

日本語母音の揺らぎはカオス的か？

Does chaotic attractor exist in Japanese vowels?

宮野尚哉*, 生内洋平*, 徳田功**, 合原一幸***

Takaya Miyano*, Yohei Ikunai*, Isao Tokuda**, Kazuyuki Aihara***

*弘前大学, **室蘭工業大学, ***東京大学大学院, CREST

*Hirosaki University, **Muroran Institute of Technology, ***The University of Tokyo

キーワード: カオス (chaos), サロゲート法 (surrogate method), 時系列解析 (time series analysis), 聴覚心理 (auditory psychology), 日本語母音 (Japanese vowels)

連絡先: 〒 036-8561 青森県弘前市文京町 3 番地 弘前大学 理工学部 知能機械システム工学科
宮野尚哉, Tel.: (0172)39-3681, Fax.: (0172)39-3681, E-mail: miyano@cc.hirosaki-u.ac.jp

1. はじめに

ヒト内耳の聴覚器官は、音を周波数パワースペクトルとして感知する。音を特徴付ける心理指標は、音の大きさ (loudness), 高さ (pitch), 音色 (timbre) である。これらの特徴量は、音のパワースペクトルにおけるピーク成分に対応するフォルマント構造から抽出される。現在の音声符号化技術では、信号を線形予測モデルで近似して得られる線形フィルターの極によってフォルマント構造が再現される。したがって、音声の特徴は線形係数で表現される。音声に固有の線形係数で特徴付けられる線形フィルターを無相関確率過程で駆動すると、音声合成される。無相関確率過程は、パワースペクトルの連続帯を形成し、フォルマント構造には影響しないので、音声品質には寄与がなく、音声の重要な特徴を形成しないと考えられている。

しかしながら、井福部らの研究に代表される最

近の研究によると¹⁾、音声には不規則な揺らぎがあり、これが音声の自然さ (naturalness), 即ち、音声の自然な響きと関係があるらしい。脳の聴覚野は、音声揺らぎを音の特徴として学習している可能性がある。著者らは、音声ダイナミクスにおける決定論的な非線形性に注目して、音声データの時系列解析を行ない、音声の不規則な揺らぎが決定論的カオスに由来する可能性があることを示唆してきた²⁾⁻⁶⁾。これを井福部らの研究成果と結び付けるとき、音声ダイナミクスのカオス性は音の自然さという第 4 の心理指標をもたらすという憶測が生まれる。パワースペクトルの連続帯には、自然さという音の重要な特徴が隠れているのかも知れない。

本研究は、上に提示した仮説の正否を明らかにするのに役立つであろう聴覚心理実験に関するものである。この聴覚心理実験は、著者らが最近報告した実験⁵⁾と同じ考え方に基づいて構成されているが、より多くのデータを用いて詳細な解析が行

なわれる．以下では，音声データの概要，時系列解析から推測される音声ダイナミックスの性質を示した後，聴覚心理実験方法と解析結果について述べる．

2. 音声データとサロゲート法

対象とする音声データは，国際電気通信基礎研究所（ATR）が所蔵する日本人男性（データ ID：mau）および日本人女性（データ ID：fsu）の日本語母音時系列 /a/, /i/, /u/, /e/, /o/ である．話者は健康な声帯を有し，発声のトレーニングを受けている．各音声は，遮断周波数 8 kHz ，サンプリング周波数 20 kHz ，サンプリング分解能 16 ビットの条件下で記録された．データ点数はいずれも 2048 である．

1 本の音声時系列について推定される統計量の誤差を評価する手法としては，サロゲート法^{7, 8)}が効果的である．この方法では，最初に，統計検定によって確かめたい帰無仮説を設定する．本研究では，「パワースペクトルの連続帯は，自己回帰過程を駆動する確率過程を表し，決定論的ダイナミックスによって生成される過程とは関係がない」という仮説を設定する．次に，帰無仮説の下でオリジナルデータと同じ統計的性質をもつデータを合成する．これをサロゲートデータと呼ぶ．サロゲートデータは，帰無仮説の下で再測定されたオリジナルデータの別の実現結果であると解釈される．本研究では，以下のようなプロセスによってサロゲート時系列を合成した．オリジナル音声 $x(t)$ のフーリエ係数 $\tilde{x}(f)$ (f は周波数) を求め，独立一様乱数 $\delta(f)$ を位相項とする係数を掛ける．

$$\tilde{x}(f) \exp[i\delta(f)]$$

次に，この係数を用いて逆フーリエ変換し，時系列データに戻す．こうして，サロゲート音声を得られる．サロゲート音声はオリジナル音声と同一のパワースペクトルを持つ．音声の大きさ，高さ，

音色はオリジナル音声と同じである．しかし，オリジナル音声のスペクトル連続帯を構成する信号成分における決定論的側面は完全に破壊されている．オリジナル音声とサロゲート音声との間で，推定される統計量に有意な差があれば，スペクトル連続帯を構成する信号成分は決定論的ダイナミックスによって生成されている可能性がある．また，聴覚心理的感覚にも有意な差があるならば，音声の大きさ，高さ，音色以外の特徴における相違が認知されている可能性がある．本研究では，Schreiber-Schmitz のアルゴリズム⁸⁾によって，サロゲート音声を 40 本合成した．フーリエ変換時のデータウィンドウには対称 Hanning 窓が用いられた．

統計量の推定値における相違の有意性は， t 検定によって定量的に評価することができる． E_o をオリジナルデータの統計量， E_s をサロゲートデータの統計量， \bar{E}_s を E_s の平均値， σ_s を E_s の分散とする． t 値を以下のように定義する．

$$t = \frac{|E_o - \bar{E}_s|}{\sigma_s} \quad (1)$$

t 値は， E_o が E_s からどのくらいかけ離れているかを定量的に示す指標である． $t > 2$ ならば，正規分布の裾の方にあるという意味で，95% の信頼性においてオリジナル時系列の統計量は帰無仮説のもとでは実現しそうにないと主張できる．こうして帰無仮説を棄却すべきかどうか判定する．もし棄却できなければ，帰無仮説を受容せざるを得ないから，オリジナル時系列の決定論的非線形ダイナミックスに関する主張は留保しなければならない．サロゲート法は手間のかかる統計検定法であるが，1 本の時系列からダイナミックスのカオス性を検定するための強力な手法である．

3. 時系列解析

著者らによるこれまでの研究によると，日本語母音のダイナミックスは非線形予測モデルによって近似することができる^{3, 6)}．母音時系列は短期

予測可能であるが、その予測可能性は予測時間が増加するにつれて急速に崩壊する。これは決定論的カオスの存在を示唆するような振舞いである。しかしながら、予測誤差と予測時間とのスケーリング性を詳細に調べると、線形相関のある確率過程に特有のスケーリング性⁹⁾が観測される場合があり、音声カオスの存在を強く主張することはできなかった⁶⁾。非線形予測に基づくカオス解析アルゴリズムは、データが観測ノイズによって汚染されている場合には、相関のある確率過程とカオス過程を識別することが困難となる¹⁰⁾。時間発展の滑らかさに着目したカオス解析アルゴリズム^{11, 12)}は、予測可能性に基づくアルゴリズムの難点を補うのに有効な手法である。

本研究では、Kaplan と Glass によって考案され¹¹⁾、Wayland らによって改良された時系列解析アルゴリズム¹²⁾を利用して、音声の不規則な揺らぎにおける決定論性を検定する。Wayland らのアルゴリズムの要点を以下に述べる。 N 個のデータ点からなる時系列 $\{x(t)\}_{t=0}^{N-1}$ について、 $x(t), x(t - \Delta t), \dots, x(t - (D - 1)\Delta t)$ を D 個の独立変数のように見て、時系列の時間発展を D 次元空間で再構成された軌道で表現する。 D は埋め込み次元と呼ばれる。 Δt は適当な時差である。時系列から再構成された埋め込みベクトル $x(t) = (x(t), x(t - \Delta t), \dots, x(t - (D - 1)\Delta t))$ から M 個のベクトル $x(t_0)$ を無作為に選択する。各々について、ユークリッド距離の意味で K 近接ベクトル $x(t_k)$ ($k = 0, 1, \dots, K$) を探索する。 $x(t_k)$ を $T\Delta t$ だけ時間推進すると $x(t_k + T\Delta t)$ が得られる。ここで、

$$v(t_k) = x(t_k + T\Delta t) - x(t_k)$$

を考える。適当な時間スケール $T\Delta t$ で時間発展を眺めると、決定論的ダイナミクスから生成される時系列は時間発展が連続で滑らかであるから、 $v(t_k)$ の方向の分散は小さいであろう。方向の分散は、以下に定義する並進誤差 E_{trans} (translation

error) で評価することができる。

$$E_{trans} = \frac{1}{K+1} \sum_{k=0}^K \frac{|v(t_k) - \bar{v}(t_k)|^2}{|\bar{v}(t_k)|^2} \quad (2)$$

$$\bar{v}(t_k) = \frac{1}{K+1} \sum_{k=0}^K v(t_k) \quad (3)$$

$x(t_0)$ の選択に依存する統計誤差を抑えるために、無作為に選択した M 個の $x(t_0)$ に関する E_{trans} の中間値 (メディアン) を求める操作を Q 回繰り返し、 Q 個の中間値の平均値で E_{trans} を表すことにする。こうして求めた E_{trans} 値は、決定論的ダイナミクスの極限では $E_{trans} \rightarrow 0$ となる。確率過程である白色ノイズでは、 D に関係なく $E_{trans} \approx 1$ となる。Wayland らのアルゴリズムは、データの観測ノイズによる汚染によって攪乱されにくい¹⁰⁾。この事実を、白色ノイズに汚染された Lorenz 時系列を例にして示す。

Lorenz 時系列は、連立常微分方程式によって表されるダイナミクスが生み出すカオスである。

$$\frac{dx(t)}{dt} = -\sigma(x(t) - y(t)) \quad (4)$$

$$\frac{dy(t)}{dt} = Rx(t) - y(t) - x(t)z(t) \quad (5)$$

$$\frac{dz(t)}{dt} = -bz(t) + x(t)y(t) \quad (6)$$

$(\sigma, R, b) = (10, 28, 8/3)$ の係数値の組はカオスの挙動を実現することが知られている。これらの値を代入し、4 次の Runge-Kutta 法を用いて、時間幅 0.001 のもとで数値計算し、 x, y, z の時系列を得た。サンプリング時間 $\Delta t = 0.06$ ごとにデータを抽出して得られた時系列 $\{x(t)\}$ を利用する。時系列 $\{x(t)\}$ に白色ノイズ $\{\xi(t)\}$ を重畳した時系列データ $\{y(t)\}$ を合成し、Wayland らのアルゴリズムを適用する。データ点数は $N = 1024$ である。

$$y(t) = x(t) + r_n \xi(t)$$

Fig. 1 は、ノイズレベル $r_n = 0, 0.1, 0.3$ の各時系列に関する計算結果である。計算条件は $K = 4, T = 5, M = 51, Q = 10$ である。ノイズレベルが上昇す

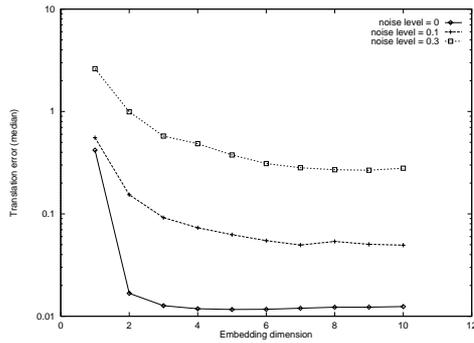


Fig. 1 白色ノイズで汚染された Lorenz 時系列に関する Wayland テスト . $r_n = 0(\diamond), 0.1(+), 0.3(\square)$

るにつれて並進誤差は増加するが, $r_n = 0.1$ では時系列の決定論的側面が十分に認められる. $r_n = 0.3$ でも, 埋め込み次元を適当に設定すると, 白色ノイズや有色ノイズに比べて並進誤差は低い値を取る.

各母音時系列とそのサロゲートデータについて, $\Delta t = 10$ のもとで, E_{trans} を $D = 2 - 15$ の範囲で推定した. mau /a/ および fsu /a/ に関する計算結果をそれぞれ Fig. 2 および Fig. 3 に示す. 各図には, オリジナルデータと任意に選んだ 10 本のサロゲートデータに対する計算結果が示されている. オリジナル音声とサロゲート音声とは決定論性において大きく異なることが認められる. このような傾向はすべての母音データについて観測された. 並進誤差の推定値における相違の統計的確からしさを t 検定によって評価した. 40 本のサロゲート音声に関する t 値の計算結果を Table 1 に示す. ただし, 埋め込み次元は, オリジナルデータの並進誤差が最小となる次元に設定されている. オリジナル音声とサロゲート音声の決定論性の相違は有意であり, オリジナルデータの統計量は, 95% 以上の信頼性において帰無仮説の下では実現しそうにないと主張することができる. 日本語母音に含まれる不規則な変動成分は, 決定論的ダイナミクスによる動的挙動であると言えそうである.

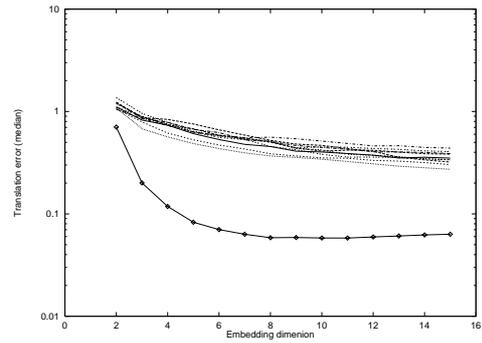


Fig. 2 mau /a/ (男性母音「あ」) に関する Wayland テスト: オリジナル時系列 (\diamond), サロゲート時系列 (—), $Q = 20, M = 301, K = 4, T = 5$

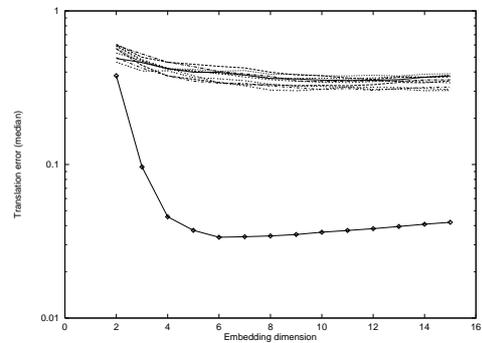


Fig. 3 fsu /a/ (女性母音「あ」) に関する Wayland テスト: オリジナル時系列 (\diamond), サロゲート時系列 (—), $Q = 20, M = 301, K = 4, T = 5$

Table 1. Wayland テストに関する t 検定結果. 埋め込み次元は, 並進誤差が最小となる次元である.

話者	/a/	/i/	/u/	/e/	/o/
mau	8.24 $D = 8$	9.90 $D = 8$	5.79 $D = 6$	7.39 $D = 7$	8.51 $D = 5$
fsu	14.08 $D = 6$	10.30 $D = 6$	3.81 $D = 6$	7.56 $D = 9$	8.91 $D = 12$

4. 聴覚心理実験

母音の不規則な揺らぎは, 決定論的ダイナミクスによって生成されていると考えられる. カオス的な揺らぎである可能性がある. このような特有の性質をもつ音声揺らぎが聴覚心理に何らかの効果をもたらすかどうか調べるために, 一対比較

法に基づく聴覚心理実験を行なった。被験者は健全な聴覚を有する男性 5 名、女性 7 名で、年齢は 19 歳～22 歳に分布し、平均年齢は 20 歳である。オリジナル音声とサロゲート音声を「あいうえお」と各母音間に数 10 msec の時間間隔をおいて連続して被験者に聞かせ、どちらの母音系列が自然に聞こえるか被験者に対比較法によって判断させる。母音系列は以下のようにラベルを付ける。

- オリジナル男性音声：oM
- サロゲート男性音声：sM
- オリジナル女性音声：oF
- サロゲート女性音声：sF

母音系列を聞かせる際には、oM, sM, oF, sF の 4 つの系列から異順不同のもとに 2 系列を選んで得られる 12 対を、乱数により無作為に配列した。被験者が判断する自然さの程度を 5 段階に分類し、各段階に以下のように点数を与えた。

- 明らかに前者の方が自然である：1 点
- どちらかといえば前者の方が自然である：2 点
- どちらとも言えない：3 点
- どちらかといえば後者の方が自然である：4 点
- 明らかに後者の方が自然である：5 点

実験結果から集計された合計点数を Table 2 に示す。これらの結果について Thurstone の比較判断の法則を適用し、得点分布を自然さの程度を表す間隔尺度に変換した。変換プロセスの概要を以下に記す。二つの刺激 S_i, S_j に対する反応、判断を表す得点を R_i, R_j とする。 R_i, R_j は、それぞれ、平均 A_i, A_j 、分散 σ_i^2, σ_j^2 で正規分布すると仮定する。刺激対の弁別距離 X_{ij} の統計的性質は、平均

$\bar{X}_{ij} = A_i - A_j$ 、分散 $\sigma_{ij}^2 = \sigma_i^2 + \sigma_j^2 - 2r_{ij}\sigma_i\sigma_j$ のように与えられる。 r_{ij} は R_i, R_j が変動するときの相関係数である。 \bar{X}_{ij} を σ_{ij} で規格化した値 z_{ij} で弁別距離を表す。

$$z_{ij} = \frac{A_i - A_j}{\sqrt{\sigma_i^2 + \sigma_j^2 - 2r_{ij}\sigma_i\sigma_j}} \quad (7)$$

本研究では、 $r_{ij} = 0$ と仮定し、 $A_i, A_j, \sigma_i^2, \sigma_j^2$ は実験結果の得点分布から推定する。こうして、二つの刺激に関する弁別の間隔尺度を z_{ij} として求めた。Table 3 および Table 4 は、それぞれ、男性音声および女性音声に関する間隔尺度の推定値である。オリジナル音声とサロゲート音声の自然さの相違は、間隔尺度において、男性母音系列では $0.21 - (-0.21) = 0.42$ 、女性母音系列では $0.23 - (-0.23) = 0.46$ である。これは、判断が正規分布するとの仮定の下で、約 60% の被験者がオリジナル音声系列の方が有意に自然であると認めたことを意味する。

Table 2. 列ラベルの母音系列が行ラベルの母音系列よりも自然であると判断する得点

	oM	sM	oF	sF
oM	0	32	24	24
sM	45	0	30	30
oF	52	45	0	29
sF	49	29	42	0

Table 3. 男性母音系列の z 値：列ラベルの母音系列が行ラベルの母音系列よりも自然であると判断する間隔尺度

	oM	sM
oM	0	-0.21
sM	0.21	0
計	0.21	-0.21

Table 4. 女性母音系列の z 値：列ラベルの母音系列が行ラベルの母音系列よりも自然であると判断する間隔尺度

	oF	sF
oF	0	-0.23
sF	0.23	0
計	0.23	-0.23

5. 結論

日本語母音の不規則な揺らぎは、決定論的な非線形ダイナミクスによって生み出されている可能性がある。無相関確率過程では期待されないような聴覚心理的效果があり、音声の自然さという聴覚心理指標と関係があるらしい。これが事実であるとすると、現在の音声合成技術によって実現される音声品質を超えて、人間による生の音声品質により近い音声を合成する技術を発見するためのヒントとなり得るかも知れない。発声のダイナミクスを研究する上で、非線形力学系という観点は採用に値するものであると言えるであろう。

参考文献

- 1) T. Ifukube, M. Hashiba, and J. Matsushima: A role of waveform fluctuation on the naturalness of vowels, *J. Acoust. Soc. Jpn.*, **47**, 903/910 (1991)
- 2) I. Tokuda, T. Miyano, and K. Aihara: Surrogate analysis for detecting nonlinear dynamics in normal vowels, *J. Acoust. Soc. Am.*, **110**, 3207/3217 (2001)
- 3) T. Miyano, A. Nagami, I. Tokuda, and K. Aihara: Detecting nonlinear determinism in voiced sounds of Japanese vowel /a/, *Int. J. Bifurcation and Chaos*, **10**, 1973/1979 (2000)
- 4) T. Miyano, I. Tokuda, and K. Aihara: Nonlinear determinism underlying voiced sounds of Japanese vowels, *Proc. Int. Symp. on Nonlinear Theory and its Applications*, **2**, 531/534 (2000).
- 5) 徳田功, 池口徹, 宮野尚哉, 合原一幸: サロゲート法に基づく音声知覚心理実験, *電子情報通信学会技術研究報告*, **NLP99-152**, 17/21 (2000)
- 6) T. Miyano: Are Japanese vowels chaotic?, *Proc. 4th Int. Conf. on Soft Computing*, **2**, 634/637 (1996)
- 7) J. Theiler, S. Eubank, A. Longtin, G. Galdrikian, and J. D. Farmer: Testing for nonlinearity in time series: The method of surrogate data, *Physica D*, **58**, 77/94 (1992)
- 8) T. Schreiber and A. Schmitz: Improved surrogate data for nonlinearity, *Phys. Rev. Lett.*, **77**, 635/638 (1996)
- 9) A. Provenzale, A. R. Osborne, and R. Soj: Convergence of the KS entropy for random noises with power law spectra, *Physica D*, **47**, 361/372 (1991)
- 10) T. Miyano: Time series analysis of complex dynamical behavior contaminated with observational noise, *Int. J. Bifurcation and Chaos*, **6**, 2031/2045 (1996)
- 11) D. T. Kaplan and L. Glass: Coarse-grained embeddings of time series: Random walks, Gaussian random processes, and deterministic chaos, *Physica D*, **64**, 431/454 (1993)
- 12) R. Wayland, D. Bromley, D. Pickett, and A. Pasamante: Recognizing determinism in a time series: *Phys. Rev. Lett.*, **70**, 530/582 (1993)