

A STOCHASTIC MODEL OF FUNDAMENTAL PERIOD PERTURBATION AND ITS APPLICATION TO PERCEPTION OF PATHOLOGICAL VOICE QUALITY

Yasuo Endo and Hideki Kasuya

Faculty of Engineering, Utsunomiya University
2753 Ishii-machi Utsunomiya, 321 Japan

ABSTRACT

This paper proposes a stochastic model of fundamental period perturbation and applies it to the perception of pathological voice quality. A fundamental period perturbation is generated by a second order auto-regressive moving average (ARMA) model which includes excitation by a random white noise. Standard deviation, pole frequency and pole bandwidth were used as parameters of the model. Sustained vowels were synthesized by systematically manipulating the model parameters and subjected to the perceptual experiment to understand roles of the parameters in the perceived quality of pathological voice. Five subjects participated in the experiment and were asked to judge whether a vowel sample was normal or pathological. It was found that perceptual impression of the pathological voice was associated not only with the magnitude but also the pole frequency and bandwidth of the fundamental period perturbation.

1. INTRODUCTION

Acoustic characteristics associated with pathological voice quality must be described from multidimensional viewpoints: perturbation in the fundamental period and peak amplitude sequence (jitter/shimmer), perturbation in the spectra of glottal airflow, laryngeal turbulence noise signal and so on. Among these characteristics, this paper focuses on stochastic modeling of fundamental period perturbation. Since Kasuya, Kobayashi, and Kobayashi [1] applied the auto-regressive (AR) model to the fundamental period sequence to make acoustic assessment of pathological voices, various models have been reported for the proper characterization of the fundamental period sequence of pathological voices [2-6]. In this paper, we extend the AR model to the auto-regressive moving average (ARMA) model to deal with the normalized period sequence in which a trend component of the sequence is removed. In order to systematically understand the contribution of individual acoustic parameters to pathological voice qualities, some synthetic method is indispensable, whereby one can perform perceptual tests under well con-

trolled experimental conditions. In most of the perceptual experiments of pathological voice qualities, the parametric (formant-type) speech synthesizer designed for producing normal voices has been used [7-9]. We propose in this paper a nonparametric harmonic synthesizer for the perceptual experiments, which requires no parametric model of a glottal airflow. Using this synthesis method, relationships between the ARMA parameters of the perturbation and perceptual impression are investigated.

2. MODELING PERTURBATION

A trend sequence included in the fundamental period sequence $\tau(m)$ is approximated by the third order polynomial,

$$\hat{\tau}(m) = \sum_{i=0}^3 \beta_i m^i, \quad (1)$$

where β_i is estimated by the least squares method. By removing the estimated trend sequence from the original sequence and dividing by the average value of the original sequence, we have a normalized sequence $x(m)$ in percent,

$$x(m) = \frac{\tau(m) - \bar{\tau}}{\bar{\tau}} \times 100 (\%), \quad (2)$$

where $\bar{\tau}$ is the average value of the original period sequence. A normalized sequence $x(m)$ is modeled by an ARMA process as follows:

$$\begin{aligned} v(m) &= -a_1 v(m-1) - a_2 v(m-2) + Ge(m), \\ x(m) &= x(m-1) + v(m), \end{aligned} \quad (3)$$

where $e(m)$ is a Gaussian white noise sequence. This model is illustrated in Fig.1.

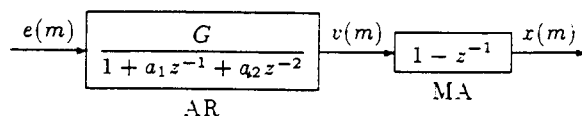


Figure 1: An ARMA model to generate the normalized sequence $x(m)$

The AR coefficients $\{a_i\}$ and the gain G are estimated using an ordinary linear prediction method. The pole frequency f_p and bandwidth Δf_p of the AR filter are obtained from the root of the AR polynomial, s_p , as follows:

$$f_p = \frac{1}{2\pi} \arg s_p \quad \Delta f_p = -\frac{1}{\pi} \log_e |s_p|. \quad (4)$$

Since the MA part has been introduced to simulate the removal of the trend sequence, the zero frequency is fixed at zero. We use the standard deviation of $x(n)$, σ , instead of the gain G , because the former appears to be more closely related to the perceived quality than the latter. We now have three ARMA parameters to represent the stochastic characteristics of the normalized sequence, i.e. σ , f_p and Δf_p . The validity of this model has been confirmed by the flatness measure [10].

3. SYNTHESIS SYSTEM

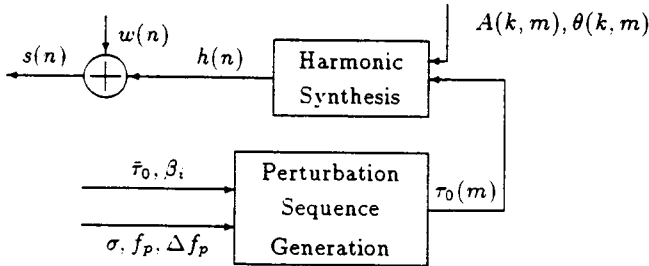


Figure 2: A synthesis system of pathological voice

A proposed synthesis system is illustrated in Fig. 2, consisting of two subsystems: generation of a perturbation sequence and harmonic synthesis. In the perturbation sequence generation, a normalized period $x(m)$ is first produced with the AR coefficients a_1 and a_2 ,

$$a_1 = -2e^{-\pi\Delta f_p} \cos 2\pi f_p, \quad a_2 = e^{-2\pi\Delta f_p}. \quad (5)$$

A gain G is also computed by σ , f_p and Δf_p . Next the fundamental period is generated by incorporating the average fundamental period $\bar{\tau}$, the trend sequence $\hat{\tau}(m)$ and the normalized period $x(m)$ as :

$$\tau(m) = \hat{\tau}(m) + \frac{\bar{\tau}}{100} x(m). \quad (6)$$

In the harmonic synthesis, the harmonic signal $h(n)$ is synthesized by using the fundamental period sequence $\tau(m)$, and the harmonic coefficients $c(k, m)$, $d(k, m)$ or harmonic amplitude $A(k, m)$ and phase $\theta(k, m)$ as

$$\begin{aligned} h_m(n) &= \sum_{k=1}^{K(m)} (c(k, m) \cos k\omega_0(m)n + d(k, m) \sin k\omega_0(m)n) \\ &= \sum_{k=1}^{K(m)} A(k, m) \cos(k\omega_0(m)n - \theta(k, m)). \end{aligned} \quad (7)$$

where

$$\begin{aligned} \omega_0(m) &= \frac{2\pi}{\tau_0(m)}, \\ A(k, m) &= \sqrt{c^2(k, m) + d^2(k, m)}, \\ \theta(k, m) &= \tan^{-1} \frac{d(k, m)}{c(k, m)}. \end{aligned} \quad (8)$$

$K(m)$ is the total number of the harmonic components to be taken into account. In order to maintain continuity of the phase of the harmonics, $\theta(k, m)$ is computed as follows :

$$\theta(k, m) = \theta(k, m-1) + 2\pi \left(1 - \frac{[\tau_0(m-1)]}{\tau_0(m-1)} \right) k. \quad (9)$$

where $[\cdot]$ is an integer operation. Finally, the voice signal $s(n)$ is computed by adding the laryngeal noise signal $w(n)$ to the harmonic signal $h(n)$:

$$s(n) = h(n) + w(n). \quad (10)$$

4. EXPERIMENT

The proposed synthesis system was applied to investigate the relationship between the stochastic nature of the fundamental period sequence and perceived quality. From a vowel /e/ sustained by a normal subject, a fundamental period sequence $\tau_0(m)$ was extracted by means of the waveform matching procedure [11]. Harmonic coefficients $c(k, m)$, $d(k, m)$ were estimated from the vowel signal data of J consecutive periods so as to minimize the following :

$$E_0 = \sum_{j=0}^{J-1} \sum_{n=0}^{\tau_0(m)-1} (s_{m+j}(n) - h_m(n))^2. \quad (11)$$

A laryngeal noise signal $w(n)$ was computed by subtracting $h(n)$ from $s(n)$

$$w(n) = s(n) - h(n). \quad (12)$$

Trend coefficients β_i were estimated by the least squares method. An average value $\bar{\tau}$ and a standard deviation σ of the period sequence were computed by the period sequence. A pole frequency f_p and a pole bandwidth Δf_p were measured by Eqs. (3) and (4).

In the synthesis, σ , f_p and Δf_p were systematically changed from 0.8 to 1.6, from 0.2 to 0.5 and from 0.02 to 0.80, respectively. 80 voices which gave reasonable quality were used for the perceptual experiment. Five subjects participated in the experiment and were asked to judge whether a voice sample was normal or pathological.

5. RESULTS AND DISCUSSION

Examples of the normalized period sequence $x(m)$ generated by the model are shown in Figs. 3 and 4 for different values of the ARMA model parameters, where the spectrum (middle) and correlogram (bottom) are also illustrated. In Fig. 3, since $f_p = 0.33$, the correlogram has a peak at a lag of 3. As compared with Fig. 3, Fig. 4 demonstrates the effect of the pole bandwidth: Δf_p of Fig. 3 is 0.02, whereas that of

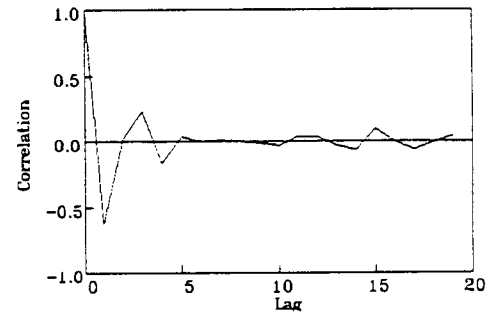
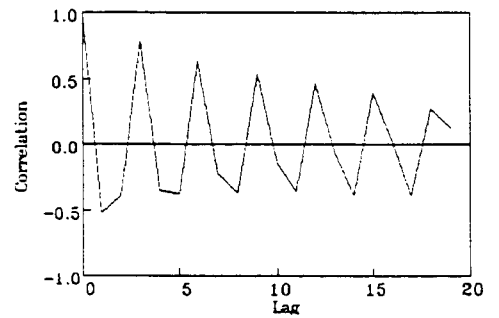
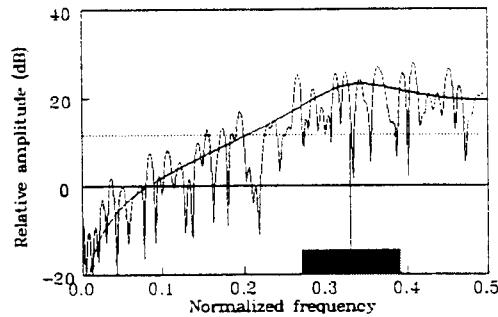
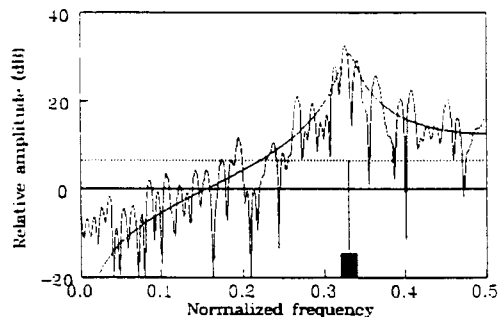
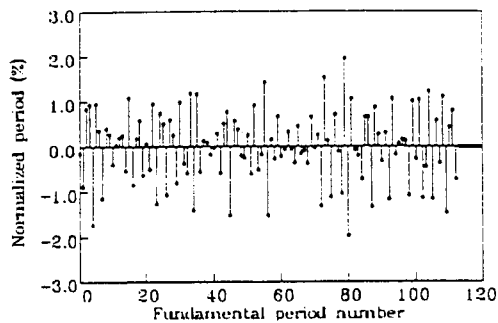
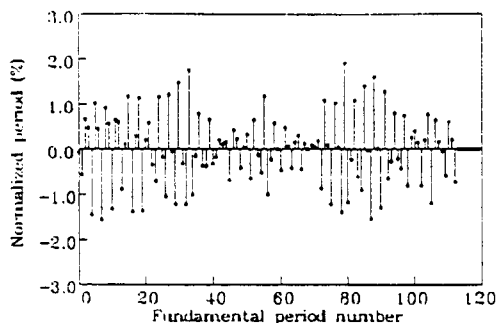


Figure 3: A normalized sequence generated by the model with $\sigma = 0.8$, $f_p = 0.33$ and $\Delta f_p = 0.02$.

Figure 4: A normalized sequence generated by the model with $\sigma = 0.8$, $f_p = 0.33$ and $\Delta f_p = 0.12$.

Fig. 4 is 0.12. This difference of the Δf_p reflects in the correlograms. The peak values of the correlogram attenuate less in Fig. 3 than in Fig. 4. To summarize the roles of the pole parameters: Pole frequency controls the frequency of a pseudo-periodic component (an inverse of the period) in the sequence, if any, and the pole bandwidth determines the amount of the component.

wide bandwidth. It can be seen in Tables 3 and 4 that the narrower the pole bandwidth, the higher the rate of pathology. This result may be caused by easiness of pulsational perception for narrow bandwidth.

Results of the perceptual experiment are shown in Table 1, 2, 3 and 4, where numbers represent rate at which the stimuli were judged as pathological. It can be seen in Table 1 that the lower the pole frequency, the higher the rate of pathology. This coincides with the result given in [8.13]. In Table 2, however, definite difference of the rate is not shown when the pole frequency is changed. This result may be caused by

6. CONCLUSION

We proposed a stochastic model of fundamental period perturbation and a nonparametric harmonic synthesizer of pathological voice. We investigated the relationship between the model parameters and perceived impression. Findings obtained from the experiment are as follows: (1) The rate of pathological judgments significantly changed when the pole frequency and bandwidth were changed, even though the standard deviation was kept constant. (2) The rate increased as the pole frequency lowered or the pole bandwidth

narrowed when the standard deviation was kept constant.

Table 1: Rate at which the stimuli were judged as pathological for $\Delta f_p=0.02$

$f_p \backslash \sigma$	0.8	1.0	1.2	1.4	1.6
0.20	50	75	100	100	100
0.25	31	50	94	100	100
0.33	0	6	81	94	100
0.40	0	6	56	81	94
0.50	6	31	69	88	94

Table 2: Rate at which the stimuli were judged as pathological for $\Delta f_p=0.8$

$f_p \backslash \sigma$	0.8	1.0	1.2	1.4	1.6
0.20	0	13	25	38	88
0.25	0	0	19	63	88
0.33	0	6	13	44	81
0.40	6	0	19	75	94
0.50	0	6	25	50	75

Table 3: Rate at which the stimuli were judged as pathological for $f_p=0.2$

$\Delta f_p \backslash \sigma$	0.8	1.0	1.2	1.4	1.6
0.02	50	75	100	100	100
0.05	31	75	100	100	100
0.12	0	69	88	100	100
0.32	0	13	44	75	100
0.80	0	13	25	38	88

Table 4: Rate at which the stimuli were judged as pathological for $f_p=0.5$

$\Delta f_p \backslash \sigma$	0.8	1.0	1.2	1.4	1.6
0.02	6	31	69	88	94
0.05	13	63	69	100	100
0.12	13	56	69	81	100
0.32	6	19	50	75	81
0.80	0	6	25	50	75

7. REFERENCES

1. H. Kasuya, Y. Kobayashi, T. Kobayashi, "Characteristics of pitch period and amplitude perturbations in pathologic voice," Proc. ICASSP, pp. 1372-1375, 1983.
2. S. Imaizumi and J. Gauffin, "Acoustical and perceptual characteristics of pathological voices : rough, creak, fry and diplophonia," ANN.BULL.RILP, No.25, pp. 109-119, 1991.
3. J. Hirama and Y. Kakita, "Effects of F_0 fluctuation on the category of two pathological voice qualities : rough and voice tremor," JJLP, Vol. 33, No. 1, pp. 2-10, 1992 (in Japanese).
4. J.Schoentgen and R. De Guchteneere, "Auto-regressive linear models of jitter", EUROSPEECH'93, pp. 2033-2036, 1993.
5. Y. Kakita and H. Okamoto, "Chaotic characteristics of voice fluctuation and its model explanation : normal and pathological voices," ICSLP94, pp. 639-642, 1994.
6. I. R. Titze, "Interference between normal vibrato and artificial stimulation of laryngeal muscles at near-vibrato rates," Journal of Voice, Vol. 8, No.3, pp. 215-223, 1994.
7. J. Hillenbrand, "Perception of aperiodicities in synthetically generated voices," J. Acoustic. Soc. Amer., pp. 2361-2371, 1983.
8. J. Gauffin and S. Granqvist, "Irregularities in the voice : some perceptual experiments using synthetic voices," ICPHS95, Vol. 2, pp. 242-245, 1995.
9. A. Alwan, P. Bangayan, J. Kreiman and C. Long, "Time and frequency synthesis parameters of severely pathological voice qualities," ICPHS95, Vol. 2, pp. 250-253, 1995.
10. Y. Endo and H. Kasuya, "Synthesis of pathological voice based on a stochastic voice source model," IC-SLP94, pp. 1991-1994, 1994.
11. H. Kasuya, S. Ogawa and Y. Kikuchi, "An adaptive comb filtering method as applied to acoustic analysis of pathologic voice," Proc. ICASSP, pp. 669-672, 1986.
12. C. Yang and H. Kasuya, "A method to detect laryngeal noise from continuous speech," Proc. ICA, pp. G2-10, 1992.
13. I. Pollack, "Detection and relative discrimination of auditory "jitter".," JASA, Vol. 43, No. 2, pp. 308-315, 1968.

