日本語母音及び子音における埋め込み次元と零点交差回数の相関

羽田真司、古瀬慶博
三菱スペース･ソフトウエア

誤り近傍点法を用いて日本語発話音声の埋め込み次元数を推定した。誤り近傍点法は、時系列データの非線形性を評価するために考案されたものである。母音と子音について従来のスペクトル解析で得られる結果との比較から、連続発話における非線形性を特徴づける指標を明らかにする。男女 4 名の被験者の音声解析から、個人認証の応用可能性についても指摘したい。

Correlation Between Embedded Dimension and Frequency of Zero-Cross of Japanese Vowel and Consonant

Shinji Haneda, Nobuhiro Furuse
Mitsubishi Space Software Co., Ltd.

Embedding dimension of the Japanese vowels and consonants was estimated using the False Nearest Neighbor (FNN) method. The FNN method was developed to evaluate the nonlinearity of time series data without the knowledge of elementary physical process of source and its propagation.

From the comparison between the results from the traditional spectrum analysis and FNN, we clarify the estimated embedded dimension becoming around 4 for the vowels. Also estimated variation of the dimension for the transition part of vowels has a strong correlation to formant variation. We suggest the applicability for personal certification based on four examinee of the volunteers.

1 はじめに

今日まで FFT や LPC（線形予測符号化）などの線形解析手法を用い、主にフォルマントについて着目した音声研究が数々なされており、音声認識や音声認証などに利用されてきた。

近年、生体システムにおいて起こる現象を非線形現象として捉え、定量化する試みが数々なされている。人間の発話音声も声帯振動を起源とし、非線形的物理過程によって生成される信号であることが知られており、近年 GP 法による相関次元推定や Lyapunov 指数等の評価など、定量的な非線形的若しくはカオス論的な解析が精力的になされている。ただし、これらの解析で対象とされてきた音声データは/a/等の単母音(V)もしくは/ka/、/ta/、/da/等の子音-母音（C-V）構造の音韻がほとんどであり（[1],[2]等）、自然発話中の音韻や、非線形性の特徴量の時間変化に関する解析の事例はほとんどない。

Lyapunov 指数は代表的なカオス性の指標であるが、その計算には埋め込み次元、埋め込み遅延時間、近傍球の大きさ等、複数のパラメータの適切な設定が必要になる。本研究では、非線形ダイナミクスの自由度とみなすことができる埋め込み次元に注目し、子音から母音の遷移過程、及び母音のわたり音（遷移音）について、誤り近傍法[3]によって推定された埋め込み次元について調査する。この結果より個人間の相違点及び共通点について着目し、普遍的な特徴量の抽出も試み、従来のフォルマントによる手法を補う方法として音声認識、個人認証等への工学的応用についても考察する。

2 手法

2.1 埋め込み

音声は非線形的多次元ダイナミクスによって生成されると考えられるが、録音によって得ることのできる音声データは、一次元時系列信号であり、一見情報が縮退していると考えられる。しかし適当な埋め込み次元と埋め込み遅延時間を指定し、高次元位相空間に埋め込みを行うことによって、系のダイナミクスの完全な再構成を行うことができることが数学的に Takens によって証明されている[4]。

埋め込みとは、N 個の一次元時系列データ列を

$\{x_i\}=\{x_0, x_1, x_2, \cdots, x_{N-1}\}$、埋め込み次元を d、埋め込み遅延時間を t_d としたとき、M 個のベクトル

$$\mathbf{y}_0 = (x_0, x_{t_d}, x_{2t_d}, \cdots, x_{(d-1)t_d})$$
$$\mathbf{y}_1 = (x_1, x_{1+t_d}, x_{1+2t_d}, \cdots, x_{1+(d-1)t_d})$$
$$\mathbf{y}_2 = (x_2, x_{2+t_d}, x_{2+2t_d}, \cdots, x_{2+(d-1)t_d})$$
$$\vdots$$
$$\mathbf{y}_M = (x_M, x_{M+t_d}, x_{M+2t_d}, \cdots, x_{M+(d-1)t_d})$$

を d 次元上の空間（位相空間）に構成することである。ただし、M は、$M+(d-1)t_d = N-1$ をみたす整数である。

このようにして位相空間上に埋め込まれた点の集合又はそれらを時系列順に結んだ曲線は、位相空間上の軌道と呼ばれる。

2.2 誤り近傍点法（FNN）

誤り近傍点法は、Kennel らによって提唱された、埋め込み次元推定法である。

具体的な手順は以下のとおりである。

i. 注目するフレームについて N 次元にて埋め込みを行い、ある埋め込み点について最近傍点との距離 L_N を求める。

ii. 同じフレームに対して $(N+1)$ 次元にて埋め込みを行い、同様にある埋め込み点の最近傍点との距離 L_{N+1} を求める。

iii. $R_{tol} \equiv \sqrt{{L_{N+1}}^2 - {L_N}^2} \Big/ L_N > 10$ を満たす場合、もしくは $R_l \equiv \sqrt{{L_{N+1}}^2 - {L_N}^2} \Big/ R_A > 2$ を満たす場合、この埋め込み点を誤り近傍点とみなす。ただし、R_A はアトラクタの大きさである。

iv. 埋め込まれたすべての点について R_{tol}, R_l を計算し、埋め込まれた全体に対する誤り近傍点の割合 FNNR(False Nearest Neighbor Ratio)を求める。

v. 次元数 N は 2 次元を初期値とし、一次元ずつ増やしていく。FNNR がゼロに収束した次元をそのフレームにおける埋め込み次元とする。

vi. 予め決めた次元の上限値(=20)においても FNNR が収束しない場合、そのフレームは次元推定が不可能（グラフ上では便宜上次元数を-1 とする）であるとする。

3 実験

本研究のデータ取得には、マイクにステレオマイク SONY ECM-M907、録音に SONY DAT WALKMAN TCD-D100 を用い、サンプリング周波数 44.1kHz でデジタル録音した。デジタル録音された音声は、Sound Forge™ を用いてデジタル端子（ SPDIF）を通じて PC に取り込み、モノラル 16 ビットの WAV フォーマットのファイルに変換した。

取得した音声は、主に単母音（/a/,/i/,/u/,/e/,/o/）を持続的に 3 秒発話したもの、/ka/を連続的に 10 回発音したもの（/kakaka…/）、自然発話的に/aiueo/を発音したものである。被験者は、健常な 20 代の男性 2 人、女性 2 人の計四人である。

時間依存性については、一般的に音韻の定常的な持続長が 25[msec]程度であるとされていることと、わたり音及び子音-母音の遷移状態を検出する時間分解能を考慮し、1 フレームを 256～1024 サンプル（約 5.80～23.2[msec]）とし、32 サンプル（約 0.723[msec]）ごとにフレームオーバーラップを行う解析を行った。また、埋め込み次元を推定するとともに、フレームの平均二乗振幅値（rms）、1 フレームあたりの零点交差回数（振幅が 0 と交差する回数、zero-cross）も計算した。

一フレームあたりの次元推定の平均計算時間はフレーム長のほぼ二乗に比例し、2GHz の Pentium4 で約 0.5[sec/FRAME]（256 サンプル）、約 2.0[sec/FRAME]（1024 サンプル）程度である。

4 結果

/aiueo/及び/kakaka…/と発音したときの解析結果の一例について示す。

4.1 /aiueo/

図 1 は、/aiueo/と自然発話したときの、パワー（縦軸対数、上段）、零点交差回数（中段）、推定された埋め込み次元（下段）の時間変化の一例、図 2 は対応するフォルマントの時間変化である。

フォルマントが各母音に移行する部分（わたり音）にて滑らかに変化しているのに対し、埋め込み次元は全母音、わたり音を通じてほぼ 4±1 次元程度の一定値をとっている。

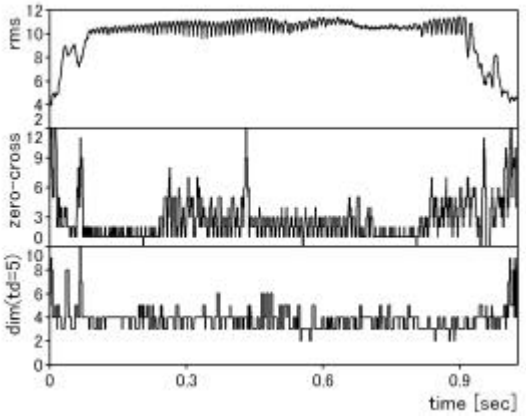


図 1　/aiueo/と発音したときの埋め込み次元の時間変化

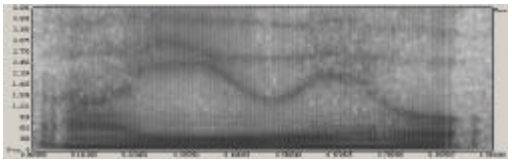


図 2　/aiueo/と発音したときのフォルマントの時間変化

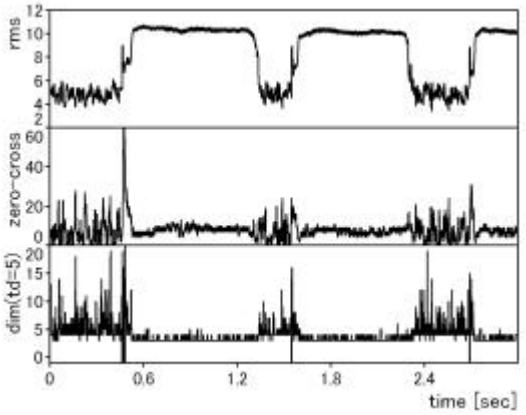


図 3　/kakaka…/と発音したときの埋め込み次元の時間変化

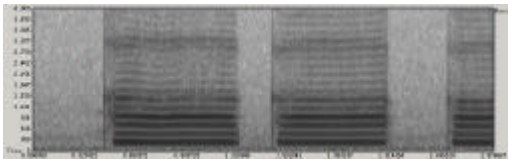


図 4　/kakaka…/と発音したときのフォルマントの時間変化

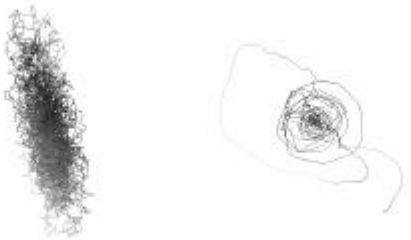
4.2　/kakaka…/

図 3 は、/kakaka…/と発話した場合の、パワー（対数軸、上段）、零点交差回数（中段）、推定された埋め込み次元（下段）の時間変化の一例、図 2 は対応するフォルマントの時間変化である。/k/の立ち上がりにおいて突発的にパワーが上昇したあとすぐ減少し、その後パワーの大きい/a/の定常状態へと移行する。

零点交差回数は、無音部と/k/の開始部においてほぼ同じ程度であるが、/a/に移行するに従って徐々に減少し、/a/の定常状態でほぼ一定値となる。

埋め込み次元数は、ノイズ部でやや暴れるものの、20 次元以下で収束している。しかし/k/から/a/の遷移部においてはほぼ確実に 20 次元では収束しないフレームが存在する。その後、定常的な/a/に移行するに従って、零点交差回数の場合と同じように徐々に値が減少しほぼ一定値（4±1 次元）に収束する。

さらに詳細に各素過程のダイナミクスを調査するため、無音部（バックグラウンドノイズ）、/k/から/a/の遷移部、/a/の持続部、/a/の減衰からノイズへの遷移部に分類した。図 5 に、バックグラウンドノイズ、/k/から/a/の遷移部、/a/の位相空間上の軌道を示す。



(1) バックグラウンドノイズ　(2) /k/から/a/の遷移部

(3) /a/

図 5　各音素の位相空間上の軌道

図 4 のスペクトル構造を見る限り、/k/はノイズライクな音素であると推察されるが、図 5 に示す位相空間上の構造からは、明らかにノイズとは異なるダイナミクスを有すると考えられる。

5 考察

/aiueo/において、零点交差回数は、被験者四人とも（変化の仕方はまちまちではあるものの）各母音によって傾向が異なるのに対し、埋め込み次元については 4 人の被験者ともほぼ 4±1 次元の間の値をとることがわかった。わたり音の部分について目立った次元の上昇がみられないことは、次に示す/kakaka…/の場合と対照的である。

/kakaka…/の解析にて分類した、ノイズ部、/k/から/a/への遷移部、/a/、/a/からノイズ部への遷移部の各音素について、推定された埋め込み次元のヒストグラムを示す（図 6～9）。

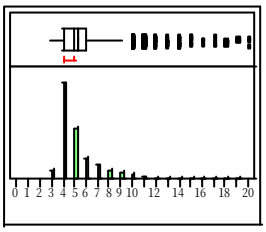


図 6　ノイズ部

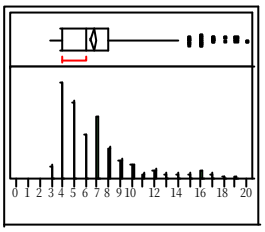


図 7　/k/ から/a/の遷移部

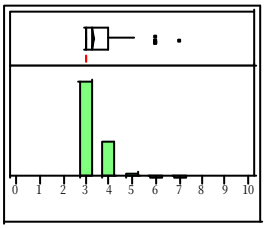


図 8　/a/

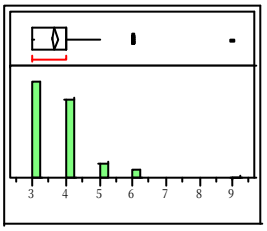


図 9　/a/からノイズへの遷移部

図 6 と図 7 の比較すると、ノイズ部のヒストグラムが対数正規分布に近い分布をとるのに対し、/k/から/a/の遷移部のヒストグラムは 7～10 次元程度の埋め込み次元の度数上昇が顕著に見られる。これは/k/から/a/に移行する際に徐々に低次元へと緩和する影響であると考えられる。また、この低次元への緩和と同時に零点交差回数も減少しており、両者には正の相関があると考えられる。これらの傾向は、4 人の被験者に対し共通に見られた。

本研究において、子音部又はノイズ部において 20 次元では収束せずに、次元推定が破綻するケースが幾つかみられた。特に、子音部においてはノイズ部から/k/に遷移する際、1 ステップ(32 サンプル)の間に約 100 倍程パワーが増加するフレームが存在するが、この部分ではほぼ確実に次元推定が破綻することが分かった。このような急激なダイナミクスの変化には、本次元推定法では対応できないことが考えられる。過渡的な信号の解析には時間分解能が必要であるが、同時にフレーム長を短くせざるを得ない。子音における次元推定の破綻が、子音が 20 次元以上のダイナミクスを有するために起こるものなのか、サンプル数が少ないために収束せず起こるものなのか、本質的に誤り近傍法による次元推定がこのような急激なダイナミクスの変化に対応できないのかを明確にするのは今後の課題としたい。

6 まとめ

本研究では、誤り近傍点法により母音のわたり音における埋め込み次元、子音から母音への遷移する過程における埋め込み次元を推定した。

母音のわたり音部分では埋め込み次元は 4±1 次元でほぼ一定であり、零点交差回数との明確な相関は得られなかった。

一方、子音から母音への遷移部において高次元（7～10 次元）から低次元（4±1 次元）への緩和がみられ、おおむね零点交差回数と正の相関が見られた。

本研究により、発話音声中の埋め込み次元推定及び零点交差回数等の統計的性質が明らかとなり、音素の抽出アルゴリズムへの応用が期待できる。推定された埋め込み次元の結果をふまえ、さらに発話音声中の Lyapunov 指数の時間変動や、位相の軌道について着目し、音声認識や音声認証等への応用可能性について探っていきたいと考えている。

参考文献

[1] Isao Tokuda, Ryuji Tokunaga, Kazuyuki Aihara: “A Simple Geometrical Structure Underlying Speech Signals Of The Japanese Vowel /a/”, *International Journal of Bifurcation and Chaos*, Vol. 6, No. 1 (1996) pp.149-160

[2] 大 聖一郎、和田 充雄、山口 明弘、広奥 暢:「日本語母音声のカオス性解析とその特徴について」、バイオメカニズム 16、バイオメカニズム学会編、東大出版、pp.285-299 (2002)

[3] Kennel, Matthew B. & Brown, Reggie: “Determining embedding dimension for phase-space reconstruction using a geometrical construction”, *Phys. Rev. A., 45,* pp.3403- (1992)

[4] Takens, F. : “Detecting strange attractors in turbulence, Lecture Notes in Mathematics”, *Springer-Verlag, Berlin,* 898, pp.366-381 (1981)