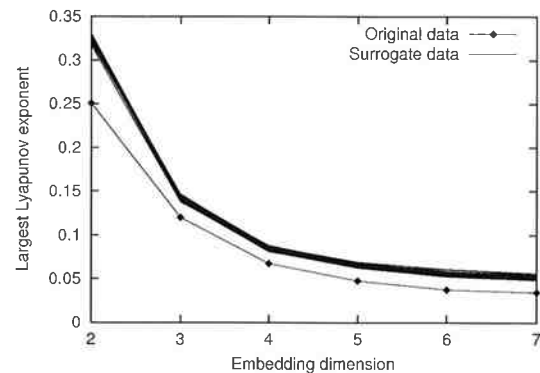
今後の課題として、重要な非線形特徴量の一つである相関次元を用いた非線形性の検定が挙げられる。また、第4節で示した応用に関する技術開発も重要な課題の一つである。

### 参考文献

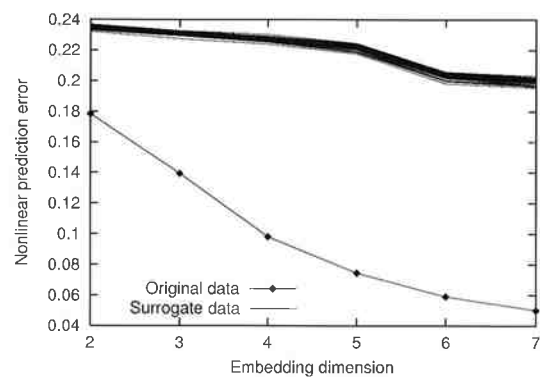
- [1] J. Theiler, S. Eubank, A. Longtin, B. Galdrikian and J. D. Farmer, "Testing for Nonlinearity in Time Series: The Method of Surrogate Data," *Physica D*, Vol. 58, pp. 77-94, 1992.
- [2] T. Schreiber and A. Schmitz, "Improved Surrogate Data for Nonlinearity Tests," *Physical Review Letters*, Vol. 77, No. 4, pp. 635-638, 1996.
- [3] R. Hegger, H. Kantz and T. Schreiber, "Practical Implementation of Nonlinear Time Series Methods: The TISEAN Package," *Chaos*, Vol. 9, No. 2, pp. 413-435, 1999.
- [4] T. Schreiber and A. Schmitz, "Surrogate Time Series," *Physica D*, Vol. 142, pp. 346-382, 2000.
- [5] F. Takens, "Detecting Strange Attractors in Turbulence," In D. A. Rand and L. S. Young, eds., *Dynamical Systems and Turbulence*, Lecture Notes in Mathematics, Vol. 898, pp. 366-381, Berlin, Springer-Verlag, 1981.
- [6] M. Sano and Y. Sawada, "Measurement of the Lyapunov Spectrum from a Chaotic Time Series," *Physical Review Letters*, Vol. 55, No. 10, pp. 1082-1085, 1985.
- [7] T. Ikeguchi, M. Takakuwa and K. Aihara, "Reconstructing Attractors with Higher Order Correlation," *Proceedings of International Symposium on Nonlinear Theory and Its Applications*, Vol. 2, pp. 1149-1152, 1997.
- [8] J. D. Farmer and J. J. Sidorowich, "Predicting Chaotic Time Series," *Physical Review Letters*, Vol. 59, No. 8, pp. 845-848, 1987.



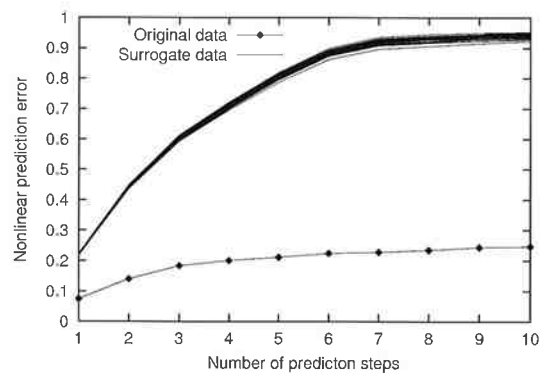
(a) 再構成アトラクタ



(b) 最大リアプノフ指数

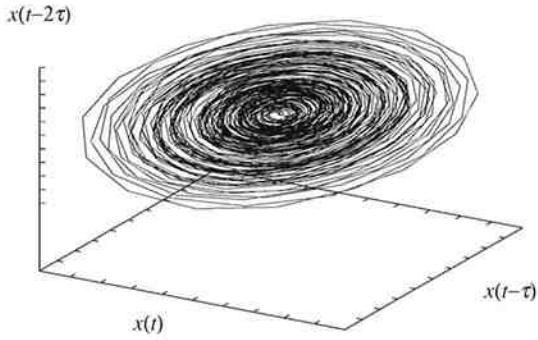


(c) 埋め込み次元に対する非線形予測誤差

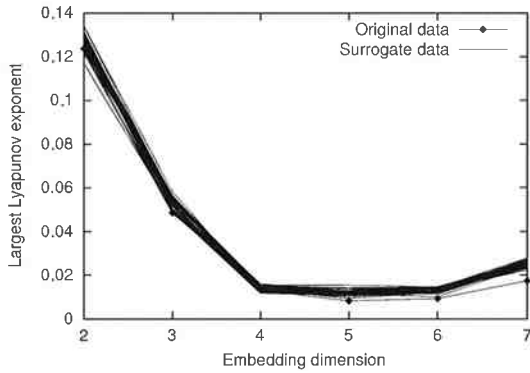


(d) 予測ステップ数に対する非線形予測誤差

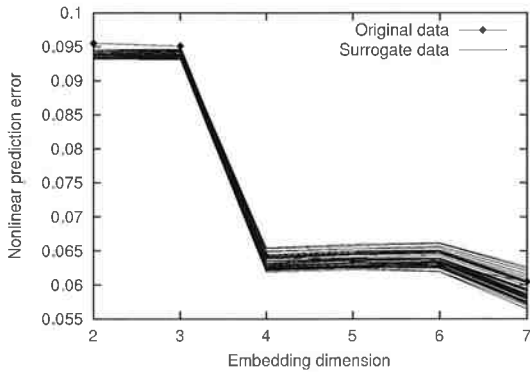
図1. 泣き声#1の結果



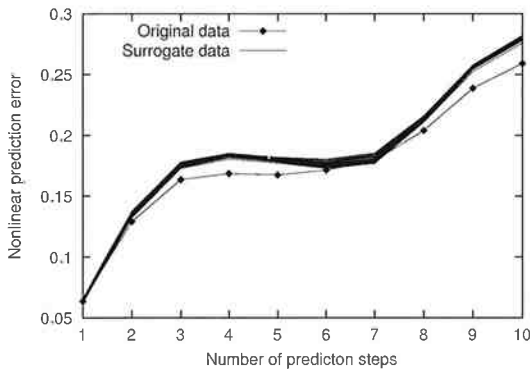
(a) 再構成アトラクタ



(b) 最大リアプノフ指数

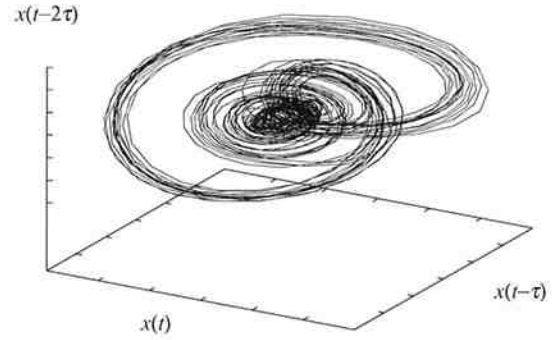


(c) 埋め込み次元に対する非線形予測誤差

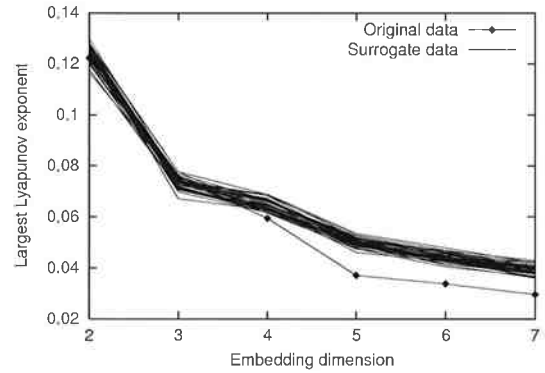


(d) 予測ステップ数に対する非線形予測誤差

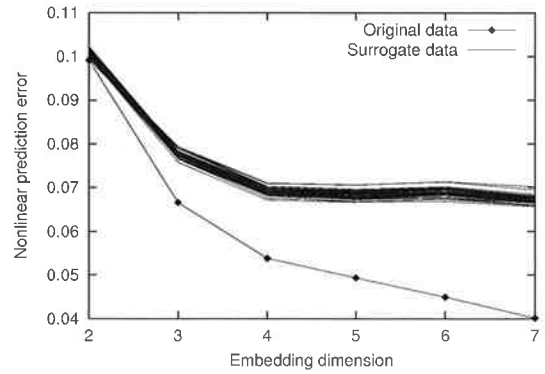
図 2. 叫び声#1 の結果



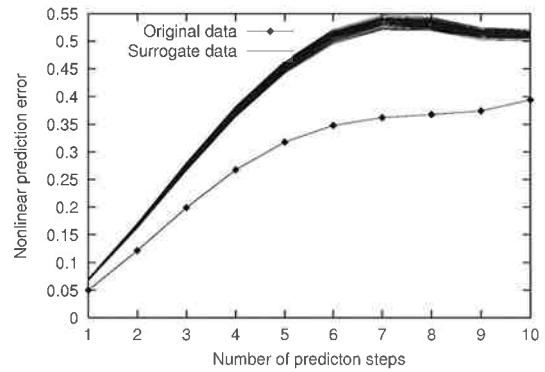
(a) 再構成アトラクタ



(b) 最大リアプノフ指数



(c) 埋め込み次元に対する非線形予測誤差



(d) 予測ステップ数に対する非線形予測誤差

図 3. 笑い声#3 の結果