Surrogate Analysis of Japanese Vowels

Mem. Muroran Inst. Tech.,50(2000)17~21

Surrogate Analysis of Japanese Vowels

Isao TOKUDA*, Takaya MIYANO** and Kazuyuki AIHARA***

(Accepted 31 August 2000)

Nonlinear deterministic dynamical structure of the normal phonation of Japanese vowels is studied by the method of surrogate. The surrogate analysis exploits Wayland translation error and nonlinear deterministic predictability as the discriminating statistics. The results imply that the vowel signals have strong nonlinear dynamical characteristics that can not be detected by conventional linear dynamical systems analyses of speech.

Keywords: Nonlinear Dynamics, Vowels, Pitch-To-Pitch Variation, Surrogate Analysis

1 Introduction

In the studies of human speech, linear dynamical systems analyses such as the power spectrum analysis and the linear predictive coding modeling are the most popular and standard methodologies [1, 2, 3]. This is because the acoustical characteristics of human speech is mainly due to the resonances of the vocal-tract, which forms the basic spectral structure of the speech signals. Although the linear dynamical systems analyses have been widely applied to speech, human speech is, strictly speaking, nonlinear dynamical phenomena which involve nonlinear aerodynamic, biomechanical, physiological, and acoustic factors. Some of the important characteristics of speech may not be detected only by linear dynamical systems analyses. For a deeper understanding of speech, nonlinear dynamical systems analyses might be indispensable. In fact, there are a variety of recent studies of analyzing nonlinear dynamics such as chaotic dynamics in human speech [4]-[10].

The aim of the present paper is to re-examine the efficiency of linear dynamical systems analyses of speech and to consider the limitation of characterizing the speech only by the linear statistical quantities. Our approach is based upon surrogate test of speech. The surrogate test [11, 12, 13] is a kind of statistical hypothesis testing which is to detect nonlinear dynamical structure in a time series data. The results imply that in the normal phonation of Japanese vowels there seem to exist some important nonlinear dynamical characteristics that can not be detected by the conventional linear dynamical systems analyses of speech.

2 Speech Data

For our analysis, speech signals of 5 Japanese vowels /a/, /i/, /u/, /e/, and /o/ recorded from 5 subjects are exploited. The subject group is com-

* Department of Computer Science and Systems Engineering
** Department of Intelligent Machine and System Engineering, Hirosaki University
***Department of Mathematical Engineering and Information Physics, University of Tokyo

posed of 3 male speakers (mau, mms, mmy) and 2 female speakers (fsu, fyn) with no evidence for any laryngeal pathology. The speech data are in the standard ATR (Advanced Telecommunications Research Institute International, www.ctr.atr.co.jp) database. The speech signal is low-pass filtered with a cut-off frequency of 8 kHz and digitized with a sampling rate of 20 kHz and with 16 bit resolution. For the analysis of speech, the initial transient phase and the final decay phase are removed from each data and the stationary part of the data is extracted.

Fig. 1 (a) shows speech signal $\{x_t : t = 1, 2, .., N_{data}\}$ ($N_{data} = 2048$) of vowel /a/ recorded from a male speaker mau.

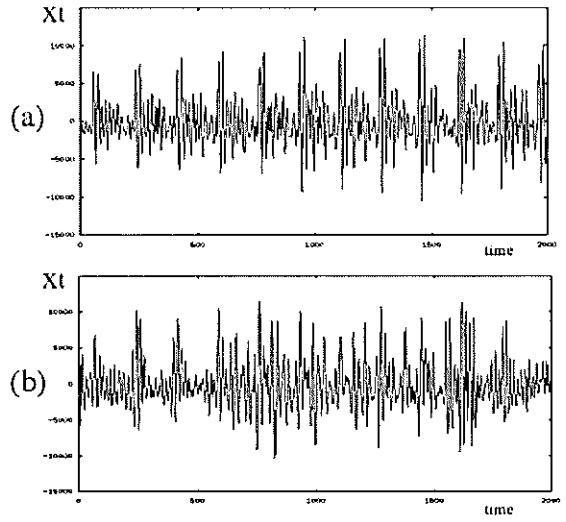

Fig. 1. (a): Original speech signal of a vowel /a/ and (b): its iterative surrogate.

The vowels are known to have irregularity in the pitch-to-pitch variation [14] and they are indispensable for the speech signals to be perceived as natural human sound [15]. Although considerable amount of works has been devoted for the analysis and synthesis of natural pitch-to-pitch variation, no satisfactory understanding of its statistical property has been obtained yet. There might be some nonlinear dynamical structure underlying the pitch-to-pitch irregularity and such nonlinear dynamics in the vowel signals is investigated in the subsequent sections.

3 Surrogate Analyses

3.1 Iterative Surrogate Algorithm

Let us study the nonlinear dynamics of speech by the surrogate test [11, 12, 13]. The surrogate test is a kind of statistical hypothesis testing. First, we set a null hypothesis H_0 that the speech signal is generated from some non-deterministic dynamical process. Then, we artificially create many sets of surrogate data which agree with the null hypothesis. By computing a discriminating statistic T of the original and surrogate data and by observing the difference between the original and surrogate statistics, we can test the null hypothesis H_0.

The surrogate test has the property of "constrained-realization," [12] which is to randomize the original data by strictly preserving some of the original statistical properties. It has been empirically known that the surrogate test is effective for statistical hypothesis testing when *non-pivotal* discriminating statistic T is utilized.

For the nonlinearity test of speech, we consider the following null hypothesis,

H_0: "The speech signal $\{x_t\}$ is generated from a linear *Gaussian* process

$$x_t = a_0 + \sum_{k=1}^{q} a_k x_{t-k} + e_t,$$

where e_t represents *Gaussian* noise."

We consider this null hypothesis, because this is the standard hypothesis when linear dynamical systems analysis is applied to speech [3].

For the null hypothesis H_0, the surrogate data is generated by the Schreiber-Schmitz iterative algorithm [13]. The iterative algorithm generates surrogate data which exactly preserves the amplitude distribution and approximately preserves the power spectrum of the original data.

Fig. 1 (b) shows an iterative surrogate signal generated from the speech signal of the vowel /a/ of Fig. 1 (a). Although the waveform structures of the original and surrogate data seem to be rather different from each other, they have exactly the same amplitude distribution and approximately the same power spectrum (see Fig. 2).

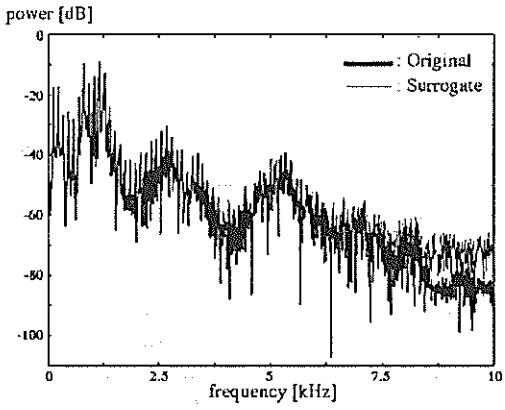

Fig. 2. Power spectra of the original speech signal of the vowel /a/ (bold line) and its iterative surrogate (thin line).

3.2 Wayland Translation Error

By computing a discriminating statistic T of the original data and the surrogate data set, we can test the null hypothesis H_0. If the T-value of the original data is far from the distribution of the T-vales of the surrogate data set, we can reject the null hypothesis H_0. As the discriminating statistic T, Wayland translation error and nonlinear deterministic prediction error are exploited.

As the first discriminating statistic T, Wayland translation error is computed as follows [16]. First, we reconstruct the geometric structure of the speech data in a delay-coordinate space [17, 18]:

$$\boldsymbol{x}(t) = (x_t,\ x_{t-\tau},\ \cdots,\ x_{t-(d-1)\tau})^T, \quad (1)$$

where T denotes transposition and d and τ stand for the reconstruction dimension and the time lag, respectively. Then, the Wayland algorithm assumes that the reconstructed data trajectory $\{\boldsymbol{x}(t) : t = 1 + (d-1)\tau, .., N_{data}\}$ is generated from a continuous map $f : R^d \to R^d$ as $\boldsymbol{x}(t+1) = f(\boldsymbol{x}(t))$ and hence "nearby" data points, e.g., $\boldsymbol{x}(t)$ and $\boldsymbol{x}(s)$, are mapped to nearby T-step future points, e.g., $\boldsymbol{x}(t+T)$ and $\boldsymbol{x}(s+T)$. With respect to the assumed determinism, the *translation error*, $e_{\rm trans}$, can be calculated as follows.

First, for a fixed data point $\boldsymbol{x}(t_0)$, we find its k-nearest neighbors as $\boldsymbol{x}(t_1)$,.., $\boldsymbol{x}(t_k)$. With respect to the *translation horizon* T, the translation vectors are computed as $v_j = \boldsymbol{x}(t_j + T) - \boldsymbol{x}(t_j)$ for $j = 0, .., k$. Then the translation error is calculated as

$$e_{\rm trans} = \frac{1}{k+1} \sum_j^k \frac{\| v_j - < v > \|^2}{\| < v > \|^2}, \quad (2)$$

with $< v > = \frac{1}{k+1} \sum_j^k v_j$.

Fig. 3 shows the result of the Wayland analysis applied to vowel /a/ (subject: mau). Translation error curves of the original speech data and 39 sets of its surrogate data are drawn with a solid line with circles and solid lines with no circle, respectively. The translation error, $e_{\rm trans}$, is averaged over 20 sets of 300-randomly chosen translation centers $\boldsymbol{x}(t_0)$, other parameters are set as $\tau = 10$, $k = 4$, $T = 10$, and the reconstruction dimension is varied as $d = 1, .., 15$.

Clear difference between the original speech data and its surrogate data is discernible in the figure, where the original data exhibits relatively lower error level than the surrogate data.

In Table 1, rejection level α is summarized for 5 Japanese vowels (/a/, /i/, /u/, /e/, /o/) and for 5 subject speakers (mau, mms, mmy, fsu, fyn). The rejection level α means that T-value of the original data is out of $100(1-\alpha)\%$ confidence range of the surrogate distribution and hence the null hypothesis H_0 can be rejected with the α-level. For all 5 vowels and for all 5 subjects, the null hypothesis H_0 is rejected with the level of $\alpha = 0.05$. This is in general a strong rejection level for statistical test and the results seem to be independent of the vowels and the subjects. Therefore the results imply that the vowels are not generated from a simple linear *Gaussian* process. There might be some nonlinear dynamics underlying the irregular structure of the vowels and such nonlinear characteristics have been destroyed by the surrogate shuffling.

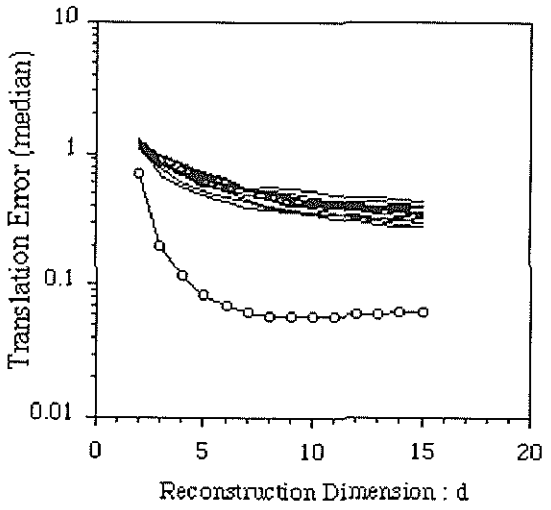

Fig. 3. Result of the Wayland analysis of vowel /a/ (subject: mau). *Translation error* curves of the original data and 39 sets of its surrogates are drawn with a solid line with circles and solid lines with no circle, respectively.

3.3 Nonlinear Prediction Error

As another discriminating statistic T, we compute nonlinear deterministic prediction error as follows.

First, we divide the time series $\{x_t : t = 1, .., N_{data}\}$ into first and second halves. From the first data, a nonlinear predictor $\tilde{f} : R^d \to R^d$ which approximates the data dynamics as $\boldsymbol{x}(t+1) \approx \tilde{f}(\boldsymbol{x}(t))$ is constructed. For the predictor, the local optimal linear-association map [19] is exploited with the embedding condition of $(d, \tau) = (6, 2)$. Then, for the latter data, nonlinear prediction is carried out. Forecasting procedure is that, for a give initial state, $\boldsymbol{x}(t)$, the s-step further state $\boldsymbol{x}(t+s)$ is predicted as $\tilde{f}^s(\boldsymbol{x}(t))$ using the s-iterate of the predictor $\tilde{f}$. For each s-prediction step, accuracy of the predictions is evaluated with the correlation coefficient r_s between the actual and prediction data.

The results of the nonlinear prediction of vowel /a/ (subject: mau) is shown in Fig. 4. Prediction curves of the original speech data and 39 sets of its surrogate data are drawn with a solid line with squares and dotted lines with no square, respectively. We clearly see that the original data exhibits better nonlinear predictability than the surrogate data.

In Table 1, rejection level α is summarized for 5 Japanese vowels and for 5 subject speakers. The prediction step is fixed at $s = 2$. For vowel /o/ from female subjects, relatively high rejection level α was computed. This might be caused by strong constriction of the vocal tract shape that dissipates nonlinear dynamics of the vocal folds and produces relatively simple harmonic speech sound. For such a harmonic signal, detection of nonlinear dynamics becomes difficult. For other vowels except for female vowel /o/, the null-hypothesis H_0 is rejected with the strong level of $\alpha = 0.05$ (for few samples, $\alpha = 0.1$). Hence the results imply that nonlinear deterministic predictability of the vowels is destroyed by the surrogate shuffling.

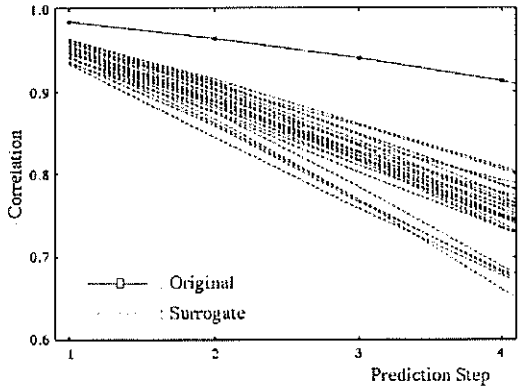

Fig. 4. Results of the nonlinear deterministic prediction of vowel /a/ (subject: mau). Prediction curves of the original speech data and 39 sets of its surrogates are drawn with a solid line with squares and dotted lines with no square, respectively.

Table 1.

Rejection level α of the surrogate test of 5 Japanese vowels (/a/, /i/, /u/, /e/, /o/) from 5 subject speakers (mau, mms, mmy, fsu, fyn). As the discriminating statistic T, Wayland translation error e_{trans} and nonlinear prediction error r_2 were exploited.

	T	/a/	/i/	/u/	/e/	/o/
mau	e_{trans}	0.05	0.05	0.05	0.05	0.05
	r_2	0.05	0.05	0.10	0.05	0.05
mms	e_{trans}	0.05	0.05	0.05	0.05	0.05
	r_2	0.05	0.05	0.05	0.05	0.05
mmy	e_{trans}	0.05	0.05	0.05	0.05	0.05
	r_2	0.05	0.05	0.05	0.05	0.10
fsu	e_{trans}	0.05	0.05	0.05	0.05	0.05
	r_2	0.05	0.05	0.05	0.05	0.20
fyn	e_{trans}	0.05	0.05	0.05	0.05	0.05
	r_2	0.05	0.05	0.05	0.05	0.40

4 Conclusions and Discussions

Nonlinear dynamical structure of the normal phonation of Japanese vowels has been tested by the method of surrogate. For a null hypothesis that the speech signal is generated from a linear *Gaussian* process, surrogate data is generated by the Schreiber-Schmitz iterative algorithm. As a discriminating statistic T, Wayland translation error and nonlinear deterministic prediction error are exploited. The surrogate analysis has shown that the null-hypothesis was rejected for almost all vowel signals with a level of $\alpha = 0.05$. The results seem to be independent of the vowels, male or female subjects, and the nonlinear discriminating statistics. This implies that there definitely exist some important nonlinear dynamical characteristics that has been destroyed by the surrogate data shuffling in the vowel signals. Nonlinear dynamical characteristics may provide us with useful information on speech signals such as individual speaker's character, speaker's emotional condition, and the laryngeal condition.

REFERENCES

(1) J.L. Flanagan, *Speech analysis, Synthesis, and Perception* (Springer, New York, 1960)

(2) D. H. Klatt & L. C. Klatt, J. Acoust. Soc. Amer. **87-2** (1990) 820

(3) B. S. Atal & S. L. Hanauer, J. Acoust. Soc. Amer. **50** (1971) 637

(4) I.R. Titze *et al.*, in *Vocal Fold Physiology*, I.R. Titze ed. (Singular, San Diego, 1993), 143

(5) W. Mende *et al.*, Phys. Lett. A 145 (1990) 418

(6) S.S. Narayanan & A.A. Alwan, J. Acoust. Soc. Am. **97-4** (1995) 2511

(7) M. Sato *et al.*, Proc. IJCNN **1** (1990) 581

(8) I. Tokuda *et al.*, Int. J. Bif. Chaos, **6-1** (1996) 149

(9) T. Ikeguchi & K. Aihara, J. Int. Fuzzy Sys., **5-1** (1997) 33

(10) T. Miyano, Proc. 4th Int. Conf. Soft Computing (1996) 634

(11) J. Theiler *et al.*, Physica D58 (1992) 77

(12) J. Theiler & D. Prichard, Physica D **94** (1996) 221

(13) T. Schreiber & A. Schmitz, Phys. Rev. Lett. **77-4** (1996) 635

(14) L. Dolanský & P. Tjernlund, IEEE Trans. AU-**16** (1968) 51

(15) T. Ifukube *et al.*, J. Acoust. Soc. Jpn. **47-12** (1991) 903

(16) R. Wayland *et al.*, Phys. Rev. Lett. **70-5** (1993) 580

(17) F. Takens, *Lecture Notes in Math.* (Springer, Berlin) **898** (1981) 366

(18) T. Sauer *et al.*, J. Stat. Phys. **65-3** (1991) 579

(19) J. Jimenéz *et al.*, Phys. Rev. A 45 (1992) 3553

日本語母音のサロゲート解析

徳田 功*, 宮野 尚哉**, 合原 一幸***

概要

日本語母音の正常発声音の決定論的非線形力学構造を Wayland のベクトル並進性と非線形予測誤差を検定量とするサロゲート法に基づいて解析する。解析結果から，日本語母音には強い非線形力学構造が存在し，そのような非線形力学特性はパワースペクトル解析や線形予測符号化法などの従来型の音声信号処理の手法では特徴付けることが難しいことが示唆された。

キーワード：非線形ダイナミクス，母音，ピッチ揺らぎ，サロゲート法

*情報工学科，**弘前大学理工学部知能機械システム工学科，***東京大学工学部計数工学科

Surrogate Analysis of Japanese Vowels

Isao Tokuda*, Takaya Miyano**, and Kazuyuki Aihara***

Nonlinear deterministic dynamical structure of the normal phonation of Japanese vowels is studied by the method of surrogate. The surrogate analysis exploits Wayland translation error and nonlinear deterministic predictability as the discriminating statistics. The results imply that the vowel signals have strong nonlinear dynamical characteristics that can not be detected by conventional linear dynamical systems analyses of speech.

Keywords: Nonlinear Dynamics, Vowels, Pitch-To-Pitch Variation, Surrogate Analysis

* Department of Computer Science and Systems Engineering
** Department of Intelligent Machine and System Engineering, Hirosaki University
*** Department of Mathematical Engineering and Information Physics, University of Tokyo